



IFC – Bank Indonesia International Workshop and Seminar on *“Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data”*

Bali, Indonesia, 23-26 July 2018

Standardised approach in developing economic indicators using internet searching applications¹

Paphatsorn Sawaengsuksant,
Bank of Thailand

¹ This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Standardized Approach in Developing Economic Indicators using Internet Searching Applications[†]

Paphatsorn Sawaengsuksant

July 2018

Abstract: This study introduces new approach of utilizing internet search data set in monitoring economic conditions. These internet search data are well-known for their ability to enhance predictive power of forecast model as well as the almost real-time availability. Yet, choosing the right searching words has always been the major hindrances for real-world application. And by raising predictive power alone is not sufficient especially for policy makers as the data could be sometimes suffered from an inconsistency as well as unintentionally sample biased from choosing the wrong words. This study introduces a more standardized approach to traverse those problematic difficulties and, at the same time, enhance reliability and economic meaningfulness of internet search data while maintaining predictive performance of indicators. In addition to the well-known Google search engine, a complementary internet application, namely the Google Correlation, is also proved to be useful in creating new economic indicators following the new introduced standardized steps. Four indicators are developed accordingly and have currently been applied in real economic condition monitoring process in the Bank of Thailand including (1) purchasing power of private household, (2) consumer confidence of private household, (3) private consumption expenditure of durable products, and (4) number of unemployed persons.

Keywords: internet search data, macroeconomic monitoring indicator, nowcast, big data

[†]Sawaengsuksant: Macroeconomic and Monetary Policy Department, Bank of Thailand (BOT), Email PaphatsS@bot.or.th. I am grateful to colleagues at BOT, especially Napat Phongluangtham, Pranee Sutthasri, Jirawat Poongam, Teerapap Pangsap, and Nutchaphol Jaroonpipatkul, for their suggestions and encouragement.

Disclaimer: The opinions expressed in this paper are those of the author and should not be attributed to BOT. All errors are my own.

1 Introduction

Google Trend is currently one of the most common analytic tools noted by various studies and applying by policy maker units. There are many reasons behind such as timeliness, broad range of applicable study fields, friendly user interface and free data access. Choi and Varian (2009a, 2009b and 2011) shows that Google Trend Index is useful for nowcasting several economic variables, for instances, retail sales, vehicle sales, home sales, travelling demand, as well as unemployment initial claims. Vosen and Schmidt (2011) introduce new indicator for private consumption using Google Trend.

Google Correlate is also useful for analytic tasks. The Google Correlate has been launched since 2011. This interface provides a word list whose searching frequency is matched with time series inserted into the interface. Instead of trial and errors, this program learns from historical searching pattern and automatically deliveries words with high correlation to the interested series. Location of search, like countries, is also available. But it is not very common in academic literature since, in most of the time, those words fail to provide meaningful economic sense. Understanding pros and cons of these data is important to prevent data misinterpretation. Advantages and disadvantages of these data are discussed as followed;

Advantages These internet search data are well-known for their ability to enhance predictive power as well as almost real-time availability. This searching program acts like internet-based survey asking people what topics they are interested in at a certain point of time. Therefore, the potential research frontier is extremely considerable as long as internet users concern and fill in the web browser. This frontier is strikingly outbound traditional surveys. And not just timeliness, these data could perform as (1) alternative indicators to previous ones as well as (2) an answer to new questions that traditional data is too expensive or too late to conduct a survey.

Disadvantages Data are unstructured. They are not originally generated for analytic purposes. In case of traditional surveys, questions are carefully designed to answer specific questions and to minimize unrelated noises as much as possible. Meanwhile, data from Google applications provide searching frequency which is actually a bundle of numerous signals: both the informative signals and unrelated noises. Sources of noises are also different from the traditional structured data, raising a challenge in further applications' validity. For examples, noises could be generated simply from technical issues of the program interface, or from human behaviors unrelated to the questions. Bortoli and Combes (2016) note that shortness of series and lack of transparency about treatments and sampling processes are weakness of Google Trend. Therefore, careful data quality assessment process is crucially needed since using data without noticing these irrelevant noises could lead to seriously inappropriate policy decisions.

Despite earlier disadvantages, this study shows that utilizing Google Trend and Google Correlate interfaces together can deliver useful words for monitoring latest economic developments. However, without proper words filtering, set of searching words might perform well in prediction but fail to provide sensible explanation and useful insights. This study suggests general criteria to filter undesirable words, in an objective that searching frequency of filtered words are suitable for macroeconomic monitoring purpose.

This study is structured as followed; the second section explains criteria of searching words to handle unfavorable and serious noises, and the third section shows study cases

applying the criteria to construct practical monitoring indicators. Limitation of data and this method is discussed in the fourth section.

2 Criteria of Searching Word

In the real-world application, information receiving from internet search data is always unstructured and untidy. Noises occur from technical issues could lead to several problems, including inconsistent patterns of series after a change in small word punctuation or change in researcher's IP address number. There are also noises from human behavior unrelated to the question but coincidentally alter patterns of series. Five criteria are introduced to filter out these noises, as followed.

2.1 First criteria: searching word is not specific to certain product or brand

This criteria objective is to confirm that the chosen words are generalized enough to cover the overall economic conditions, not to certain names. This is to avoid problems that such words are unintentionally tied to unrelated events. Although some specific names of popular products and brands might sound sensible to monitor economic conditions. However, search of those names actually represents specific shocks to the brand or product. For example, searching frequency of famous brand might be affected by competition among firms, temporary promotions or news of the brand. Some brands might even be coincidentally similar to name of songs, locations, movies, or related events, which are not associated to the real economic conditions.

Figure 1

This figure shows searching frequency of the recently popular e-business brand in Thailand. The series is too short to confirm that it can capture developments of e-business in Thailand.



Source: Google Trend Interface, searching “บัตรเรบบิท” in Thailand, data as of June 2017

2.2 Second criteria: searching word covers sufficient large sample size

This criteria objective is to confirm that, the real searching number of the chosen words are sufficiently large. This is to prevent significant revision of the whole series which is arisen by a simply change in timing of using the Google Trend interface or change in researchers' IP

address number. Reactions of searching series to important economic events are also observable given that the chosen series is long enough.

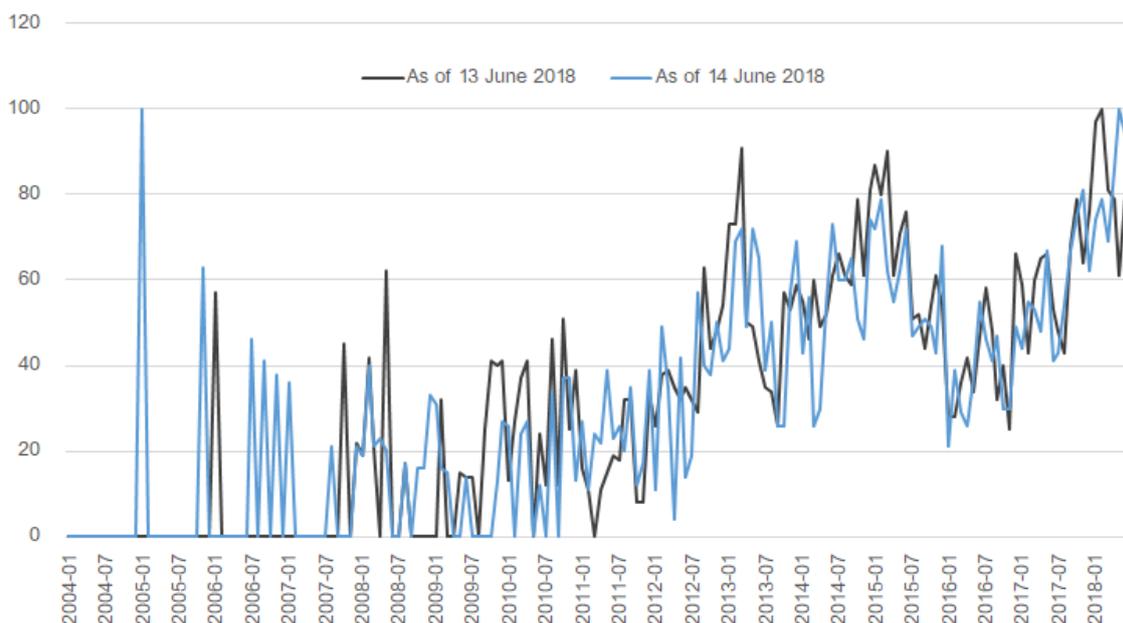
In Google Trend interface, searching frequency is normalized and ranged from zero to a hundred. Zero value means that almost no user putting this word into his web browser while a hundred value shows the most frequent search period. Some words, for examples, newly created words or those words which have becoming popular just recently, could show zero value for long time before spiking in a certain period. Performance evaluation of these kinds of words is impossible as the series is too short. This study uses only Google Trend Index which contains no zero value since 2008. The reasons are that (1) words which have been searched for at least 10 years should cover sufficiently large number of searches (2) the length of 10 years is long enough to capture major cyclical economic development, such as the Global Financial Crisis (2008 – 2010), the big flood in Thailand (2011), and the political uncertainty (2013-2014).

Moreover, the Google Trend Index is automatically calculated from sample, not the total population. Consequently, merely change in timing of using the Google Trend interface or researcher’s IP address number result in “revision” of series. In case that popularity of search is not sufficiently large, a revision could significantly alter a story suggested from the series. But when total population is sufficiently large, the revision is insignificant as shown in Figure 2. Alternatively, central value, like average, median, and mode, of series can be applied to reduce this inconsistent from data revision.

Figure 2

Google Trend Index using “ซื้อกองทุนรวม”, meaning “purchase mutual fund”

Every time the date or researcher’s IP address number is changed, searching frequency changes significantly and returns inconsistent story.



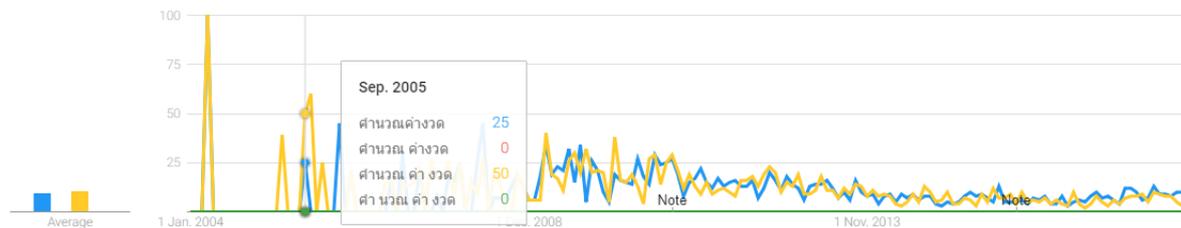
Source: Google Trend Interface, data as of June 2018

2.3 Third criteria: change in punctuation provides consistent searching series

This criteria objective is to confirm that, pattern of the whole series behave consistently after a small variation in word pattern¹. For instances, space between two words means that searching number of those two are included, regardless of their ordering. Punctuation, however, normally appears in Thai language as separation of sentences, not words. But sometimes an insignificant punctuation cause a large change in Google Trend Index as shown in Figure 3 and 4. Different searching language may encounters this kind of noises differently.

Figure 3

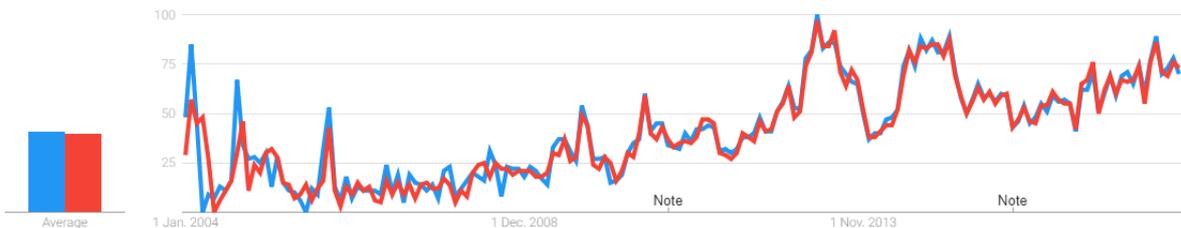
Google Trend Index searching for “คำนวณค่างวด”, meaning “calculate payment”, in different but insignificant punctuation, namely “คำนวณค่างวด” “คำนวณ ค่างวด” “คำนวณ ค่า งวด” and “คำนวณ ค่า งวด”. However, Google Trend Index of these words behave differently.



Source: Google Trend Interface, data as of June 2017

Figure 4

Google Trend Index searching for “ข่าวหุ้น”, meaning “stock news”, in different but insignificant punctuation, namely “ข่าวหุ้น” and “ข่าว หุ้น”. In this case, Google Trend Index behave accordingly especially since 2008 where number of total real searches seems to be sufficiently large compared to earlier period.



Source: Google Trend Interface, data as of June 2017

2.4 Fourth criteria: searching series are statistical significant with the reference series

The objective of this criteria is to make sure that, the chosen series statistically correlated to the reference series, such as the traditional survey-based data. Simple linear

¹ For more details, <https://support.google.com/trends/answer/4359582?hl=en>

correlation is applied in this study. The Google Trend Index which is at least 50 percent correlated to the reference series passes this filter. Reason of the threshold 50 percent is simple, the chosen series perform better in prediction than a random guess, like using a coin toss. For example, Google Trend Index searching for “ข่าวหุ้น”, meaning “stock news”, results in 85 percent correlation with private consumer confident index in Thailand.

2.5 Fifth criteria: searching word provides economic meaningfulness.

The objective of this criteria is to test validity of the series, whether it provides meaningful insights related to an interested question. In other words, this criteria leaves room for economic judgment in a complementary to earlier statistical criteria. For example, according to Figure X, the word “stock news” could measure a change in household economic expectation. Private households would generally seek for investment return whenever they perceive a better economic condition or expect future favorable growth (Zatlin, 2016.)

3 Use case

Finding a set of rational words for Google Trend is one of the most challenging tasks. Practically, only weekly and monthly can be plugged into the Google Correlate interface. In this study, these five criteria are applied to create four macroeconomic indicators including (1) purchasing power of private household, (2) consumer confidence of private household, (3) private consumption expenditure of durable products, and (4) number of unemployed persons. These indicators are commonly used by policy maker in monitoring macroeconomic conditions.

3.1 Private household purchasing power

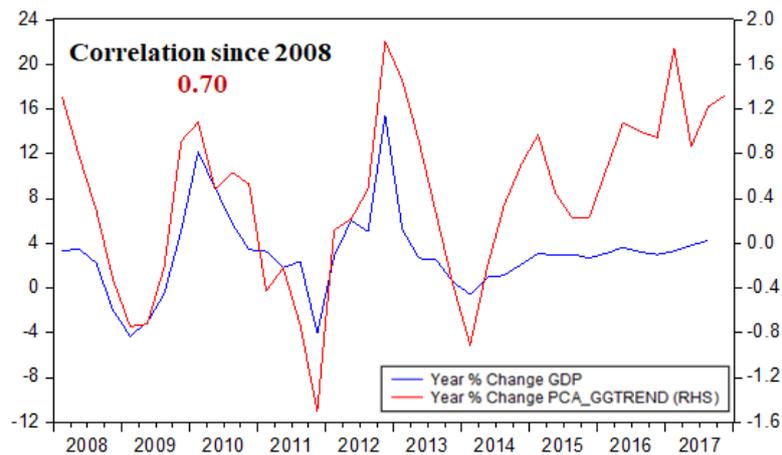
Private household purchasing power is normally correlated with Gross Domestic Product (GDP) since around 60 percent of Thailand GDP is distributed to private household. This is correspondent to the fact that, Private Consumption Expenditure (PCE) always contributes the largest share in GDP in several economies. However, both PCE and GDP are reported quarterly and usually lag for 6 - 7 weeks. In Thailand, earning of employees from Labor Force Survey (LFS) are alternatively used as private household purchasing power. These data are survey-based data collected by the National Statistical Office of Thailand (NSO). The survey is rich in both sample and data features. However, earning of employee covers only a half of total household income. The other half, such as earning of self-employed and small household businesses, are not collected.

Since GDP is available only in quarterly basis, Google Correlate is not able to suggest additional word in this case. In order to figure out potential word lists, this study utilizes basic economic framework: household income should be correlated to consumption expenditure as well as household demand to investment choices. Four words have passed all five criteria, namely, (1) “LTF + RMF” which is the most common tax-deductible and investment product for salaried employee in Thailand, (2) “ภาษี รายได้”, which means “income tax”, (3) “ลงทุน” which means “invest” and (4) “ซื้อ” which means “buy or purchase.”

For monitoring purpose, data dimensionality reduction techniques are required to squeeze only important signals embedded in set of Google Trend Indexes. In fact, there are various tools for dimensionality reduction. This study applies Principal Component Analysis (PCA), extracting only the main common signals from all series. Correlation between annual growth of GDP and the PCA is 70 percent. (Figure 5)

Figure 5

Annual growth of GDP and PCA of Google Trend Index of four chosen words



Source : Office of the National Economic and Social Development Board, Google Trend Index using four chosen words, own calculation.

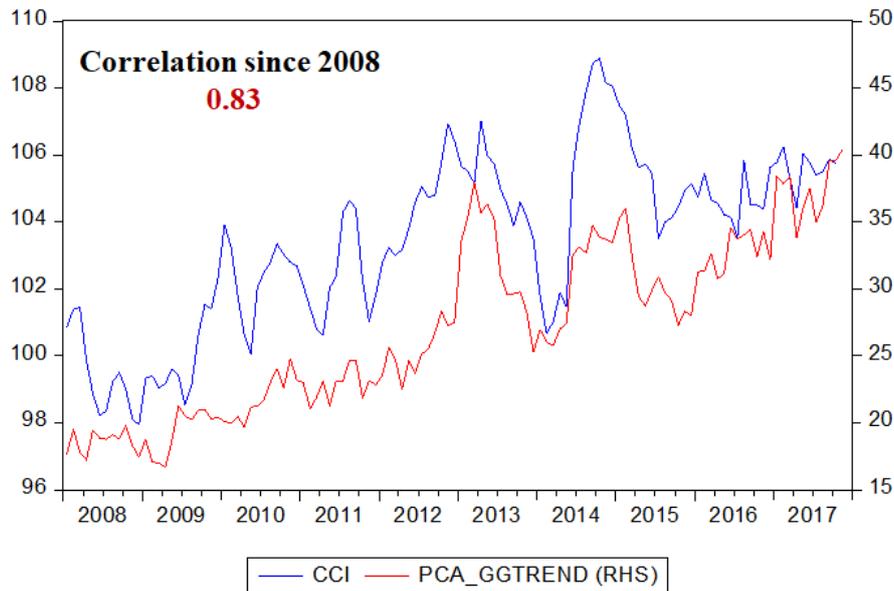
3.2 Consumer confidence of private household

In Thailand, Consumer Confidence Index (CCI) is collected from around 3,500 samples across every provinces. This survey questions private households about their perspective about current and future economic conditions, labor market conditions, as well as their current and expected future income. In addition to CCI, Indicator of Human Well Being (HHI) is surveyed by University of the Thai Chamber of Commerce asking private households about their social circumstances and change in their happiness. Both indicators are monthly reported.

According to Google Correlate, words related to investment news and stock market analysis are suggested. This is correspondent to Zatlin (2016), households' sentiment goes in line with their interest in seeking return from their assets. And nine words have passed all five criteria, namely, (1) “ข่าว หุ้น” which means “stock news”, (2) “กราฟ หุ้น” which means “graph stock”, (3) “กองทุน ปันผล” which means “dividend fund”, (4) “หุ้น ราคา” which means “stock price”, (5) “วิเคราะห์ หุ้น” which means “stock analysis”, (6) “หุ้น กองทุน” which means “stock fund”, (7) “ซื้อขาย หุ้น” which means “stock trade”, (8) “นำ ลงทุน” which means “recommended investment”, (9) “การ เล่น หุ้น” which means “invest in stock market”. After dimensionality reduction, correlation between CCI and the PCA is 83 percent. (Figure 6)

Figure 6

Consumer Confidence Index (CCI) and PCA of Google Trend Index of eight chosen words



Source : Ministry of Commerce, Google Trend Index using eight chosen words, own calculation.

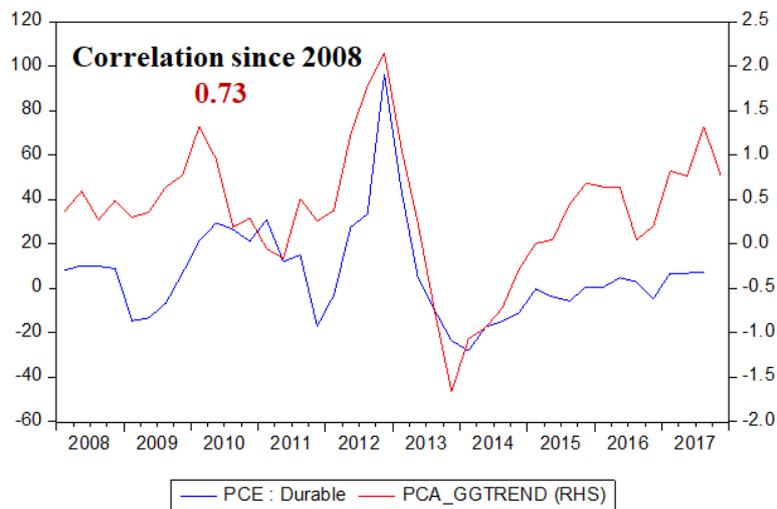
3.3 Private consumption expenditure of durable products

Share of private consumption spending for durable goods is only 10 percent of total PCE. Fast indicator of this product category, mostly vehicles, is useful because it is applicable to indicate cyclical change of macroeconomic condition (Black and Cusbert, 2010) than spending on non-durable goods and services as the latter can be more readily postponed in times of economic slowdown. Also, durable goods are expensive and their payment relies on future income flow, households tend to delay purchases of durable goods during weak business cycle.

Sales of vehicles was monthly reported by the Federation of Thai Industries (FTI) but no meaningful words are able to pass all five criteria. Therefore, this study uses two words including (1) “วัน รับ รถ” which means “the date to take a purchased car” and (2) “ทะเบียนรถ” which means “vehicle license plate”. These two words are chosen because, after the purchase of vehicle, households normally wait for certain periods, around 1-3 weeks, before signing purchasing agreement and getting their vehicles. The process of registering the vehicle license plate is detailed and comprehensive and households take this time to see what they need to inspect when they receive the car. Frequency of searching these two words therefore scopes down real demand for vehicle, and filters out short-term interest of household. After dimensionality reduction, correlation of the Private spending on durable products (PCE: durable) and the PCA is 70 percent.

Figure 7

Annual growth of Private Consumption Expenditure on durable products (PCE : durable) and PCA of Google Trend Index of two chosen words



Source : Office of the National Economic and Social Development Board, Google Trend Index using two chosen words, own calculation.

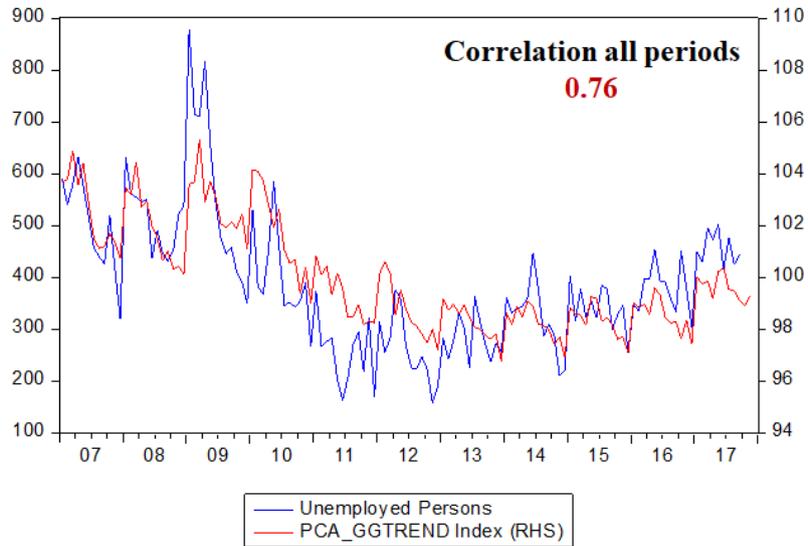
3.4 Number of unemployed persons

Number of unemployed persons is a very common macroeconomic indicators. Lekfuangfu, et al (2017) suggests three potential words that can represent unemployed persons in Thailand including (1) “หางาน” which means “finding a job”, (2) “ประกันสังคม” which means “social security”, and (3) “สมัครงาน” which means “register for a job”. This study finds that additional five words suggested from Google Correlate Interface are also passed the five criteria including (4) “เรียน ต่อ” which means “studying higher degree”, (5) “เขียน resume” which means “write resume”, (6) “resume example”, (7) “ปริญญาโท” which means “master degree” and (8) “ตัวอย่าง resume” which means “resume example”.

These five words reflects many unemployed persons’ behaviors, for example, when they are unemployed, they would search for new jobs, prepare some documents like job resume, review their social security benefit during unemployed periods, or alternatively getting higher education degree. After dimensionality reduction, number of unemployed persons and the PCA is 76 percent (Figure 8).

Figure 8

Unemployed persons and PCA of Google Trend Index of eight chosen words



Source : National Statistical Office of Thailand, Google Trend Index using eight chosen words, own calculation.

4 Limitation

The four new indicators is useful for both prediction and in-time monitoring purposes. However, some limitations are worth mentioned as followed.

4.1 These indicators informs direction, not magnitude of development. This is because the fourth criteria is based on simple correlation value, which simply measures corresponding direction of two series. Other aspects than direction would require additional statistical tools like regression model. Also, it is quite usual that magnitude of Google Trend index is higher than the real behavior suggested from the reference series.

4.2 Searching frequency of the chosen words might be affected by other sources than economic reasons. This could seriously lead to an invalid interpretation. For example, word related to flu, such as “fever” and “cough”, might not genuinely represent flu cases but news instead. In fact, the underlying algorithms may also cause a change in outturn series as well. Continuing quality assessment of word lists are required to confirm that change in searching frequency is not disturbed by other factors than the interested sources.

4.3 Frequency of search is calculated from sample not the population data. Therefore, change in date of using Google Trend interface or researcher’s IP address number would result in different Google Trend Index. However, given that searching word covers sufficiently large number of total searches, the change would simply be a small revision and implications informed by the series would be consequently consistent.

4.4 Process of developing new indicators especially choosing the right word, always takes time and requires specialized insights in certain area. Number of words suggested from the Google Correlate Interfaces are almost a hundred per single input series. And number of candidate

words could possibly reach around a thousand. Although four criteria have already straightforward statistical tools, prudential economic judgment is still necessary to assess validity of economic reasons. Complementary information, such as traditional indicators or knowledge of related topics, is also required to ensure validity of the indicator.

5 Conclusion and Discussion

Google data is proved to be useful in many cases and in many countries. But improving prediction performance alone cannot guarantee validity of conclusion. With different types of prediction model, including an autoregressive model, adding Google Trend would not always improve model performance (Bortoli and Combes, 2016.) Ones may try to harness the richness of Google data, by bundling all words suggested from Google Correlate interface or several words categories in Google Trend interfaces, and let the data speak. This approach creates a serious risk to the policy makers as there is totally no useful insight to support decisions. Strong prediction performance can also arise simply because of an overfitting problem. Instead, the most important task is to extract new and useful insights from new data source as much as possible. Not only that the outturn series could suggest an accurate story, additional knowledge gained from the data can considerably support policy decisions especially in the case of unexpected and unprecedented shocks. This study suggests a guideline for filtering words which are generalized enough for macroeconomic monitoring purpose: total search numbers is large enough and meaning of word tends to reflect overall condition, not a specific or unrelated event. Statistical test and prudential assessment are also suggested to enhance validity and economic meaningfulness of the chosen word.

Reference

Andrew Zatin (2016), Google Data Points to Strong Consumer Sentiment, URL <http://www.moneyballeconomics.com/google-data-points-to-strong-consumer-sentiment/>

Black and Cusbert (2010), Durable Goods and the Business Cycle, Reserve Bank of Australia Bulletin September, pp 11-18, URL <https://www.rba.gov.au/publications/bulletin/2010/sep/pdf/bu-0910-2.pdf>

Greg Tkacz (2013), Predicting Recession in Real-Time: Mining Google Trends and Electronic Payments Data for Clues, C.D. Howe Institute Commentary

Hyunyoung Choi and Hal Varian (2009a), Predicting the Present with Google Trends, Technical report, URL http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf

Hyunyoung Choi and Hal Varian (2009b). Predicting initial claims for unemployment insurance using Google Trends. Technical report, URL <http://research.google.com/archive/papers/initialclaimsUS.pdf>

Hyunyoung Choi and Hal Varian (2011), Predicting the Present with Google Trends, URL <http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf>

Iren Onder (2017), Forecasting Tourism Demand with Google Trends: Accuracy Comparison of Countries Versus Cities, International Journal of Tourism Research

Konstantin A. Kholodilin, Maximilian Podstawski, and Boriss Siliverstovs (2010), Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US, DIW Berlin Discussion Paper No.997

Lynn Wu and Erik Brynjolfsson (2015), The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales, URL <http://www.nber.org/chapters/c12994>

Nick McLaren and Rachana Shanbhogue (2011), Using Internet Search Data as Economic Indicators, DIW Berlin Discussion Paper No.899

Rawley Z. Heimer, Timothy Stehulak, and Daniel Kolliner (2015), Assessing Consumer Confidence with Google Search Terms, Federal Reserve Bank of Cleveland

Simeon Vosen and Torsten Schmidt (2011), Forecasting private consumption: survey-based indicators vs. Google trends, Journal of Forecasting, Volume 30, Issue 6 URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/for.1213>

Stephanie Combes and Clemen Bortoli (2016), Nowcasting with Google Trends, the more is not always the better, First International Conference on Advanced Research Methods and Analytics, URL <http://www.carmaconf.org/carma2016/wp-content/uploads/pdfs/4226.pdf>

Xinyuan Li (2016), Nowcasting with Big Data: is Google useful in Presence of other Information?, URL https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IAAE2016&paper_id=215



Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *“Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data”*

Bali, Indonesia, 23-26 July 2018

Standardised approach in developing economic indicators using internet searching applications¹

Paphatsorn Sawaengsuksant,

Bank of Thailand

¹ This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Macroeconomic Indicators from Google Trend and Google Correlate

Advantages:

.... Almost Real time: monthly, weekly, and daily basis

.... Reflecting Real Activities: comparable to internet-based survey asking topics which citizens are interested in

.... Applicable for several research areas: Household and Business Sentiment, Private Household Purchasing Power, Private Consumption Expenditure, Unemployment, Demand for tourism sectors, Property Price, E-Commerce, Popularity of certain policy tools etc.

.... User friendly and free access



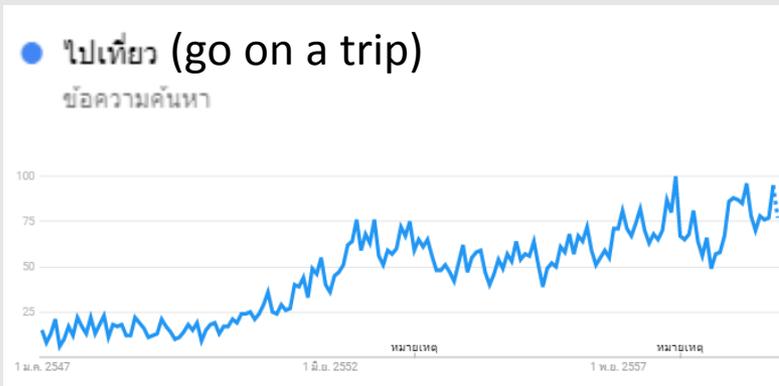
Google Trends and Correlate Interface

Google Trends

Input : Reference words

Output : Time Series

Searching Frequency Index



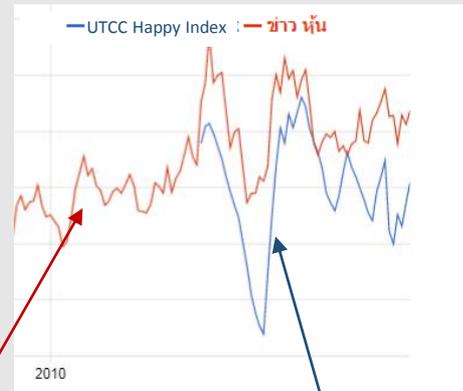
Searching Frequency of “ข่าวหุ้น” or “Stock News”

Google Correlate

Input : Reference Time Series

Output : Correlated Words

List of highly correlated (but mostly spurious) words



Reference Series: Household Happiness Index

Correlated with UTCC Happy Index

0.7817	ข่าวหุ้น (Stock News)
0.7529	www.mybycat.com
0.7512	พร บ ตำรวจ แห่งชาติ (Cop Act)
0.7509	หุ้น nusa
0.7471	กองทุน หุ้น (Stock and Equities)
0.7411	gorilla glass (Screen Protector)
0.7394	5 chord (Guitar Chord)
0.7370	หุ้น หุ้น (Stock News)
0.7367	รูป รถยนต์ (Car Image)
0.7347	ไซโคปัส (Anime Name)
0.7343	หุ้น cash balance

New Data = New Challenges

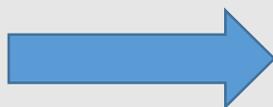
.... Data is unstructured: not initially collect for analysis purpose

.... Garbage In = Garbage Out \Rightarrow **Policy Decision??**

.... Choosing the **right searching word** is very crucial to construct **reliable economic indicators**, especially for policy maker.

(1) Words suggested from the Google Correlate are mostly spurious; not always provide meaningful insights.

Household
Sentiment
Index



(My Dog) (Screen Protector) (Stock News) (Counterfeit)
(Car Image) (Watch Movie) (Guitar Chord) (Cop Act)
(Public Officer Job Title) (My Pet) (Migration) (Anime Name)
(My Cat) (Prove of Citizen) (Music Name) (Alienate worker)



(???)



New Data = New Challenges

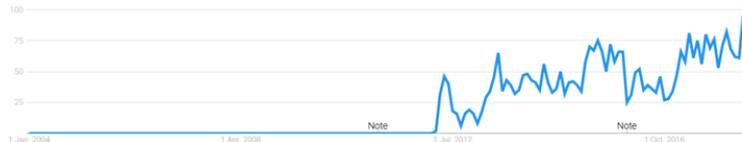
.... Garbage In = Garbage Out \Rightarrow **Policy Decision??**

.... Choosing the **right searching word** is very crucial to construct **reliable economic indicators**, especially for policy maker.

(2) Words inserted in the Google Trend might not be sufficiently general, for example, the word is too new, too specific, or covers too small number of searches.

Figure 1

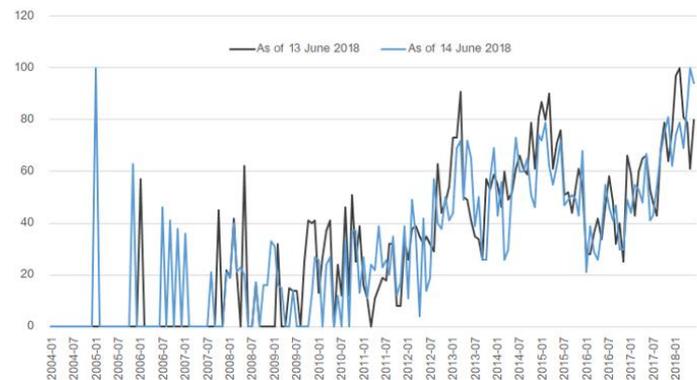
This figure shows searching frequency of the recently popular e-business brand in Thailand. The series is too short to confirm that it can capture developments of e-business in Thailand.



Source: Google Trend Interface, searching “บัตรแม่หมู” in Thailand, data as of June 2017]

Figure 2

Google Trend Index using “ซื้อกองทุนรวม”, meaning “purchase mutual fund”



Source: Google Trend Interface, data as of June 2018

Word Filtering

for Generalization of words (1st-3rd filters)
and Economic Intuition (4th-5th filters).

(1) Not specific to certain brands



(2) Sufficiently popular words

google trend returns non-zero value since 2008 so that (2.1) series are not too sensitive to change in timing of search and (2.2) series' behaviors during all significant cycles can be observed



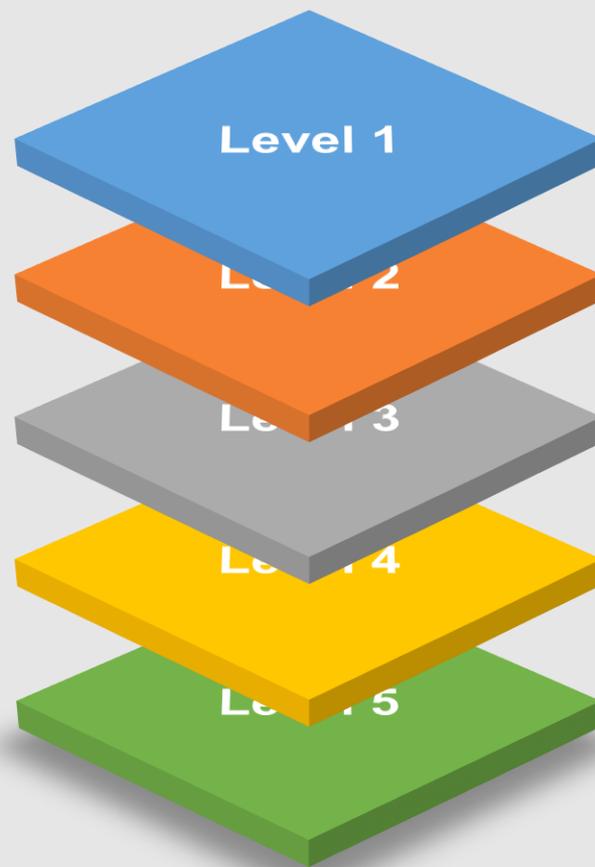
(3) Not sensitive to insignificant word change, like word ordering or innotation



(4) Correlation with Alternative Series > 0.5



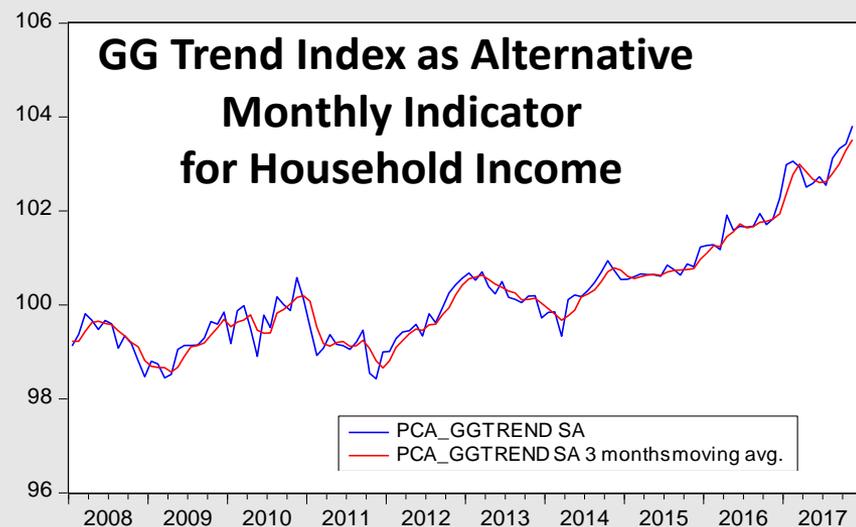
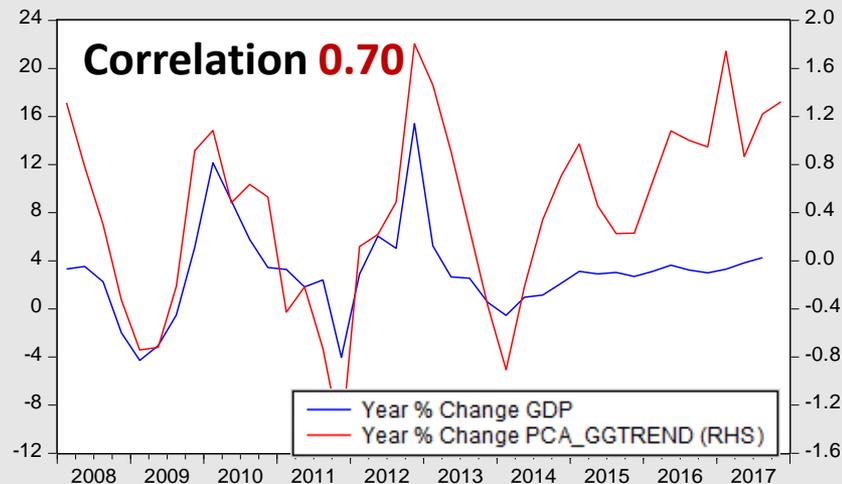
(5) Strongly Intuitive



Application :

I. Household's Income (Quarterly Report)

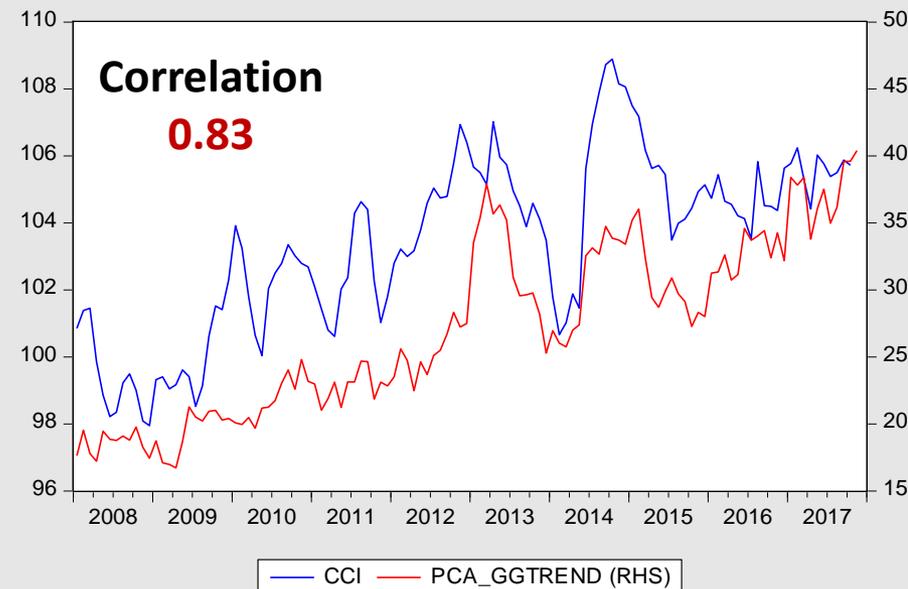
Reference Series	4 Word Lists	Correlation with Real GDP
-	LTF + RMF (Popular Income Tax Reduction Products in Thailand)	0.74
-	ภาษี รายได้ (Income Tax)	0.56
-	ลงทุน (Investment)	0.61
-	ซื้อ (Purchase)	0.50



Application :

II. Household's Sentiment (Monthly Report)

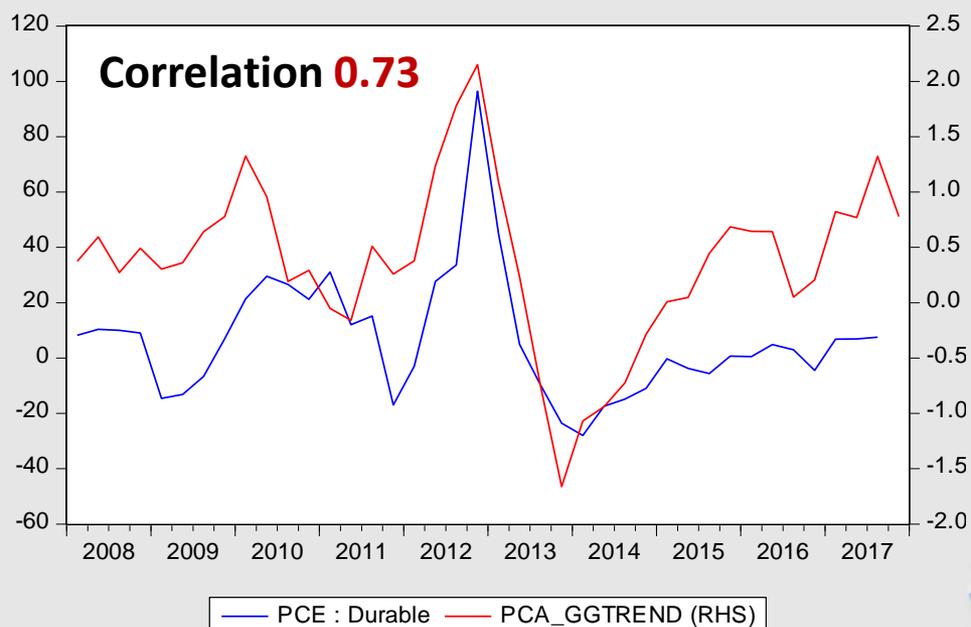
Reference Series	9 Word Lists	Correlation with CCI
Consumer Confident Index: CCI <small>(Source: Ministry of Commerce, Thailand)</small>	ข่าว หุ้น (Stock News)	0.85
	กราฟ หุ้น (Stock Chart)	0.83
	กองทุน ปันผล (Dividend Fund)	0.83
	หุ้น ราคา (Stock Price)	0.82
	วิเคราะห์ หุ้น (Stock Analysis)	0.81
	หุ้น กองทุน (Purchase Fund)	0.76
	ซื้อขาย หุ้น (Stock Purchase)	0.70
	นำ ลงทุน (Profitable Investment Products)	0.68
	การ เล่น หุ้น (Stock Investment)	0.55



Application :

III. Private Consumption Expenditure (PCE) : Durable Goods (Quarterly Report)

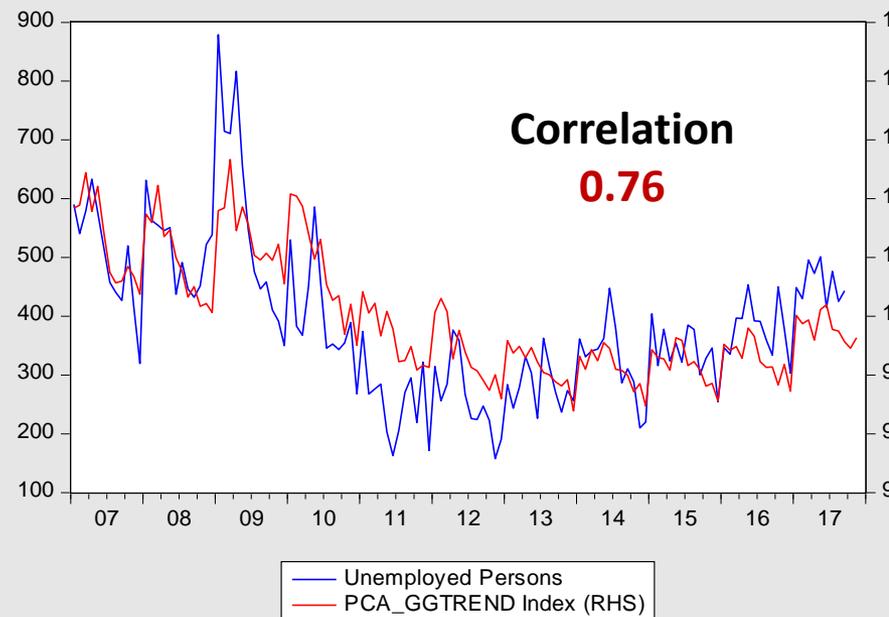
Reference Series	2 Word Lists	Correlation with PCE Durable Goods
PCE: Durable Goods	วัน รับ รถ (the date to get a purchased car)	0.63
-	ทะเบียนรถ (vehicle license plate)	0.66



Application :

IV. Unemployment Rate (Monthly Report)

Reference Series	8 Word Lists	Correlation with Unemployed Persons
Unemployed Persons	หางาน (Finding Job)	0.77
	เรียน ต่อ (Further Study)	0.66
	เขียน resume (Writing Resume)	0.64
	resume example	0.63
	ปริญญาโท (Master Degree)	0.56
	ประกันสังคม (Social Security)	0.56
	ตัวอย่าง resume (Example of Resume)	0.54
	สมัครงาน (Job Application)	0.48



Source : Labor Force Survey, “Labor Market Insights: The Power of Internet-Based Data” Presented at Bank of Thailand Symposium (15 Sep 2016).



Summary :

Application : Early and Alternative Indicators	Alternative and Early Indicators for	Searching Words List	Performance: Corr. between PCA Series and Alternative Series
HH Income	Gross Domestic Products Source: NESDB	LTF + RMF, ภาษี รายได้, ลงทุน, ซื้อ	0.70
HH Sentiment	Consumer Confidence Index Source: UTCC and MOC	ข่าวหุ้น, กราฟหุ้น, กองทุน ปันผล, หุ้น ราคา, วิเคราะห์ หุ้น, กองทุน หุ้น, ซื้อ ขาย หุ้น, นำ ลงทุน, การเล่นหุ้น	0.83
Consumer's Expenditure on Durable Goods	Private Consumption Expenditure : Durable Goods Source: NESDB	วัน รับ รถ, ทะเบียน รถ	0.73
Unemployment	Unemployment Source: Labor Force Survey	หางาน, เรียนต่อ, เขียน resume, resume example, ปริญญาโท, ประกันสังคม, ตัวอย่าง resume, สมัครงาน	0.76

Limitation and Future Development :

Google Trend Information is subjected to changes in households' searching behavior. Instead of consumption/income determinants, change in series might be contaminated by certain changes in searching behavior. Close monitoring of individual series is required.

Google Series perform well as indicators of growth momentum,

not magnitude. Additional methods/filters, like bivariate model, are specifically needed to capture other desirable aspects.

Series Index is subject to timing of search.

Like other data, google trend is drawn from samples, not population. Different timing of search will therefore returns slightly different or "revised" series index.