國立政治大學社會科學學院經濟學系

碩士學位論文

Department of Economics

College of Social Sciences

National Chengchi University

Master Thesis

Google 搜尋趨勢與認知金融：從臺灣股票市場學到的新知

Google Trends and Cognitive Finance: Lessons Gained from the Taiwan Stock Market

指導教授：陳樹衡 博士

研究生：沈佩璇 撰

中華民國 107 年 1 月

January, 2018

Department of Economics

College of Social Sciences

National Chengchi University

Master thesis

Google Trends and Cognitive Finance: Lessons Gained from the Taiwan Stock Market

Shen Pei Hsuan

Advisor: Prof. Chen Shu Heng

January 2018

Abstract

Behavioral finance is the study of the influence of psychology on the behaviors of financial practitioners and the subsequent effect on the markets. Although behavioral finance theory has been popular for many years, empirical studies only become possible recently, thanks to the advancement of technology and the availability of data and tools. This research adopts an empirical approach to investigate how investors' attention and interview sentiments influence Taiwan stock market. In particular, we identify the psychological factors that have an impact on Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX).

In addition to TAIEX data, including TAIEX prices and trading volume, two other data sources have been used in this study: (1) Investor Confidence Index interview data provided by J.P. Morgan Asset Management, representing investors' interview sentiments and (2) search volume data from Google Trends, symbolizing investors' attention. We first analyzed weekly data from January 5, 2014 to November 6, 2016, and then ran regression on the data, under the Newey-West correction of standard errors method, to identify the effects of investors' attention and interview sentiments on TAIEX.

We have found many interesting results. First, we discovered the investors in the Taiwan stock market normally use company names, not ticker symbols, to conduct Google search for information related to investment decisions. Second, investors' attention based on the Google Search Volume Index (SVI) searched by company names is significantly and positively correlated with the average returns of TAIEX, which agrees with the attention hypothesis of Barber and Odean (2007). Third, we verified the hypothesis of Barber and Odean (2007) that the positive trend of SVI is an indication of investors' intention of purchasing a stock. Fourth, investors' interview sentiment of Taiwan Stock Price Index is negatively correlated with the average returns of TAIEX, which supports the overconfident hypothesis proposed by De Bondt and Thaler (1995). By contrast, their interview sentiment of Taiwan Economic Situation Index is positively correlated with the average returns of TAIEX. Finally, trading volume is positively related to the average returns of TAIEX, which aligns with that reported in Chuang, Ouyang, and Lo (2010).

Table of Contents

List of Tables

List of Figures

Introduction

1. Research Motivation

According to classical economists, people have sufficient information before making a decision to gain the maximum benefit with the minimum cost. By contrast, behavioural economists adopt psychological viewpoints and believe that people don't always make the best decision. Human behaviours, emphasized by behavioral economists, are influenced by their sentiments and their mental ability. The two rationales to support this belief are *insufficient information* and *bounded mental ability*.

First, in reality, people cannot gather all information before making a decision; hence their behaviour will be under the influence of insufficient information. For example, McCall (1970) claimed that in the labour market, both the information obtained and the uncertainty about the job market will influence a job seeker in accepting an offer. He explained that if a job seeker receives sufficient information, he would turn down a job offer requiring high skills with a low salary. However, if the job seeker receives insufficient information, he tends to accept the job because he worries that he may not find other satisfying jobs if he turns it down.

Second, even though people have enough information about the stock market, Russo (1974) showed that the relationship between the volume of the information and the help that the information provides in decision-making is not totally positively correlated. This is because once the amount of information is beyond the scope of a person's mental ability, they will be frustrated by the information and cannot utilize the information to help them make correct decisions.

In financial stock markets, investors' behaviours are also subject to the influence of insufficient information and bounded mental ability. They can't gather all information they need to analyze. Even though they collection all information, they can't be always rational in making a decision. In other words, they are limited by their psychological states. To understand what psychological factors have an impact on investors' decision making, this research adopts an empirical approach using Taiwan stock market data, including stock prices and trading volume. We also used Investor Confidence Index interview data provided by J.P. Morgan Asset Management as investors' interview sentiments and search volume data from Google Trends as their attention. Under a sequence of experiments, we found that investors' attention and interview sentiments, in addition to trading volume, have effects on the Taiwan stock market.

2. Contributions

This thesis has made the following two contributions:

- We are the first to investigate the impact of Google search volume index (SVI) on the Taiwan stock market using both company names and ticker symbols as the search keywords. We found that they give very different results. The SVIs searched by company names of all ranks are significantly and positively correlated with the average returns of TAIEX. By contrast, the SVIs searched by ticker symbols are not, except that of companies ranked higher than the top-100. This gives strong evidence that the investors in the Taiwan stock market normally use company names, not ticker symbols, to conduct Google search for information related to investment decisions.
- Since Da, Engelberg, and Gao (2009), there have been various works using Google search volume data as investors' attention to study different stock markets. Meanwhile, there are other works using market sentiments to study financial markets. However, we are the first to integrate J.P. Morgan interview data as investors' interview sentiments and the Google search volume data as investors' attention to investigate their impacts on the average returns of TAIEX. Our empirical results support the attention hypothesis of Barber and Odean (2007), the overconfident research of De Bondt and Thaler (1995) and the trading volume study reported in Chuang, Ouyang, and Lo

(2010).

3. Organizations

The remainder of the thesis is organized as follows.

In chapter 2, we present the research of behaviour finance that is related to this thesis. In particular, we summarize the psychological factors that our empirical study investigates. Next, we review empirical studies of behaviour finance related to this work. We also compare their results with ours.

In chapter 3, we explain the research methodology used in this thesis. We first explain the sources of the data and the methods used to process the data. Next, we describe the algorithms used to analyze the data.

In chapter 4, we present the research results and give analysis. In particular, we discuss if our empirical results support the hypotheses proposed in behaviour finance. We also highlight the new discovery about the Google search behaviours of the investors in the Taiwan Stock Market.

In chapter 5, we summarize our research findings, provide application of the study, and discuss the limitations of the work.

Background and Related Works

In behaviour finance, various research works have reported that investors' attention and interview sentiments can affect their decision-making. In Section 1, we discuss about these psychological phenomena associated with investors' behaviours. It explains the variables we have chosen in our empirical study to investigate these phenomena. In Section 2, we review related empirical studies using Google SVI as investors' attention, trading volume as the past stock market performance, and various confidence indexes as investors' interview sentiments to investigate their impacts on different stock markets. Then, we summarize the chapter in Section 3.

1. Investors' Psychology and Behaviours

People are not only influenced by the limitation of their mental ability, but also by their sentimental changes. Miller (1977) stated that even if investors have received the same information about a stock, they might hold different opinions on the stock because they have different feelings about the information. Baker and Wurgler (2006) also pointed out that investors' sentiments have an effect on stock returns.

1.1 Investors' Attention

Barber and Odean (2007) noted that individual investors would buy stocks that attract more attention. They added that a stock could attract more attention through appearing on newspapers/ websites or having abnormal trading volume, or having the highest return in a day. Furthermore, they hypothesized that the phenomenon is due to the fact that it is hard for individual investors to collect information of all stocks, so they only buy the stocks receiving high attention. Barber and Odean (2007) added that when investors sell their stocks, however, this phenomenon doesn't exist, because investors only pay attention to the stocks they own.

1.2 Investors' Overconfidence and Over Optimism

Taylor and Brown (1988) pointed out that people tend to be overconfident and overoptimistic in their decision-making. Later, De Bondt and Thaler (1995) demonstrated that the phenomenon exists in the financial field. This happening can be used to explain why when investors are optimistic about the price of a stock, they will lose money in the end. This is because they are under the illusion of other factors, causing them to make wrong judgements about the price of the stock.

1.3 Herding

Barber, Odean, and Zhu (2009) showed that individual investors are highly influenced by one another on their buying and selling behaviours. When they invest in a stock market, they move together to buy or to sell stocks, which is one of the irrational behaviours. Moreover, Barber et al. (2008) added that when retail investors buy certain stocks in crowd to raise the stock prices, the price of the stocks would fall later. Andrade, Chang, and Seasholes (2008) reported that the herding phenomenon exists in the Taiwan stock market.

1.4 Short-term Price Momentum and Long-term Price Reversal

Daniel, Hirshleifer, and Subrahmanyam (1998) proposed a hypothesis that irrational investors could influence short-term stock prices, which can go above or below their fundamental value. They believed that the phenomenon will last for a long time but eventually the stock prices will reverse back to their fundamental value.

1.5 Stock Trading Volume

A number of empirical studies have documented that there is a positive correlation between stock trading volume and the stock's absolute price changes (see Karpoff, 1987). Lee and Swaminathan (2000) pointed out that the trading volume of a stock could represent the past performance of the stock. They added that if the past performance of a stock is good, the stock would attract more attention and make investors buy the stock more, leading to an increase in trading volume.

2. Related Works

There are other works that also used an empirical approach to study the impact of investors' attention and sentiments on stock markets. First, we focus on works using Google SVI to study different markets, and then compare their results with ours. Next, we discuss about works using trading volume and various market interview sentiment indexes to predict stock performance.

2.1 Google Search Volume Index

Da, Engelberg, and Gao (2009) is the first to conduct research using Google SVI data on the U.S. stock market. They found there is a positive relationship between the SVI and the returns of Russell 3000. They pointed out that the increase of the SVI would increase the turnover of Russell 3000 for two weeks. For the Taiwan stock market, our investigation indicates that the change of SVI can predict the average returns of TAIEX from the first to the fifth week, except the second week. Another study using Google SVI data is on the largest 30 stocks traded in NYSE by Vlastakis and Markellos (2012). They showed that SVI is positively related to the stock trading volume and the stock return volatility of the U.S. stock market. Moreover, they reported that there is a positive link between the information investors received and the risk aversion, based on the expected variance risk premium.

Focused on the German stock market, Bank, Larch, and Peter (2011) also paid attention to risk liquidity and reported that the higher the Google search volume is, the higher the trading volume and the more improved stock liquidity are, leading to higher returns of the German stocks in the short run. They believed that attention-grabbing stocks are the subjects of temporary buying interests, causing price pressure and then leading to higher stock prices.

For the Japan stock market, Takeda and Wakao (2014) focused on the relationship between the intensity of SVI, the returns of Nikkei 225, and their trading volume. They reported that the relationship between the SVI and the trading volume is strong but the relation between the SVI and the Nikkei 225 return is not significant, probably because the sampling period included major negative economic shocks, such as the 2008 world financial crisis and the 2011 Great East Japan Earthquake.

For the France stock market, Aouadi, Arouri, and Teulon (2013) conducted a research on the relationship between investors' attention, based on Google search volume, stock liquidity and volatility, using the CAC 40 index data. The research pointed out that more attention is given to larger sized firms, leading to more liquidity. Although our research focus on the average returns of TAIEX, not on the stock liquidity or volatility, we decide to apply their research results and used the company size as the weight to compute the weighted sum of SVI to represent investors' attention.

For the Taiwan stock market, Fan, Liao, and Chen (2014) showed that using the top-50 company names as keywords for Google Search, the search volume could predict the average returns of TAIEX. In our research, we found that search volume based on company names of all ranks, not just the top-50, have the power to predict the average returns of TAIEX.

There are many studies that used company names or ticker symbols as keywords in Google search to represent investors' attention to study different stock markets. Among them, Fan, Liao, and Chen (2014), Vlastakis and Markellos (2012), Bank, Larch, and Peter (2011), and Takeda and Wakao (2014) used company names as the search keywords. They reported that SVI is significantly correlated with the stock returns in their studied markets. By contrast, Da, Engelberg, and Gao (2011) used ticker symbols as the search keywords. They reported their SVI is also significantly correlated to the Russell 3000 index. For the Taiwan stock market, we find that SVIs based on company names of all ranks have the power to predict the average returns of TAIEX. Additionally, SVIs based on the ticker symbols of companies that are ranked higher than 100 also have a similar predictive power.

2.2 Trading Volume and Market Sentimental Indexes

There are also works focusing on the impact of investors' sentiments on the financial stock markets. For example, Chuang, Ouyang, and Lo (2010) used much theory to support the use of trading volume as an investors' sentiment index and found that trading volume is a suitable proxy as sentiments. So, trading volume reflects investors' expectation of the stock prices; hence influences the average returns of the stocks in the Taiwan stock market. Another work is by Chung and Yeh (2009), who did research on the U.S. stock market using many sentiment indexes, such as consumer confidence level[1], the VXO[2] (the old VIX (Volatility Index)), Baker and Wurgler's orthogonal sentiment index[3] to capture consumers' and investors' sentiments. They reported that sentiments could be used to predict the stock returns in the U.S. stock market. Our research uses J.P. Morgan confidence indexes as investor's interview sentiments and also found that two particular investors' interview sentiments can help predict the average returns of TAIEX.

3. Summary

Based on the established behaviour finance research findings, we have identified several investors' psychological factors that might have an impact on the average returns of the Taiwan stock market and will incorporate them in our empirical study. These factors include investor's interview sentiments, their attention, herding,

1 http://www.sca.isr.umich.edu/
2 http://www.cboe.com/VXO
3 http://people.stern.nyu.edu/jwurgler/

overconfidence and over optimism. The investors' attention and herding are reflected on the Google SVI. The investors' overconfidence and over optimism are reflected on the J.P. Morgan data of Taiwan Stock Price Index. Moreover, we include trading volume in our study to investigate if the past performance of a stock has an impact on its present performance.

We will explain these data and the research methods used in this empirical study in the following chapter.

Research Methods

To investigate how investors' attention and interview sentiments influence the average returns of TAIEX, we selected eight independent variables and one dependent variable from three different data sources. Section 1 describes the data from the Taiwan Economic Journal; Section 2 explains the data from the J.P. Morgan Asset Management, and Section 3 discusses the search volume data from Google Trends. In section 4, we explain the econometric method we choose to run the regression. Section 5 summaries the whole chapter.

1. Taiwan Economic Journal Data

Taiwan Economic Journal (TEJ)[4] is a database that contains historical financial data and information in the major financial markets in Asia. We downloaded weekly average opening price of TAIEX and weekly total trading volume from January 5, 2014 to November 6, 2016 for this research.

2. J. P. Morgan Asset Management Confidence Indexes

We obtained six confidence indexes form J.P. Morgan Asset Management[5]. J.P.

[4] http://www.tej.com.tw/twsite/
[5] https://www.jpmrich.com.tw/wps/portal/!ut/p/b0/04_Sj9CPykssy0xPLMnMz0vMAfGjzOK9AkIDjEJcDQz83XycDIxczIyd3TzcDAycTfULsh0VAXrBmJ0!/?WCM_PORTLET=PC_Z7_JPUP2TE008CAC02

Morgan Asset Management started investigating Taiwan investors' sentiments change in 2004. They interview investors from time to time by asking six questions. The interviews are designed to evaluate investors' confidence in Taiwan economics, politics, and stock market. The results are used to compose six interview sentiment indexes. Moreover, they set the threshold used to evaluate these indexes as 100. If an index is higher than 100, it means that investors are optimistic about the market. The higher the score is, the more optimistic the investors are. If an index is below 100, it means the investors are pessimistic about the market. The lower the score is, the more pessimistic the investors are. The six confidence indexes are defined as follows:

1. Taiwan Stock Price Index: What is the possibility that TAIEX rises in the future?
2. Taiwan Economic Situation Index: What is the possibility that Taiwan economic situation becomes better in the future?
3. Taiwan Political Index: What is the possibility that the political situation between Taiwan and China becomes more stable in the future?
4. Taiwan Investment Environment Index: What is the possibility that the investment situation in Taiwan becomes better in the future?

DL2KSHG28U1000000_WCM&WCM_GLOBAL_CONTEXT=/wps/wcm/connect/jpmrich/eportal/lb_b2c_l01/b2c_l01p220/b2c_l01p220_01/at_b2c_l01p220_01-1_00002

5. Global Economic Index: What is the possibility that the global economic situation becomes better in the future?

6. Possibility of the Value of Your Portfolio Increases within next six months: What is the possibility that the value of your portfolio increases in the future?

We downloaded the confidence indexes data from 2014 to 2016. The data were mostly quarterly. To work with other data in weekly format, we used frequency conversion method by Litterman (1983) to raise the frequency of these interview sentiment indexes to become weekly using the statistical software Eviews.

The method assumes that an interview sentiment index this week will affect the index next week and their residuals are correlated. These functions are described as follows:

$$x_{it} = x_{it-1} + \varepsilon_{it},$$

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + e_{it},$$

where $\varepsilon_{it} \sim N(0, V)$ and $i = 3, 4, 5, 6, 7, 8$.

We used $x_{3t} \sim x_{8t}$ to represent the six interview sentiment variables. The variable x_{1t} was reserved for SVI while the variable x_{2t} was reserved for trading volume. The initial condition is $x_{i0} = 0$. The function is an ARIMA (1,1) model. The last weekly data series was November 6, 2016.

3. Google Trends Data

Google Trends[6] is a public web facility of Google Inc. (original Google Search) that shows how often a particular search-term (keyword) is entered, relative to the total search volume across various regions of the world, and in various languages. The SVI values represent search interest relative to the given region and time. The highest search number during the downloading period is given SVI value of 100. The weekly SVI is calculated by dividing the weekly search volumes with the highest search volume assigned SVI value 100. We can find the details in Google Trends help[7].

We downloaded two sets of SVI data from Google Trends using two sets of search terms. The first set consisted of the company names of the all companies[8] listed in TAIEX and the second set contained the ticker symbols of these companies. The following subsections explain the procedures to obtain these data.

3.1 Company Names

A company listed in TAIEX has a full name. However, in the stock market, investors often call these companies by their abbreviated company names. For example, “台積電” is the abbreviated company name of “台灣積體電路製造股份有限公司”.

---

6 https://trends.google.com.tw/trends/

7 https://support.google.com/trends/answer/4365533?hl=zh-Hant&ref_topic=6248052

8 http://wiki.mbalib.com/zh-tw/%E5%8F%B0%E6%B9%BE%E8%AF%81%E5%88%B8%E4%BA%A4%E6%98%93%E6%89%80%E5%8F%91%E8%A1%8C%E9%87%8F%E5%8A%A0%E6%9D%83%E8%82%A1%E4%BB%B7%E6%8C%87%E6%95%B0

For brevity, we will use company names instead of abbreviated company names in the rest of the thesis. We retrieved the SVI in the Taiwan region using the company name of each company listed in the TAIEX as the search term from January 5, 2014 to November 6, 2016. However, we found some small capital companies do not have any SVI information. Also, some company names are common terms that may be used to conduct Google search by non-investors. For these two kinds of company, we replaced the search results with that obtained using ticker symbols. The total number of companies whose SVI have been replaced under the process was 49.

To sum the search volume data up as a single index, we used a weighted sum approach, where the weight was the company size, represented by their relative percentage of market value on November 18, 2016, downloaded from Taiwan Futures Exchange[9]. This approach is based on the following assumptions:

- Each search volume is independent. Increased attention on one stock will not influence others.
- The higher a company's market value is, the more attention the company receives and hence the higher the search volume.
- The companies that constitute TAIEX remain unchanged.

After we summed up the weighted SVI of all companies, the time series contained

[9] http://www.taifex.com.tw/chinese/9/9_7_1.asp

146 weeks of data.

3.2 Ticker Symbols

A company listed in TAIEX also has a ticker symbol. For example, “2330” is the ticker symbol of “台積電”. We first used ticker symbols to retrieve their SVI from Google Trends. Next, we used the similar procedures described in the previous section to obtain the weighted SVI of all companies. The time series also had 146 weeks of data.

4. Newey-West Correction of Standard Errors

Newey-West correction of standard errors method[10] is a method to estimate the coefficients of a linear regression model applied to time series data. It is used to correct autocorrelation (also called serial correlation) and heteroskedasticity in the error terms in the regression model. We used the statistical software SAS to run the regression.

Below are the variables names and their meaning.

y：average returns of TAIEX

x_1: Google SVI

x_2: trading volume of TAIEX

x_3: Taiwan Stock Price Index

[10] Please refer to the appendix to know the details.

x_4: Taiwan Economic Situation Index

x_5: Taiwan Political Index

x_6: Taiwan Investment Environment Index

x_7: Global Economic Index

x_8: Possibility of the Value of Your Portfolio Increases within six months

We first converted all time series data into log10, and then used the difference between adjacent weeks ($x_{it} - x_{it-1}$) to run the regression. In this way, the interpretation of the regression is easier: 1 percentage change of an independent variable will change a certain percentage of the dependent variable, specified by the coefficient of the independent variable in the regression model. Moreover, we can reduce the scale difference of the variables; hence increase predicting accuracy.

The linear regression model is as follows:

$$\mathrm{y}_t = \beta_0 + \sum_{i=1}^{8} \beta_i x_{it} + \mathrm{e}_t,$$

and where t = 1, 2, ... ,146

We used the regression results to estimate what variables are significant in effecting the average returns of TAIEX and how long the effect lasts. Because the residuals had heteroscedasticity and autocorrelation, the method used HAC (Heteroskedasticity and Autocorrelation Consistent) estimators to correct them. We used Parzen kernel, an HAC estimator, to run the regression. The relationship between Newey-West correction of standard errors method and Parzen kernel is shown in the appendix.

5. Summary

In the chapter, we have explained the data sources and the methods we used to conduct our research. We downloaded the data from TEJ, Google Trends, and J.P. Morgan Asset Management to help us test different behaviour hypothesis. After that, we used the frequency conversion method to process J.P. Morgan Asset Management data and the weighted sum approach to process Google Trends SVI data. Finally, we applied Newey-West correction of standard errors method to run the regression on the processed data. We present our results with analysis in the following chapter.

Research Results and Analysis

This chapter presents the results and provides our analysis. In Section 1, we report the regression results based on the SVI by ticker symbols, various investor interview sentiment indexes and trading volume to identify which of them have influence on the average returns of TAIEX. In Section 2, we show the regression results based on the SVIs searched by company names and by ticker symbols of companies of different ranks. Additionally, we evaluated if the SVI based on the top-50 companies' ticker symbols helps predict the average returns of TAIEX. In Section 3, we analyzed the p-value of the SVI searched by companies with different ranks in TAIEX to obtain deeper insights about investors' behaviour in the Taiwan stock market. In Section 4, we ran regression, using the SVI with an increased trend to test the attention hypothesis of Da, Engelberg, and Gao (2011). In Section 5, we used data in different time lag to evaluate whether SVI can predict TAIEX average returns for a long period of time or not. In section 6, we summarise the chapter.

1. Variables with a Significant Impact on Average Stock Returns

We first ran regression of all 7 variables on the average TAIEX returns. Following Da, Engelberg, and Gao (2009), we used the SVI searched by all companies' ticker symbols to run the regression. The results are given in Table 1.

Table 1

Regression of all 7 variables on the average TAIEX returns

	Parameter Estimates			
Parameter	Estimate	Approx. Std Err	T-value	Approx. Pr $> \|t\|$
Intercept	6.455E-6	0.000884	0.01	0.9942
x_1	0.021949	0.0140	1.57	0.1195
x_2	0.029167	0.00946	3.08	0.0025**
x_3	-0.01201	0.00541	-2.22	0.0282*
x_4	0.005001	0.00329	1.52	0.1306
x_5	0.001493	0.00192	0.78	0.4378
x_6	0.006546	0.00521	1.26	0.2107
x_7	-0.00268	0.00314	-0.85	0.3941
x_8	-0.0013	0.00371	-0.35	0.7267

Note. t = 146. $l(n) = 4$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

y： average returns of TAIEX x_1: Google SVI

x_2: trading volume of TAIEX x_3: Taiwan Stock Price Index

x_4: Taiwan Economic Situation Index x_5: Taiwan Political Index

x_6: Taiwan Investment Environment Index x_7: Global Economic Index

x_8: Possibility of the Value of Your Portfolio Increases within six months

Table 1 indicates that *trading volume* and *investors' interview sentiment of Taiwan Stock Prices* are significantly related to the average TAIEX returns, while the other six variables are not (p > 0.05). Although Taiwan Economic Situation Index isn't shown to have an impact on the average returns, this result is likely to be caused by other confounding interview sentimental variables. To test the hypothesis, we removed other interview sentimental variables $x_5, x_6. x_7,$ and x_8 and then ran another regression. The results are shown in Table 2.

Table 2

Regression of two interview sentiment indexes, the SVI and trading volume on average TAIEX returns

Parameter Estimates				
Parameter	Estimate	Approx. Std Err	T-value	Approx. Pr $> \lvert t \rvert$

Intercept	0.000382	0.000767	0.50	0.6188
x_1	0.024569	0.0144	1.71	0.0895
x_2	0.031812	0.00949	3.35	0.001**
x_3	-0.00613	0.00231	-2.66	0.0088**
x_4	0.002872	0.00136	2.12	0.0361*

Note. t = 146. $l(n) = 4$. $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

y： average returns of TAIEX x_1: Google SVI

x_2: trading volume of TAIEX x_3: Taiwan Stock Price Index

x_4: Taiwan Economic Situation Index

Table 2 shows that investors' interview sentiments of both Stock Price Index (p = 0.0088) and Taiwan Economic Situation Index (p = 0.0361) affect the average returns of TAIEX. Additionally, trading volume (p = 0.001) is significantly related to the TAIEX returns. This supports the research of Lee and Swaminathan (2000) that trading volume is an indication of a stock past performance and can be used to predict the stock future performance. It also agrees with that reported in Chuang, Ouyang, and Lo (2010), that trading volume is positively related to the average returns of TAIEX.

Among the three variables that impact TAIEX, trading volume and the investors' interview sentiment of Taiwan Economic Situation are positively correlated to the average TAIEX returns: 1 % increases in trading volume and Taiwan Economic

Situation will increase 0.031812% and 0.002872% in the average returns of TAIEX. We gauged that investors' opinions about the economic of Taiwan in the future come from their observations of the changes in daily living, which reflect the reality. It therefore has positive correlation with the TAIEX returns.

By contrast, investors' interview sentiment of Taiwan Stock Price is negatively correlated to the average TAIEX returns: 1% increase in Taiwan Stock Price Index will decrease 0.00613 in the average returns. This can be explained by the overconfidence and over-optimism hypothesis of De Bondt and Thaler (1995) in that while investors' confidence about Taiwan Stock Price is positive, the actual TAIEX stocks returns are not so good as they believe.

2. Ticker Symbols vs. Company Names: SVIs for Investors Attention

Table 1 shows that the SVI based on the ticker symbols of all companies is not correlated to the average TAIEX returns. Since we used SVI to represent investors' attention and companies with bigger market capital normally receive more attention, the SVI of companies with higher ranks in TAIEX might better reflect investors' attention. To verify this hypothesis, we used a top-down approach by separating the SVI of all companies' ticker symbols into various ranks (top-800, top-700, top-600, top-500, top-400, top-300, top-200, top-100). After that, we ran regressions using each of the

SVI series. The results show that only the SVI by ticker symbols of the top-100 companies have an impact on the average TAIEX returns while all others have not. Table 3 presents the results of SVI based on ticker symbols of the top-100 companies.

Table 3

Regression of two interview sentiment indexes, trading volume and the SVI based on ticker symbols of the top-100 companies on average TAIEX returns

Parameter Estimates				
Parameter	Estimate	Approx. Std Err	T-value	Approx. Pr ＞|t|
Intercept	0.000336	0.000756	0.44	0.6576
x_1	0.016388	0.00736	2.23	0.0276*
x_2	0.030368	0.00915	3.32	0.0012**
x_3	-0.00563	0.00228	-2.47	0.0147*
x_4	0.002768	0.00135	2.05	0.0422*

Note. t = 146. $l(n) = 4$. *p < 0.05. **p < 0.01. ***p < 0.001.

y： average returns of TAIEX x_1: Google SVI

x_2: trading volume of TAIEX x_3: Taiwan Stock Price Index

x_4: Taiwan Economic Situation Index

The result triggered our curiosity about whether the same pattern also exists in the SVI by company names for different ranks. To answer that question, we also separated the SVI of all company names into various ranks (top-800, top-700, top-600, top-500, top-400, top-300, top-200, top-100). We then ran regressions using each of the SVI series.

To our surprise, we found that all SVIs are significantly and positively related to the TAIEX average returns. Table 4 shows the regression result of the SVI searched by the top-100 company names. Regression of other SVI series has similar results. This result supports the attention hypothesis of Barber and Odean (2007) that individual investors would buy stocks that attract more attention where attention is measured by the Google search volume in our study. We will provide more in depth analysis of the p-value of SVIs under these regressions in the next Section.

Table 4

Regression of two interview sentiment indexes, trading volume and the SVI based on company names of the top-100 companies on average TAIEX returns

Parameter Estimates				
Parameter	Estimate	Approx. Std Err	T-value	Approx. Pr ＞|t|

Intercept	0.000275	0.000760	0.36	0.7183
x_1	0.038288	0.0132	2.89	0.0044**
x_2	0.02717	0.00811	3.35	0.0010**
x_3	-0.00606	0.00230	-2.63	0.0096**
x_4	0.003069	0.00138	2.22	0.0282*

Note. t = 146. $l(n) = 4$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

y： average returns of TAIEX x_1: Google SVI

x_2: trading volume of TAIEX x_3: Taiwan Stock Price Index

x_4: Taiwan Economic Situation Index

In Fan, Liao, and Chen (2014), the authors reported that the SVI based on the top-50 company names is significantly related to the average returns of TAIEX. To compare our results with theirs, we made two more SVIs based on the top-50 companies. One used ticker symbols and the other used company names. We then ran two regressions using each of the series. The results are given in Tables 5 & 6.

Table 5

Regression of two interview sentiment indexes, trading volume and the SVI of ticker symbols for the top-50 companies on average TAIEX returns

Parameter Estimates				
Parameter	Estimate	Approx. Std Err	T-value	Approx. Pr > |t|
Intercept	0.000311	0.000756	0.41	0.6810
x_1	0.016982	0.00693	2.45	0.0155*
x_2	0.029725	0.00923	3.22	0.0016**
x_3	-0.00566	0.00225	-2.52	0.0130*
x_4	0.002779	0.00133	2.08	0.0391*

Note. t = 146. $l(n) = 4$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

y： average returns of TAIEX x_1: Google SVI

x_2: trading volume of TAIEX x_3: Taiwan Stock Price Index

x_4: Taiwan Economic Situation Index

Table 6

Regression of two interview-sentiment indexes, trading volume and the SVI based on top-50 company names on average TAIEX returns

Parameter Estimates				
Parameter	Estimate	Approx. Std Err	T-value	Approx. Pr > |t|
Intercept	0.000232	0.000764	0.30	0.7618
x_1	0.033182	0.0109	3.05	0.0028**

x_2	0.028716	0.00868	3.31	0.0012**
x_3	-0.00597	0.00231	-2.59	0.0107*
x_4	0.003052	0.00139	2.20	0.0292*

Note. t = 146. $l(n) = 4$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

y： average returns of TAIEX x_1: Google SVI

x_2: trading volume of TAIEX x_3: Taiwan Stock Price Index

x_4: Taiwan Economic Situation Index

As the above two tables show, both SVIs searched by the top-50 company names and by the top-50 ticker symbols are significantly related to the average TAIEX returns, which agrees with that reported in Fan, Liao, and Chen (2014).

3. Analysis of Investors' Behaviours

In the previous section, we have presented the regression results using the SVIs of company names and ticker symbols of various ranks. To better understand the relationship between these SVIs and investors' attention, we plotted the p-value of SVIs for company names and ticker symbols under various ranks on Figure 1. The lower the p-value is, the more significant the SVI is in impacting the TAIEX average returns.

Figure 1

p-value of the SVI searched by companies with different ranks in TAIEX

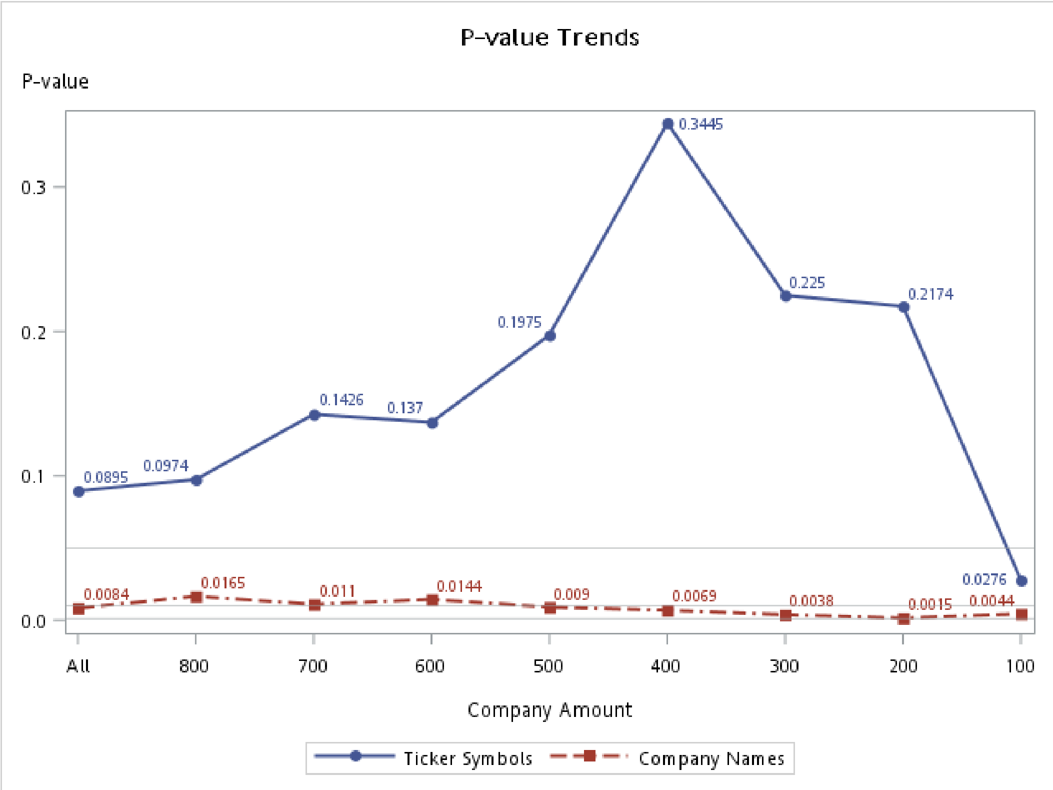

Figure 1 shows that all SVIs by ticker symbols have p-value indicating they are not significant in impacting TAIEX average returns, except the one from the top-100 companies. It suggests that investors in the Taiwan stock market normally don't use ticker symbols to conduct Google search for stock information. This makes sense, as the ticker symbols of Taiwanese stocks are 4-digit numerical values, which might be ambiguous and can be confused as product item number, specific year, phone extension and others by the Google search engine. The only exception is the top-100 ranked companies, which are traded more often; hence their ticker symbols are easily associated with the company name by the Google search engine. Moreover, investors tend to pay attention to the companies that have more capital, and likely to remember their ticker symbols. These explain why search volume using ticker symbols for the top-100 company is significantly related to TAIEX average returns.

By contrast, SVIs by company names of all ranks have p-values that indicate they are significantly related to the TAIEX average returns. This indicates that investors mostly use company names, not ticker symbols, to search for stock information to invest in the Taiwan stock market. This discovery endorses Google's Chief Economist Hal Varian's claim that Google search data can provide insights into people's interests, intentions and future actions (see Varian 2011).

4. Verification of Investors Buying Attention

According to Barber and Odean (2007), investors pay attention to stock information when planning to purchase stocks. However, when planning to sell stocks, investors only pay attention to the stocks they own. Using Google search volume as the proxy for investors' attention, this means that an increased search volume is a sign of buying intention, which leads to price pressure and a possible price increase. By contrast, selling intention has no impact on the search volume. Hence, a decreased trend in search volume of a certain stock is not directly connected to the decreased stock returns.

To test their hypothesis, we selected SVI indexes that have an increased trend, i.e. $I_t > I_{t-1}$, from the SVI with all company names, which consists of 67 weeks of data. The other 79 weeks of data become the SVI with a decreased trend. The reason why we used the data from the SVI with all company names is because we believe the SVI with all company names represents investors' attention in the Taiwan stock market the best. Here, we expect the SVI with an increased trend will have an even more significant impact on the average TAIEX returns than the original SVI while the SVI with a decreased trend will have no impact on the TAIEX returns.

We first present the regression results from the SVI with all company names as the base line in Table 7. Then, we show the results from two regression results, one on the SVI with an increased trend and the other on the SVI with a decreased trend in Tables 8 and 9. We included trading volume and two interview sentimental variables when running the two regressions.

Table 7

Regression of two interview sentiment indexes, trading volume and SVI based on all *company names on average TAIEX returns*

Parameter Estimates				
Parameter	Estimate	Approx. Std Err	T-value	Approx. Pr $> \|t\|$
Intercept	0.000355	0.000748	0.47	0.6360
x_1	0.042557	0.0159	2.68	0.0084**
x_2	0.027507	0.00821	3.35	0.0010**
x_3	-0.00576	0.00226	-2.55	0.0120*
x_4	0.002885	0.00135	2.14	0.0345*

Note. $t = 146$. $l(n) = 4$. $*p < 0.05$. $**p < 0.01$. $***p < 0.001$.

y： average returns of TAIEX x_1: Google SVI

x_2: trading volume of TAIEX x_3: Taiwan Stock Price Index

x_4: *Taiwan Economic Situation Index*

Table 8

Regression of an increased trend SVI with all company names, trading volume and interview sentimental variables on average TAIEX returns

Parameter Estimates				
Parameter	Estimate	Approx. Std Err	T-value	Approx. Pr $> \|t\|$
Intercept	-0.00046	0.00142	-0.32	0.7469

x_1	0.077757	0.0277	2.80	0.0068**
x_2	0.020013	0.0104	1.93	0.0584
x_3	-0.00877	0.00490	-1.79	0.0783
x_4	0.005084	0.00331	1.54	0.1298

Note. t = 67. $l(n) = 1$. *p < 0.05. **p < 0.01. ***p < 0.001.

y： average returns of TAIEX　x_1: Google SVI

x_2: trading volume of TAIEX　x_3: Taiwan Stock Price Index

x_4: Taiwan Economic Situation Index

Table 9

Regression of a decreased trend in SVI with all company names, trading volume and interview sentimental variables on average TAIEX returns

Parameter Estimates				
Parameter	Estimate	Approx. Std Err	T-value	Approx. Pr $> \|t\|$
Intercept	-0.0275	0.0380	-0.72	0.4718
x_1	-0.07284	0.2297	-0.32	0.7521
x_2	0.649021	0.8200	0.79	0.4312
x_3	-0.04488	0.0460	-0.98	0.3325
x_4	0.028232	0.0309	0.91	0.3633

Note. t = 79. $l(n) = 1$. *p < 0.05. **p < 0.01. ***p < 0.001.

y： average returns of TAIEX x_1: Google SVI

x_2: trading volume of TAIEX x_3: Taiwan Stock Price Index

x_4: Taiwan Economic Situation Index

In Table 8, we see that the SVI series with an increased trend is more significant in impacting the average TAIEX returns (smaller p-value) than that of the original SVI (see Table 7). Moreover, the SVI series with a decreased trend has no impact on the average returns of TAIEX (see Table 9). This supports the attention hypothesis of Barber and Odean (2007) that increased attention is connected to the increased stock prices while the decreased attention is not connected to the decreased stock prices.

5. Short-Term Average Returns Prediction.

In Da et al. (2009), the authors reported that SVI has the predictive power for Russell 3000 over two-week time. To compare our results on TAIEX with theirs, We used different lag period (in weeks) to run regression for 5 weeks. For example,

$$\text{if lag} = 1,\ y_t = \beta_0 + \sum_{i=1}^{8} \beta_i x_{it} + e_t,$$

and where t = 1, 2, ... ,146,

$$\text{if lag} = 2,\ y_t = \beta_0 + \sum_{i=1}^{n=8} \beta_i x_{it-1} + e_{t-1},$$

and where t = 2, 3, ... ,146,

and so on. The results are given in Table 10.

Table 10

Regression of two sentiment indexes, trading volume and SVIs with all company names with different time lags on TAIEX average returns

Lag period \ Parameter	Intercept	x_1	x_2	x_3	x_4
1	0.000355	0.042557	0.027507	-0.00576	0.002885
p-value	0.636	0.0084**	0.001***	0.012*	0.0345*
2	0.000466	-0.02834	-0.0355	-0.00618	0.00318
p-value	0.5793	0.1742	0.0032**	0.0204*	0.0425*
3	0.000838	0.053154	0.026307	-0.00239	0.001552
p-value	0.2246	0.003**	0.0069*	0.3188	0.3161
4	0.000698	-0.02902	-0.02396	-0.00351	0.002232
p-value	0.3465	0.0229*	0.0254*	0.1369	0.1331
5	0.000784	0.038865	0.01987	-0.00203	0.001666
p-value	0.2875	0.0115*	0.0327*	0.4556	0.3596

Note. $l(n) = 4$. $*p < 0.05$. $**p < 0.01$. $***p < 0.001$.

y： average returns of TAIEX x_1: Google SVI

x_2: trading volume of TAIEX x_3: Taiwan Stock Price Index

x_4: Taiwan Economic Situation Index

Table 10 shows that only trading volume can predict the average returns of TAIEX for 5 weeks while investors' sentiments on Taiwan Stock Index and Taiwan Economic Situation can be used to predict the returns for two weeks. Although Google SVI has predictive power for TAIEX returns on week 1, 3, 4 and 5, its p-value for the second week is not significant. This result is different from that of Da et al. (2009) on Russell 3000, which indicated that the SVI has continuously predicative power over two weeks.

6. Summary

Our empirical study found that trading volume, Google SVIs, and investors' interview sentiments of Taiwan Stock Price Index and Taiwan Economic Situation Index are significant in predicting the TAIEX average returns. These results support behavior finance hypothesis that investors' psychological states can affect financial stock markets. In particular, their attention, and overconfidence are significantly related to the TAIEX average returns.

We also discover investors in the Taiwan Stock Market normally use company names, not ticker symbols, to conduct Google search for information related to investment decisions. This finding endorses Google's Chief Economist Hal Varian's claim that Google search data can provide insights into people's interests, intentions and future actions (see Varian 2011).

In the next chapter, we will conclude the thesis and provide suggestions for future research.

Conclusions and Outlooks

1. Review of Research Findings

Our empirical study of how investors' attention and interview sentiments influence Taiwan stock market has produced many interesting results. First of all, we discover the investors in Taiwan stock market mostly use company names, not ticker symbols, to conduct Google search for information related to investment decisions. This finding endorses Google's Chief Economist Hal Varian's claim that Google search data can provide insights into people's interests, intentions and future actions (see Varian 2011).

Next, we found that only two of the six inventors' interview sentiments provided by the J.P. Morgan confidence indexes are significant in influencing Taiwan stock market. The first one is investors' sentiment about Taiwan Stock Price, which is shown to be negatively correlated to the average TAIEX returns. This result supports the overconfidence hypothesis of De Bondt and Thaler (1995) in that while investors' confidence about Taiwan Stock Price Index is positive, the actual TAIEX stocks returns are not so good as they expected. The second investors' sentiment that has an impact on TAIEX is their opinions about Taiwan Economic Situation, which is shown to be positively correlated with the average TAIEX returns. When investors are optimistic about the economic situation of Taiwan, this view is likely to come from their daily life in reality, not an illusion projected by other resources, like media, hence is a reliable indication of the performance of TAIEX.

Third, investors' attention measured by Google search volume of all company names listed in TAIEX is significantly and positively related to the TAIEX average returns. This supports the attention hypothesis of Barber and Odean (2007). In terms of how long this impact lasts, our analysis shows that the correlations between the SVI and the TAIEX average returns are significant from the first to the fifth week except the second week. Moreover, we verified the hypothesis of Barber and Odean (2007) that an increased attention, measured by the increased SVI, is an indication of investors' intention to purchase a stock.

These empirical results support the research reported in behaviour finance that psychological states of investors have an impact on stock markets. In particular, their attention, and overconfidence are significantly related to the TAIEX average returns.

Trading volume is also found to be significantly related to the TAIEX average returns. This supports the research of Lee and Swaminathan (2000) that trading volume is an indication of a stock past performance and can be used to predict the stock's future performance. The result also aligns with that reported in Chuang, Ouyang and Lo (2010) about Taiwan stock market.

2. Applications of the Study

This research has identified 4 indicators (Google SVIs, J.P. Morgan Asset Management interview sentiments about Taiwan Stock Prices and Economic Situation, and TAIEX trading volume) that have an impact on the average returns of TAIEX. Since these indicators are public information that can be accessed freely, they can be incorporated into any financial systems to forecast the Taiwan Stock Market, for trading and other business applications.

3. Limitations of the Study

There are three limitations of the study. First, investors' sentiments are based on interviews from a selected sample group, which might not represent the opinions of all investors in Taiwan. There are better sources to obtain investors sentiments, such as social media, bloggers, news, whose rich text data reflect investors' opinion timely; hence allow us to capture the influence of investor's sentiments more precisely. Text mining for sentimental analysis is an active research area in machine learning. This research can be improved by incorporating sentimental analysis of text data from social media and other investment related websites.

Second, the six confidence indexes from J.P. Morgan Asset Management are quarterly data. To work with the weekly data from Google Trends and from the Taiwan Economic Journal, we have transformed the indexes into weekly data, which might not reflect investor's sentiments precisely. As shown in this research, some of the indexes are not significantly related to the TAIEX average returns. This result can be improved when higher frequency data become available.

Last, it is not clear to us why SVI has predictive power for the TAIEX average returns of the first, third, fourth and fifth week but not the second week. There might be other factors of SVIs we need to consider when using it to represent investors' attention.

4. Recommendations for Future Research

We outline three areas of research for future works:

- The incorporation of sentimental analysis of text data from social media, such as Facebook, Twitter, and other investment related websites to represent investors' sentiments. Not only are these data easier to obtain through web crawler, they also provide more precise and timely information about investors' sentiments. Consequently, more accurate results might be obtained with less time delay.

- The investigation of investors' Google search behaviors in other financial markets. While we have identified investors in the Taiwan Stock Market mostly use company names, not ticker symbols, to conduct Google search for stocks related information, it is not clear if this behavior is unique to the Taiwan Stock Market. This is an interesting research question to be explored in the future.
- The exploration of investors' buying and selling intentions revealed from their Google search behaviours. In particular, we will investigate expanded Google search keywords, in addition to company names or ticker symbols, and will use other tools such as Google Correlate to study investors behaviours in the Taiwan Stock Market (see Stocking & Matsa, 2017).

Appendix

1. Newey-West Correction of Standard Errors Method Explained

1. Below is the proof of kernel HAC from Kuan (2008).

"Consider the linear specification $y_i = x_i'\beta + e_i$, and the OLS estimator $\hat{\beta}_T(k \times 1)$. We shall review some basic asymptotic theory for $\hat{\beta}_T$ and the Wald test of regression parameters. In what follows, we let $[c]$ denote the integer part of c, $\overset{\mathbb{P}}{\rightarrow}$ convergence in probability, $\Rightarrow$ week convergence (of associated probability measures), $\overset{D}{\rightarrow}$ convergence in distribution $\overset{d}{=}$ equality in distribution, W_k a vector of k independent, standard Wiener processes and B_k the Brownian bridge obtained from W_k such that $B_k(r) = W_k(r) - rW_k(1)$, 0≤ r ≤ 1. When k = 1, we simply write W_1 as W and B_1 as B.

We impose the following "high level" conditions on data.

[A1] For some β_0, $\varepsilon_t = y_t - x_t'\beta_0$ such that $\mathbb{E}(x_t\varepsilon_t) = 0$ and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{|T_r|} x_t\varepsilon_t \Rightarrow S_0 W_k(r),\ r \in [0,1],$$

where S_0 is the nonsingular, matrix square root of

$$\Sigma_0 = \lim_{T\rightarrow\infty} var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t\varepsilon_t)\ ,$$

i.e., $\Sigma_0 = S_0 S_0'$.

[A2] $M_T = T^{-1}\sum_{t=1}^{|T_r|} x_t x_t' \overset{\mathbb{P}}{\rightarrow} M_0$ uniformly in $r \in (0,1]$ such that M_0 is nonsingular.

By [A1],

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t \varepsilon_t \xrightarrow{D} S_0 W_k(1) \stackrel{\text{def}}{=} \mathcal{N}(0, \Sigma_0).$$

By [A2], $M_T = T^{-1}\sum_{t=1}^{T} x_t x_t' \xrightarrow{\mathbb{P}} M_0$, and hence

$$\sqrt{T}(\hat{\beta}_T - \beta_0) = M_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t(y_t - x_t'\beta_0) \xrightarrow{D} M_0^{-1}S_0 W_k(1)$$

$$\xrightarrow{D} M_0^{-1}S_0 W_k(1) \tag{1}$$

$$\stackrel{d}{=} \mathcal{N}\left(0, M_0^{-1}\Sigma_0 M_0^{-1}\right).$$

This is the well-known asymptotic normality result for the OLS estimator.

With the results in (1), the limiting distributions of the well-known large-sample tests are easily obtained. Consider the null hypothesis H_0: $R\beta_0 = r$ with R a $q \times k$ matrix with full row rank. Under the null hypothesis, (1) implies

$$\sqrt{T}(R\hat{\beta}_T - r) \xrightarrow{D} \mathcal{N}(0, R\mathrm{M}_0^{-1}\Sigma_0\mathrm{M}_0^{-1}R'). \tag{2}$$

Replacing M_0 and Σ_0 with their respective consistent estimators M_T and $\hat{\Sigma}_T$, we have

$$(R\mathrm{M}_\mathrm{T}^{-1}\hat{\Sigma}_T\mathrm{M}_\mathrm{T}^{-1}R)^{-1/2}\sqrt{T}R(\hat{\beta}_T - \beta_0) \xrightarrow{D} \mathcal{N}(0, I_q).$$

It follows that the Wald test of this hypothesis is

$$W_T = T(R\hat{\beta}_T - r)'(R\mathrm{M}_\mathrm{T}^{-1}\hat{\Sigma}_T\mathrm{M}_\mathrm{T}^{-1}R')^{-1}(R\hat{\beta}_T - r) \xrightarrow{D} \mathcal{X}^2(q). \tag{3}$$

Note that W_T would not have a limiting $\mathcal{X}^2$ distribution if $\hat{\Sigma}_T$ is not a consistent estimator for Σ_0. For other large-sample tests, such as the LM test and Hausman test, it is also crucial to have a consistent estimator of the asymptotic variance-covariance matrix.

2. HAC Estimators

For consistent estimation of Σ_0, first note that, by definition,

$$\Sigma_0 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}(x_t \varepsilon_t \varepsilon_s x_t') .$$

For notation simplicity, we write

$$\Sigma_0 = \lim_{T\to\infty} \Sigma_{\mathrm{T}} = \lim_{T\to\infty} \sum_{j=-T+1}^{T-1} \Gamma_T(\mathrm{j}), \tag{4}$$

with

$$\Gamma_T(j) \begin{cases} \frac{1}{T} \sum_{t=j+1}^{T} x_t \hat{e}_t \, \hat{e}_{t-j} x_{t-j}', \ \mathrm{j} = 0,1,2, \ldots \ldots \\ \frac{1}{T} \sum_{t=-j+1}^{T} x_{t+j} \hat{e}_{t+j} \, \hat{e}_t x_t', \ \mathrm{j} = -1, -2, \ldots \ldots \end{cases}$$

When $x_t \varepsilon_t$ are covariance stationary, $\Gamma_T(j) = \Gamma(j) = \mathbb{E}\big(x_t \varepsilon_t \varepsilon_{t-j} x_{t-j}\big)$, and the spectral density of $x_t \varepsilon_t$ at frequency ω is

$$\mathrm{f}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma(\mathrm{j})\, e^{-iw},$$

where $i = \sqrt{-1}$. In this case, Σ_0 is $2\pi \times \mathrm{f}(0)$ and hence also known as the long-run variance of $x_t \varepsilon_t$.

3. Kernel HAC Estimators

It is clear that exact form of Σ_0 depends on data characteristics. When $x_t \varepsilon_t$ are serially uncorrelated, all the autocovariances in (4) vanish, so that

$$\Sigma_0 = \lim_{T\to\infty} \Gamma_T(0) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\varepsilon_t^2 x_t x_t'). \tag{5}$$

The variance-covariance matrix can be consistently estimated by White's heteroscedasticity consistent estimator:

$$\hat{\Sigma}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_t^2 x_t x_t',$$

with $\hat{e}_t$ the OLS residuals; see White (1984). The matrix (5) can be further simplified when ε_t are conditionally homoscedastic: $\mathbb{E}(\varepsilon_t^2 \mid \mathcal{F}^{t-1}) = \sigma_0^2$, where $\mathcal{F}^{t-1}$ denotes the σ-algebra generated by $\{(x_i, \varepsilon_i), i \leq t-1\}$. In this case, (5) is simplified as

$$\Sigma_0 = \sigma_0^2 (\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(x_t x_t')) = \sigma_0^2 M_0,$$

which can be consistently by $\hat{\Sigma}_T = \hat{\sigma}_T^2 M_T$, with $\hat{\sigma}_T^2 = T^{-1} \sum_{t=1}^{T} \hat{e}_t^2$.

When $x_t \varepsilon_t$ exhibit serial correlations, it is still possible to estimate Σ_0 consistently. Letting $l(T)$ denote a function of that diverges with T we have

$$\Sigma_T^{+} = \sum_{j=-l(T)}^{l(T)} \Gamma_T(j) \to \Sigma_0,$$

as T tends to infinity. It is then natural to estimate Σ_T^{+} by its sample counterpart:

$$\Sigma_T^{+} = \sum_{j=-l(T)}^{l(T)} \Gamma_T(j),$$

with the sample autocovariances:

$$\hat{\Gamma}_T(j) \begin{cases} \frac{1}{T} \sum_{t=j+1}^{T} x_t \hat{e}_t \, \hat{e}_{t-j} x_{t-j}', \ \mathrm{j} = 0,1,2, \ldots \ldots \\ \frac{1}{T} \sum_{t=-j+1}^{T} x_{t+j} \hat{e}_{t+j} \, \hat{e}_t x_t', \ \mathrm{j} = -1, -2, \ldots \ldots \end{cases}$$

The estimator $\hat{\Sigma}_T^{+}$ would be consistent for Σ_0 provided that $l(T)$ does not grow too fast with T; see the discussion in Section 3.2.

A problem with $\hat{\Sigma}_T^{+}$ is that it need not be a positive semi-definite matrix and hence cannot be a proper variance-covariance matrix. A consistent estimator that is also positive semi-definite is the following estimator of the spectral density:

$$\hat{\Sigma}_T^k = \Sigma_{j=-T+1}^{T-1} \kappa\left(\frac{j}{l(T)}\right) \hat{\Gamma}_T(j), \tag{6}$$

where κ is a proper kernel function and $l(T)$ is its bandwidth, which jointly determine the weights assigned to $\hat{\Gamma}_T(j)$. Typically, κ is required to satisfy: $|\kappa(x)| \leq 1,\ \kappa(0) = 1, \kappa(x) = \kappa(-x)$ for all $x \in \mathbb{R}$, $\int|\kappa(x)|\, dx < \infty$, κ is continuous at 0 and at all but a finite number of other points in $\mathbb{R}$, and

$$\int_{-\infty}^{\infty} \kappa(x)\, e^{-ix\omega} dx \geq 0, \forall \omega \in \mathbb{R}.$$

Note that the last condition ensures positive semi-definiteness; see Andrews (1991).

Below is Parzen kernel (Gallant, 1987):

$$\kappa(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & |x| \leq \frac{1}{2} \\ 2(1-|x|)^3, & \frac{1}{2} \leq |x| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

”

2. Code

Data paper;

input date log_search_volume_change log_AvgPrice_change log_volume_change

mood_price_change mood_Tweconomic_change mood_government_change

mood_possibility_change mood_global_change mood_investment_change;

datalines;

```
… …

;

proc model data=paper;

exog log_search_volume_change log_volume_change mood_price_change

                 mood_Tweconomic_change ;

instruments _exog_;

parms b0 b1 b2 b3 b4 ;

log_AvgPrice_change = b0 + b1*log_search_volume_change +

b2*log_volume_change + b3*mood_price_change +

b4*mood_TWeconomic_change;

fit log_AvgPrice_change / gmm kernel=(parzen,4,0) vardef=n;

run;

quit;
```

References

Aouadi, A., Arouri, M., & Teulon, F. (2013). Investor attention and stock market activity: Evidence from France. *Economic Modelling*, *35*, 674-681.

Andrade, S. C., Chang, C., & Seasholes, M. S. (2008). Trading imbalances, predictable reversals, and cross-stock price pressure. *Journal of Financial Economics*, *88*(2), 406-423.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, 817-858.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross‐section of stock returns. *The Journal of Finance*, *61*(4), 1645-1680.

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial markets and portfolio management*, *25*(3), 239-264.

Barber, B. M., & Odean, T. (2007). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785-818.

Barber, B. M., Odean, T., & Zhu, N. (2008). Do retail trades move markets?. *The Review of Financial Studies*, 22(1), 151-186.

Barber, B. M., Odean, T., & Zhu, N. (2009). Systematic Noise. *Journal of Financial Markets*, volume 12, issue 4, pages 547-569, Elsevier

Chuang, W. J., Ouyang, L. Y., & Lo, W. C. (2010). The impact of investor sentiment on excess returns a Taiwan stock market case. *International Journal of Information and Management Sciences*, 21(1), 13-28.

Chung, S. L., & Yeh, C. Y. (2009). Investor Sentiment, Regimes and Stock Returns. (February 13, 2009). Available at SSRN: https://ssrn.com/abstract=1342588.

Da, Z., Engelberg, J., & Gao, P. (2009). In Search of Attention. AFA 2010 Atlanta Meetings Paper. Available at SSRN: http://ssrn.com/abstract=1364209

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5), 1461-1499.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A.(1998) Investor Psychology and Security Market Under- and Over-Reactions, *Journal of Finance*, 53(6):1839-1885.

De Bondt, W. F. & Thaler, R. H. (1995). Financial decision-making in markets and firms: A behavioral perspective. *Handbooks in operations research and management science*, *9*, 385-410.

Fan, M. H., Liao, E. C., & Chen, M. Y. (2014, December). A TAIEX forecasting model based on changes of keyword search volume on Google Trends. *In IEEE*

International Symposium on Independent Computing (ISIC). IEEE.

Gallant, A. R. (2009). *Nonlinear statistical models* (Vol. 310). John Wiley & Sons.

Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and quantitative Analysis*, 22(1), 109-126.

KUAN, C. M. (2008). LECTURE ON Robust Tests with and without Consistent Estimation of Asymptotic Covariance Matrix. Institute of Economics Academia Sinica. Retrieved from http://homepage.ntu.edu.tw/~ckuan/pdf/Lec-HAC_0802.pdf

Lee, C., and Swaminathan, B. (2000), Price momentum and trading volume, *Journal of Finance*, Vol. 55, pp.2017-2069.

Litterman, R. B. (1983). A random walk, Markov model for the distribution of time series. *Journal of Business & Economic Statistics*, *1*(2), 169-173.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of finance*, *32*(4), 1151-1168.

McCall, J. J. (1970). Economics of information and job search. *The Quarterly Journal of Economics*, 113-126.

Russo, J. E. (1974). More information is better: A reevaluation of Jacoby, Speller and Kohn. *Journal of Consumer Research*, *1*(3), 68-72.

Stocking, G., & Matsa, K. E. (2017). Using Google Trends data for research? Here are 6 questions to ask. Retrieved from https://medium.com/@pewresearch/using-google-trends-data-for-research-here-are-6-questions-to-ask-a7097f5fb526

Takeda, F., & Wakao, T. (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. *Pacific-Basin Finance Journal*, *27*, 1-18.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin*, *103*(2), 193.

Varian H (2011), Predicting the present, Think with Google Newsletter. https://www.thinkwithgoogle.com/marketing-resources/predicting-the-present/, September 2011.

Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance*, 36(6), 1808-1821.