

RESEARCH

Open Access



# Baidu index and predictability of Chinese stock returns

Dehua Shen<sup>1,2</sup>, Yongjie Zhang<sup>1,3\*</sup>, Xiong Xiong<sup>1,2</sup> and Wei Zhang<sup>1,3</sup>

\* Correspondence: yjz@tju.edu.cn

<sup>1</sup>College of Management and Economics, Tianjin University, Tianjin 300072, People's Republic of China

<sup>3</sup>Key Laboratory of Computation and Analytics of Complex Management Systems (CACMS), Tianjin 300072, People's Republic of China

Full list of author information is available at the end of the article

## Abstract

A number of studies have investigated the predictability of Chinese stock returns with economic variables. Given the newly emerged dataset from the Internet, this paper investigates whether the Baidu Index can be employed to predict Chinese stock returns. The empirical results show that 1) the Search Frequency of Baidu Index (*SFBI*) can predict next day's price changes; 2) the stock prices go up when individual investors pay less attention to the stocks and go down when individual investors pay more attention to the stocks; 3) the trading strategy constructed by shorting on the most *SFBI* and longing on the least *SFBI* outperforms the corresponding market index returns without consideration of the transaction costs. These results complement the existing literature on the predictability of Chinese stock returns and have potential implications for asset pricing and risk management.

**Keywords:** Stock return predictability, Baidu index, Trading strategy, Financial Big data analytics, Chinese stock market, Investor inattention

## Background

The predictability of stock returns has long been the focus in financial economics with both cross-sectional and time series analysis. Voluminous variables have been put forward to predict the stock returns, including book-to-market ratio (Pontiff and Schall, 1998), inflation rate (Nelson, 1976), dividend-price ratio (Fama and French, 1988) and term structure of the interest rates (Campbell, 1987), etc. As for the predictability of Chinese stock market, Lee and Rui (2000) show that US stock return helps predict returns of Shanghai A and B stocks. Chen et al. (2010) document that only 5 out of 18 firm-specific variables can predict returns in Chinese stock market and attribute this weak predictability to the low informativeness and less heterogeneously distributed of stock returns. Goh et al. (2013) find that US economic variables have statistically significant predictive power for Chinese stock returns after China entering into the WTO. Specifically, the joint variables of China and US economic variables have superior predictive power. Jordan et al. (2014) find that the returns of economically-linked economies can predict the aggregate Chinese stock market. Chen et al. (2016) show that the U.S. economic variables can strongly forecast the monthly volatility of Chinese stock returns.

Recently, scholars begin to employ the information extracted from Internet as the new source of information for financial studies, including the online stock message boards (Tumarkin and Whitelaw, 2001; Antweiler and Frank, 2004), Twitter (Bollen et

al., 2011; Zhang et al., 2016a), Google Trends (Da et al., 2011; Da et al., 2015), Baidu Index (Zhang et al., 2013), Baidu News (Zhang et al., 2014; Shen et al., 2016), online stock commentary column (Zhang et al., 2016c) and Sina Weibo (Jin et al., 2016). In particular, Da et al. (2011) show that the search frequency from Google Trends can predict stock returns in the next weeks. Joseph et al. (2011) consider the online ticker search as the proxy for investor sentiment and find that the sentiment can predict abnormal stock return and trading volume at weekly horizon. Dimpfl and Jank (2016) find that the internet search queries can predict next day's volatility.

This paper connects with the above-mentioned two streams of literature and contributes to the existing literature in two aspects. On the one hand, we give the first empirical study of the predictability of Baidu Index, the newly emerged internet information, on Chinese stock returns, while other studies mainly focus on the market variables (Goh et al., 2013; Chen et al., 2010; Chen et al., 2016). Although Zhang et al. (2013) advocate the search frequency of stock name in Baidu Index as the proxy for investor attention, they only focus on the explanatory power of this proxy on stock return and do not investigate its predictability. On the other hand, we complement the existing studies on Chinese stock returns (Goh et al., 2013; Jordan et al., 2014) in the sense that our predictive variable is at daily horizon. The causes of the changes in daily returns, volatility of trading volume is significantly different from the monthly movement of the market performance (Admati and Pfleiderer, 1988; Amihud and Mendelson, 1987). Therefore, we can gain more insights into the pricing mechanism and our study has potential important implications for pricing mechanisms, asset allocation, and risk management.

The remainder of this paper is organized as follows. "Data description" section describes the Baidu Index and capital data. "Empirical analysis" section performs the empirical analysis, including the lead-lag relationships, the cross-sectional analysis and the trading strategy. "Conclusions" section sets forth the conclusions.

## Methods

The sample focuses on both the Shanghai Stock Exchange and the Shenzhen Stock Exchange with 30 stocks in each board respectively, i.e., the ChiNext, the SME Board and the Main Board, covering the calendar days from March 1st 2011 to March 30th 2012.<sup>1</sup> Totally, there are 90 stocks with 267 trading days. In particular, we obtain the daily trading volume, daily stock return rate after dividend reinvestment and the market return from the China Stock Market and Accounting Research (CSMAR) database. The 1-min prices are retrieved from the RESSET Database to calculate the intraday volatility. In a pooled analysis, the daily trading volume ranges from 33700 to  $1.5 \times 10^8$ , the firm capitalization ranges from  $6.5 \times 10^5$  to  $5.8 \times 10^9$ , the PE ratio ranges from -667 to 3119 and the turnover ranges from 0.0398 to 58.94. Due to the diversified distribution in different boards as well as the distinct characteristic of the stocks, our sample can be viewed as a parsimonious representation of Chinese stock market.

## Baidu index

Baidu Index is a keyword-searching tool launched by Chinese largest search engine and its main customers are Chinese language users. We search the stock name in Baidu

Index and record the Search Frequency of Baidu Index (*SFBI*) for each stock during the sample period. To make the *SFBI* comparable across firms, we calculate the standardized *SFBI* (*SSFBI*) for each individual stock.

$$SSFBI_t = \frac{SFBI_t - AV_{SFBI}}{SD_{SFBI}} \quad (1)$$

where  $AV_{SFBI}$  is the average value in the sample period and  $SD_{SFBI}$  is the standard deviation of the *SFBI* time series.

### Variables

We calculate the absolute value of the difference between individual stock return and market return as the abnormal return (*AbRet*). This measurement captures absolute price changes, rather other the upward and downward directions. In a similar way, we also calculate the cumulative abnormal return (*CAR*) for future analysis.

$$AbRet_t = |Ret_{I,t} - Ret_{M,t}| \quad (2)$$

$$CAR(-30, +30) = \sum_{t=-30}^{+30} (Ret_{I,t} - Ret_{M,t}) \quad (3)$$

where  $Ret_{I,t}$  is the daily stock return rate after dividend reinvestment and  $Ret_{M,t}$  is the corresponding market index return. In particular, we choose the return of Chinese Stock Index 300 (CSI 300) as the market return for the reason that the CSI 300 is the first index launched by both the Shanghai and Shenzhen Stock Exchanges and thus it represents the whole market.

Following Barber and Odean (2008), we calculate the excess trading volume (*ETV*) for each stock as the daily ratio of the stock's trading volume that day to its average trading volume in the whole sample period.

$$ETV_t = \frac{TV_t}{AV_{TV}} \quad (4)$$

where  $AV_{TV}$  is the daily trading volume on each trading day and  $AV_{TV}$  is the average trading volume in the whole sample period.

The intraday volatility is calculated as the standard deviation of the 1-min prices.

$$Volatility_t = \sqrt{\frac{\sum_1^N (P_t - AV_P)^2}{N}} \quad (5)$$

where  $P_t$  is the 1-min stock price and  $AV_P$  is the average stock price.  $N$  denotes the number of observation and  $N = 240$ .

### Results and discussion

This section provides empirical analysis on the lead-lag relationships between *SFBI* and stock returns, the market reactions around the *MSD* and *LSD* events as well as the performance of the trading strategy.

### Lead-lag relationships

In line with other recent studies on the lead-lag relations (Schmeling, 2009 and Siganos et al., 2014), we use five lags for *SFBI* and stock return as the explanatory variables and formulate the following ordinary regression model.

$$AbRet_{i,t} = \alpha_1 + \sum_{j=1}^5 \beta_{i,j} SSFBI_{i,t-j} + \sum_{j=1}^5 \gamma_{i,j} AbRet_{i,t-j} + \varepsilon_{i,t} \quad (6)$$

Table 1 reports the regression results of model (6). Since we employ the stock by stock regression, the Bonferroni correction method is used to correct the  $p$ -values for the multiple comparisons problem. We observe that the one lagged *SFBI* is significantly related to next trading day's abnormal return with positive correlation coefficient across different subsamples. These findings suggest that the online search behavior from individual investors can predict price changes in the next trading day.

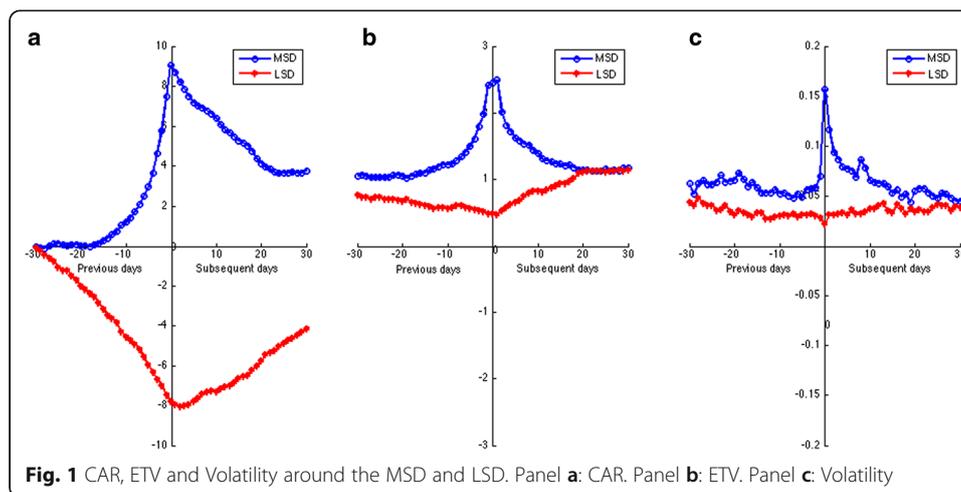
### MSD and LSD

To give a meticulous observation of the market reaction, for each individual stock, we sort the trading days based on the corresponding *SFBI* and select the highest 10 trading days as the Most Searching Days (*MSD*) and the lowest 10 trading days as the Least Searching Days (*LSD*). We can then consider the *MSD* and *LSD* as event day and observe the market reactions around *MSD* and *LSD* with the event study methodology. Figure 1 illustrates the *CAR*, *ETV* and Volatility around the *MSD* and *LSD*. In particular, Panel A plots the changes of *CAR* around the *MSD* and *LSD*, we find that the stock price goes down after the *MSD* and goes up after the *LSD*. These results are inconsistent with the findings in the US stock market (Da et al., 2011), which support higher search volume predicts future higher stock return. This inconsistency may be driven by the "big position construction" by institutional investors. After constructing the position for certain stocks, the institutional investors release some news about the companies and attract the individual

**Table 1** Summary of the regression results This table reports the regression results of model (6). As the analysis is based on the stock by stock regression, we employ the Bonferroni correction method to correct the  $p$ -values for the multiple comparisons problem. The significance cut-off is set to  $\alpha/n$  ( $\alpha = 0.05$  and  $n = 90$ ). Since not all the variables are standardized, the reported coefficients are not the basis points but depend on the scale of the variables

Variables	ChiNext	Main board	SME board	Full sample
Intercept	0.0173 <sup>a</sup>	0.0153 <sup>a</sup>	0.0157 <sup>a</sup>	0.0161 <sup>a</sup>
<i>SSFBI</i> <sub><i>t</i>-1</sub>	0.0023 <sup>a</sup>	0.0019 <sup>a</sup>	0.0029 <sup>a</sup>	0.0024 <sup>a</sup>
<i>SSFBI</i> <sub><i>t</i>-2</sub>	-0.0008	-0.0002	-0.0013	-0.0008
<i>SSFBI</i> <sub><i>t</i>-3</sub>	0.0003	0.0003	0.0005	0.0004
<i>SSFBI</i> <sub><i>t</i>-4</sub>	0.0002	0.0000	-0.0005	-0.0001
<i>SSFBI</i> <sub><i>t</i>-5</sub>	0.0001	0.0000	0.0001	0.0001
<i>AbRet</i> <sub><i>t</i>-1</sub>	0.0162	0.0138	0.0103	0.0134
<i>AbRet</i> <sub><i>t</i>-2</sub>	-0.0041	0.0128	0.0128	0.0072
<i>AbRet</i> <sub><i>t</i>-3</sub>	-0.0041	-0.0000	0.0077	0.0012
<i>AbRet</i> <sub><i>t</i>-4</sub>	-0.0128	0.0046	0.0049	-0.0011
<i>AbRet</i> <sub><i>t</i>-5</sub>	0.0027	-0.0072	0.0031	-0.0004

Note: <sup>a</sup>denotes the significant at 1% level



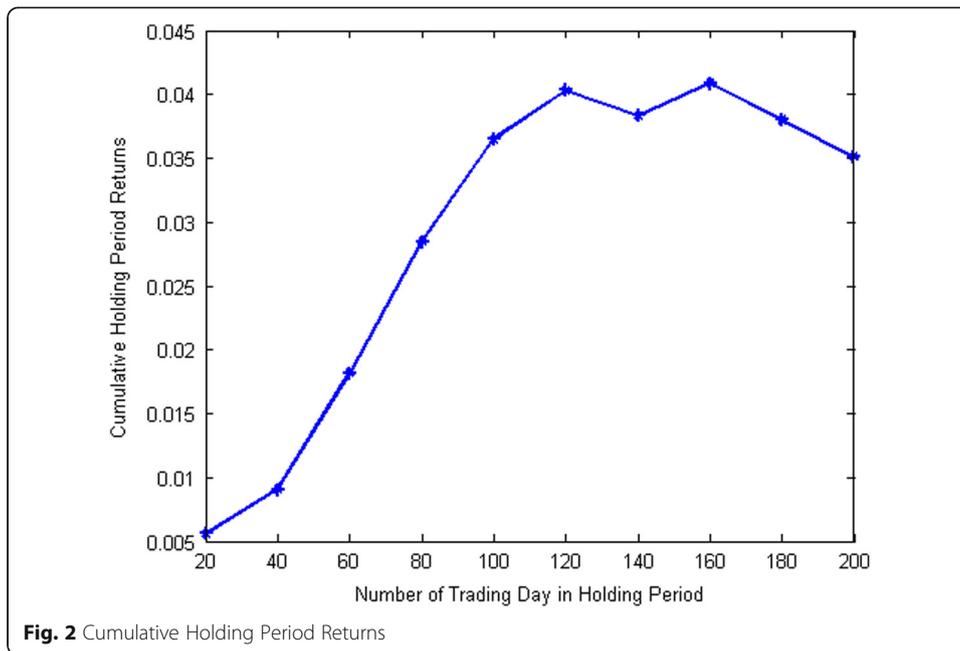
investors to buy these stocks. Meanwhile, the institutional investors sell out their holdings to gain profit. This argument reconciles with report conducted by the Shanghai Stock Exchange showing that the individual investors account for 93.20% in A shares at the end of 2012 as well as the some scholars claiming that Chinese stock market are dominated by irrational individual investors who are subject to strong psychological bias and thus resulting in speculation (Feng and Seasholes, 2008 and Zhang et al., 2016b). Besides, Fig. 1 also show that there are significantly larger *ETV* and *Volatility* on the *MSD*.

### Trading strategy

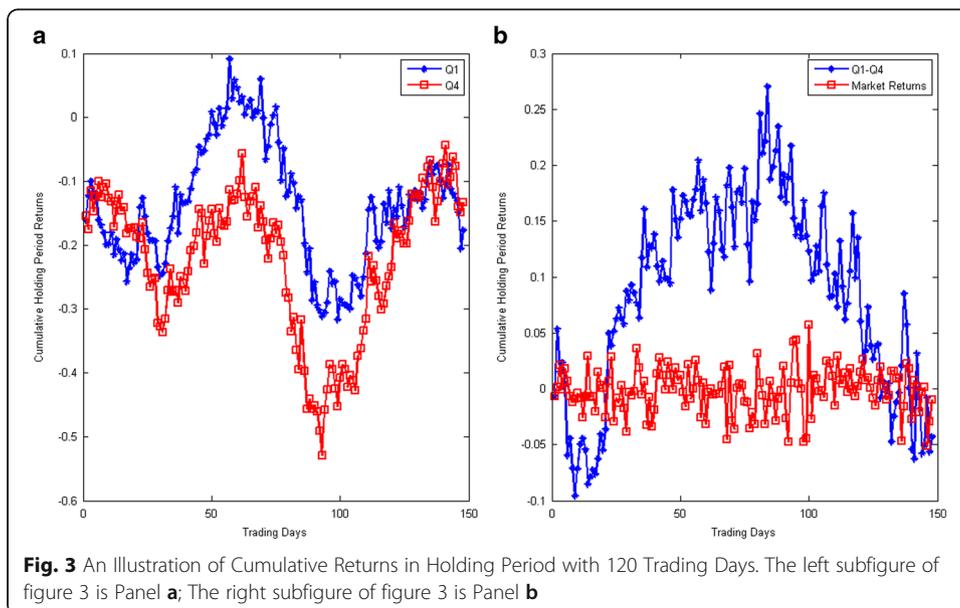
In this section, we further investigate the economic significance of our findings by constructing a long-short trading strategy based on the *SFBI*. The long-short trading strategy is formed as follows: on each trading day, the 90 firms are sorted into quantiles (Q) based on the *SFBI* in previous trading day. Q1 contains the firms with the least *SFBI* and Q4 contains the firms with the most *SFBI*. The firms are held in their portfolios for the entire sample period with different holding periods, e.g., 20, 40, 60, 80, 100 and 120 trading days, respectively. We then obtain a portfolios that consists of a long position in the least quantile of firms (Q1) and a short position in the most quantile of firms (Q4), i.e., the returns of the portfolio are the Q1 minus Q4. Figure 2 illustrates the cumulative profit with different holding periods. As is plotted, this trading strategy has positive return for all the holding periods. Specifically, Fig. 3 illustrate the strategy with 120 holding period. Panel A of Fig. 3 plots the returns of the Q1 and Q4, we can find that the return is Q1 is significantly larger than that of the Q4 at 1% level with  $p$ -value = 0.0017. The blue line in Panel B of Fig. 3 is the difference between the Q1 and Q4 and the red line is the corresponding market returns. We can find that our trading strategy outperforms the market returns significantly at 1% level with  $p$ -value = 0.0000.

### Conclusions

This paper employs the Baidu Index as a predictive variable and investigates its predictability for Chinese stock returns. The empirical findings show that Baidu



Index can predict the price changes on the next trading day. After constructing the *MSD* and *LSD*, we mainly find that the stock prices go up when individual investors pay less attention to the stocks and go down when individual investors pay more attention to the stocks. Besides, we also construct a trading strategy by shorting on the most *SFBI* and longing on the least *SFBI*. The trading strategy outperforms the corresponding market index returns. However, we must caution scholars in adopting our trading strategy in real investment. Because the transaction costs associated with the rebalancing are not considered.



## Endnotes

<sup>1</sup>Baidu Index does not provide the data anymore. Therefore, the data used in this paper is from Zhang et al. (2013), but it addresses different research questions. We believe that this paper has some managerial implications for practitioners who may have private access to the Baidu Index.

## Acknowledgements

This work is supported by the National Natural Science Foundation of China (71320107003 and 71532009).

## Authors' contributions

DS and YZ contributed to study design, data collection and provided the first draft of the manuscript. DS, WZ, YZ and XX participated in the empirical analysis, interpretation of the results, final draft and proof reading of this manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>College of Management and Economics, Tianjin University, Tianjin 300072, People's Republic of China. <sup>2</sup>China Center for Social Computing and Analytics (CCSCA), Tianjin University, Tianjin 300072, People's Republic of China. <sup>3</sup>Key Laboratory of Computation and Analytics of Complex Management Systems (CACMS), Tianjin 300072, People's Republic of China.

Received: 5 September 2016 Accepted: 13 March 2017

Published online: 24 March 2017

## References

- Admati AR, Pfleiderer P (1988) A Theory of Intraday Patterns: Volume and Price Variability. *Rev Financ Stud* 1(1):3–40
- Amihud Y, Mendelson H (1987) Trading Mechanisms and Stock Returns: An Empirical Investigation. *J Financ* 42(3):533–53
- Antweiler W, Frank MZ (2004) Is all that talk just noise? The information content of internet stock message boards. *J Financ* 59:1259–1294
- Barber BM, Odean T (2008) All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Rev Financ Stud* 21:785–818
- Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. *J Comput Sci* 2:1–8
- Campbell JY (1987) Stock returns and the term structure. *J Financ Econ* 18:373–399
- Chen X, Kim KA, Yao T, Yu T (2010) On the predictability of Chinese stock returns. *Pac Basin Financ J* 18(4):403–25
- Chen J, Jiang F, Li H, Xu W (2016) Chinese stock market volatility and the role of U.S. economic variables. *Pac Basin Financ J* 39:70–83
- Da Z, Engelberg J, Gao P (2011) In Search of Attention. *J Financ* 66:1461–1499
- Da Z, Engelberg J, Gao P (2015) The Sum of All FEARS Investor Sentiment and Asset Prices. *Rev Financ Stud* 28:1–32
- Dimpfl T, Jank S (2016) Can Internet Search Queries Help to Predict Stock Market Volatility? *Eur Financ Manag* 22:171–192
- Fama EF, French KR (1988) Dividend yields and expected stock returns. *J Financ Econ* 22:3–25
- Feng L, Seasholes MS (2008) Individual investors and gender similarities in an emerging stock market. *Pac Basin Financ J* 16:44–60
- Goh JC, Jiang F, Tu J, Wang Y (2013) Can US economic variables predict the Chinese stock market? *Pac Basin Financ J* 22:69–87
- Jin X, Shen D, Zhang W (2016) Has microblogging changed stock market behavior? Evidence from China. *Physica A* 452:151–156
- Jordan SJ, Vivian A, Wohar ME (2014) Sticky prices or economically-linked economies: The case of forecasting the Chinese stock market. *J Int Money Financ* 41:95–109
- Joseph K, Babajide Wintoki M, Zhang Z (2011) Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *Int J Forecast* 27:1116–1127
- Lee CF, Rui OM (2000) Does trading volume contain information to predict stock returns? Evidence from China's stock markets. *Rev Quant Finan Acc* 14:341–360
- Nelson CR (1976) Inflation and rates of return on common stocks. *J Financ* 31:471–483
- Pontiff J, Schall LD (1998) Book-to-market ratios as predictors of market returns. *J Financ Econ* 49:141–160
- Schmeling M (2009) Investor sentiment and stock returns: Some international evidence. *J Empir Financ* 16:394–408
- Shen D, Zhang W, Xiong X, Li X, Zhang Y (2016) Trading and non-trading period Internet information flow and intraday return volatility. *Physica A* 451:519–524
- Siganos A, Vagenas-Nanos E, Verwijmeren P (2014) Facebook's daily sentiment and international stock markets. *J Econ Behav Organ* 107(Part B):730–743
- Tumarkin R, Whitelaw RF (2001) News or noise? Internet postings and stock prices. *Financ Anal J* 57:41–51
- Zhang W, Shen D, Zhang Y, Xiong X (2013) Open source information, investor attention, and asset pricing. *Econ Model* 33:613–619

- Zhang Y, Feng L, Jin X, Shen D, Xiong X, Zhang W (2014) Internet information arrival and volatility of SME PRICE INDEX. *Physica A* 399:70–74
- Zhang W, Li X, Shen D, Tegli A (2016a) Daily happiness and stock returns: Some international evidence. *Physica A* 460: 201–209
- Zhang W, Li X, Shen D, Tegli A (2016b)  $R^2$  and idiosyncratic volatility: Which captures the firm-specific return variation? *Econ Model* 55:298–304
- Zhang Y, Song W, Shen D, Zhang W (2016c) Market reaction to internet news: Information diffusion and price pressure. *Econ Model* 56:43–49

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---