The Correlation and Causality Research Between Baidu Index and Stock price: An Empirical Study in China Solar Energy industry

Jian Huang, Yingqi Zeng, Siyan Chen, Shuhui Tang, Xueting Gong, Chengyu Li

Jian Huang (King)

Finance Major (Student ID: 1530002046)

Division of Business and Management

Beijing Normal University-Hong Kong Baptist University United International College

Address: 2000 Jintong Road, Tangjiawan, Zhuhai, Guangdong Province, People's Republic of China

810474480@qq.com Phone: +86 18664549728

Yingqi Zeng (Alyssa)

Finance Major (Student ID: 1530002166)

Division of Business and Management

Beijing Normal University-Hong Kong Baptist University United International College

Address: 2000 Jintong Road, Tangjiawan, Zhuhai, Guangdong Province, People's Republic of China

k530002166@mail.uic.edu.hk

Siyan Chen (Christina)

Finance Major (Student ID: 1530007003)

Division of Business and Management

Beijing Normal University-Hong Kong Baptist University United International College

Address: 2000 Jintong Road, Tangjiawan, Zhuhai, Guangdong Province, People's Republic of China

k530007003@mail.uic.edu.hk

Shuhui Tang (Deemo)

Finance Major (Student ID: 1530002124)

Division of Business and Management

Beijing Normal University-Hong Kong Baptist University United International College

Address: 2000 Jintong Road, Tangjiawan, Zhuhai, Guangdong Province, People's Republic of China

k530002124@mail.uic.edu.hk

Xueting Gong (Gracie)

Finance Major (Student ID: 1530002031)

Division of Business and Management

Beijing Normal University-Hong Kong Baptist University United International College

Address: 2000 Jintong Road, Tangjiawan, Zhuhai, Guangdong Province, People's Republic of China

k530002031@mail.uic.edu.hk

Chengyu Li (Cheryl)

Finance Major (Student ID: 1530013014)

Division of Business and Management

Beijing Normal University-Hong Kong Baptist University United International College

Address: 2000 Jintong Road, Tangjiawan, Zhuhai, Guangdong Province, People's Republic of China

k530013014@mail.uic.edu.hk

Date: 2018.5.24

Abstract

Chinese investors are easily affected by hot topics on the media and Internet and likely take irrational behavior. They tend to consume rationally and invest emotionally. This research aims to analyze the Baidu index's effect on share price of solar power industry in China. Three main models are established to identify the correlation between Baidu index and stock price of solar industry. The media index and searching index of Baidu are chosen as an independent variable to measure the influence from media and Internet, and stock price as the dependent variable. To improve and complete the analysis, Hexun index is taken into consideration as mediator variable. This study applied Multiple Linear Regression Model and VAR (Vector auto-regression) Model. The research mainly uses the tools of SPSS, Eviews and Excel to conduct data analysis. The test is based on the theories of Granger Causality, unit root test, F-test, and t-test. The research proves the relationship between the close price of stock and Baidu index, and it also created a model for further forecasting.

Keywords: Baidu index, media index, Hexun index, solar power industry, close price

Table of Content

1. Introduction

From the prospect of development of high-tech and big data, in the days of overwhelmingly dominant mass media, Chinese investors are easily affected by hot topics on the media on the Internet and are likely to take irrational actions. On one hand, with the power of mass media, they tend to consume rationally but invest emotionally. Additionally, from the view of sustainable development, faced with the problems of resource shortage and environmental pollution, many countries have proposed a low-carbon economy to achieve sustainable development. They have adopted renewable energy as an important part of their energy development. The renewable energy is an important part of sustainable development in China. On the other hand, China is rich in solar energy resources and now it has become the world's largest producer of solar energy. Solar energy has caught the attention by larger amounts of people in various industries in China.

In order to help the investors and companies behave better in the process of investment, the research carried out in order to give them a way to figure out the correlation between searching volume in the Baidu index and perform better in the stock market. The research chose Baidu index as our research object to measure the influence of mass media because media information has good reliability, accuracy, and accessibility. The searching index in Baidu can reflect the amount of attention in a comprehensive and objective way. Searching Index and Media Index are two major parts of Baidu index. These two Baidu indexes will be chosen as a measurement for the Internet's influence. Plus, Hexun index will be also taken into consideration as the mediator variable (MV) in the analysis of Baidu index's effect in stock market. After completing the research, we hope we could know the connections between the dynamics of stock price and Baidu index and then help both investors and companies better understand the correlation between stock market and solar industry. With a better understanding of their connections, both investors and companies can take wiser and

more rational plans and strategies when making their decisions of their purchasing and selling decisions.

The concept of *Hexun Index*

Hexun index is one of the most complete fund database in China's fund market with the longest operation time. It enjoys a high reputation in the industry due to its strong financial background and research strength. Since the first open-end fund issued in 2001, Hexun index has compared all domestic open-end funds and the continuous data tracking. Thus, it established a better basic database and more humanistic multi-function value-added database. Through the integration and extraction, a lot of data information and data-oriented services can be more logically provided to investors.

The concepts of *Searching Index* and *Media Index* of Baidu

Baidu is China's largest search engine. Search Index and Media Index are two major parts of Baidu index. The Search Index uses search volume of netizens in Baidu as data basis, keyword as statistical object, and it calculates the weighted sum of keyword search frequency through a scientific analysis. The Media Index is related to the keyword that included in the Baidu news channels and news coverage of various Internet media.

Baidu indexes will be chosen as a measurement for the Internet's influence and Chinese Internet users' attention to certain events. Baidu index is processed by algorithm, which can basically reflect the frequency of keyword search. Baidu index shows the Internet users' attention to certain keywords and the continuous changes, the search size of a certain keyword, the ups and downs within a period, and changes of related news public opinion.

2. Objective and Significance of This Study

Faced with the problems of resource shortage and environmental pollution, many countries have proposed a low-carbon economy to achieve sustainable development. Many countries have adopted renewable energy as an important part of their energy

development. The renewable energy is an important part of sustainable development in China. China is rich in solar energy resources and now it has become the world's largest producer of solar energy. Solar energy is being paid more and more attention to various industries in China.

In order to find out whether there is correlation between the sectors of stock market and Baidu index. We will take the stock sector of solar industry as an object of our analysis. Baidu Index comes from the source of Internet media Big Data. We hope to study the influence of media on the stock market performance and investors investment behavior though the tool of Baidu Index. In addition, the research will further study about which index's effect is most significant.

After completing the research, we could help both investors and companies to predict or better understand about the trends in the stock market. Besides, the stock market may be controlled within a limited range if we understand the influence of external factors that in addition to the factors of the stock itself. For the further study, they can do research based on the influence of searching index to learn more about the behavior of stock investors and the performance of stock market.

3. Literature Review

A large and growing body of literature has investigated the correlation between investors' attention and stock markets and measured the attention with internet search index. Internet search query index includes a firm's name, alternative words, and correlative words.

3.1. Foreign Research of Subject

Much of the current literature analyzed the influence of search volume on stock market. Google, as a free, easy access website and leading search engine provider, it helps researchers to capture a large number of search interests in specific areas (Dietzel, Braun and Schäfers, 2014). In the theory of Efficient Markets Hypothesis, which delivered by Eugene Fama in 1970s, the stock price will sharply rise when the positive

7

information appears. The stock market is not only affected by the performance of each listed company itself, but also will be influenced by Internet search query data. With the increase of the searching volume, there will be some change and special performance in the stock market. For example, when the stock market volatile, the searching volume will increase and as the search volume increased, the volatile of stock market also improved (Dimpfl & Jank, 2016).

On the one hand, search volume data can be a leading indicator of short-term trends and can even be a turning point for the stock market. For instance, Google search data is suitable for measurement of sentiment in commercial real estate market. (Dietzel, Braun and Schäfers, 2014). On the other hand, the change in the search volume of a company's name has a lot to do with trading activity and investor awareness (Bank, Larch and Peter, 2011).

3.2. Domestic Research of Subject

In recent years, there has been an increasing amount of literature about the influence of investors' attention on the stock market in China. Baidu is the biggest search engine in China so most of the scholars researched based on Baidu as their research object which represented media influence and investors' attention in China (Li, Zhang, Fu, & Wan, 2013). Chen, Mi, Li, and Zheng (2016) claimed that Baidu Index is an effective quota for the research.

Several studies have found that media have a positive correlation stock market, like its liquidity, and joint positive with stock returns when the media information is positive to the investors or the hottest topic at that moment (Fang, Jiang & Qian, 2014; Wang & Hu, 2017; Wang, Long & Wei, 2018). After expanding the theoretical model of Barber and Odean (2008), the Price Pressure Hypothesis, Shi, Huang, Chen (2012) and Li, Zhang, Fu and Wan (2013) have also found that the return of stock investment will decrease when the stock got raising attention. It shows that media can be a forecasting factor to the stock price because it has influence on stock market.

To more accurately determine how the media can affect the stock market, Wang and Hu (2017) compared stocks closing prices under different periods: before the event, at the event and after the event based on the Baidu Index. They found the correlation between

these two variables that when time passed and the attention on the topics declined, stock prices volatile and volumes of stock decreased at the same time. This proved that nowadays investors invest rationally. They also stated that during the period that the topic is hottest and optimistic, the excess return of related stocks will increase while the average abnormal return will present periodic trends and decreased as time passed (Wang & Hu, 2017). What's more, several studies have found that the media correlation to stock rate of returns and the average rate of returns was only for a short-term and after that, the rate of returns would decline and performed bad for a long period (Rao, Peng & Cheng, 2010; Yu & Zhang, 2012; Fang, Jiang & Qian, 2014).

Overall, the proof of previous research is not adequate to explain the effect of media and searching index to the fluctuation of stock price, because firstly some research focus on the industry which has both positive news and scandal, and they cannot fully represent the sustainable hot topic. For example, a scandal also increases the attention of public yet it decreases the stock price. For the second, most past research has not concerned about other variables of stock searching index like Hexun index. Thus, this research will focus on the sustainable positive topic—solar energy industry and form the multiple-regression model to test the effect of Baidu index with the concerning on moderator variable of Hexun index. Third, most previous research just analyzed the correlation but did not focus on the causal relationship between the search index and stock market. Therefore, our research will use Granger Causal Relation Test to figure out which is the cause to the other. Besides, as many previous papers didn't have reliable supporting about the future estimation, this research concentrate on how to apply the result to get the reliable future estimation. Our research is to fill these gaps in this field.

4. Assumption

Assumption 1: All the data's processing and calculating is based on the 4 categories of assumptions according to the basic theory in applied econometric:

MLR.1 Introduce the population model and interpret the population parameters which we hope to estimate. This research assume that the samples can represent the whole population of Baidu indexes and the closing stock prices of solar energy. And the

parameter can estimate the correlation between searching indexes in Baidu and the close price of solar industry

MLR.2 Introduce random sampling from the population and describe the data that we use to estimate the population parameters. In our research, the sample size is very big and we have enough power to make the results accurate without many coincidence factors due to the small sample size.

MLR.3 Add the assumption on the explanatory variables that allows us to compute the estimates from our sample. More assumption provides chances to make the research strict and free from error.

MLR.4 In the population, the mean of the unobservable error does not depend on the values of the explanatory variables. Although in one formula we can't find all the dependent factors, we can assume the unobserved factors don't have a big impact on the dependent variable.

Assumption 2: The sample data approximately follows the rule of normal distribution and then based on this assumption, we build the model of multiple linear regression,

Assumption 3: There' s no cycle for the time series data.

5. Methodology

5.1. Hypothesis, Models and Variables

● *The OLS multiple linear regression model of correlation*

H1: the stock price of solar industry is associate with Baidu Media index and Baidu Searching index.

H2: the stock price of solar industry is associate with Baidu Media index, Baidu Searching index and Hexun index.

H3: the growth of stock price of solar industry is associate with the change of Baidu Media index, Baidu Searching index and Hexun index.

● *The models for Regression Analysis*

Model 1: $Price = \beta_0 + \beta_1 * \text{Media index} + \beta_2 * \text{Searching index}$

Model 2: $Price = \beta_0 + \beta_1 * \text{Media index} + \beta_2 * \text{Searching index} + \beta_3 * \text{Hexun index}$

Model 3: $Ln(Price) = \beta_0 + \beta_1 * Ln(\text{Media index}) + \beta_2 * Ln(\text{Searching index}) + \beta_3 * Ln(\text{Hexun index})$

- ***The Granger causal relation test***

H1: DCLOSE does not cause SEARCHING

H2: DCLOSE does not cause DSEARCHING

H3: SEARCHING does not cause DCLOSE

H4: DESEARCHING does not cause DCLOSE

- ***The VAR (Vector auto-regression) model***

Model 1: $DCLOSE = c + \sum_i^i bi * DCLOSE(-i) + \sum_i^i hi * DSEARCHING(-i)$

For example,

DCLOSE= c + b1* DCLOSE (-1) + b2* DCLOSE (-2) + b3*DCLOSE (-3) + h1* DSEARCHING (-1) + h2* DSEARCHING (-2) + h3* DSEARCHING (-3)

[**Hints:** DCLOSE (-1) represents the DCLOSE of 1 day before]

Model 2: $DCLOSE = c + \sum_i^i bi * DCLOSE(-i) + \sum_i^i hi * SEARCHING(-i)$

Model 3: $DSEARCHING = c + \sum_i^i bi * DCLOSE(-i) + \sum_i^i hi * DSEARCHING(-i)$

Model 4: $SEARCHING = c + \sum_i^i bi * DCLOSE(-i) + \sum_i^i hi * SEARCHING(-i)$

Variables	Measurement	Labels
Dependent variables		
The stock price of solar industry	The stock index of solar industry from Oct. 20, 2011 to April 4, 2018	P
Independent variables		
Media index	The Baidu Media index of solar industry from Oct. 20, 2011 to April 4, 2018 Baidu media index of solar industry on day T $= \sum_1^I Baidu\ media\ index\ of\ Company\ i\ on\ day\ T$ $(i = 56, which\ is\ the\ total\ number\ of\ the\ companies\ in\ solar\ industry)$	M
Searching index	The Baidu Searching index of solar industry from Oct. 20, 2011 to April 4, 2018 Baidu searching index of solar industry on day T $= \sum_1^I Baidu\ searching\ index\ of\ Company\ i\ on\ day\ T$ $(i = 56, which\ is\ the\ total\ number\ of\ the\ companies\ in\ solar\ industry)$	S or SEARCHING
Hexun index	The Hexun index of solar industry from Oct. 20, 2011 to April 4, 2018 (moderating variable)	H
The first-order difference of Searching index	$Dsearching_t = searching\ index_t - searching\ index_{t-1}$ $Dsearching_t$ represents the change of seearching index on day T	DSEARCHING
The first-order difference of close price	$Dclose_t = close\ price_t - close\ price_{t-1}$ $Dclose_t$ represents the change of close price on day T	DCLOSE

5.2. Research Design

The research design includes two steps, which begin with the examination of correlation among the stock price of the solar energy industry, Baidu media index, searching index with the moderating of Hexun index. Plus, the result of this step can figure out the most essential factor that is most correlated to the close price of the solar energy industry. The second step is conducting the Granger causal relation test to clarify which one is the casual factor and which one is the result between the close price and the Baidu index. Moreover, a VAR model is applied in this process to establish the prediction and forecasting.

The first step of the research is to test the correlation of the stock price of the solar energy industry, Baidu media index, searching index with the moderating of Hexun index via multiple linear regression model. We consider that there may not be a defined correlation between the three variables and the stock price and there is even a mixed relationship. Hence, we hope to establish a certain formula so that the relationship between the variables can be inferred by the multi-regression model. This is a regression model of the dependent variable for multiple independent variables. We regarded Baidu searching index and media index as the measurement of Internet factors, and Hexun index as a stock market factor. Baidu searching index and media index are independent variables, and stock price of solar industry is dependent variable. In order to improve the model, Hexun index will be chosen as a moderating variable.

First, we need to determine if there is a correlation between these variables and, if exist, find the right mathematical expression between them. So, we put forward 3 hypotheses and established 3 models respectively. The first hypothesis is used to test the correlation between the Internet factors and the stock price, the second hypothesis is added to the stock market factor, the Hexun index, and the third hypothesis is based on converting the model into a possible linear one when the nonlinear model is encountered. That is, to tests the rate of change of the variables. We used OLS method to obtain a linear regression equation with EXCEL and SPSS, and then conduct statistical tests on it, including the regression coefficient and significance test, and accept or reject the hypothesis. Finally, we conducted factor analysis, identified principle and secondary factors, and analyzed what was the relationship between them and stock price.

13

The second step is to examine the causality between close price and Baidu index especially the searching index. Before the application of Granger causal relation test, a Unit-root test should be done to test the stationary characteristic of the time-series data. If the variable is stationary, it is acceptable for the Granger causal relation test, yet if one variable is not stationary, a first-order difference of that variable should be calculated and checked via the Unit-root test again.

After making sure the station of the data, it was easy to find the Granger cause-result relationship with EVIEWS. Then, in order to make the applicable prediction and forecasting, the VAR model was used to form the optimal model and make the forecasting test. Plus, the VAR model was examined by Q-test to make sure whether the residuals are White Noise, which means the error is random and have no serial correlation with zero mean.

5.3. Data Collection and Description

The database includes Baidu index, Hexun index, TONGDAXING financial planform. For all the 71 companies in the solar industry, we collected data from 56 companies of them. This is because we removed the companies whose company abbreviation or stock code was not included in Baidu Index or some company's abbreviation contains multiple meanings such as "Sunflower" and "Solar energy." It may affect the accuracy of our research subjects.

Each company corresponds to 2 nouns, that is an abbreviation and a stock code which as one set. Then we will collect 56 sets of nouns from the 56 companies' daily data of Baidu searching index, media index and Hexun Index. At the same time, the corresponding daily closing price of the stock were collected. The data are time-series daily data and it covers from Oct. 20, 2011 to April 4, 2018 and the total number group of observation is 1571 because it omits the data of holiday and weekend.

The preliminary calculation is summing up two Baidu indexes of 56 companies and using them to estimate the searching index and media index of the solar industry.
For example:
Baidu searching index of solar industry on day i:

$$= \sum_{1}^{I} Baidu\ searching\ index\ of\ Company\ i\ on\ day\ T$$

$(i = 56, which\ is\ the\ total\ number\ of\ the\ companies\ in\ solar\ industry)$

Baidu media index of solar industry on day T

$$= \sum_{1}^{I} Baidu\ media\ index\ of\ Company\ i\ on\ day\ T$$

$(i = 56, which\ is\ the\ total\ number\ of\ the\ companies\ in\ solar\ industry)$

6. Results

6.1. Empirical Results

Overall Time Series Data Analysis

The overall sample is consisting of 1571 observations from Oct. 2011 to Apr. 2018. The following tables show the regression results of the closing price of solar industry shares, Baidu index (including Searching index and Media index), and HEXUN index. To identify the correlation between Baidu index and stock price, this study used solar industry stock closing price as a dependent variable and two Baidu index as independent variable.

The regression analysis between searching index, media index, and closing price was performed by OLS (ordinary least square). This study also looked at HEXUN index as a **moderator variable** to analyze the outcome between price and index. The purpose is to identify whether there is relation between the stock price and the overall state China's fund market, except the influence of searching engine.

For the overall sample, the regression results shown that **searching index** had a **high correlation with stock price**. Besides, **HEXUN index** also has a **significant correlation with stock price**. However, the **media index** is **less relevant** to the stock price change.

Regression Analysis:

Using regression analysis, we obtained 3 regression equations based on 3 hypotheses:

15

Model 1: $Price = 792.18 + 0.001 * Media\ index + 0.005 * Searching\ index$

Table 1: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.679[a]	.461	.460	299.80741

a. Predictors: (Constant), search index, media index

In Table 1, the R-square (=0.46) shows that the regression of Model 1 can explain 46% of the total variation in stock price outcome.

Table 2: ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	119386622.487	2	59693311.244	664.111	.000[b]
	Residual	139770369.983	1555	89884.482		
	Total	259156992.471	1557			

a. Dependent Variable: close

b. Predictors: (Constant), search index, media index

The Null hypothesis (H_0) of **F-test** is that all the parameters of regression model are zero. In Table 2, the F-test of the regression Model 1 is 664.11 with the **significant less than 5%**. Thus, the Model 1 passed the F-test, and we can **reject H_0**. It means that there is a **significant relationship** between stock price and the other Baidu index.

Table 3: Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	792.180	14.621		54.182	.000
	media index	.001	.001	.037	1.994	.046
	search index	.005	.000	.678	36.397	.000

a. Dependent Variable: close

In Table 3, all coefficients are positive with the p-value less than 0.05. It shows that media and searching index have significant positive correlation with stock price. In addition, searching index has higher coefficients, which means there is a **greater correlation between indices and share prices**.

The coefficient of media index is 0.001 with 0.046 significant. It means the media index has less significant correlation with stock closing price based on 95% confident level.

The coefficient of search index means that each unit change in searching index is associated with 0.005 units change in the stock price of solar industry.

Thus, we can accept the **Hypothesis 1**:

The stock price of solar industry is positively associate with Baidu Media index and Baidu Searching index.

Model 2: $\text{Price} = 946.467 + 0.001 * \text{Media index} + 0.005 * \text{Searching index}$

$-0.003 * \text{Hexun index}$

Table 4: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.826[a]	.682	.681	230.39768

a. Predictors: (Constant), HEXUN, media index, search index

In Table 4, after adding an additional explanatory variable, **HEIXUN index**, the value of **R-square increase** from **0.46 to 0.68**. It means that the **Model 2** has more explanatory power than previous **Model 1** as R-square increase. It shows that the regression of Model 2 can explain 68% of the total variation in stock price outcome.

Table 5: ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	176665870.732	3	58888623.577	1109.367	.000[b]
	Residual	82491121.739	1554	53083.090		
	Total	259156992.471	1557			

a. Dependent Variable: close
b. Predictors: (Constant), HEXUN, media index, search index

In Table 5, the F-test of the regression Model 2 is 1109.367 with the significant much less than 5%. Thus, the **Model 2 passed the F-test**, and shows that there is a significant relationship between stock price and the three independent variables: searching index, media index, and HEXUN index.

Table 6: Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		

17

1	(Constant)	946.467	12.178		77.720	.000
	media index	.001	.001	.027	1.887	.059
	search index	.005	.000	.772	52.891	.000
	HEXUN	-.003	.000	-.480	-32.849	.000

a. Dependent Variable: close

The results in Table 6 shows that **expect for media index** (p-value=0.059), other independent variables are significantly correlate with the stock price.

It shows that **searching index** has significant **positive correlation** with stock price. In addition, **HEXUN index** has significant **negative correlation** with stock price.

The coefficient of searching index means that each unit increase in searching index is associated with 0.005 units increase in the stock price of solar industry. Also, the coefficient of HEXUN index means that each unit increase in HEXUN index is associated with 0.003 units decrease in price.

Because the media index does not pass the test, we **cannot** accept the **Hypothesis 2**:

The stock price of solar industry is associate with Baidu Media index, Baidu Searching index and HEXUN index.

Model 3:

$$Ln(Price) = 2.379 \quad 0.015 * Ln(Media\ index) + 0.584 * Ln(Searching\ index)$$
$$- 0.175 * Ln(Hexun\ index)$$

Because the Model 2 is not accepted, we improve the model and adjust the Hypothesis. In Model 3, we take the **logarithm of both sides** of the equation based on Model 2.

Table 7: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.957[a]	.915	.915	.09802

a. Predictors: (Constant), Ln(S), Ln(H), Ln(M)

In Table 7, after taking the logarithm of each variable, the value of **R-square increase** from **0.68 to 0.91**. It means that the **Model 3** has more explanatory power than previous **Model 2** as R-square increase. It shows that the regression of Model 3 can explain 91% of the total variation in stock price outcome.

Table 8: ANOVA[a]

18

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	160.586	3	53.529	5571.055	.000[b]
	Residual	14.931	1554	.010		
	Total	175.518	1557			

a. Dependent Variable: Ln(N)

b. Predictors: (Constant), Ln(S), Ln(H), Ln(M)

In Table 8, the F-test of the regression Model 3 is 5571.055 with the significant much less than 5%. Thus, it shows that **Model 3 passed the F-test**, and there is a significant relationship between the percentage change of stock price and the percentage change of three independent variables: searching index, media index, and HEXUN index.

Table 9: Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.379	.072		33.182	.000
	Ln(M)	-.015	.003	-.047	-5.597	.000
	Ln(H)	-.175	.002	-.591	-73.528	.000
	Ln(S)	.584	.006	.769	97.658	.000

a. Dependent Variable: Ln(N)

Table 9 shows that **all the three independent variables** are significantly correlated with the stock price. The results show that **Ln(searching index)** has significant **positive correlation** with stock price. In addition, **Ln(HEXUN index)** and **Ln(media index)** have significant **negative correlation** with stock price.

The searching index has **the largest coefficient**, which means that it has the strongest relationship with stock price. Searching index (beta=0.584) means **each percentage increase** in searching index is associated with **0.584% units increase** in the stock price of solar industry.

Thus, Model 3 pass the test and we can accept the **Hypothesis 3**:

The growth of stock price of solar industry is associate with the percentage change of Baidu Media index, Baidu Searching index and HEXUN index.

6.2. The Result of Causality

The result of previous part show that the searching index is the most significant variable that is correlated to close price of solar industry. Thus, in this part, the research focus on the causality among the level of searching index(SEARCHING), the change of searching index(DSEARCHING), the change of close price(DCLOSE)

● **Precondition: the data is stationary and pass the unit root test**

With the reason that the time-series data- "Close price", cannot pass the unit root test, it should be transformed to the first-order different of close price

Null Hypothesis: Unit root (individual unit root process)
Series: SEARCHING, DCLOSE, DSEARCHING
Date: 05/21/18   Time: 01:38
Sample: 10/10/2011 4/03/2018
Exogenous variables: Individual effects
Automatic selection of maximum lags
Automatic lag length selection based on SIC: 0
Total number of observations: 3023
Cross-sections included: 3

Method	Statistic	Prob.**
ADF - Fisher Chi-square	342.498	0.0000
ADF - Choi Z-stat	-17.1692	0.0000

** Probabilities for Fisher tests are computed using an asymptotic Chi
   -square distribution. All other tests assume asymptotic normality.

Intermediate ADF test results UNTITLED

Series	Prob.	Lag	Max Lag	Obs
SEARCHING	0.0000	0	0	1225
DCLOSE	0.0000	0	0	899
DSEARCHING	0.0000	0	0	899

The result of above table shows that the p-value of Unit-root test which examines three variables-SEARCHING, DCLOSE and DSEARCHING are all far less than 0.05, which means that we have 95% confidence to conclude that these 3 variables are stationary data and can be used to conduct the Granger causal relation test and VAR model.

● *Granger causal relation test*

Pairwise Granger Causality Tests

Date: 05/21/18   Time: 01:15

Sample: 10/10/2011 4/03/2018

Lags: 1

Null Hypothesis:	Obs	F-Statistic	Prob.
DCLOSE does not Granger Cause SEARCHING	899	5.57504	0.0184
SEARCHING does not Granger Cause DCLOSE		3.91439	0.0482
DSEARCHING does not Granger Cause DCLOSE	899	1.48273	0.2237
DCLOSE does not Granger Cause DSEARCHING		12.2487	0.0005

Pairwise Granger Causality Tests

Date: 05/21/18   Time: 01:16

Sample: 10/10/2011 4/03/2018

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
DCLOSE does not Granger Cause SEARCHING	583	9.00152	0.0001
SEARCHING does not Granger Cause DCLOSE		3.09882	0.0459
DSEARCHING does not Granger Cause DCLOSE	583	2.18109	0.1138
DCLOSE does not Granger Cause DSEARCHING		11.4602	1.E-05

Pairwise Granger Causality Tests

Date: 05/21/18   Time: 01:16

Sample: 10/10/2011 4/03/2018

Lags: 3

Null Hypothesis:	Obs	F-Statistic	Prob.
DCLOSE does not Granger Cause SEARCHING	282	7.77118	5.E-05
SEARCHING does not Granger Cause DCLOSE		1.46184	0.2252
DSEARCHING does not Granger Cause DCLOSE	282	1.86551	0.1357
DCLOSE does not Granger Cause DSEARCHING		5.44199	0.0012

These four tables above reflect that:

1. DCLOSE causes SEARCHING, which means the change of close price is the Granger cause of the level of searching index for 3 days, and the change of close price possibly can be used to predict the level of searching index.

2. DCLOSE causes DSEARCHING, which means the change of close price is the Granger cause of the change of searching index for 3 days, and the change of close price possibly can be used to predict the change of searching index.

3. SEARCHING causes DCLOSE, which means the level of searching index is the Granger cause of the change of close price for 2 days, and the level of searching index possibly can be used to predict the change of close price.

4. DESEARCHING does not cause DCLOSE, which means the change of searching index is not the Granger cause of the change of close price, and the change of searching index cannot be used to predict the change of close price.

- **VAR model**

According to Granger causal relation test, Model 1 is rejected and the rest of them are possible workable for doing the VAR model.

Model 1: $DCLOSE = c + \sum_i^i b_i * DCLOSE(-i) + \sum_i^i h_i * DSEARCHING(-i)$

For example,

DCLOSE= c + b1* DCLOSE (-1) + b2* DCLOSE(-2)+ b3*DCLOSE(-3)+ h1* DSEARCHING(-1)+ h2* DSEARCHING (-2)+ h3* DSEARCHING (-3)

[**Hints:** DCLOSE (-1) represents the DCLOSE of 1 day before]

Model 2: $DCLOSE = c + \sum_i^i b_i * DCLOSE(-i) + \sum_i^i h_i * SEARCHING(-i)$

Model 3: $DSEARCHING = c + \sum_i^i b_i * DCLOSE(-i) + \sum_i^i h_i * DSEARCHING(-i)$

Model 4: $SEARCHING = c + \sum_i^i b_i * DCLOSE(-i) + \sum_i^i h_i * SEARCHING(-i)$

The Result of Model 2:

Model 2: $DCLOSE = c + \sum_i^i bi * DCLOSE(-i) + \sum_i^i hi * SEARCHING(-i)$

Dependent Variable: DCLOSE

Method: Least Squares

Date: 05/21/18   Time: 02:25

Sample (adjusted): 10/28/2011 3/30/2018

Included observations: 282 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DCLOSE(-1)	0.181901	0.056634	3.211851	0.0015
DCLOSE(-2)	-0.141086	0.063158	-2.233864	0.0263
DCLOSE(-3)	-0.004709	0.048782	-0.096521	0.9232
SEARCHING(-1)	0.000344	0.000191	1.801745	0.0727
SEARCHING(-2)	-8.75E-05	0.000176	-0.498261	0.6187
SEARCHING(-3)	-0.000244	0.000142	-1.718459	0.0868
R-squared	0.089113	Mean dependent var		0.742943
Adjusted R-squared	0.072612	S.D. dependent var		22.87424
S.E. of regression	22.02812	Akaike info criterion		9.043564
Sum squared resid	133925.8	Schwarz criterion		9.121051
Log likelihood	-1269.143	Hannan-Quinn criter.		9.074637

The result of this model suggests that the p-value of searching index (SEARCHING) are all more than 0.05, which means the null hypothesis should not be rejected with 95% confidence and the level of the searching index cannot be used to predict the change of close price (DCLOSE).

The Result of Model 3

Model 3: $DSEARCHING = c + \sum_i^i bi * DCLOSE(-i) + \sum_i^i hi * DSEARCHING(-i)$

Dependent Variable: DSEARCHING

Method: Least Squares

Date: 05/21/18   Time: 02:10

Sample (adjusted): 10/27/2011 3/30/2018

Included observations: 583 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.

23

DSEARCHING(-1)	-0.224316	0.036894	-6.080038	0.0000
DSEARCHING(-2)	-0.194797	0.037978	-5.129174	0.0000
DCLOSE(-1)	66.84804	14.35882	4.655539	0.0000
C	-2983.648	333.8502	-8.937088	0.0000

R-squared	0.096085	Mean dependent var	-3077.274
Adjusted R-squared	0.091401	S.D. dependent var	8445.249
S.E. of regression	8050.049	Akaike info criterion	20.83158
Sum squared resid	3.75E+10	Schwarz criterion	20.86155
Log likelihood	-6068.406	Hannan-Quinn criter.	20.84326
F-statistic	20.51555	Durbin-Watson stat	1.849635
Prob(F-statistic)	0.000000		

The result of this model suggests that the p-value of the change of prior searching index[DSEARCHING(-1) and DSEARCHING(-2)] and close price[DCLOSE(-1)] are all less than 0.05, which means the null hypothesis should be rejected with 95% confidence and the change of close price can be used to predict the change of searching index(DSEARCHING). However, both the R-squared and adjusted R-squared are nearly equal to 0.09, which means the VAR model only can explain 9% data of DSEARCHING. Therefore, the model is not good enough because of its fitness.

The Result of Model 4

Model 4: $\text{SEARCHING} = c + \sum_1^i bi * \text{DCLOSE}(-i) + \sum_1^i hi * \text{SEARCHING}(-i)$

Dependent Variable: SEARCHING

Method: Least Squares

Date: 05/23/18   Time: 23:03

Sample (adjusted): 10/28/2011 3/30/2018
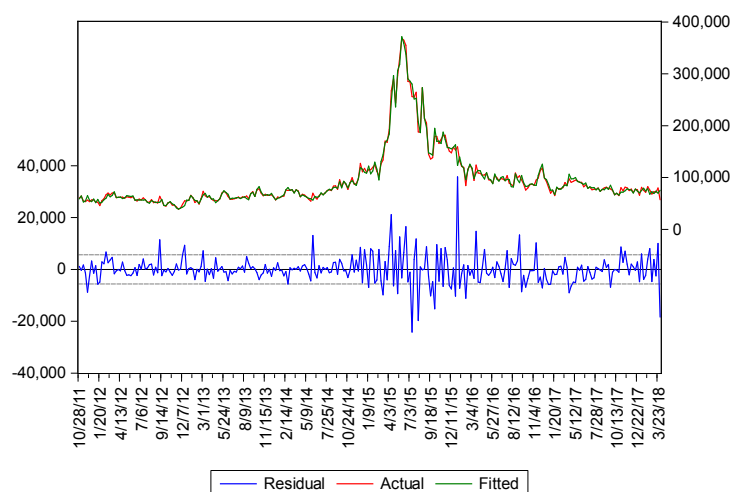
Included observations: 282 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SEARCHING(-1)	0.828765	0.034853	23.77858	0.0000
SEARCHING(-3)	0.099028	0.034716	2.852516	0.0047
DCLOSE(-1)	33.26763	14.22716	2.338319	0.0201
DCLOSE(-2)	-46.35703	16.13735	-2.872654	0.0044
DCLOSE(-3)	28.24141	12.48031	2.262877	0.0244
C	1895.761	646.4376	2.932627	0.0036

24

R-squared	0.990307	Mean dependent var	94343.58
Adjusted R-squared	0.990132	S.D. dependent var	56682.17
S.E. of regression	5630.714	Akaike info criterion	20.13091
Sum squared resid	8.75E+09	Schwarz criterion	20.20839
Log likelihood	-2832.458	Hannan-Quinn criter.	20.16198
F-statistic	5639.912	Prob(F-statistic)	0.000000

The result of this model reflects that the p-value of the prior searching index[SEARCHING(-1) and SEARCHING(-3)] and the change of close price[DCLOSE(-1), DCLOSE(-2) and DCLOSE(-3)] are all less than 0.05, which means the null hypothesis should be rejected with 95% confidence and the change of close price can be used to predict the level of searching index(SEARCHING). In addition, both the R-squared and adjusted R-squared are nearly equal to 0.99, which means this VAR model can explain 99% data of SEARCHING.

Moreover, based on the graft shown below, the fitted line is really close to actual curve, which means the model is good enough to explain the change.

Date: 05/23/18 Time: 23:21
Sample: 10/10/2011 4/03/2018
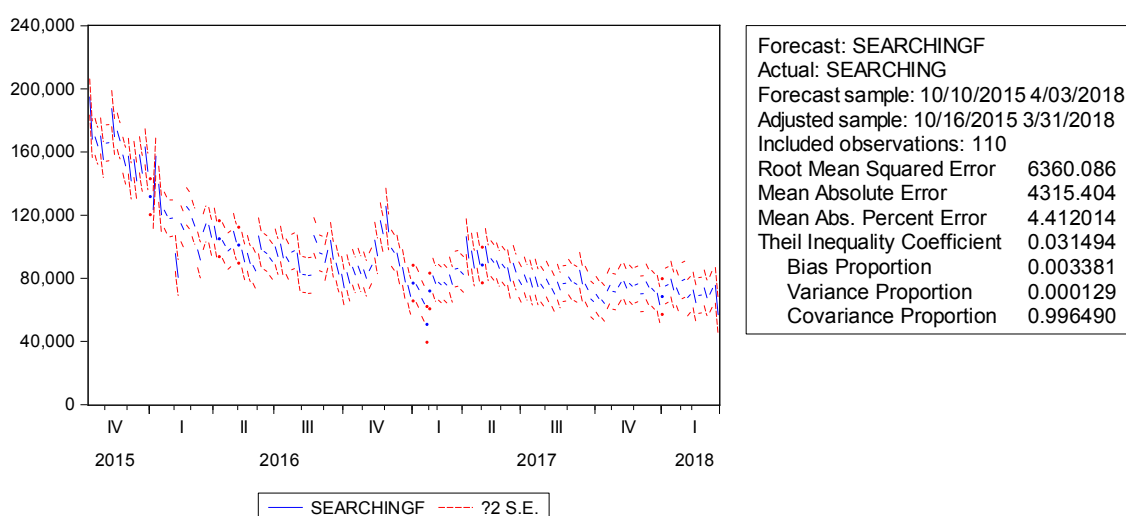Included observations: 282
Q-statistic probabilities adjusted for 2 dynamic regressors

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.000	0.000	0.0000	1.000
		2	0.000	0.000	0.0000	1.000
		3	0.000	0.000	0.0000	1.000
		4	0.000	0.000	0.0000	1.000
		5	0.000	0.000	0.0000	1.000
		6	0.000	0.000	0.0000	1.000
		7	-0.177	-0.177	9.1084	0.245
		8	0.000	0.000	9.1084	0.333
		9	0.000	0.000	9.1084	0.427
		10	0.000	0.000	9.1084	0.522
		11	0.000	0.000	9.1084	0.612
		12	0.000	0.000	9.1084	0.694
		13	0.000	0.000	9.1084	0.765
		14	-0.021	-0.054	9.2383	0.815
		15	0.000	0.000	9.2383	0.865
		16	0.000	0.000	9.2383	0.903
		17	0.000	0.000	9.2383	0.932
		18	0.000	0.000	9.2383	0.954
		19	0.000	0.000	9.2383	0.969
		20	0.000	0.000	9.2383	0.980
		21	0.068	0.056	10.651	0.969
		22	0.000	0.000	10.651	0.979
		23	0.000	0.000	10.651	0.986
		24	0.000	0.000	10.651	0.991
		25	0.000	0.000	10.651	0.994
		26	0.000	0.000	10.651	0.997
		27	0.000	0.000	10.651	0.998
		28	-0.118	-0.101	15.063	0.978
		29	0.000	0.000	15.063	0.985
		30	0.000	0.000	15.063	0.989
		31	0.000	0.000	15.063	0.993
		32	0.000	0.000	15.063	0.995
		33	0.000	0.000	15.063	0.997
		34	0.000	0.000	15.063	0.998
		35	-0.026	-0.064	15.282	0.998
		36	0.000	0.000	15.282	0.999

*Probabilities may not be valid for this equation specification.

Then, a Q-test also is conducted to tested whether the residual is White Noise that the expected value of mean is equal to 0 and variance is constant. According to the result, the p-value are large than 0.05 and AC and PAC are equal to zero in most of condition, which means that the residual of data are White Noise, which data is stationary.

Lastly, a foresting-test is done by using the data from Oct. 10, 2011 to Oct. 15, 2015 as estimator and the data from Oct. 16, 2015 to Mar. 31, 2018 as testing sample. The graph shows that the actual fluctuation of searching index is within the 95% confidence interval of VAR model.

Forecast: SEARCHINGF	
Actual: SEARCHING	
Forecast sample: 10/10/2015 4/03/2018	
Adjusted sample: 10/16/2015 3/31/2018	
Included observations: 110	
Root Mean Squared Error	6360.086
Mean Absolute Error	4315.404
Mean Abs. Percent Error	4.412014
Theil Inequality Coefficient	0.031494
Bias Proportion	0.003381
Variance Proportion	0.000129
Covariance Proportion	0.996490

Overall, the change of close price can be used to predict the level of searching index(SEARCHING) and the model is applicable for forecasting.

7. Discussion and Recommendation

The Discussion on the Correlation

Correlations

		Close price	Searching index	Ln(C)	Ln(S)
Close price	Pearson Correlation	1	.678**	.990**	.750**
	Sig. (2-tailed)		.000	.000	.000
	N	1558	1558	1558	1558
Searching index	Pearson Correlation	.678**	1	.632**	.954**
	Sig. (2-tailed)	.000		.000	.000
	N	1558	1558	1558	1558
Ln(C)	Pearson Correlation	.990**	.632**	1	.738**
	Sig. (2-tailed)	.000	.000		.000
	N	1558	1558	1558	1558
Ln(S)	Pearson Correlation	.750**	.954**	.738**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	1558	1558	1558	1558

**. Correlation is significant at the 0.01 level (2-tailed).

The above table shows the Pearson correlation of closing price with searching index, searching index with closing price, the searching index with Ln(closing price), and the closing price with ln(searching index). And according to the concept of Pearson Correlation, it is better for the correlation to be larger than 0.5. From the number of these four pairs we got, they were all larger than 0.5, so the correlation is significant at the 0.01% level, which means it is 99% certain to conclude that these four pairs have strong correlation.
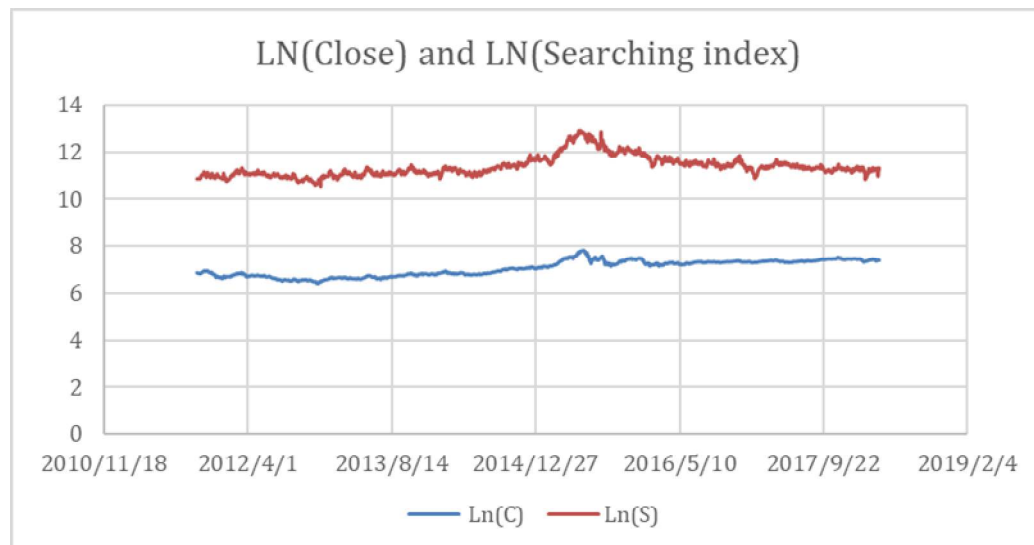
Especially, the correlation between these two variables shows in the above graph, ln(c) and ln(s) had almost same trend, which means closing price and searching index are strongly correlative.

Based on our hypothesis, we assumed that there is a certain correlation between the rate of change of Baidu Index and the rate of change of the stock price of solar energy industry. However, after using Granger causality tests to explore the hypothesis, it surprised us that the change of closing price is the cause; on the contrary, the change of searching both Baidu and Hexun index is the effect. And it further explains that using searching index to predict the stock price is wrong. Therefore, we launched an inference as the following graph shows to describe their relationships, since there is still some correlation between them. And these four stages are interlocked by each other.

```
                    Change rate of
                    Searching Index
                           │
            3              │            2
         Scandal           │        Bull Market
       (Bear Market)       │
                           │                 Change rate of
      ─────────────────────┼────────────────  closing price of
                           │                      industry
                           │
          Recession        │         Opportunity
            4              │            1
                           │
```

In view of the above reasoning, we had two assumptions to have a better explanation of this graph. One is we assume that always individual households use Baidu and Hexun index when buying and selling stock (especially the speculation action of those who find a stock price rise and quickly use Baidu Index or Hexun Index to search for the stock). In contrast, institutional households are hardly doing like what individuals did. Then we made a further assumption that the institutional households are ultimately earning more than individual households. Therefore, we utilize solar energy industry to demonstrate these four stages.

In the first stage, when the searching index of the solar energy industry is quite low, the price of the stock has already risen and the growth rate of it has increased, which means the institutional households start to buy the stock; however, the individual households do not be pay attention to it. So, the solar energy industry is in a condition of "**Opportunity**", that is a good chance to invest in the stock.

In the second stage, the individual households are becoming to be aware of the growth of the solar energy industry stock, then they start to search for it and the searching index of the stock rises, which lead to the price of the stock rises sharply. The solar energy stock market changes to "**Bull Market**" at this moment. Therefore, the proper way is to underweight the stock appropriately at the point.

Next, in the third stage, as the result of the second stage, the stock price starts to go down, which means the institutional households gradually reduce the share of the stock; still the searching index keeps in a popular state, there might be two possible cause

of the phenomenon, one is for the market changes to "**Bear Market**", the other might because there has been some negative news in the industry.

In the fourth also the last stage, the price of the stock continues to fall, the heat of searching index of the stock back to normal. The solar energy industry is experiencing the investment period of **recession**.

The Discussion on the Causality Analysis
Based on Granger causality test and VAR model, we can draw the following relationship:

a. The level of the search index cannot predict the change of the close price.

b. The change of the close price can predict the level of the search index.

- **Granger causality test**

In these results, the Granger causality test shows that DCLOSE causes SEARCHING and the change of close price is the Granger cause of the change of searching index. And the results of the VAR model show that the previous change of the close price change with a later increase in search volume.

For this result, here is a possible reason: It explains that the stock market contains irrational factors, that is, investors are emotional rather than completely rational. Investors increased their attention to the stock after the stock price changed and prompted them to search on Baidu. Therefore, higher stock price changes can promote the search volume of relevant keywords.

- **Prediction**

In the VAR model, based on the optimal model, model 4, we can effectively predict the future search volume through the historical data of DCLOSE.

a. Both the R-squared and adjusted R-squared are nearly equal to 0.99, which means this VAR model can explain 99% data of SEARCHING.

b. The variance of the model is very low, and the Q-test shows that the residual of data is White Noise, which means that our model is very stable.

c. The foresting-test shows that the model is acceptable and applicable for forecasting.

- **Application**

The results show that the level of the search index cannot be used to predict the change in the closing price, but through the change in the closing price, we can

30

effectively predict the level of the search index in the future. And this result can be used in the application in advertising with the forecasting to search index.

The "advertising" here represents the "search advertising". It is an online marketing method that displays online advertisements on search engine results pages. The system that makes profits through this method is the search engine advertising system. It is designed to allow users to see the advertisements provided by advertisers by matching the query words that inputted by the user and the keyword used by the advertiser for advertising and to get advertising fees.

Our results show that the search index cannot predict the stock price, but the stock price can predict the search index. Thus, changes in the search volume of related keywords can be predicted through changes in the stock price. And the search volume represents the search behavior of the Internet users, so that this prediction can be applied to the placement of search advertising.

According to the results of our research, the change in the close price before can be used to predict the change in the search volume of keywords associated with the stock at the probability of more than 99%. And the behavior of one user's searching for the stock related data also shows that the user is interested in stock-related information. It includes the behavior preferences of user. As a result, this application is feasible.

● **Problems**

However, there are some problems that need to be solved for the application in advertising:

a. What is the proportion of searches for stock-related keywords in total search volume?

b. How does the type of ad match the type of user?

c. For other stocks, does stock price change have the same ability to predict?

Because this forecasting method is based on the stocks of the solar energy industry. If we want to apply it in a wider range of fields, we need to analyze the same type of stocks in other industries. In order to predict the search volume through the changes in stock close prices, that is, a prediction of network hotspots and to create value from it.

8. Limitations

1. Although we assume that the sample data follows the rule of normal distribution, after doing the calculation and drawing the graph, we found that the data slightly deviate from the normal distribution. Instead, the value in the middle range has a trend of concavity. In other words, the shape of distribution has two peaks which appeared on two sides of the axis of symmetry. However, the assumption of dealing with the data with normal distribution is acceptable since it's a time series data.

2. The research has no detrending since we want our models have better significance of applications. This model is lack of the process of detrending because we assume that there's no cycle for the time series data.

3. In our research we only consider about the closing price of solar industry stock. The conclusions and results of this study might be limited since all the assumptions and analysis were restricted in the range of solar industry in China.

9. Conclusion

The main conclusions from regression analysis are:

● The stock price of solar industry is positively associate with Baidu Media index and Baidu Searching index.

● The growth of stock price of solar industry is associate with the percentage change of Baidu Media index, Baidu Searching index and HEXUN index.

The major contribution done after the research can be summarized into three steps from primary to mature stage:

● The research results show that the value of Baidu searching index and the close price are highly correlated with each other. The change rate of these values also carries the same regulations of big correlation.

● In the primary level we can't decide which variable is dependent and which is independent. However, with the fundamental theory of Granger causality test, we find out that the change of close price cause searching indexes' fluctuation. This symptom makes a precise implication that investors in the market use the search engine of Baidu to pay close attention to the stocks' price due to the fluctuation of them. It also proves that both the macro trend of market and the investors inside the

market are emotional to behave while making decisions of buying and selling.

- This research has created a new model which can be used to predict the Baidu searching index (In other words, the industry's degree of attention) by observing the change of the close price of stocks with EAR model.

- For further study, we provide to directions to develop the implications for case. In the discussion part, we use a model about four stages interlocked by each other. For deeper research, we can find other methods to prove the conclusion in the model about how these stages of market interlock or interaction with each other. We can also use the same methodology to examine that whether other industries can correspond with the theory in solar industry and then find out its range of application.

References

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.

Chen, Z. -Y., Mi, Y. -X., Li, Y. -J. & Zheng, J. -J. (2016)   [The Empirical Analysis Based on Baidu Index's Investors' Attention and Stock Market Performance]. *Journal of Statistics and Decision,* No. 23, 155-157.

Dietzel, A., Braun, M., & Schäfers, W. (2014). Sentiment-based commercial real estate forecasting with Google search volume data. *Journal of Property Investment & Finance, 32*(6), 540-569.

Dimpfl, T., & Jank, S. (2016). Can internet search queries help to predict stock market volatility?. *European Financial Management, 22*(2), 171-192.

Fang, X. -M., Jiang, Y., & Qian Z. J. (2014). The effects of individual investors' attention on stock returns: evidence from the chinext market. *Emerging Markets Finance & Trade, 50*(sup3), 158-168.

Rao, Y. L., Peng, D. F., & Cheng, D. -C. (2010). Does media attention cause abnormal return? evidence from china's stock market. Systems Engineering—Theory & Practice, 30(2), 287-297.

Shi, R., Huang, J., Xu, Z., & Chen, Z. (2012). Does attention affect individual investors' investment return? *China Finance Review International, 2*(2), 143-162.

Wang, B., Long, W., & Wei, X. (2018). Investor attention, market liquidity and stock return: a new perspective. *Asian Economic & Financial Review, 8*(3), 341-352.

Wang, X. -Y., Wang, X. -F., Hu D. (2017).   [The popularity of artificial intelligence and its relationship with the stock performance of AI industries' index]. *South China Finance*, No. 42-53.

Yu Q. -J., & Zhang B. (2012).   [Investor limited concern and stock returns--an empirical study based on Baidu index]. *Journal of Finance Research* (8), 152-165.