



Quantifying buying behavior in the personal car market using google trends

Introduction

Cars are major purchases that draw influence and affect various aspects of life. We suggest that **data** sources resulting **from consumer researching about cars from the internet** may offer a new perspective on the **behavior of buyers in the market.**

By **analyzing the changes in Google query volumes for search terms related to car purchasing** from 2004-2018, we find patterns that correlate with actual car sales. Our results illustrate the combining behavioral data sets offer a better understanding about Thai car-buying behavior

Sources of Data

Conventional Data

Personal car sales data from 2004-2018 From Monthly Economic indicator by Bank of Thailand

Alternative Data

Retrieved from Google trends

- ซื้อรถ (buy car)
- ขายรถ (sell car)
- หางาน (find job)

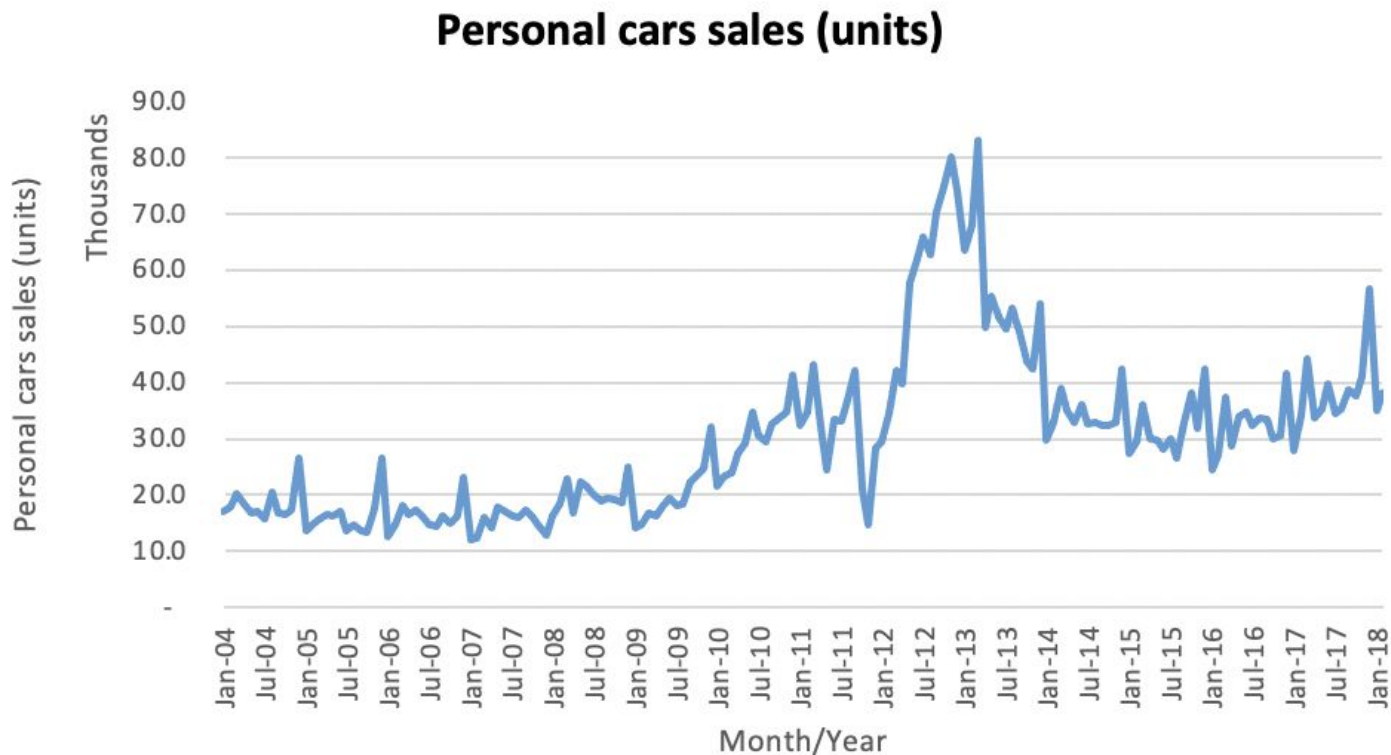
Additional Data

Research about Thai Consumer Behavior on Car Purchasing and related data about the first car policy

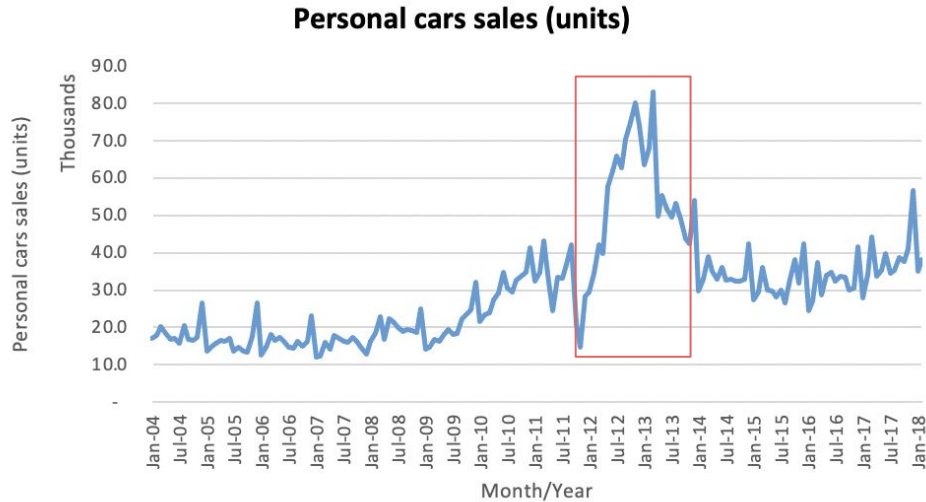


Historical trends of conventional indicators

Personal cars sales (units) as the **conventional indicators** officially produced and distributed by Office of Industrial Economic, Ministry of Industry



The trend of **personal cars sales (units)** dramatically **increased** since the end of 2011 and reached a **peak** in March 2013 expected effect of the **first- car scheme**



- The graph shows information about Personal cars sales (units) in Thailand over a 15-year period from 2004-2018 CE (2547-2561 BE)
- The trend of personal cars sales (units) is steady surge from 2004 to the beginning of 2012.
- The trend of personal cars sales (units) dramatically increased since the end of 2011 and reached a peak in March 2013 expected effect of the first- car scheme
- the trend is decreased after the end of the policy for a while to the leveling off.

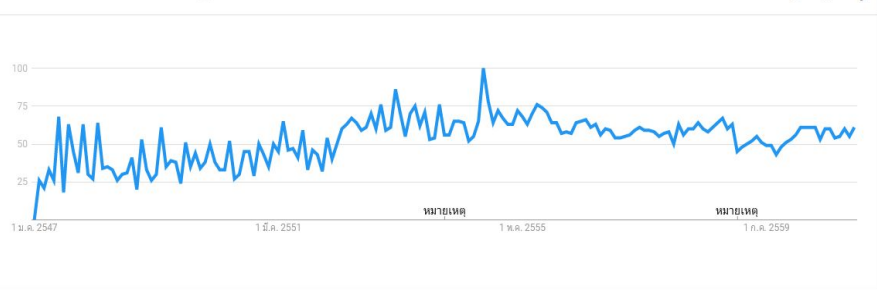
Historical trends of alternative indicators

ซื้อรถ

ข้อความค้นหา

(Buy car)

ระดับความสนใจตลอดเวลาที่ผ่านมา

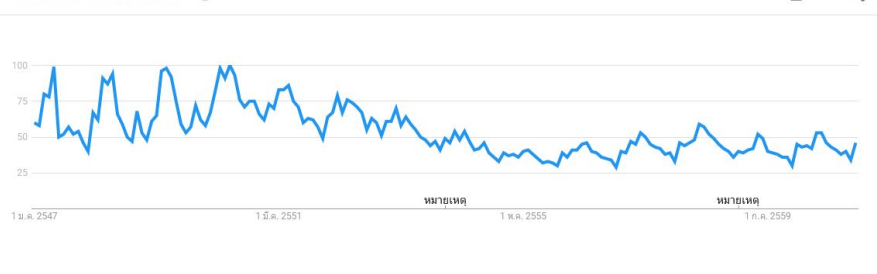


หางาน

ข้อความค้นหา

(Find job)

ระดับความสนใจตลอดเวลาที่ผ่านมา

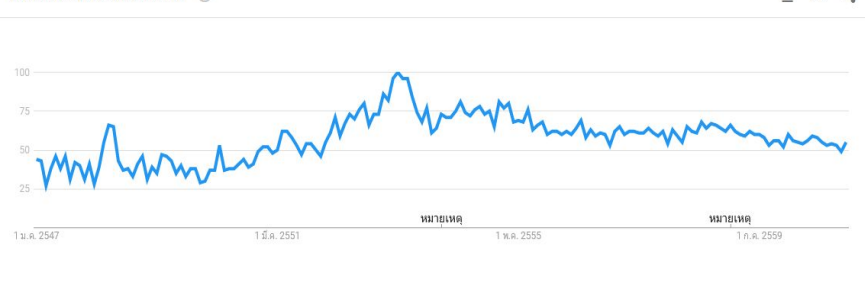


ขายรถ

ข้อความค้นหา

(Sell car)

ระดับความสนใจตลอดเวลาที่ผ่านมา



- The pattern of all line graph are fluctuated over time.
- Buy car and sell car keywords term are positively correlated with the conventional indicator.
- We spotted that line of Buy car will rise up when line of find job is decline.

Methodology

Excel and R studio are used to investigate the relationship between conventional and alternative indicator from google trends



Excel

- Clean data by matching data's year
- Normalized data
- Make visual illustrations



R studio

- Run normal regression and lagging regression to see the relationship between conventional and alternative indicators

Regression result with no lagging show low result of R square

No lagging result

```
Call:
lm(formula = log(frame$`Personal cars sales`) ~ frame$`buy car` +
    frame$`find job` + frame$`sell car`, data = frame)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.08620	-0.16277	-0.02523	0.14385	0.92057

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	10.499048	0.145893	71.964	< 2e-16	***
frame\$`buy car`	0.081160	0.021998	3.689	0.000304	***
frame\$`find job`	-0.015796	0.001481	-10.665	< 2e-16	***
frame\$`sell car`	0.010686	0.007293	1.465	0.144758	

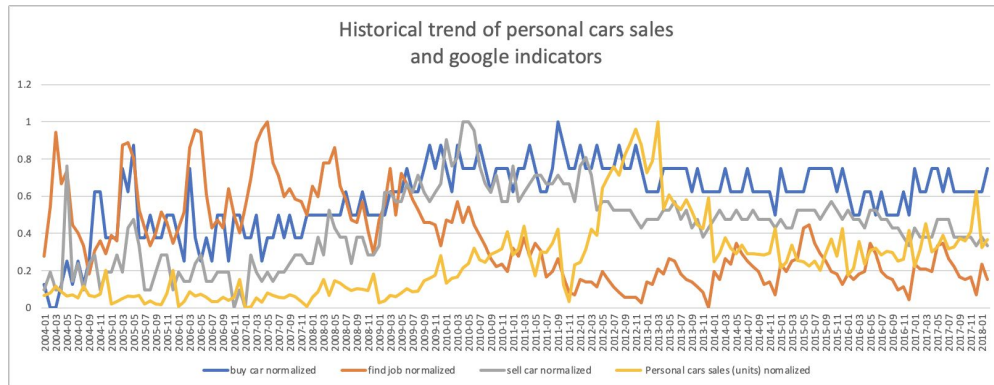
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2904 on 166 degrees of freedom

Multiple R-squared: 0.6092, Adjusted R-squared: 0.6021

F-statistic: 86.26 on 3 and 166 DF, p-value: < 2.2e-16

From the result, there are 2 variables that are statistically significant, and multiple R-squared is around 0.6



Regression result with lagging show higher result of R square

9 months lag result

```
Call:
lm(formula = log(frame$`Personal cars sales`[9:170]) ~ frame$`buy car`[1:162] +
  frame$`find job`[1:162] + frame$`sell car`[1:162], data = frame)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.85932	-0.15639	-0.04001	0.15341	0.76531

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	9.989185	0.134743	74.135	< 2e-16	***
frame\$`buy car`[1:162]	0.076407	0.020094	3.802	0.000204	***
frame\$`find job`[1:162]	-0.012531	0.001363	-9.192	< 2e-16	***
frame\$`sell car`[1:162]	0.034403	0.006703	5.132	8.31e-07	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2629 on 158 degrees of freedom
Multiple R-squared: 0.6826, Adjusted R-squared: 0.6766
F-statistic: 115.3 on 3 and 158 DF, p-value: < 2.2e-16

10 months lag result

```
Call:
lm(formula = log(frame$`Personal cars sales`[10:170]) ~ frame$`buy car`[1:161] +
  frame$`find job`[1:161] + frame$`sell car`[1:161], data = frame)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.9133	-0.1727	-0.0314	0.1573	1.0762

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	9.970677	0.139245	71.605	< 2e-16	***
frame\$`buy car`[1:161]	0.080844	0.020765	3.893	0.000146	***
frame\$`find job`[1:161]	-0.012011	0.001409	-8.525	1.2e-14	***
frame\$`sell car`[1:161]	0.032694	0.006927	4.720	5.2e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2716 on 157 degrees of freedom
Multiple R-squared: 0.6609, Adjusted R-squared: 0.6544
F-statistic: 102 on 3 and 157 DF, p-value: < 2.2e-16

Lagging month of the data to 9 and 10 months and regress, it show a **higher** number of R square. With higher R square means that the regression model fits the observed data. From 9 months lagging, the result reveals that X variable can explain Y variable up to 68.2 percents

Regression result with lagging show higher result of R square

11 months lag result

```
Call:
lm(formula = log(frame$`Personal cars sales`[11:170]) ~ frame$`buy car`[1:160] +
    frame$`find job`[1:160] + frame$`sell car`[1:160], data = frame)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.99822 -0.15376 -0.02204  0.14851  0.75426
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    10.055345   0.132632   75.814 < 2e-16 ***
frame$`buy car`[1:160]  0.042078   0.019840    2.121  0.0355 *
frame$`find job`[1:160] -0.013296   0.001342   -9.908 < 2e-16 ***
frame$`sell car`[1:160]  0.043378   0.006604    6.568 7.18e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2587 on 156 degrees of freedom
Multiple R-squared:  0.692,    Adjusted R-squared:  0.6861
F-statistic: 116.8 on 3 and 156 DF, p-value: < 2.2e-16
```

12 months lag result

```
Call:
lm(formula = log(frame$`Personal cars sales`[12:170]) ~ frame$`buy car`[1:159] +
    frame$`find job`[1:159] + frame$`sell car`[1:159], data = frame)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.9902 -0.1676 -0.0164  0.1506  0.7350
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     9.923822   0.134897   73.566 < 2e-16 ***
frame$`buy car`[1:159]  0.059503   0.020258    2.937  0.00382 **
frame$`find job`[1:159] -0.011883   0.001365   -8.708 4.34e-15 ***
frame$`sell car`[1:159]  0.041872   0.006743    6.210 4.66e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2629 on 155 degrees of freedom
Multiple R-squared:  0.6821,    Adjusted R-squared:  0.6759
F-statistic: 110.8 on 3 and 155 DF, p-value: < 2.2e-16
```

This two model showed the 11 and 12 months lagging result. All of 3 alternative indicators are all statistically significant and it also showed higher multiple R-squared compared to normal model (same period model)

Summary and Policy recommendations

Results

- From our regression analysis without lagging Google trends data cannot represent economic indicator accurately but after we lagged the data, **R square value increases to 68% which mean it is highly correlated with conventional data (personal car sales)**
- Results from our data which collected in 2004-2018 suggest that Thai car buyers have **decision making period around 9-12 months** before purchasing a car.
- This results can be beneficial for car dealers to **predict the behavior of their customers**, this concept can also be applied to other big purchases such as house or real estate.

Summary and Policy recommendations

Policy recommendation

- Result from **lagging model** also **indicate liquidity** in each country as well. For example, the US citizen take 2-3 months to make a decision before purchasing a car but Thai citizen take around 9-12 months to decide.
- Since the result show lagging period of 9-12 months If the government want to **shorten the decision making period** and spike up the car sales volume. They can implement a policy such as “รถเก่าแลกใหม่”
- Car dealers can use the results to conduct a marketing promotion to match with the predicted customer decision making period.

Group members

Lalita Techaphunpisut 6104641268

Prima Donmuang 6104640401

Pattera Rattanatanatevilai 6104640211

Piyachat Kliniam 6104640880