

Full Length Research Paper

## Can web-searching index help to predict renminbi exchange rate?

Xuan Wang<sup>1\*</sup>, Suxiao Li<sup>1</sup>, Haizhen Yang<sup>1</sup> and Lingxiao Cui<sup>2</sup>

<sup>1</sup>School of Management, University of Chinese Academy of Sciences, Beijing, China.

<sup>2</sup>Research Center of Fictitious Economy and Data Science, Chinese Academy of Sciences (CAS), Beijing, China.

Received 31 August, 2015; Accepted 1 October, 2015

The stability of exchange rates plays a decisive role in a country's economic development and internal and external equilibrium. Therefore, the prediction of short-term exchange rate is of vital importance to maintain a country's economic stability and financial security. Traditional time series forecasting models are only based on historical data, which cannot reflect other important factors, like investors' currency exchange expectations and their emotions. This study aims to build a web-searching index to predict the short-term exchange rate, which is based on web search data, the natural language processing and information retrieval sharing platform (NLPIR) which is a Chinese segmentation technique and the TextRank keywords extraction system. In this way, the study establishes a Conditional Autoregressive Model (CAR) model integrated with our web-searching index to include the investor's expectations and emotions in the prediction, and therefore enhance prediction accuracy. The outcome shows that the accuracy of the CAR model based on search data can be significantly higher than other models. Besides, compared with traditional exchange rate prediction models, the integrated CAR model has a better fitting effect and a lower prediction error.

**Key words:** Exchange rate prediction, web-searching index, big data, TextRank, NLPIR.

### INTRODUCTION

Since the exchange rate reform in 21 July, 2005, China's exchange rate regime has become a managed floating exchange rate system, which tracks a package of currencies. From July, 2005 to December, 2014, China Yuan or Renminbi (CNY) nominal exchange rate cumulatively appreciated by 34%, and CNY real exchange rate cumulatively appreciated to 46.21%, which illustrates that the regime of CNY exchange rate tends to be more flexible and market-oriented than before (Data source: Wind Database). The continuous appreciation and greater volatility of CNY not only lead to a relative

devaluation of foreign exchange reserves, but also result in increased exchange rate risk of financial and trading activities that are not CNY denominated. Therefore, it is of vital importance to predict short-term exchange rate to manage exchange risks. Besides, the CNY exchange rate formation mechanism has become more market-oriented, and therefore the difficulty of exchange rate forecasting increases significantly. It becomes increasingly relevant to explore a precise and effective approach to predict the short-term exchange rate.

Previous forecasting methods generally relied on time

\*Corresponding author. E-mail: wang.xuan@hotmail.com.

series models to forecast short-term exchange rates, like autoregressive (AR), autoregressive integrated moving average models (ARIMA) and generalized autoregressive conditional heteroskedasticity (GARCH) model. More recent approaches employ neural networks and chaos theory. However, these models are under an assumption that market transactions have contained all the relevant information, which makes it possible to do accurate prediction which only relies on historical data. But the reality is that price itself can not contain all the information, especially in developing countries. The investors' expectations and their emotions cannot be reflected promptly in historical price data. Therefore, if we want to get an accurate prediction by time series models, we have to take the expectations and emotions factors into consideration.

With the advent of the internet era, it becomes possible to obtain people's expectations and their anticipated trading behaviors by analyzing data from Internet. When the two trading parties' psychological expectations change, it will be reflected in their network behaviors. The expectations can be presented in multiple way, such as the relevant comments written by economists, the news and financial columns related to the fluctuations of exchange rates, "weibo" messages and forums. We can catch the changes of expectations and emotions through the web-searching index, and make some pre-judgments of the fluctuations of exchange rates, and therefore make accurate prediction. In 2015, the number of internet users exceeded 668 million people (Data source: China Internet Network Information Center statistics). The behavior of such huge amount of Internet users has constructed an unprecedented scale of big data, including their hobbies, expectations, concerns and doubts. This data can help us to do prediction for exchange rate accurately.

Based on big data context therefore, this paper constructs a web-searching index to predict exchange rate. Firstly, we use web-searching data, the NLPIR Chinese segmentation technique and the TextRank keywords extraction system to obtain the keywords which are highly related to Chinese Yuan (CNY) exchange rate. Then, the study built a keywords web-searching index system and developed a qualitative measure of internet users' expectations and judgments from big data. Afterwards, the study constructs a comprehensive web-searching index based on the keywords web-searching index. The study employ Conditional Autoregressive Model (CAR) model integrated with web-searching index to forecast the CNY short-term exchange rate. The result illustrates that the CAR model integrated with the index has lower prediction error than traditional exchange rate forecasting models.

## LITERATURE REVIEW

### Literature related to exchange rate prediction

The fluctuations of exchange rate play a decisive role in a

country's economic development and internal and external equilibrium. Therefore, researchers pay a lot of attention on the prediction of exchange rate accurately.

This study classifies this research into three strands. The first strand is the exchange rate forecasting method based on macro-economy theory. But this method cannot explain exchange rate volatility, forex-trading volume zoomed out effect and the separate of market exchange rate from equilibrium exchange rates, which are all common phenomenon in markets. The second strand is the exchange rate forecasting method based on market micro-structure. Responding to the existing problems in the first method, scholars established financial market microstructure theory. The theory points out that information are private and it is the key to price formation (Lyons, 2001). Forex traders who possess different information tend to play games while trading, and therefore leads to the fluctuations of exchange rates. The third strand is the time series forecasting method based on historical data. After years of development, this method has yielded fruitful results. Box et al. (1994), put forward the famous forecasting model—autoregressive moving average (ARMA) model. Based on previous research, Engle et al. (1982) improved the model and created autoregressive conditional heteroskedasticity (ARCH) model. Based on Engle's study, Bollerslev (1986) considered more likelihood estimation and testing and established generalized autoregressive conditional heteroskedasticity model. These forecasting models have profound impacts on exchange rate predicting research.

With the development of exchange rate prediction models, more and more scholars use time series models, like AR, ARIMA, ARCH, and GARCH, to predict short-term CNY exchange rates. Compared with the previous predicted models, time series models can make better use of historical data, and also provide very accurate short-term forecasts. However, one assumption of time series method is that market transactions contain all the relevant information, and traders make decisions based on this information. In fact, price itself dose not contain all the information, especially in developing countries. The expectations of traders cannot be reflected in the price promptly. Therefore, traditional time series models, which are only based on historical data, cannot reflect factors affecting exchange rates such as traders' expectations and emotions. The models' forecasting accuracy is not high enough.

### Literature related to web-searching system

With the development of internet, the information, which is not covered by market transactions, can reflect in the web big data. Some web-searching platforms launched web-searching index query based on keyword, and provide an approach for scholars to use web big data.

The keywords web-searching index, which is based on searching platforms, has many advantages. For example,

searching platforms, has many advantages. For example, it contains a large volume of information, and it covers a wide field and multiple samples. Forecasting with web-searching index therefore, has rich data advantages and tends to be more accurate. The reason is that people use search engine to inquire information to satisfy their requirements (Wilson, 2000), and this behavior is the response to the environment changes (Moe and Fader, 2004). Internet users' keywords searches on search engine are therefore; inevitably have close and natural links to external conditions and the environment changes (Sun and Lv, 2011). That is why lots of scholars use web-searching data to research and forecast economic issues.

Forecasting based on web-searching was first used on health issues. Scientists discovered that the search index and the incidence of influenza have a long-term stable relationship. They used the search index to predict the incidence of influenza and influenza mortality in advance (Ginsberg et al., 2009). From the index study based on Google keywords, they found that the partial keywords web-searching index have a long-term positive correlation relationship. In other studies, scientists also built a prediction model and predicted the flu outbreak trend with two weeks notice. Scholars gradually discovered the results of binding the traditional forecasting model and keywords index were better than the original traditional forecasting model's results.

In the macro-economic studies, researchers found that monitoring and early-warning based on web-searching data shows better prediction result, such as the monitoring of cost performance index (CPI), gross domestic product (GDP), unemployment rate, private consumption index and consumer confidence index. Many research studies shows that the precision of a predicted model integrated with web-searching index is significantly higher than traditional models (Francesco et al., 2012; Askitas et al., 2009; Vosen and Schmidt, 2011; Nicolás et al., 2009; Kholodilin, Konstantin, 2010). Web-searching index is also commonly used in the prediction of micro-market. Preis et al. (2013), found web-searching index has relevance to the quantified trading behaviors in stock markets. Besides, taking advantage of the web-searching index to grasp investors' preference and the characteristics of their behaviors, this study can effectively predict stock market performance. In the mean time, making full use of web-searching index to conduct and build the investing strategy in stock markets will reduce stock markets risks and gaining stable profits (Kristoufek, 2013). Via internet search index, we can not only predict economic issues but also forecast social problems, that is because the search index show us a behavioral data sets combined extensive which can help us to get a better understanding of human behavior, emotion and expectation (Preis et al., 2013).

Internet search engine is the most popular tool to catch information. It is also the channel and carrier for users to

obtain news. It has evolve into a central source of information which has been closely linked to people's day-to-day decisions, and of course the internet search data will reflects peoples' behavior and patterns which is hidden behind their searching action (Curme et al., 2014). This study integrate web-searching index into time series models to forecast the CNY short -term exchange rate. The study is expected to acquire better results than the traditional models.

### Prediction mechanism

Traditional time series model only rely on historical data to forecast. For instance, traditional ARIMA model and even the improved ARIMA models is the autoregressive moving average of univariate variables, under the assumption that market transactions have contained all the relevant information, including the expectations of the traders. Hence, someone believes that traditional time series models could make accurate prediction based on historical data. However, the reality is that price itself does not contain all the information especially in developing countries. The expectation of buyers and sellers can not reflect promptly in market prices. Under this situation, traditional time series models cannot reach a high precision, because it does not contains all the information such as traders' emotions and their expectations of exchange rates.

Combining traditional time series models with web-searching index would solve the problem caused by insufficient information. Web-based big data contains a lot of news including traders' behaviors and psychological information, which have been reflected in the current exchange rates but not contained in the prediction of traditional time series models. Therefore, the study compiles the network index as the proxy of web information, which could characterize the factors that influence exchange rates such as psychological expectations and traders' emotions. This study constructed a comprehensive web-searching index and integrated it into CAR model to make a short-time prediction of exchange rates. As is shown in Figure 1, the right side of the figure illustrates the logic of traditional time series model, and the left side presents the integration of web-searching data. The study also extracts the key words by NLPIR system and TextRank algorithm, and constructs a comprehensive web-searching index to give short-term prediction of exchange rates.

Figure 2 presents the influential process of web-searching index on exchange rates. Literature shows that the fluctuations of exchange rates are mainly influenced by three factors: macro-economic factors like monetary policy and inflation, political factors like trade disputes and political games, and micro-economic factors like hot money flows. The fluctuations of exchange rates caused

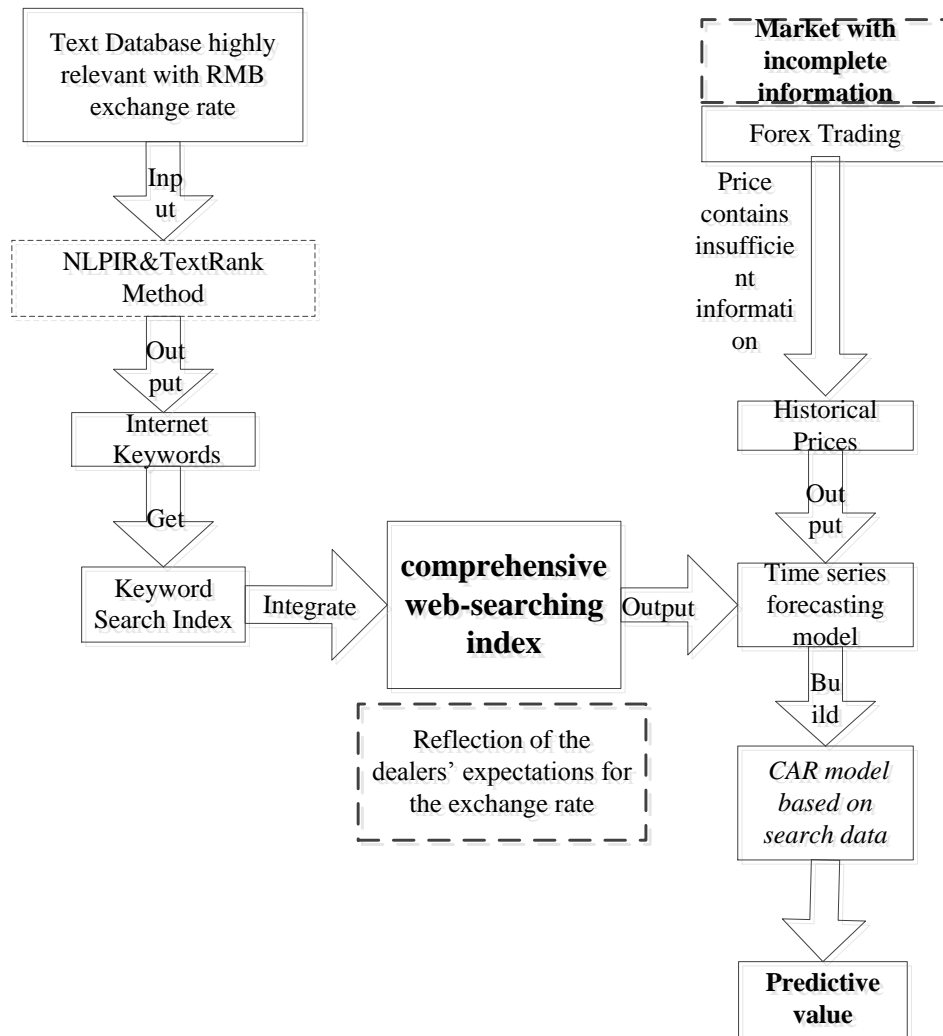


Figure 1. Mechanism of prediction.

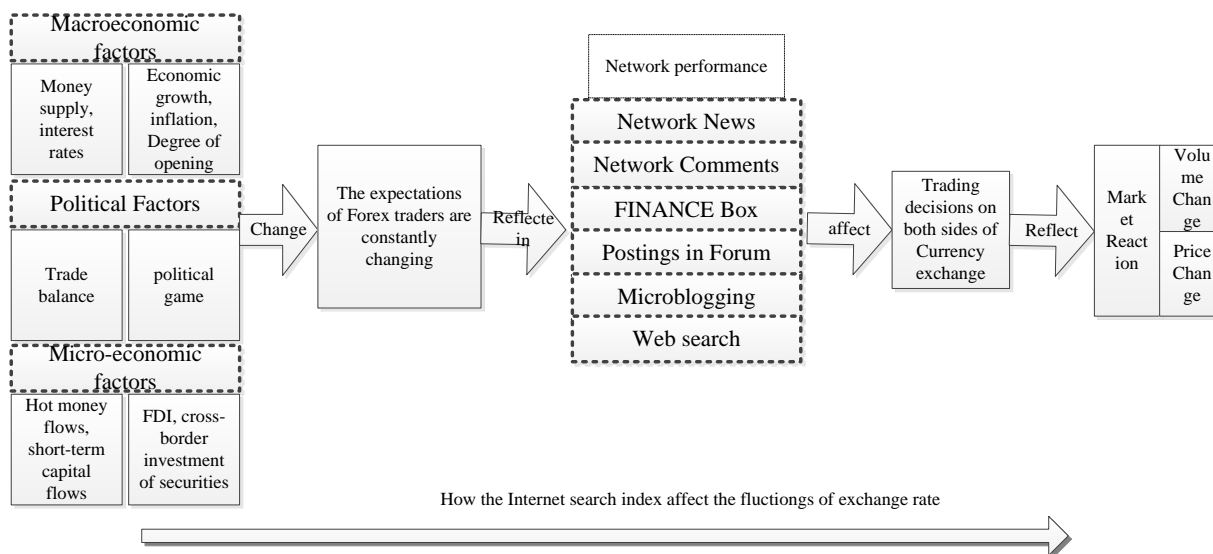
by these factors would change the two trading parties' psychological expectations and reflect in their network behaviors. The expectations can be shown in the relevant changes of expectations and emotions through the web-searching index, we are able to predict the next step of traders and make pre-judgment of the fluctuations of exchange rates, and therefore make accurate prediction. For this reason, this study hope to integrate web-searching index that contains large amount of information with the traditional time series model, will, unlike the old model that used to only use a single argument (historical prices), become more diverse in its accept of arguments. This new type of multivariate model, in other words, the traditional time series model integrated with web-searching index, would be multivariate time series model. There are many methods to build a multivariate time series models, example, continuous-time autoregressive-moving-average model (CARMA) and recursive-identification of multivariable CAR model. But considering

comments written by economists, the news and finance columns related to the fluctuations of exchange rates, weibo messages and forums. When we catch the modeling CARMA is very complex and ineffective, in this experiment, the study chose to create the CAR model (Bollerslev 1986). The study model is like this:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_n y_{t-n} + b_0 \text{Index}_t + b_1 \text{Index}_{t-1} + b_2 \text{Index}_{t-2} + \dots + b_n \text{Index}_{t-n} + \epsilon_t$$

**The construction of comprehensive web-searching index**

There are five steps to construct the comprehensive web-searching index: the construction of corpus, which contains huge amount of text information related with exchange rates; segmentation of Chinese words;



**Figure 2.** Linkage between web-search index and exchange rate fluctuations.

keywords screening based on TextRank algorithm; obtaining Keywords web-searching index; and the construction of comprehensive web-searching index.

### **The construction of corpus**

This study research on the CNY to US dollars exchange rates. Large amount of text materials from financial reports, financial columns, forums and weibo of financial commentators were collected. The time span is from 2011 to 2014. Tens of thousands words in the corpus after filtering and duplicate removal were gotten.

### **Segmentation of Chinese words**

Then, the study did semantic segmentation of the corpus by employing NLPIC Chinese words segmented system. Based on Chinese semantics and tones, the study split large amounts of text information into the separate words using computer programs, which is named Chinese segmentation. For instance, if we do semantic segmentation for a sentence "The exchange rate of the CNY to US dollars has appreciated", we would get these words: "CNY", "US dollars" and "appreciated".

Chinese segmentation system NLPIC (former ICTCLAS) was designed by Zhang Huaping, a doctor in the department of computer science and technology at Chinese Academy of Sciences, in 2000. After 14 years developments and improvements, this system is employed by more than 300 thousands of people. It was also awarded the first prize of Qian Weichang Chinese Information Processing Science and Technology Award in 2010 and the first prize of Sighan segmentation

competition in 2003. The 2014 edition NLPIC system could recognize English word prototypes automatically, tag the part of speech, and name the entities and keywords. Besides, based on Chinese lexical analysis, the system can accomplish complete and semantic analyses of documents, including names, places, institution names, article authors, published medias, keywords and abstracts. Therefore, this study made use of open-source NLPIC Chinese segmentation system to do word splitting work.<sup>1</sup>

### **Keywords screening based on TextRank algorithm**

Mihalcea (2004), proposed TextRank<sup>2</sup> model in 2004. They firstly applied PageRank algorithm to the keywords screening work, and named the model TextRank. Using TextRank algorithm to screen, keywords has many advantages. Firstly, TextRank algorithm does not need numerous corpus to train, which saves time and costs. Secondly, TextRank algorithm chooses un-supervised way of learning, making it more applicable to any contents, subjects and any length of text. Thirdly, TextRank has fast Convergence rate. If you use TextRank algorithm to calculate matrix, it has faster speed and better results (Liang, 2010) .

<sup>1</sup> The introduction of the NLPIC Chinese segmentation system is from "NLPIC natural language processing and information retrieval platform", <http://ictclas.nlpic.org/>

<sup>2</sup> The idea of TextRank model is from PageRank. It thinks that any text is a complexity network combined by groups of words. The central node of the network is the keywords we are looking for. Those nodes adjacent to the central node, although they have very low frequencies of appearance, they may still be the candidates of keyword.

By segmentation, the study decomposes the corpus and got preliminary vocabularies. Then we screen the keywords, which are closely related to exchange rate fluctuations, from a large number of vocabularies by TextRank algorithm. Finally, the study got 281 keywords, including “CNY exchange rates”, “investments”, “capital flows”, “trading” and “US dollars”. After removing the keywords without web-searching index, the study got 88 effective keywords. These keywords not only contain the words which are closely related with CNY to US dollars exchange rates such as “the CNY”, “US dollars” and “exchange rates”, but also contain some unrelated words like “markets”, “programs”, “space”, “stocks”, “data” and “processing”. But these seemingly unrelated vocabularies are screened by statistical analysis, and may have some hidden relationship with exchange rates. These words were kept in this research.

### ***Obtaining the web-searching index of keywords***

The study obtained the web-searching index of keywords through search engine. In the past researches, scholars usually use Google search engine. Chinese scholars use Baidu search engine or Micro index. However, these search engines have their own disadvantages. Google search engine is sometimes blocked in China, and few people use it. Hence, there are significant differences between the Google search index with economic reality. Baidu search engine is not open-source, and the study cannot get the precise daily web-searching index for model. Sina weibo provides micro-index responding to weibo’s contents and keywords, but it can only reflect the weibo users’ concern degree about exchange rate rather than the whole internet concern degree. In the mean time, there are few weibo users who analyze exchange rate fluctuations professionally. Therefore, the study selected Haosou search engine to obtain the web-searching index of each keywords. Haosou search engine provides open-source search index and it is in a leading position of the searching field. In order to acquire data as much as possible, the study downloaded the web-searching index for 88 keywords from 1 November, 2013, and 66 thousands data in total were obtained.

### ***The construction of comprehensive web-searching index***

The study calculated the Pearson correlated coefficient of all the keywords web-searching index and CNY exchange rate, and pick up the highly correlated keywords. The study also gave sufficient consideration on the setting of threshold. If the threshold is too high, the study would lose the keywords that are very important to exchange rates. If the threshold is very low, the study would get multiple noisy keywords. Ginsberg is the earliest researcher on web-searching index. In his paper in 2009

he chose the threshold as follows: Firstly, he added the key words into the comprehensive web-searching index one by one, and during that time he observed the correlation coefficient between comprehensive web-searching index and target benchmark. When the correlation coefficient cannot enhance after adding to a new key word, he believe that at that number of keywords the coefficient comes to its threshold. Scholars usually apply this method to determine the threshold. This study also follows this rule, and when the threshold rise from 0.1 to 0.4, it was found that the correlation between the search index sequence and the target sequence appeared as a substantial growth if it reached 0.84. When the threshold was set for 0.5 the correlation coefficient was 0.86, the correlation coefficient declined to 0.85 when threshold was setting for 0.6, which means 0.5 is the best threshold value. But due to the correlation of large number of keywords concentrated in the range 0.4 to 0.6, and if the study insist to increase the threshold from 0.4 to 0.5, it will lost a large number of key words so as to reduce the prediction accuracy.

So in this study, the threshold of Pearson correlated coefficient is 0.4, and therefore 40 keywords in total were gotten. Part of them is listed in Table 1. In the experiment, the web-searching index of keywords presents multiple outliers and violent volatility. Therefore, the study combine multiple keywords searching indexes into a robust one by synthesis method, in order to exhibit a clear trend of web big data and include the traders’ behavior into the prediction model. The weighted average of keywords web was calculated.

In order to calculate the comprehensive web-searching index, weight of each keyword first should be figured out. In the related research, scholars usually apply weighted average or arithmetic average method to construct the index. They choose the method mainly based on the performance on accuracy in the prediction model. In this research, it was found that weighted average method weighted by the correlation coefficient of each network search keywords can produce better predictions. Therefore, this study choose this method, and take the correlation coefficient of each network search keywords as its weight. The formula is as follows:

$$\text{Index} = a_1 \text{keywords}_1 + a_2 \text{keywords}_2 + \dots + a_{40} \text{keywords}_{40}$$

Where  $a_n$  means the weight of each keywords in the web-searching index.

### **Data and model**

#### ***Data and stationary test***

In this study, daily CNY to US dollars central parity rate was selected as the proxy of exchange rate. Data in non-trading days as the arithmetic average price of the

**Table 1.** The correlation coefficient of part network search keywords.

S/N	Key word	The correlation coefficient	No.	Keyword	The correlation coefficient
1	United States	0.808	11	Exchange rate	0.602
2	Center	0.76	12	Purchase	0.601
3	Global	0.75	13	Drop out	0.598
4	Dollar	0.707	14	Status	0.586
5	International	0.68	15	Exit	0.55
6	Import	0.67	16	Index	0.529
7	Affect	0.632	17	Accounts	0.528
8	EUR	0.623	18	Market	0.526
9	Renminbi	0.617	19	Aisle	0.515
10	Analysis	0.607	20	Effect	0.505

**Table 2.** ADF test of variable.

Variable	ADF test (P value)	MacKinnon Threshold			ADF test results
		1%	5%	10%	
y	0.0944	-2.568151	-1.941260	-1.616406	Non-stationary
index	0.7298	-2.56815	-1.941260	-1.616406	Non-stationary
Log (y)	8.483e-38	-2.568151	-1.941260	-1.616406	Smooth
Log (index)	5.255e-16	-2.568151	-1.941260	-1.616406	Smooth

adjacent two trading days was set. Since the 360 platform started to provide the web-searching index from 11 January, 2013, the sample interval is from 11 January, 2013 to 21 January, 2015. Among them, data from January 11st, 2013 to January 14th, 2015 were used as samples to construct models, and forecast the exchange rate for next 7 days. The rest of the data is employed as testing data to calculate deviations of models.

The traditional time series model requires the input data to be stable. Therefore, the unit root test for CNY exchange rate and the comprehensive web-searching index were used. ADF test is employed and the result is shown in Table 2. Both CNY exchange rate and comprehensive web-searching index are non-stationary. The ADF test of logarithm of two variables reject the null hypothesis at 1% confidence level, which illustrates that the two variables are first-order single integer series. By the ADF test, it was found out find that the logarithm of variables  $y_t$  and  $index_t$  are stable. Hence,  $\log_{y_t}$  and  $\log_{index_t}$  to construct models instead of the variables  $y_t$  and  $index_t$  will be used.

In order to check if there is a long-term stable relationship between the variables, the study also conducts the co-integration test. There are two kinds of methods for co-integration test. One method is based on the co-integration test of regression coefficients, and the other one is based on the co-integration test of regression residuals. The study applied the Engle-Granger two-step method to test if there is a long-term co-integration relationship between the CNY exchange rate and web-searching index.

After the regression, a unit root test for the regression residuals was conducted. If the independent and dependent variables have a co-integration relationship, the residuals series should be stable and vice versa. The results of co-integration test illustrates that the CNY exchange rate has a long-term co-integration relationship with the web-searching index, which shows that the web-searching index has long-term influence on the fluctuations of exchange rates.

### ***CAR model integrated with comprehensive web-searching index***

As the variables  $y_t$  and  $index_t$  are not stable, the study made use of the logarithm to construct models, labeled as  $\log_{y_t}$  and  $\log_{index_t}$ .

There are two approaches to construct the CAR model: one is recursive least square estimation, the other one is the least square equation of constructing a combined regression model. In this study, the second method to construct the model was used. Firstly, by calculating the Pearson correlated coefficient between web-searching index in different lag-order and exchange rate series (Table 3a and b), the study found out find that Index series is in three orders leading the exchange rate series. Therefore, the study set the leading-order of web-searching index  $\log_{index_t}$  as third order. Meanwhile, from the graph of autocorrelation and partial autocorrelation, thye study found out that the partial correlation graph of explanatory variable  $\log_{y_t}$  shows

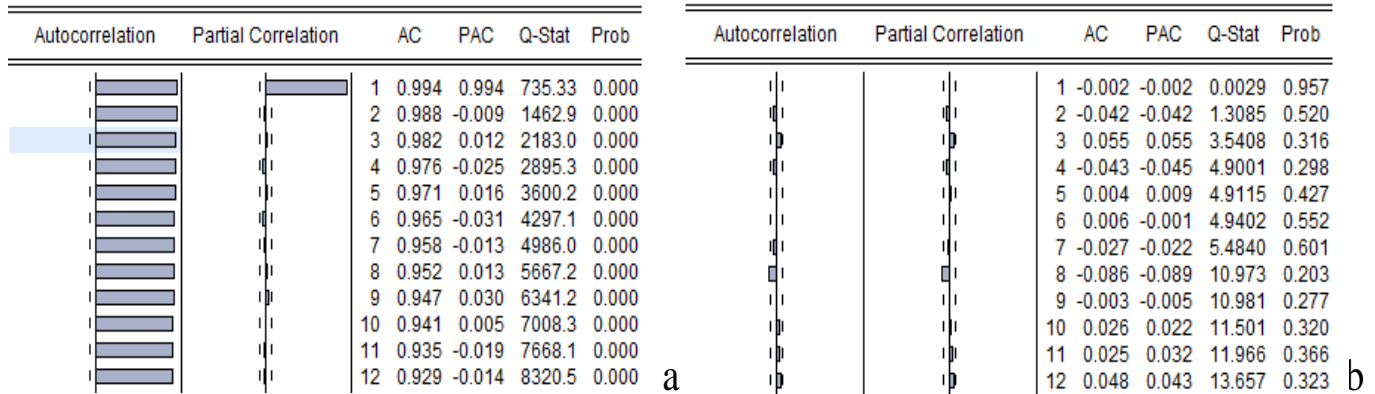


Figure 3. Graph of autocorrelation and partial autocorrelation.

Table 3. Modeling analysis of CAR model.

Variable	Coefficient	P-value	
C	0.019491*	0.0038	
Log_index(-3)	-4.65E-05*	0.0577	
Log_Y <sub>t</sub> (-1)	0.989323***	0.0000	
AR(8)	-0.087177**	0.0197	
R-squared	0.995247	Adjusted R-squared	0.995227
F-statistic	50184.73	Prob (F-statistic)	0.000000
Akaike info criterion	-12.25369	Schwarz criterion	-12.22833
Durbin-Watson stat	2.016464	-	-

Note: \*, \*\* and \*\*\* are representing the significance level of 10, 5 and 1%.

that it has first-order autocorrelation. Therefore, the study included  $\log_{y_{t-1}}$  in the model (Figure 3 left). The autocorrelation graph of residuals was also analyzed as well, and the result shows that the residual has eight-order autocorrelation (Figure 3b). The CAR model is as follows:

$$\log_{y_t} = 0.019491 - 0.0000465 * \log_{index_{t-3}} + 0.989323 * \log_{y_{t-1}} + u_t$$

$$u_t = -0.087177 * u_{t-8} + \varepsilon_t$$

Among them,  $\log_{y_t}$  is the logarithm of daily CNY to US dollars central parity rate,  $\log_{index_{t-3}}$  is the third-order lag web-searching index series,  $u_t$  is the auto-regressive item. The model with EVIEWS was estimated, and the result is shown in Table 3.

In the regression result, it can see that  $R^2$  is 0.995, which means our model is fitted well to the fluctuations of exchange rates. DW statistics is equal to 2, denoting that there is no serial correlation. The coefficient of 3 lagged web-searching index is significantly negative related with CNY exchange rate, illustrating that higher web-searching index is related with exchange rate appreciation. Figure 4 exhibits this model's fitting effect and residual errors.

The study further observe the regressive residual plot and its autocorrelation coefficient, partial autocorrelation coefficient and the Q-stat statistics, which show that each order of the autocorrelation coefficient and partial autocorrelation coefficient is not significantly different from zero. In other words, there is no sequence autocorrelation in residual series.

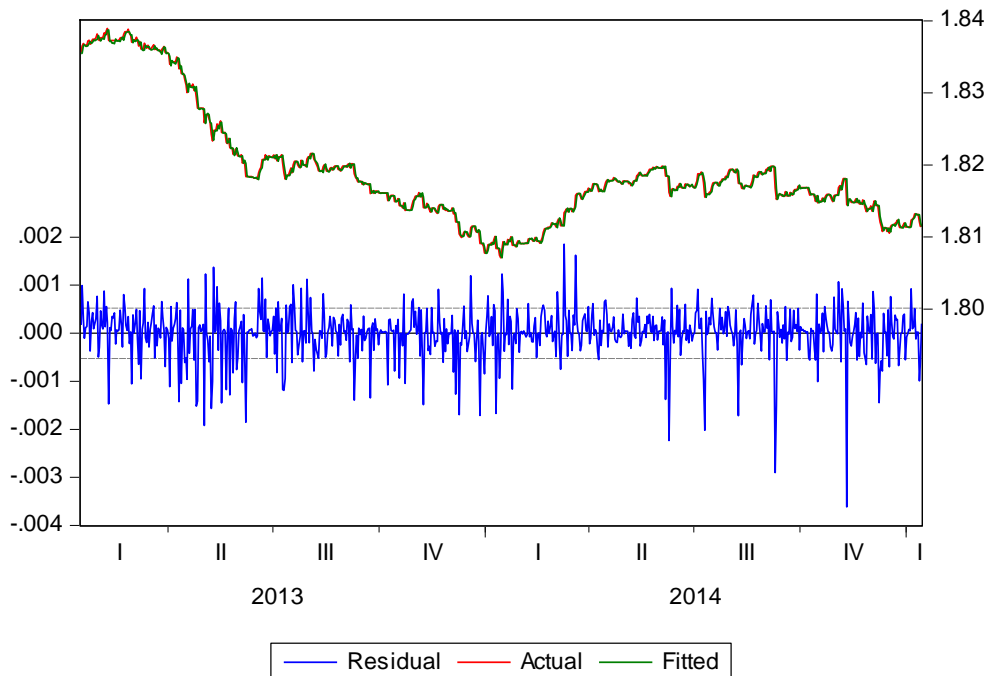
## DISCUSSION

### Comparison of different models under normal circumstances

As is mentioned previously, the study made use of the data from 11 January, 2013 to 13 January, 2015 as training data, and also data from 14 January, 2015 to 21 January, 2015 were also used as testing data. The study compared the results of CAR model integrated with web-searching index with other models, including AR model, ARIMA model and GARCH model. The comparison of results is listed in Table 4.

In order to measure the prediction precisions accurately and objectively, the study apply the error indicators,





**Figure 4** Fitting renderings of Search index and exchange

**Table 4.** Different time series model.

Model	CAR model based on search data	ARIMA model	AR model	GARCH model
C	0.019491 <sup>*</sup> (2.899744)	1.813961 <sup>***</sup> (401.4886)	1.812691 <sup>***</sup> (362.0053)	1.811655 <sup>***</sup> (237.8172)
LOG_INDEX (-3)	-4.65E-05 <sup>*</sup> (-1.900932)	-	-	-
LOG_Y <sub>t</sub> (-1)	0.989323 <sup>***</sup> (269.429)	-	-	-
AR (1)	-	0.993455 <sup>***</sup> (312.6730)	0.995032 <sup>***</sup> (403.2602)	0.996202 <sup>***</sup> (411.6845)
AR (8)	-0.087177 <sup>**</sup> (-2.336841)	-	-	-
MA (2)	-	0.237488 <sup>***</sup> (931.5233)	-	-
C	-	-	-	2.34E-08 <sup>*</sup> (1.698615)
RESID (-1) <sup>2</sup>	-	-	-	0.014272 <sup>*</sup> (0.0952)
GARCH (-1)	-	-	-	0.902016 <sup>***</sup> (16.16296)
R-squared	0.995247	0.995150	0.995482	0.995481
Log likelihood	4433.708	4510.215	4536.465	4538.891
AIC	-12.25369	-12.18166	-12.25531	-12.25376
SC	-12.22833	-12.16299	-12.24286	-12.22263
DW	2.016464	2.029912	1.998579	2.018485

Note: \*, \*\* and \*\*\* are representing the significance level of 10, 5 and 1%.

searching index and other traditional time series models to compare the differences. As it is shown in Table 6, the new CAR model integrated with web-searching index has the smallest absolute and relative error compared with other forecasting models. As the absolute error in total, integrated CAR model illustrates significant advantage over other models. As to the mean absolute percentage error (MAPE) indicator, other models' MAPE are higher than the integrated CAR model.

#### The comparison of different models under sudden fluctuation of exchange rate

In order to verify the accuracy and robustness of the CAR mode in the prediction of exchange rate fluctuations, another seven days prediction from 8 August, 2015 to 15 August, 2015 was done. It is because during this period the RMB exchange rate against the dollar appeared, and sustained depreciation for several days. Therefore, the

**Table 5.** Forecasting result.

Date	The actual value	CAR model based on search data		ARIMA model		AR model		GARCH model	
		Predictive value	Relative error (%)	Predictive value	Relative error (%)	Predictive value	Relative error (%)	Predictive value	Relative error (%)
1/14/2015	1.811644	1.811492	0.0084	1.811256	0.02	1.81282	0.06	1.811451	0.01
1/15/2015	1.811448	1.81165	0.0112	1.81112	0.02	1.812817	0.08	1.811421	0
1/16/2015	1.811366	1.811426	0.0033	1.811134	0.01	1.812814	0.08	1.811391	0
1/17/2015	1.811366	1.811399	0.0018	1.811148	0.01	1.81281	0.08	1.811361	0
1/18/2015	1.811366	1.811392	0.0014	1.811162	0.01	1.812807	0.08	1.811331	0
1/19/2015	1.812052	1.811392	0.0364	1.811176	0.05	1.812804	0.04	1.811302	0.04
1/20/2015	1.811987	1.812158	0.0094	1.81119	0.04	1.8128	0.04	1.811272	0.04
1/21/2015	1.812673	1.812057	0.034	1.811203	0.08	1.812797	0.01	1.811242	0.08

Note: Actual and predicted exchange rate listed in the table are the values being logarithmic.

study can test the robustness of the CAR model by making a prediction for this time. Figure 5 exhibits the trend of search index and prediction value of CAR model under sudden fluctuation of exchange rate. The exchange rate of RMB against the U.S. dollar began to appear substantial devaluation in 10 August, 2015. The web-searching index which is ahead of the exchange rate changes in 2 to 3 days significantly was seen, and the prediction effect of CAR model is very accurate. The study further compared the results of CAR model integrated with web-searching index with other models under the sudden fluctuation of exchange rate, including AR model, ARIMA model and GARCH model (Table 7). The CAR model can make a very accurate prediction for the exchange rate during the sudden fluctuations, and the major error indicators, which are commonly used to measure precision, are much lower than other models. Therefore, the study verify the accuracy and robustness of the CAR model under both the normal circumstances and the sudden fluctuation situations.

## CONCLUSIONS

The traditional time series model, which is commonly used in the prediction, ignore a fact that historical prices cannot reflect all the influential factors of change rates, and therefore, their prediction is not accurate. In order to reduce the prediction errors caused by insufficient information, the study construct a comprehensive web-searching index to forecast the short-term fluctuations of CNY to US dollars exchange rates. With the CAR model integrated with web-searching index, the study include large amount of traders' behaviors and expectations in the prediction.

It is expected to further reduce errors in the short-term exchange rate prediction.

In the empirical research, this study employ the CAR integrated with web-searching index to predict the short-term CNY exchange rate, and compare the prediction result with traditional time series models, including AR, ARIMA and GARCH model. The MAPE) of the new CAR model is

1.32%, while the MAPE for ARIMA, AR, GARCH model is 3.11, 1.92, 2.19% respectively, which indicates that the prediction accuracy of the new CAR model is significantly higher than other models. It also exhibits advantage in other prediction indicators. The study also mase some prediction for sudden fluctuation situation for exchange rates (from 8 August, 2015 to 15 August, 2015), and conclude that the accuracy of the CAR model integrated with web-searching index is still higher than other models in sudden fluctuation situations. It means the prediction results of the CAR are robust.

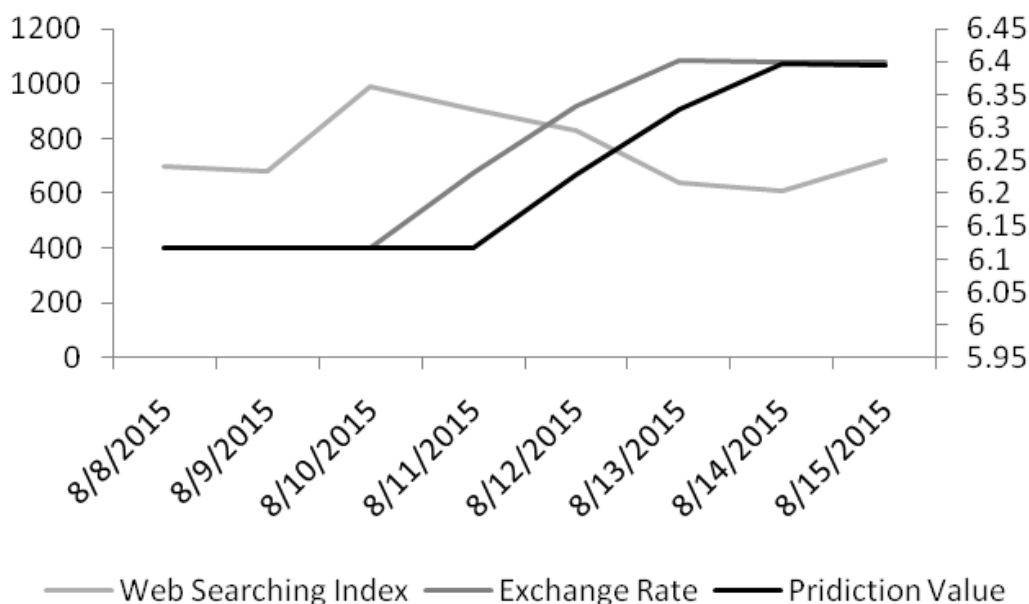
Although the CAR model integrated with web-searching index can overcome the problem of insufficient information, it still has drawbacks. For example, there are a few explanatory variables in the model, which makes it difficult to explain the mechanism of prediction. Therefore, each prediction model has its own drawback.

The study also tries to composite these models together into an integrated one, and further enhances the forecasting precision by synthesis

**Table 6.** Comparison of prediction accuracy between different models.

Model	Total absolute error	MAPE (%)	MAE	Thiel
CAR model integrated with web-searching index	0.001920	1.32	0.000240	0.0000932
ARIMA	0.004513	3.11	0.000564	0.000194
AR	0.008567	1.92	0.000348	0.000137
GARCH	0.003181	2.19	0.000398	0.000173

Note: The absolute error = | predicted values - Actual observations |; relative error = | predicted values - Actual observations | / actual observed value \* 100%; MAPE (mean absolute percentage error) = sum [| predictive value - the actual value | \* 100 / actual value] / sample size; MAE (mean absolute error) = sum [| predictive value - actual value |] / sample size; Thiel (not equal coefficient) is between (0,1) of the a coefficients, if the coefficient is equal to 1 indicates that the model predicts poor; when the coefficient is 0, that the model predicted and actual values of a complete match a good predictive power.

**Figure 5.** Search index and prediction value under sudden fluctuation of exchange rate.

different prediction models.

Specifically, this study would explore the new methods to combine models, such as arithmetic averages, weighted averages, and principal component analysis. The study, also compare their difference so as to obtain more accurate combination-forecasting model.

### Conflict of interests

The authors have not declared any conflict of interests.

### Acknowledgements

This research is supported by National Natural Science Foundation of China (NSFC Grant Numbers: 71273257 and 71532013).

### REFERENCES

- Askitas N, Zimmermann KF (2009). Google Econometrics and Unemployment Forecasting. *Soc. Sci. Electro. Publish* 55:107-120.
- Bollerslev T (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econ.* 31(3):307-327.
- Box GEP, Jenkins GM, Reinsel GC (1994). *Time Series Analysis: Forecasting and Control (Revised Edition)*. *J. Market. Res.* 14 p.
- China Internet Network Information Center (2015). The 36rd Statistical Report on Internet Development in China. CNNIC. [http://www.cnnic.net.cn/hlwfzyj/hlwzxbg/hlwjtjbg/201507/t20150722\\_52624.htm](http://www.cnnic.net.cn/hlwfzyj/hlwzxbg/hlwjtjbg/201507/t20150722_52624.htm).
- Curme C, Preis T, Stanley HE (2014). Quantifying the semantics of search behavior before stock market moves. *Proc. Natl. Acad. Sci. USA* 111(32):11600-11605.
- Engle RF (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica. J. Econ. Soc.* 987:1007.
- Francesco D, Juri M (2012). The predictive power of Google searches in forecasting unemployment. *Temi Di Discussione*.
- Ginsberg J, Mohebbi MH, Patel RS (2009). Detecting influenza epidemics using search engine Query data. *Nature* 457:1012-1014.
- Kholodilin KA (2009). "Google Searches as a Means of Improving the

- Nowcasts of Key Macroeconomic Variables." Social Science Electronic Publishing.
- Kristoufek L (2013). Can Google Trends search queries contribute to risk diversification?. *Sci. Rep.* 3:2713.
- Liang W (2010). Chinese keyword extraction technology (Master's thesis), Shanghai Jiaotong University).
- Lyons RK (2001). New perspective on FX markets: Order-flow analysis. *Int. Finance.* pp. 303-320.
- Mihalcea R, Tarau P (2004). TextRank: Bringing Order into Texts. *Unt Scholarly Works.*
- Moe WW, Fader PS (2004). Dynamic conversion behavior at e-commerce sites. *Manage. Sci.* 50(3):326-335.
- Nicolás D, Penna H, Huang (2009). Constructing consumer Sentiment Index for US Using Internet Search Patterns. *Dept. Econ.* 2009-26.
- Preis T, Moat HS, Stanley HE (2013). Quantifying trading behavior in financial markets using Google Trends. *Sci. Rep.* 3 p.
- Sun Y, Lv B (2011). A Review of Researches on the Correlation between Internet Search and Economic Behavior. *Manage. Rev.* 7(7):72-77.
- Vosen S, Schmidt T (2011). Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends. *J. Forecast.* 30(6):565-578.
- Wilson T D (2000). Human information behavior. *Inf. Sci.* 3(2):49-56.