Finance Research Letters xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

The impact of Baidu Index sentiment on the volatility of China's stock markets

Jianchun Fang[a], Giray Gozgor[b], Chi-Keung Marco Lau[c], Zhou Lu[d,*]

[a] *Zhejiang University of Technology, Hangzhou, China & Chinese Academy of Social Sciences, Beijing, China*
[b] *Istanbul Medeniyet University, Istanbul, Turkey*
[c] *University of Huddersfield, Huddersfield, United Kingdom*
[d] *Tianjin University of Commerce, Tianjin, China*

ARTICLE INFO

Keywords:
investor sentiment
stock market volatility
GARCH models
Baidu-Index
China
search engine

ABSTRACT

This paper examines the relationship between investor sentiment and the volatility of China's stock markets. We use the data from Baidu, China's leading search engine, for information on investor sentiment. In two different Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, the baseline model and the Baidu-Index extended model, we forecast the return volatility of China's stock markets. We find that the Baidu-Index extended model performs better than the baseline model. This paper shows that the search volume of relevant key words from Baidu Index improves volatility forecasting of China's stock markets.

1. Introduction

Reliable information is valuable for stock market investors. Nowadays, search engines provide investors with an easy and low-cost means of information access (Dastgir et al., 2018). Hence, keywords search volume, which might reflect investor sentiment, can be considered as a potential indicator for stock market forecasting. This paper is aimed to examine the effects of investor sentiment on the volatility of China's stock market.

Caporin and Poli (2017) and Dimpfl and Jank (2016) examine the relationship between the daily queries and the trading volume and price volatility of a given stock and suggest that autoregressive models with search queries included are shown to perform better than traditional time series model. Ackert et al. (2016) and Mao et al. (2011) conclude that the Google search volume data provide a better prediction than the traditional survey-based investor sentiment data. Pyo (2017) finds a negatively correlation between stock market returns and search frequency. Bank et al. (2011) find that search volume is positively related to stock returns in the short run. Kim et al. (2018) find that Google search data can predict volatility of individual stocks. Aalborg et al. (2018) examine this prediction issue for Bitcoin, but conclude that Google searches do not predict Bitcoin volatility. Bijl et al. (2016) and Rochdi and Dietzel (2015) show that Google Trends search keywords are positively related to an increase in investment return and diversification of portfolio risk. Heiberger (2015) finds that public concern about negative corporation news serves as a warning signal for systematic financial risk. Ramos et al. (2017) find an inverse relationship between financial markets and Google's search volume in Brazil.

Most of the existing literature on the relationship between search volume and stock market is based on Google Trends and concentrated on developed countries. However, little has been done in using non-Google data to investigate the impact of investor sentiment on the stock markets of emerging economies. The main contribution of our paper is to use the keyword search data from

* Corresponding author.
E-mail addresses: fangjch@cass.org.cn (J. Fang), giray.gozgor@medeniyet.edu.tr (G. Gozgor), c.lau@hud.ac.uk (C.-K.M. Lau), luzhou59@tjcu.edu.cn, luzhou59@gmail.com (Z. Lu).

https://doi.org/10.1016/j.frl.2019.01.011
Received 16 August 2018; Received in revised form 17 December 2018; Accepted 26 January 2019
1544-6123/

Please cite this article as: Fang, j., Finance Research Letters, https://doi.org/10.1016/j.frl.2019.01.011

Baidu, China's leading search engine, to examine whether investor sentiment reflected in the Baidu Index data can be used to improve the prediction of the Chinese stock market returns. To the best of our knowledge, our article is the first study to use the data from Baidu Index in examining the relationship between investor sentiment and stock market volatility.

The remainder of the paper is structured as follows. Section 2 describes the data collection and the keyword selection. Section 3 discusses the methodology and conducts the estimation and forecasting procedures. Section 4 presents the empirical findings and discussion. Section 5 concludes.

2. Data and keywords selection

The data on closing prices of China's stock markets are all from the Wind Info. The following formula is used to calculate the daily returns of stock markets:

$$Return_{i,t} = \frac{(\text{Closing}price_{i,t} - \text{Closing}price_{i,t-1})}{\text{Closing}price_{i,t-1}} \tag{1}$$

In the above formula, $\text{Closing}price_{i,t}$ represents the closing price of the stock market i on the t day.

To construct indicators of investor sentiment from search engine data, we considered stock market relevant search keywords that might reflect investor sentiments and use keyword search volume data from Baidu Index instead of Google Trends.[1] The search volume time series data are obtained from the website (https://index.baidu.com).[2] According to the research by Baker and Wurgler (2006), Abugri (2008), Kasman et al. (2011), and Javed and Ullah (2014), stock market volatility tends to be driven by factors including interest rates, exchange rates, money supply, macroeconomic variables and investor sentiment. We divide these factors into two categories, the positive and the negative. In selecting keywords for study, we consider "Niushi (Bull Market)" and Fenhong (Dividend) as positive phrases and "Xiongshi (Bear Market)" and "IPO" as negative ones. We also examine other seemingly relevant phrases, such as Huilv (Exchange rate), Lilv (Interest rate), Fangjia (Housing prices), Tongzhanglv (Inflation rate), Shiyelv (Unemployment rate), etc. However, as none of them are shown to have significant effects we exclude them from our analysis.

We report the descriptive characteristics (Jarque-Bera statistics, kurtosis, and skewness) of the time-series in Table 1. SHANGHAI and SHENZHEN represent the closing prices of Shanghai Stock Exchange Composite Index (SSE Index) and ShenZhen Stock Exchange Composite Index (SZSE Index) respectively. CSI300 and CSI100 represent the closing prices of CSI300 Index and CSI100 Index respectively.[3] SHANGHAIRET, SHENZHENRET, CSI300RET, and CSI100RET represent the return rates of SHANGHAI, SHENZHEN, CSI300, and CSI100 respectively. LNBULL, LNBEAR, LNIPO, and LNDIV represent the natural logarithm of Baidu Index search volumes for "Niushi" (Bull Market), "Xiongshi" (Bear Market), IPO, and "Fenhong" (Dividend) respectively. Stock market returns and search data are shown to be not subject to the normal distribution as the null hypothesis is rejected for all tests. At the 1% significance level, the returns are negatively skewed, while the search volume data are positively skewed. The kurtosis is leptokurtic for of most variables.

We also use the unit root tests to examine the stationarity of the data. The results show there exists a unit root in the closing price time series. However, the existence of differentials in stock market returns takes care of this stationarity issue. There does not exist unit root in the search volume data. Table 2 presents the results of unit root test on stock market returns and Baidu Index search volumes. As can be seen from Table 2, all the time series in the sample period are stationary at the 1% significance level. As Baidu Index search volumes are based on standardized data, they are expected to be stationary when there is an intercept term in test equation.

We examine Granger causality between the search volumes of investor sentiment phrases and stock market returns (see Tables 3 and 4). Table 3 shows that the search volumes of "IPO" are all shown to drive stock market returns and the search volumes of "BULL Market" are mostly shown to be the cause of stock market returns. By contrast, no reverse causality is identified from stock market returns to the search volume of "IPO" (See Table 4).

3. Methodology, estimation, and forecasting

In time series financial models, disturbance variance is often found less stable. The conditional variance of the error term usually varies with time and depends on the magnitude of previous errors. As a result, prediction errors take place in groups and result in heteroscedasticity. To examine ARCH effect in residuals, we use two different tests, the Ljung-Box Q squared residual correlation diagram and the ARCH Lagrange Multiplier (LM). We use Ljung-Box Q to investigate the autocorrelation and partial correlation for the squared residuals of the mean equation. Then, we tested the null hypothesis that the data are independently distributed, that is to say the overall correlation coefficient is zero and any observed correlation is attributed to random sampling error only. We use 10 lags to calculate the results. Fig. 1 shows that the coefficients are non-zero significantly, and Q statistic is significant, indicating that a significant ARCH effect in the residuals of Eq. (2). The fluctuation in the squared residual (Fig. 2) shows significant time-varying and

[1] Since Google withdrew from China in 2010, Baidu has risen dramatically in market share and secured its leading position in China's search engine market.

[2] Although the Baidu Index does not provide data on query results, we managed to use a web crawler to obtain the information.

[3] CSI100 Index and CSI 300 are capitalization-weighted stock market indices designed to replicate the performance of selected top stocks traded in both Shanghai and Shenzhen Stock Exchange. CSI100 Index is a sub-index of CSI 300 Index.

Table 1
The descriptive statistics of alternative stock prices, returns and Baidu Index.

	Mean	Median	Max.	Min.	Std. Dev	Skewness	Kurtosis	Jarque-Bera
SHANGHAI	2872.01	2834.89	5410.86	2040.67	639.23	0.95	4.04	326.14***
SHENZHEN	1506.27	1290.80	3287.57	766.34	529.25	0.60	2.35	128.42***
CSI300	2974.68	2956.38	5353.75	2086.97	643.47	0.81	3.54	198.97***
CSI100	2789.34	2735.55	4773.52	1903.48	594.94	0.66	2.91	121.26***
SHANGHAIRET	0.02	0.06	5.76	−8.49	1.40	−0.88	9.27	2934.80***
SHENZHENRET	0.04	0.13	6.54	−8.26	1.67	−0.78	6.25	895.98***
CSI300RET	0.02	0.03	6.72	−8.75	1.49	−0.62	8.23	1994.93***
CSI100RET	0.03	0.01	6.93	−8.94	1.48	−0.43	8.60	2219.69***
LNBULL	6.01	5.965	8.40	5.22	0.49	1.74	6.22	1549.20***
LNBEAR	5.83	5.697	8.52	5.04	0.43	2.45	10.71	5753.41***
LNIPO	8.22	8.151	11.40	7.02	0.43	1.89	11.06	5474.85***
LNDIV	5.89	5.872	7.67	5.34	0.20	0.93	7.41	1580.35***

Note: ***indicates the 1% significance level.

Table 2
The unit root tests on alternative stock returns and Baidu Index.

Variable	ADF	Included observations: after adjustments	PP	Included observations: after adjustments
SHANGHAIRET	−38.80136***	1657	−38.77931***	1657
SHENZHENRET	−37.29437***	1657	−37.45583***	1657
CSI300RET	−39.35662***	1657	−39.34078***	1657
CSI100RET	−39.79164***	1657	−39.78265***	1657
LNBULL	−3.488482***	1654	−4.843818***	1658
LNBEAR	−4.237866***	1651	−7.655978***	1658
LNIPO	−9.659468***	1655	−15.83961***	1658
LNDIV	−5.722207***	1654	−19.49118***	1658

Note: ***indicates the 1% significance level.

Table 3
Causality test from Baidu Index sentiment to returns.

Variables			Lags	Obs	Coefficients
LNBULL	Does not Granger Cause	SHANGHAIRET	6	1652	2.21718**
LNBEAR	Does not Granger Cause	SHANGHAIRET	6	1652	0.92146
LNIPO	Does not Granger Cause	SHANGHAIRET	6	1652	2.42153**
LNDIV	Does not Granger Cause	SHANGHAIRET	6	1652	1.24358
LNBULL	Does not Granger Cause	SHENZHENRET	6	1652	0.76730
LNBEAR	Does not Granger Cause	SHENZHENRET	6	1652	0.82049
LNIPO	Does not Granger Cause	SHENZHENRET	6	1652	3.14785***
LNDIV	Does not Granger Cause	SHENZHENRET	6	1652	0.93197
LNBULL	Does not Granger Cause	CSI300RET	3	1655	4.47998***
LNBEAR	Does not Granger Cause	CSI300RET	3	1655	2.05016*
LNIPO	Does not Granger Cause	CSI300RET	3	1655	3.11598**
LNDIV	Does not Granger Cause	CSI300RET	3	1655	1.90694
LNBULL	Does not Granger Cause	CSI100RET	5	1653	3.40501***
LNBEAR	Does not Granger Cause	CSI100RET	5	1653	1.36832
LNIPO	Does not Granger Cause	CSI100RET	5	1653	2.41836**
LNDIV	Does not Granger Cause	CSI100RET	5	1653	1.48028

a clustering characteristic.

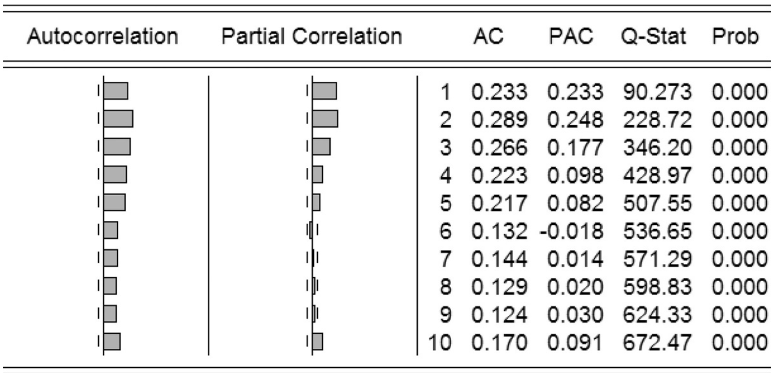
The ARCH Lagrange Multiplier LM test is calculated by an auxiliary test regression. To test the null hypothesis that there is no ARCH effect in the residual series for up to the p-order, the following regression is used:

$$\hat{u}_t^2 = \beta_0 + (\sum_{s=1}^{p} \beta_s u_{t-s}^2) + \varepsilon_t \tag{2}$$

$\hat{u}_t$ is the residual. Eq. (4) represents the regression of squared residual $\hat{u}_t^2$ over a constant and the sum of residual squared up to pth order lag. The test regressions include F-test and $T \times R^2$, where T is the number of samples for assistance regression. The LM test usually follows $\chi^2(\mathbf{p})$ distribution. We conduct a conditional heteroscedasticity ARCH LM test for Eq. (2) for lag order $p = 2$. As Table 5 shows, the probability value is zero, which rejects the null hypothesis indicating the existence of an ARCH effect in the residual sequence of Eq. (2). Hence, the Generalized ARCH (GARCH) model is appropriate to use.

Table 4
Causality test from returns to Baidu Index sentiment.

Variables			Lags	Obs	Coefficients
SHANGHAIRET	Does not Granger Cause	LNBULL	6	1652	2.67291**
SHANGHAIRET	Does not Granger Cause	LNBEAR	6	1652	7.76930***
SHANGHAIRET	Does not Granger Cause	LNIPO	6	1652	0.50340
SHANGHAIRET	Does not Granger Cause	LNDIV	6	1652	1.65889
SHENZHENRET	Does not Granger Cause	LNBULL	6	1652	3.76290***
SHENZHENRET	Does not Granger Cause	LNBEAR	6	1652	8.61674***
SHENZHENRET	Does not Granger Cause	LNIPO	6	1652	0.81165
SHENZHENRET	Does not Granger Cause	LNDIV	6	1652	2.21208**
CSI300RET	Does not Granger Cause	LNBULL	3	1655	2.79044**
CSI300RET	Does not Granger Cause	LNBEAR	3	1655	8.92551***
CSI300RET	Does not Granger Cause	LNIPO	3	1655	0.40602
CSI300RET	Does not Granger Cause	LNDIV	3	1655	1.35736
CSI100RET	Does not Granger Cause	LNBULL	5	1653	2.43287**
CSI100RET	Does not Granger Cause	LNBEAR	5	1653	5.80139***
CSI100RET	Does not Granger Cause	LNIPO	5	1653	0.51095
CSI100RET	Does not Granger Cause	LNDIV	5	1653	1.95085*



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.233	0.233	90.273	0.000
		2	0.289	0.248	228.72	0.000
		3	0.266	0.177	346.20	0.000
		4	0.223	0.098	428.97	0.000
		5	0.217	0.082	507.55	0.000
		6	0.132	-0.018	536.65	0.000
		7	0.144	0.014	571.29	0.000
		8	0.129	0.020	598.83	0.000
		9	0.124	0.030	624.33	0.000
		10	0.170	0.091	672.47	0.000

Fig. 1. Autocorrelation and partial correlation diagrams of the residual square of the mean equation.

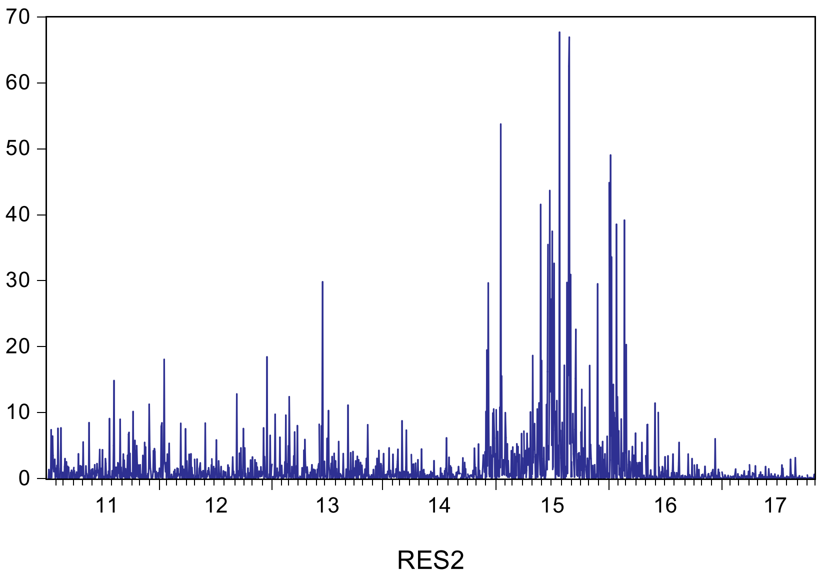


Fig. 2. Squared residuals.

Table 5
Results of the Lagrange multiplier (LM) Test.

Heteroskedasticity Test: ARCH			
F-statistic	104.6688	Prob. F(2,1650)	0.0000
Obs*R-squared	186.1066	Prob. Chi-Square(2)	0.0000

The general form of a GARCH (p, q) model can be written as follows (Bollerslev, 1986):

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 \tag{3}$$

Where $\omega > 0$, $\beta_j \geq 0$, $j = 1,\ldots, q$, $\alpha_i \geq 0$, $i = 1, \ldots, p$.

These model assumptions are quite common in the financial literature. Traders tend to predict variance of current period from the constant variance, variances of previous periods, and a weighted average of interest rates of previous periods. If changes in the rate of return are substantial, traders make upward adjustment on the estimates of next-period variance. This assumption on variance implies a clustering in the fluctuations of return on financial assets. To determine the values of p and q, we estimate different GARCH (p, q) models and consider the Akaike Information Criterion (AIC) value of the variance equation. The results show that GARCH (0,2) is more appropriate. Therefore, the variance in Eq. (5) can be simplified as the following model (Dyhrberg, 2016):

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \alpha_1 u_{t-2}^2 \tag{4}$$

This baseline model will be extended with a term of Baidu Index.

To determine the optimal lag term for the Baidu-Index extended model, we estimate the hysteresis effects for up to 4 lags and find that a lag term of 1 has a significant effect on the variance at the significance level of 5%. This result implies that Baidu Index tends to reflect real-time investor sentiment as it is shown to have an immediate impact on the stock market return. Therefore, the extended model that contains Baidu Index can be written as follows:

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \alpha_1 u_{t-2}^2 + \lambda LnBaiduindex_{t-1} \tag{5}$$

Based on the above model, we take a one-step-forward approach to the prediction of the Shanghai Stock Exchange returns. Our sample period is January 4, 2011 through October 31, 2017. We first estimate the model with the data January 4, 2011 through December 30, 2011 and then use the results to predict the variance for January 4, 2012. By repeating this process, we make prediction of the variance for the entire sample period. We also assess out-of-sample estimates. This ensures the investigation of each observation period.

4. Empirical results and discussion

The prediction error value is defined to be the difference between the actual and predicted conditional variance. According to the forecasting results shown in Table 6 and Fig. 3, if the keyword is “Niushi”, then the stock market return for next-period would be positive. By contrast, if the keyword is “Xiongshi” and “IPO”, then stock market return would be negative. However, if the keyword is “Fenhong”, then stock market would be negative as well. This is opposite to common sense and might be related to the fact that most Chinese companies distribute little or no dividends.

Overall, the Baidu-Index extended model has a smaller prediction error than the basic model, indicating that incorporating search data into the model can improve forecasting performance. An interesting observation is that the prediction errors are significantly higher during the high volatility periods than other periods for both the basic model and the extended model. High volatility refers to the periods when the conditional variance exceeds the mean (October 2012 through May 2013 and mid-November 2014 to mid-May

Table 6
Estimation of the basic and extended models for Shanghai Stock Exchange returns.

	Model(1)	Model(2)	Model(3)	Model(4)
SHANGHAIRET(−2)	−0.043***	−0.063***	−0.061***	−0.065***
LNBULL(−1)		0.446***		0.572***
LNBEAR(−1)		−0.456***		−0.396***
LNIPO(−2)			−0.160**	−0.220***
LNDIV(−1)			−0.033	−0.393*
GARCH(−1)	1.858***	1.854***	1.846***	1.855
GARCH(−2)	−0.998***	−0.998***	−0.999***	−0.998
R-squared	0.002***	0.009	0.004	0.015
S.E.	1.402	1.397	1.401	1.394
AIC	3.512	3.508	3.510	3.504
DW	1.901	1.953	1.901	1.954
Observations	1656	1656	1656	1656

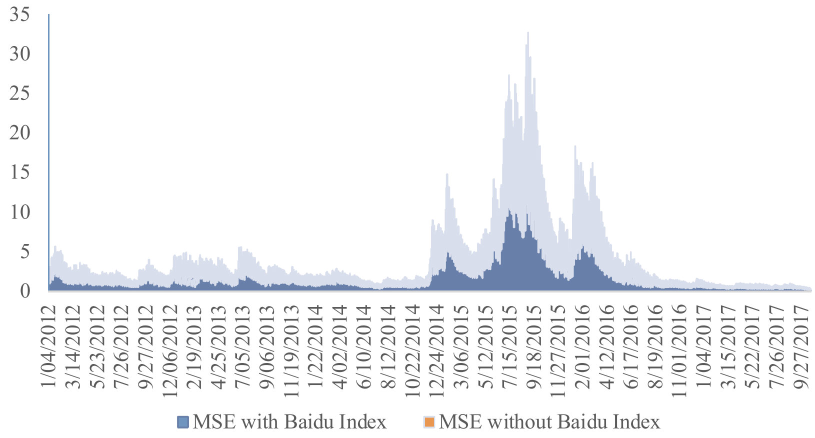


Fig. 3. MSE over the sample period for the baseline and the extended models.

2016) and low volatility the periods below the mean. However, during the high volatility regimes, the prediction error for the Baidu-Index extended model is significantly smaller than that of the basic model. This shows that prediction accuracy improvement by the extended model over the baseline model tends to be larger for high volatility periods than low. This is similar to the results of existing studies such as Horpestad et al. (2018). During market turbulence periods investors tend to seek additional market information to avoid risks and cause an increase in search volume.

To find out the extent to which the extended model improves the prediction relative to the baseline model, we calculate two measures of prediction accuracy, namely the mean squared error (MSE) and the Theil (*U1*) parameter. The Theil (*U1*) parameter takes a value between 0 and 1. A smaller value in the Theil parameter represents a higher level of prediction accuracy. Table 7 displays the values of MSE and U1 parameters for the baseline model and the extended model respectively. The Baidu-Index extended model is shown to be 0.55% smaller than the baseline model in MSE and 8.2% smaller in the U1 parameter.

We divide the sample into high and low volatility sub-periods. In addition, we divide the sample evenly into six sub-periods with each year as a separate sub-period.

Table 8 shows the extents to which the Baidu-Index extended model improves the prediction accuracy for different sub-periods. The Theil parameter reduced by 13.0% and MSE 1.1% for the period of low volatility. Improvements are more significant for the periods of high volatility such as period 3 (2014) and period 4 (2015).

To check the robustness of the improvement in prediction accuracy by the extended models, we also use alternative stock market indices including CSI 100, the CSI 300, and the Shenzhen Stock Exchange Index. The results of the robustness test are shown in Table 9–12. The findings turn out to be robust to different stock market indices in China.

Overall, Baidu Index search data can be used to help improve stock market forecasting. In addition, the Baidu-Index extended model performs better for high volatility periods than low volatility periods, which implies the Baidu Index as an early warning indicator for prediction of market trends.

5. Conclusion

This paper uses the search volume data from Baidu Index to predict the volatility of China's stock market returns. We considered relevant search keyword that reflect investor sentiments and include them into the baseline GARCH model to predict the stock market volatility. Similar to the findings from existing literature based on Google Trends data, the Baidu-Index extended model significantly improves the accuracy in forecasting stock market volatility. And this result is robust to different stock market indices, years within the sample, and volatility regimes.

Nowadays, investors take advantage of search engines such as Baidu to gather information relevant to investment decisions. We argue that the Baidu Index data can be used not only as an effective indicator of investment sentiment but also an effective tool to improve market trend forecasting. In particular, during periods of high volatility, investors use search engine for investment information more intensively as volatility in stock market returns stays high. Overall, using the data from Baidu Index improves the forecasting of stock market trends, especially during the periods of economic fluctuations.

Table 7
Prediction accuracy comparison.

Models	U1 Theil	U1 Theil Reduction	MSE	MSE reduction
Basic Model	0.950591		0.922881	
Baidu-Index Extended Model	0.805726	15.24%	0.894504	3.07%

Table 8
Improvement by the extended model in different sub-periods.

Sub-periods	Without Baidu Index	With Baidu Index	Reduction in U1 Theil (%)	Without Baidu Index	With Baidu index	Reduction in MSE (%)
High volatility	0.980321	0.853239	13.0	2.165596	2.142348	1.1
Low volatility	0.986618	0.886860	10.1	1.009030	1.001059	0.8
Period 1	0.993169	0.913937	8.0	1.095387	1.093707	0.2
Period 2	0.948396	0.931449	1.8	1.158009	1.155108	0.3
Period 3	0.859114	0.753864	12.3	1.086029	1.060084	2.4
Period 4	0.973676	0.864821	11.2	2.441405	2.413399	1.2
Period 5	0.978356	0.888999	9.1	1.453714	1.436527	1.2
Period 6	0.921867	0.858556	6.9	0.521951	0.516836	1.0

Table 9
Estimation of the basic and extended models for Shenzhen Stock Exchange returns.

	Model(1)	Model(2)	Model(3)	Model(4)
SHENZHENRET(−2)	−0.028121*	−0.049548***	−0.027305*	−0.045051***
LNBULL(−1)		0.417393***		0.523463***
LNBEAR(−1)		−0.432779***		−0.374832***
LNIPO(−2)			−0.185277**	−0.238011***
LNDIV(−1)			−0.084442	−0.367986
GARCH(−1)	−1.647052***	1.864899***	−0.084442	−0.057941
GARCH(−2)	−1.000030***	−1.008121***	0.875816	0.877881
R-squared	0.000556	0.004753	0.003029	0.009567
S.E.	1.672912	1.670405	1.671852	1.667370
AIC	3.866776	3.864237	3.871587	3.867483
DW	1.820791	1.855024	1.822667	1.855184
Observations	1657	1657	1657	1657

Table 10
Estimation of the basic and extended models for CSI300 stock returns.

	Model(1)	Model(2)	Model(3)	Model(4)
CSI300RET(−2)	−0.044265***	−0.064441***	−0.051133***	−0.066727***
LNBULL(−1)		0.415714***		0.553082***
LNBEAR(−1)		−0.383318***		−0.319553***
LNIPO(−2)			−0.156503**	−0.243514***
LNDIV(−1)			−0.047354	−0.420189*
GARCH(−1)	1.854797***	1.854613***	1.859263***	1.855069***
GARCH(−2)	−0.998011***	−0.997675***	−0.999086***	−0.998302***
R-squared	0.002329	0.007962	0.004714	0.014556
S.E.	1.490666	1.487352	1.489785	1.483298
AIC	3.636998	3.633814	3.636609	3.629263
DW	1.932371	1.972440	1.933865	1.975458
Observations	1656	1656	1656	1656

Table 11
Estimation of the basic and extended models for CSI100 stock returns.

	Model(1)	Model(2)	Model(3)	Model(4)
CSI100RET(−2)	−0.047183***	−0.063400***	−0.045586***	−0.064374***
LNBULL(−1)		0.392044***		0.511050***
LNBEAR(−1)		−0.293401***		−0.226546**
LNIPO(−2)			−0.118291	−0.183276**
LNDIV(−1)			−0.022561	−0.498126*
GARCH(−1)	1.829379***	1.829331***	1.882515***	1.859043***
GARCH(−2)	−0.999758***	−0.999751***	−0.999516***	−0.999111***
R-squared	0.002355	0.008029	0.003634	0.013919
S.E.	1.480402	1.477080	1.480348	1.473580
AIC	3.620082	3.616850	3.619273	3.613490
DW	1.954476	1.985713	1.955166	1.990155
Observations	1656	1656	1656	1656

Table 12
Robustness analysis in alternative stock market indices.

Stock market index	Without Baidu Index	With Baidu Index	Reduction in U1 Theil (%)	Without Baidu index	With Baidu index	Reduction in MSE (%)
Shenzhen	0.976323	0.918753	5.90	1.671941	1.665929	0.36
CSI300	0.977749	0.902315	7.72	1.491132	1.483508	0.51
CSI100	0.976904	0.905288	7.33	1.480917	1.473683	0.49

Acknowledgments

The authors would like to thank the anonymous reviewers and the editor for their helpful comments and suggestions. The authors acknowledge the financial supports from the Philosophy & Social Science Fund of Tianjin City, China (TJLJ17-005).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.frl.2019.01.011.

References

Aalborg, H.A., Molnár, P., de Vries, J.E., 2018. What can explain the price, volatility and trading volume of Bitcoin? Finance Res. Lett 1–11.
Abugri, B.A., 2008. Empirical relationship between macroeconomic volatility and stock returns: evidence from Latin American markets. Int. Rev. Financ. Anal. 17, 396–410.
Ackert, L.F., Jiang, L., Lee, H.S., Liu, J., 2016. Influential investors in online stock forums. Int. Rev. Financ. Anal. 45, 39–46.
Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. J. Finance 61 (4), 1645–1680.
Bank, M., Larch, M., Peter, G., 2011. Google search volume and its influence on liquidity and returns of German stocks. Financ. Market Portfolio Manage. 25 (3), 239–264.
Bijl, L., Kringhaug, G., Molnar, P., Sandvik, E., 2016. Google searches and stock returns. Int. Rev. Financ. Anal. 45, 150–156.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econ. 31 (3), 307–327.
Caporin, M., Poli, F., 2017. Building news measures from textual data and an application to volatility forecasting. Econometrics 5 (3) paper 35, 1-46.
Dastgir, S., Demir, E., Downing, G., Gozgor, G., Lau, C.K.M., 2018. The causal relationship between Bitcoin attention and Bitcoin returns: evidence from the Copula-based Granger causality test. Finance Res. Lett forthcoming, 1-5.
Dimpfl, T., Jank, S., 2016. Can internet search queries help to predict stock market volatility. Eur. Financ. Manage. 22 (2), 171–192.
Dyhrberg, A.H., 2016. Bitcoin, gold and the dollar – A GARCH volatility analysis. Finance Res. Lett. 16, 85–92.
Heiberger, R.H., 2015. Collective attention and stock prices: evidence from Google Trends data on Standard and Poor's 100. PloS One 10 (8) paper e0135311, 1-14.
Horpestad, J.B., Lyócsa, Š., Molnár, P., Olsen, T.B., 2018. Asymmetric volatility in equity markets around the world. North Am. J. Econ. Finance.
Javed, U.Md., Ullah, Naeem, 2014. Dividend Policy and Stock Price Volatility throughout Pakistan. Int. J. Current Res. 6 (issue 2), 5298–5303.
Kasman, S., Vardar, G., Tunç, G., 2011. The impact of interest rate and exchange rate volatility on banks' stock returns and volatility: evidence from Turkey. Econ. Model. 28 (3), 1328–1334.
Kim, N., Lučivjanská, K., Molnár, P., Villa, R., 2018. Google searches and stock market activity: evidence from Norway. Finance Res. Lett.
Mao, H., Counts, S., & Bollen, J. (2011). *Predicting financial markets: cComparing survey, news, Twitter and search engine data*, arXiv preprint arXiv:1112.1051.
Pyo, D-J., 2017. Can big data help predict financial market dynamics? Evidence from the Korean stock market. East Asian Econ. Rev. 21 (2), 147–165.
Ramos, H.P.R., Mendes, K.K., Perlin, S.M., 2017. The forecasting power of internet search queries in the Brazilian financial market. Revista de Administração Mackenzie 18 (2), 184–210.
Rochdi, K., Dietzel, M., 2015. Outperforming the benchmark: online information demand and REIT market performance. J. Prop. Invest. Finance 33 (2), 169–195.