

# Scaling and Predictability in Stock Markets: A Comparative Study

Huishu Zhang, Jianrong Wei, Jiping Huang\*

Department of Physics and State Key Laboratory of Surface Physics, Fudan University, Shanghai, China

## Abstract

Most people who invest in stock markets want to be rich, thus, many technical methods have been created to beat the market. If one knows the predictability of the price series in different markets, it would be easier for him/her to make the technical analysis, at least to some extent. Here we use one of the most basic sold-and-bought trading strategies to establish the profit landscape, and then calculate the parameters to characterize the strength of predictability. According to the analysis of scaling of the profit landscape, we find that the Chinese individual stocks are harder to predict than US ones, and the individual stocks are harder to predict than indexes in both Chinese stock market and US stock market. Since the Chinese (US) stock market is a representative of emerging (developed) markets, our comparative study on the markets of these two countries is of potential value not only for conducting technical analysis, but also for understanding physical mechanisms of different kinds of markets in terms of scaling.

**Citation:** Zhang H, Wei J, Huang J (2014) Scaling and Predictability in Stock Markets: A Comparative Study. PLoS ONE 9(3): e91707. doi:10.1371/journal.pone.0091707

**Editor:** Tobias Preis, University of Warwick, United Kingdom

**Received:** August 11, 2013; **Accepted:** February 14, 2014; **Published:** March 14, 2014

**Copyright:** © 2014 Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work is supported by the National Natural Science Foundation of China under Grant Nos. 11075035 and 11222544 (<http://www.nsf.gov.cn/>); the Fok Ying Tung Education Foundation under Grant No. 131008 (<http://www.moe.edu.cn/>); the Program for New Century Excellent Talents in University (NCET-12-0121) (<http://www.moe.edu.cn/>); the Shanghai Rising-Star Program (No. 12QA1400200) (<http://www.stcsm.gov.cn/>); and the Chinese National Key Basic Research Special Fund under Grant No. 2011CB922004 (<http://www.most.gov.cn/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [jphuang@fudan.edu.cn](mailto:jphuang@fudan.edu.cn)

## Introduction

Stock markets provide an opportunity for people to increase their wealth. Most of people who invest in stock markets want to earn excess profits [1]. So various technical trading strategies based on moving averages, Bollinger lines and so on have been set up [2–4]. However it is shown that even the experienced fund managers can't ensure the abnormal returns [5], let alone the individual investors [6]. Biondo (2013) compared four mainstream technical strategies including the random one, and obtained the following results: the average percentages of wins for these strategies are similar, but the risk of the random one is surprisingly the lowest [7]. We all know that the stock markets belong to the family of complex systems. Not only many real systems in physics [8,9] but also lots of others like geophysical, social and biological systems are part of them. Agent-based modeling is an important way to analyze the behavior of agents in complex systems [10,11], which has been also widely used in analyzing stock markets [12,13].

The strength of predictability of time series is a significant property for one to design a trading strategy, so many methods like Box-Jenkins methodology [14] and Bayesian model averaging [15] have been created to analyze the predictability of time series. So far several relevant works have been done. For example, Fama and French (1989) found that conditional expected returns vary with the business cycles because rational agents smooth consumption over time [16]. Burton (2004) reviewed the success of past studies to predict future equity returns, and found that although some predictability of returns exists, there is no evidence of any

systematic inefficiency that would ensure investors to earn excess returns [17]. Venkat (2004) found that annual excess returns on the stock market index are negatively related to the returns of glamour stocks during the 2001–2004 period, which means there existed some predictability [18]. In 2012, Grönlund, Yi and Kim found that the local maxima in the profit landscape are spread in the form of a fractal structure, that it shows the stock market has low predictability by technical analysis [19].

It is well known that the US stock market has a long history, which can be traced back to the period of the War of Independence. Nowadays, the US stock market has grown up as the most developed and mature stock market in the world. Comparing with the US stock market, the Chinese stock market is much more emerging and just precedes a quarter of a century. [20]. According to the Efficient Markets Hypothesis, the more developed market which is more efficient has weaker predictability by technical analysis [21]. Is the emerging market of China really more predictable than the developed market of US? In this work, we will use the profit landscape to compare the predictability of indexes and individual stocks in Chinese and US stock markets.

## Methods

To proceed, we resort to the profit landscape method [19] for the following data processing. Meanwhile, one of the most basic trading strategies will be utilized to make the profit landscape. The trading decision at time  $t$  only depends on the log return  $R(t, t') = \ln[x(t)/x(t')]$ , in which  $x(t)$  and  $x(t')$  is the daily closing

price at time  $t$  and  $t'$  with  $t-t'=d$  ( $d \geq 1$ ). Next is this simple trading strategy.

- If  $R(t,t') \geq p/K$ ,  $f_s$  fraction of stocks in account are sold. And  $K$  is the scale factor.
- If  $R(t,t') \leq -q/K$ , one spends  $f_c$  fraction of cash buying the stocks at current price.

Here  $p$  and  $q$  are two parameters that determine in which condition one should buy or sell. For convenience, we just define that  $p, q \in [0, 1]$ . Then the detailed parameters for performing our simulations are as follows: (i) At the first trading day, one gain initial cash  $m(1)$  and initial number  $s(1)$  of stocks. We set initial money  $m(1)$  as  $10^7$  and initial stocks  $s(1)$  as 0. Moreover,  $d=1$  (the value of  $d$  hardly influence the results [19]); (ii) In real stock markets, the minimum trading volume exists. We set this minimum volume as 1 for convenience. One can imagine that no difference will appear if we increase this volume and proportional increase the initial cash as well. (iii) We also can't ignore the transaction tax in real market, thus we assume that this tax rate is 0.1%, the same as the Chinese securities transaction tax rate; (iv) The two proportion parameters,  $f_s$  and  $f_c$ , rarely affect the exponent which we care about in our study. Thus in order to facilitate the model, they can just be set as 0.5. (v)  $K$  is the scale factor that will be adjusted in different data series and explained concretely soon after.

For given values of  $p$ ,  $q$ , the strategy is repeatedly executed till the last day  $T$ . For the sake of simplicity, we assume that the risk-free interest rate is zero, and the cash dividends provided by companies are neglected. Then, we can assess the performance of strategy only by final profit defined by  $P(p, q) = [m(T) + s(T)x(T) - x(1)]/x(1)$ . In order to plot the profit landscape, we'd like to traverse almost all possible  $P(p, q)$ . Therefore, We define parameter  $N$  as the resolution to determine the minimum variation amount  $p$  and  $q$ . Then the  $p = (1/N, 2/N, \dots, 1)$  and  $q = (1/N, 2/N, \dots, 1)$ . According to the various combinations of  $p$  and  $q$ , we can gain  $N^2$  values of  $P(p, q)$ . Then we can plot a three-dimensional figure called profit landscape through  $p$ ,  $q$ , and  $P(p, q)$ .

When the resolution  $N$  changes, the appearance and scale of profit landscape also changes. So we can obtain the relationship between  $N$  and the number of local maxima,  $M$ , on landscape. The local maxima are these points whose values are larger than the four neighboring ones. If one wants to optimize the investment strategies, he or she will establish strategies within a certain range of peaks on profit landscape. Hence, nobody hopes that the locations and amounts of peaks change a lot when  $N$  increases. It is quite clear that if a price series is completely periodic or regular, the best strategies will be determinate. The peaks will be certain points or small platforms on profit landscape and their locations will never change when  $N$  increases. In this case, one will have algebraic form of  $M \propto N^0$  ( $M$  does not vary with  $N$ ), so the technical strategies are highly easy to be formulated. If  $M$  increases with  $N$  following  $M \propto N^1$ , it means that the new local maxima only appear along horizontal or vertical axis on the landscape when  $N$  increases, so it will not be too difficult for one to optimize strategies (either  $p$  or  $q$  is more likely to be optimized). And in case we choose the unpredictable random series (say, geometric Brownian motion [22]) to do the above analysis, we will obtain algebraic form of  $M \propto N^2$  [19]. In this situation, no parameters can be optimized and the technical strategies are invalid. To sum up, one can easily see that the predictability of price series is stronger when the exponent of power-law is smaller. In this article, we gain the relation between  $M$  and  $N$  following

the power-law distribution with exponents all ranging from 1 to 2. These results are consistent with the fact that real price movements cannot be totally described by random series [23,24].

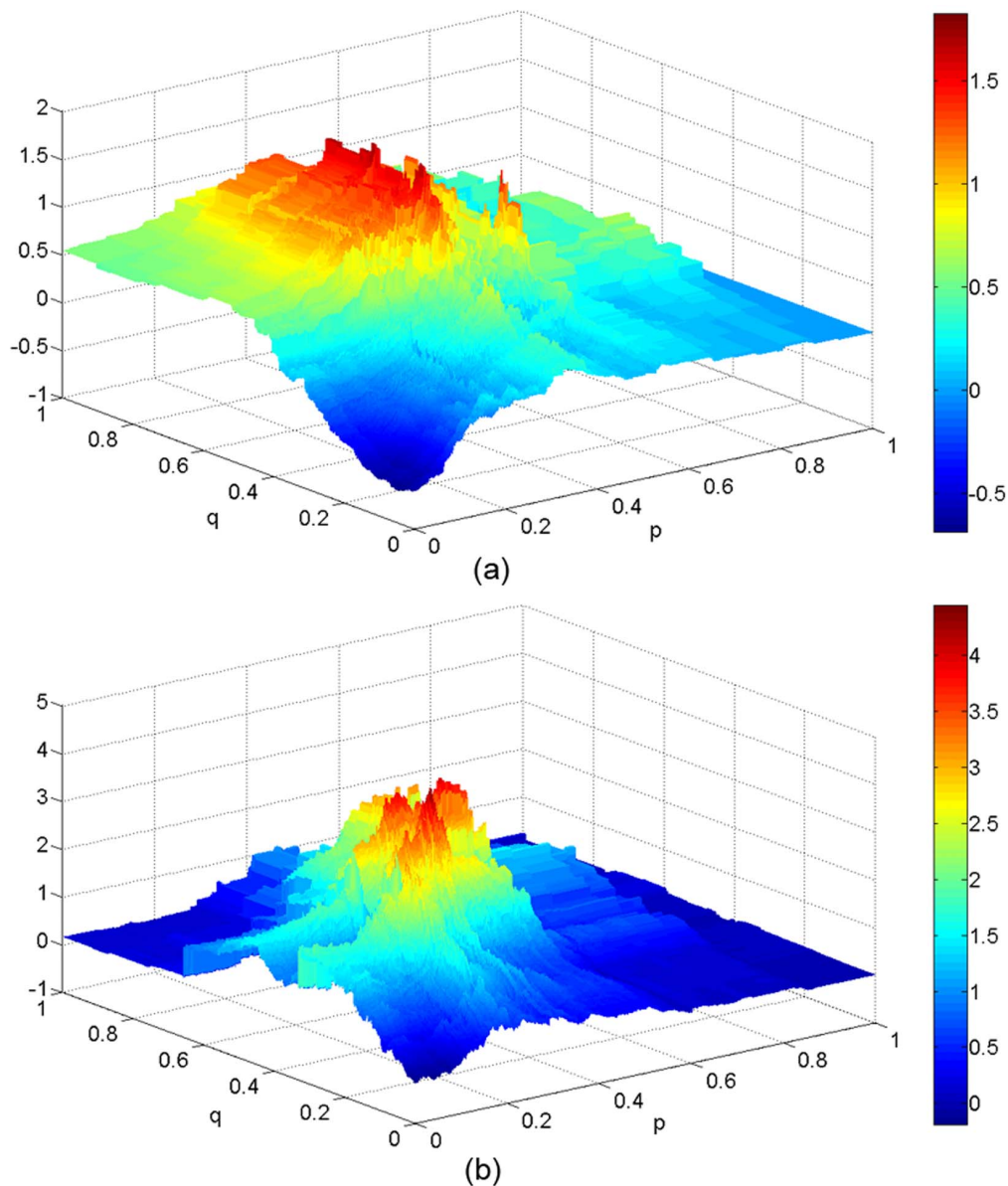
Then we'd like to explain the meaning about scale factor  $K$ . It can be easily seen that the amount of computation for simulations is directly proportional to  $N^2$ . So when  $N$  is large, the amount of the computation is extremely tremendous. Parameter  $K$  is the scale factor for zooming in the profit landscape as if we use a magnifier to observe it. Given example, if  $K=10$ , then the  $p/K = (0.1/N, 0.2/N, \dots, 0.1)$ . As a result, this parameter can help to increase precision even  $N$  is not large enough. In addition, if  $R > \max(p/K)$ , some strategies will fall out the landscape resulting the imperfectly effective optimization. Consequently,  $K$  should be adjusted to confirm that  $\max(p/K) \geq R \geq \min(-q/K)$ , at the same time the landscape contains all the possible strategies. Fig. 1 shows two figures of the profit landscape with certain value of  $N$  and  $K$ .

## Results

In this work, we use daily closing price data of Standard & Poor's 500 Index (S&P500 Index; a representative stock index in US stock markets) and the Shanghai Composite Index (a representative stock index in Chinese stock markets) from 1997 to 2012. We also choose daily closing price data of the 80 earliest listed stocks from Standard & Poor's 500 during 1997 to 2012. The chief reasons why we select these companies as symbol of US stocks involve two factors. One attribute to their long histories, and the other is that these companies belong to various industries. At the same time, the daily closing price data of all the 34 stocks (listed before 1997) from the Shanghai Stock Exchange during the same period are chosen for analysis.

We first analyze the indexes of both China and US. There is no doubt that the Shanghai composite index floats within plus or minus 10% in one day. Simultaneously, though the S&P500 Index does not have such limitation, the maximum daily price fluctuation from 1997 to 2012 is only a little more than 10%. In order to enhance the accuracy of the result, we can set  $K$  as 10. In this situation, the maximum  $p/K$  and  $q/K$  is 0.1, which is approaching to the maximum value of  $R$  ( $\ln[x(t)/x(t')] = \ln[1.1] \approx 0.095$ ). Therefore, the profit landscape can has completeness and high precision as well. Fig. 2(a,b) shows the relation between  $M$  and  $N$  of the landscape from two indexes. Clearly, we obtain the power-law distribution of  $M$  versus  $N$  with the exponent  $\gamma$  equaling 1.48 for the Shanghai Composite Index and  $\gamma$  equaling 1.42 for S&P500 Index, respectively. In view of the presence of fitting error, we can conclude that the predictability of the Shanghai Composite Index is approximately equal to that of S&P500 index.

Then we analyze the individual stocks which are components of these two indexes. Owing to the daily price fluctuation of several stocks in S&P500 are enormous, we just set  $K$  as 1 to ensure that the landscape contains almost all the possible conditions. First of all, we use the above-mentioned method to get every  $M$  of component stocks. After that, we calculate the average of  $M$  and utilize it to analyze the result for individual stocks; see Fig. 2(c,d) what shows relation between  $M$  and  $N$  of the landscape from individual stocks. Also, we observe power-law distribution of  $M$  versus  $N$  with the exponent  $\gamma$  equaling 1.79 for the stocks of Chinese companies and  $\gamma$  equaling 1.56 for US ones. That is, the predictability of individual stocks of the Shanghai Stock Exchange is lower than that of S&P500. On the other hand, comparing Fig. 2(a,b) with Fig. 2(c,d), we can also see that the value of exponent of index [in Fig. 2(a,b)] is smaller than that of individual



**Figure 1. Two model profit landscapes in the two dimensional space for (a) the Shanghai Composite Index and (b) one component stock in the S&P500 Index.** Parameters:  $K = 10$  and  $N = 1000$ .  
doi:10.1371/journal.pone.0091707.g001

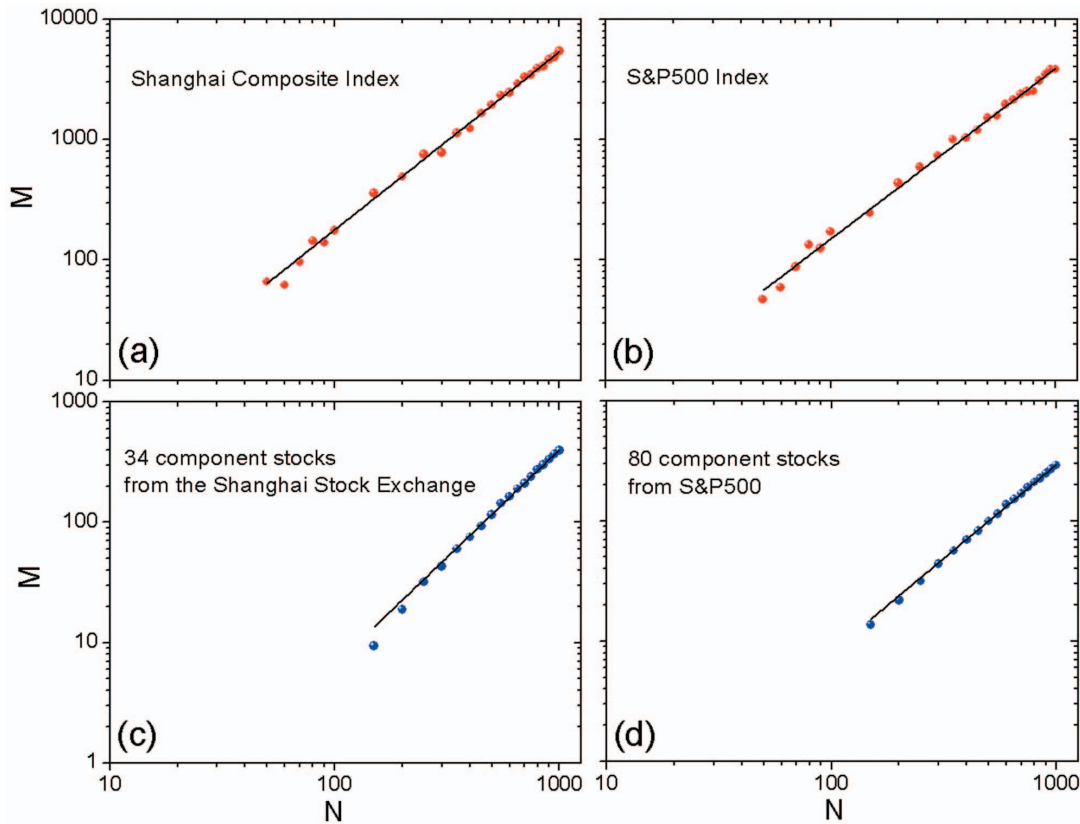
stocks [in Fig. 2(c,d)] in both two markets. Therefore, it appears to be more difficult for one to predict the individual stocks than indexes in both Chinese and US stock markets.

## Discussion

We have studied the profit landscape defined by some straightforward investment strategies in this work. The results show that the relation between resolution  $N$  and the number of local maxima  $M$  on the profit landscape follows the power-law distribution, and the value of power-law exponent can be used to characterize the predictability of the price series. From the results, we can conclude that the predictability of Shanghai Composite Index is roughly the same as that of S&P500 Index. However, the predictability of individual stocks does not give the similar results: the predictability of individual stocks in Shanghai

Stock Exchange is lower than that in S&P500. It is well known that the Chinese government sometime has a great influence on stock markets, and the black-box operation exists in the exchanges of some individual stocks. So we suppose these two reasons play a crucial role in this result.

On the other hand, we can also conclude that the predictability of stock indexes is higher than individual stocks in both the Shanghai Composite Index and S&P500 Index. We know that the index is calculated by the individual stocks. In a price-weighted index such as the Dow Jones Industrial Average, the price of each component stock is the only consideration when determining of the index. S&P500 Index or the Shanghai Composite Index is a capitalization-weighted index whose components are weighted according to the total market capitalizations of their circulating shares. So, if the price of one stock has some fluctuations for unknown reason, these changes will hardly influence the index.



**Figure 2.**  $M$  versus  $N$  for (a) the Shanghai Composite Index ( $K = 10$ ), (b) the S&P500 Index ( $K = 1$ ), (c) 34 component stocks from Shanghai Stock Exchange ( $K = 1$ ), and (d) 30 component stocks from S&P500 ( $K = 1$ ). The relation between  $M$  and  $N$  follows the power-law distribution (as indicated by the straight line in each panel) with exponent (a)  $\gamma = 1.48$  for  $R^2 = 0.998$ , (b)  $\gamma = 1.42$  for  $R^2 = 0.994$ , (c)  $\gamma = 1.79$  for  $R^2 = 0.999$ , and (d)  $\gamma = 1.56$  for  $R^2 = 0.999$ . Here  $R^2$  denotes the regression coefficient that represents the degree of fitting with the power law: the perfect fitting corresponds to  $R^2 = 1$  [26].  
doi:10.1371/journal.pone.0091707.g002

On the contrary, if the price of index varies greatly someday, lots of the individual stocks follow the similar trend with the index. Therefore, the weighted averaging operation of index enhances its predictability. We know that the price series in long time scales are approximate as a random process [25], so the predictability is not high as expected. We plan to study the profit landscape of time series in short time scales like high frequency price series and so on.

Finally, the comparative study reported in this work is of potential value not only for conducting technical analysis, but also

for understanding physical mechanisms of different kinds of markets in terms of scaling, being beyond the efficient markets hypothesis.

### Author Contributions

Conceived and designed the experiments: HSZ, JRW. Performed the experiments: HSZ, JRW. Analyzed the data: HSZ. Contributed reagents/materials/analysis tools: HSZ. Wrote the paper: HSZ. Modified the manuscript: JPH, JRW.

### References

- Garvey R, Murphy A (2005) The profitability of active stock traders. *Journal of Applied Finance* 15: 93–101.
- Blume M, Easley D, O'hara M (1994) Market statistics and technical analysis: The role of volume. *The Journal of Finance* 49: 153–181.
- Brown DP, Jennings RH (1989) On technical analysis. *Review of Financial Studies* 2: 527–551.
- Park CH, Irwin SH (2004) The profitability of technical analysis: A review. *AgMas Project Research Report No.* pp 2004–04.
- Chen HL, Jegadeesh N, Wermers R (2000) The value of active mutual fund management: An examination of the stockholdings and trades of fund managers. *Journal of Financial and Quantitative Analysis* 35: 343–368.
- Barber BM, Odean T (2000) Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance* 55: 773–806.
- Biondo AE, Pluchino A, Rapisarda A, Helbing D (2013) Are Random Trading Strategies More Successful than Technical Ones? *PLoS ONE* 8(7): e68344.
- Gammaitoni L, Hanggi P, Jung P, Marchesoni F (1989) Stochastic Resonance. *Reviews of Modern Physics*, 70 (1): 223–287.
- Caruso F, Huelga SF, Plenio MB (2010) Noise-Enhanced Classical and Quantum Capacities in Communication Networks. *Phys Rev Lett* 105(198): 190–501.
- Horiuchi S, Takasaki H (2012) Boundary nature induces greater group size and group density in habitat edges: an agent-based model revealed. *Population Ecology* 54: 197–203.
- Dosi G, Fagiolo G, Napoletano M, Roventini A (2013) Income distribution, credit and fiscal policies in an agent-based Keynesian model. *Journal of Economic Dynamics and Control* 37: 1598C1625.
- Wei JR, Huang JP, Hui PM (2013) An agent-based model of stock markets incorporating momentum investors. *Physica A: Statistical Mechanics and Its Applications* 392: 2728C2735.
- Feng L, Li B, Podobnik B, Preis T, Stanley HE (2012) Linking agent-based models and stochastic models of financial markets. *Proc Natl Acad Sci* 209: 8388–8393.
- Harvey AC, Todd PHJ (1983) Forecasting Economic Time Series With Structural and Box-Jenkins Models: A Case Study. *Journal of Business & Economic Statistics* 1: 299–307.

15. Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: A tutorial. *Statistical Science* 14: 382–401.
16. Fama EF, French KR (1989) Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25: 23–49.
17. Malkiel BG (2004) Models of stock market Predictability. *Journal of Financial Research* 27: 449–459.
18. Eleswarapu VR, Reinganum MR (2004) The Predictability of Aggregate Stock Market Returns: Evidence Based on Glamour Stocks. *The Journal of Business* 77: 275–294.
19. Grönlund A, Yi IG, Kim BJ (2012) Fractal Profit Landscape of the Stock Market. *PLoS ONE* 7: e33960.
20. Masih AMM, Masih R (1999) Are Asian stock market fluctuations due mainly to intra-regional contagion effects? Evidence based on Asian emerging stock markets. *Pacific-Basin Finance Journal* 7: 251–282.
21. Malkiel BG, Fama EF (1970) Efficient Capital Markets: A Review Of Theory And Empirical Work. *The Journal of Finance* 25: 383–417.
22. Black F, Scholes M (1973) The Pricing of options and corporate liabilities. *J Polit Econ* 81: 637–654.
23. Mantegna RN, Stanley HE (2000) *An introduction to econophysics*. England: Cambridge University Press.
24. Mantegna RN, Stanley HE (1995) Scaling behaviour in the dynamics of an economic index. *Nature* 376: 46–49.
25. Carbone A, Castelli G, Stanley HE (2004) Time dependent Hurst exponent in financial time series. *Physica A* 344: 267–271.
26. Clauset A, Shalizi C, Newman M (2009) Power-law distributions in empirical data. *SIAM Review* 51: 661–703.