Fitting of Russell 2000 Index for the First 20 Years in the 21st Century with Random Walk – Application in Big Data and Digital Economy

Shaomin Yan and Guang Wu(✉)

National Engineering Research Center for Non-Food Biorefinery, State Key Laboratory of Non-Food Biomass and Enzyme Technology, Guangxi Academy of Sciences, 98 Daling Road, Nanning 530007, Guangxi, China
{shaominyan,hongguanglishibahao}@gxas.cn

Abstract. Without any question, the stock markets are the engine to generate the big data, which become an active part of digital economy. The big data provide rich opportunities through internet to combine various technologies with the issues raised in the economic developments in modern society. Whether a stock index can be described by the random walk is an important approval to the efficient market hypothesis (EMH). Of well-known stock indices, the Russell 2000 index, which includes 2000 small-cap companies, has a smaller capitalization compared with other well-known indices. It therefore draws less attention from investors and is less subject to manipulations. Hence, it potentially behaves more randomly than other heavily manipulated indices. As a result, it might be more suitable for the random walk statistical tests and simulation/fitting. In this study, the Russell 2000 index from 2001to 2020 was fitted by the random walk method in five segments of time. In general, the results showed that the random walk method can fit the Russell 2000 index for different periods of time. Thus, the results add a piece of evidence to support EMH. However, how to reconcile the random walk fitting with the random walk statistical tests still requires more studies in the future. More importantly how to apply the artificial intelligence (AI) to the random walk model to fit the stock index in order to decide whether EMH is valid demands many new studies.

Keywords: Big Data · Digital Economy · Russell 2000 Index · Random Walk · Stock Market

1 Introduction

Without any question, the stock markets are the engine to generate the big data, which become an active part of digital economy. The big data provide rich opportunities through internet to combine various technologies with the issues raised in the economic developments in modern society. The Russell 2000 index, which includes 2000 small-cap companies, gauges the US small cap market segment. It was created in 1984, and becomes

N. Radojević et al. (Eds.): ICAID 2022, AHIS 7, pp. 238–247, 2023.
https://doi.org/10.2991/978-94-6463-010-7_25

an indicator of wellbeing of small-cap companies accounting for around 85% of small cap assets in the US equity market.

Although the Russell 2000 index has a smaller capitalization compared with Dow Jones Industrial Average, Nasdaq and S&P 500 indices in the US equity market, it is widely studied by academic researchers and industry practitioners [1, 4, 5, 9, 10, 13].

Because the small-cap companies draw far less attention from big investors, the Russell 2000 index appears far less to be manipulated. Yet, each year the Russell 2000 index undergoes the annual reconstitution in May and June, this makes the heavy manipulation impossible. Moreover, the large coverage of the Russell 2000 index offers a better investment opportunity than the indices, which cover narrower ranges of stocks. Additionally, the Russell 2000 index could be considered as a counterpart of Dow Jones Industrial Average and S&P 500, which track the large-cap stocks.

Thus, this means that the Russell 2000 index has a better random characteristic than the other indices. This feature provides us with an opportunity to simulate/fit the Russell 2000 index with the random walk method. This type of simulation and fitting is important because the efficient market hypothesis (EMH) [3, 15] is supported by the random walk method to some degree [12]. The supporting evidence mainly comes from statistical tests, which include the unit root test, variance ratio test, autocorrelation test, and run test [2, 6, 7, 11]. With the advance in computational power, it is already possible to directly simulate/fit any stock index using the random walk method [17–21].

The advantage of random walk simulation/fitting is that the simulated results are highly visible whereas the statistical tests are invisible. The disadvantage of random walk simulation/fitting is that it is hard to conduct a statistical test.

In any case, the random walk method has yet to apply to the Russell 2000 index according to the literature search conducted prior to this study. Hence, the aim of the current study is to apply the random walk method to fit the Russell 2000 index for the first 20 years in the 21st century.

2 Data and Methods

2.1 Russell 2000 Index

The Russell 2000 index is downloaded from the Yahoo Finance [16], which includes the daily open, high, low, close, adjusted close prices, and volume. The Russell 2000 index close is used as a target for random walk fitting.

The Russell 2000 index for first 20 years in the 21st century includes 5032 trading days. To be consistent with our previous studies [17–21], we fragment these 20 years into five segments: (1) the segment from 2001 to 2020 includes 5032 trading days, (2) the segment from 2006 to 2020 includes 3776 trading days, (3) the segment from 2011 to 2020 includes 2517 trading days, (4) the segment from 2016 to 2020 includes 1259 trading days, and (5) the segment for 2020 includes 253 trading days.

2.2 Random Walk Method

The random walk method [8] comes from the tossing of a single coin. It was suggested that the coin tossing can construct two types of probabilities: (i) a single throw of many coins, which results in roughly around 0.5 probabilities for one face of coins, and (ii) continuous throws of a single coin, which results in a random walk, that is, many throws result in the same face of a coin up or down. In computational era, the process that a program continuously produces random numbers is a random walk. Graphically, a random walk is easily plotted in the *x, y* coordinates, where the *x*-axis is the throwing of a coin and *y*-axis is the accumulation of results of throwing a coin. In the case of random numbers generated by a computer program, the *x*-axis is a sequence, and the *y*-axis is the accumulation of random numbers. Because the classical random walk deals with 1/–1 representing each face of a coin, the generated random numbers by a computer program is usually ranged between 1 and –1, which are rounded to 1 and –1 for the random walk.

2.3 Russell 2000 Index as a Random Walk

In accordance with the random walk, the Russell 2000 index can be arranged in the same manner. Methodologically, we can compare the Russell 2000 index day by day, when the close value is larger than its close value in preceding day, we assign 1, otherwise –1. Thereafter, we add these 1s or –1s along the *x*-axis, which is the sequence of trading days, to construct the values for *y*-axis. If the up and down in the Russell 2000 index holds the random characteristic, then this sequence would be a random walk.

To take a step forwards, we can use a computer program to generate the random numbers in the range that is equal to the range of differences between trading days in the Russell 2000 index, and then add these random numbers along the time course, which will result in a random walk in the range of movement of the Russell 2000 index.

Table 1 explains how we work the Russell 2000 index along the above mentioned thoughts. The first two columns are the Russell 2000 index close values from January 2, 2020 to January 15, 2020. Column 3 is the comparison of the Russell 2000 index closes between two trading days in light of whether it is larger or smaller than the value in the preceding day in form of 1/–1. Column 4 is the addition of each cell in Column 3, and constructs a random walk of the Russell 2000 index. Column 5 is the random numbers generated by SigmaPlot (SPSS Inc., 1986–2001) with the seed of 3.09996 in the range between 1 and –1. Column 7 is the addition of each cell in Column 6, and constructs a random walk of a series of random numbers. Column 8 is the random numbers generated by SigmaPlot with any one of seeds ranged from 7.39101 to 7.39110 with the upper/lower ranges of standard deviations of the Russell 2000 index close in 2020. Column 9 is the addition of each cell in Column 8, and constructs a random walk of a series of random numbers together with the addition of the Russell 2000 index close values in Column 2.

In this way, we generate 100 000 random walks with different seeds to compare with Russell 2000 index close.

Table 1. The Russell 2000 index and random walk construction.

Date	Russell 2000 index Close	Comparison with Preceding Close	Random Walk	Generated Random Number	Comparison with Preceding Random Number	Random Walk	Generated Random Number	Random Index Close
Jan 2, 2020	1666.77			0.33946		0	1.712794	1666.77
Jan 3, 2020	1660.87	−1	−1	−0.523253	−1	−1	−27.737663	1639.03
Jan 6, 2020	1663.26	1	0	−0.650768	−1	−2	−19.2728791	1619.76
Jan 7, 2020	1653.31	−1	−1	−0.373050	1	−1	10.648337	1630.41
Jan 8, 2020	1663.59	1	0	−0.927587	−1	−2	−28.934176	1601.47
Jan 9, 2020	1664.99	1	1	0.648761	1	−1	−20.255107	1581.22
Jan 10, 2020	1657.64	−1	0	−0.987383	−1	−2	−31.467961	1549.75
Jan 13, 2020	1669.61	1	1	−0.731706	1	−1	21.265340	1571.02
Jan 14, 2020	1675.74	1	2	0.801500	1	0	−27.335864	1543.68
Jan 15, 2020	1682.4	1	3	−0.583511	−1	−1	−19.813709	1523.87

3 Results and Discussion

To be consistent with our previous studies [17–21], we limited this study to the Russell 2000 index for the first 20 years of the 21st century, but did not extend the data back to the 20th century. The consideration is that our studies actually fragment each index into five segments. In general, the more the data are involved, the worse the simulation/fitting is. This is understandable because the more the data are involved, the more the seeds should be selected for simulation/fitting. But we limited our study to one hundred thousand seeds, which naturally prevent the better performance for a longer period of time.

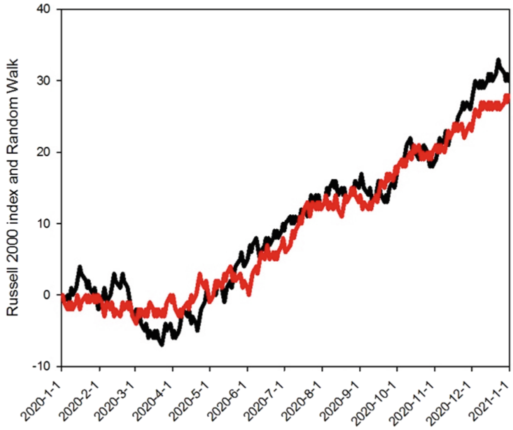

Fig. 1. The Russell 2000 index in 2020 in the form 1/–1 (black line) and its fitting (red line) generated by the random walk using the seed of 3.09996.

Figure 1 is the comparison between the Russell 2000 index in the form of 1/–1 and random walk fitting for 2020. As shown, the fitting is not only similar to but also very close to the trend of the Russell 2000 index in the form of 1/–1. This suggests that the random walk at least can approximately answer the question of whether the Russell 2000 index will be up or down in the following trading days retrospectively. However, one should be aware that a simple question of whether the Russell 2000 index will be up or down in the following trading days does not mean that we can actually and practically follow it because this form of 1/–1 movement does not consider the magnitude of each movement.

Figure 2 is a more realistic picture of the fitting of the Russell 2000 index in 2020. Comparing with Fig. 1, the fitting is quite good for the second half of 2020. This should implicate that the upper and lower ranges in the command for generation of random numbers are closely defined in terms of standard deviations of the Russell 2000 index in 2020. It is interesting to note that ten seeds from 7.39101 to 7.39110 generated the same result. This implies that we would not expect any improvement in fitting by furthermore increasing the digits in seed selection for this algorism. In Fig. 2, the directional departure between two lines just suggests that there are still non-random factors in the Russell 2000 index in the given period of time.

Figure 3 expands the fitting for another four years than what were done in Figs. 1 and 2. As demonstrated, the fitting clearly is not able to accommodate two extreme events in 2018 and 2020, which graphically generated two downhill sharp declines in the Russell 2000 index. Yet, the fitting also fails to catch up with the sharp uphill trend at the end of 2020. However, it is also notable that the fitting is quite reliable for the first two years. In this fitting, once again, we can find that several seeds generated the same result, which confirms what we observed in Fig. 2, and suggests no need to furthermore increase the number of seeds in a certain range. In fact, if we take a closer look at the Russell 2000 index over this period of time, we could say that it actually has two segments for the sake of the magnitude of its fluctuations. These two segments perhaps

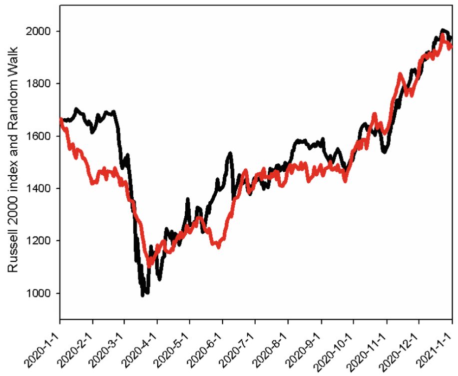

Fig. 2. The Russell 2000 index in 2020 (black line) and its fitting (red line) generated by the random walk using any one of ten seeds ranged from 7.39101 to 7.39110 with standard deviation of the Russell 2000 index in 2020 as upper and lower ranges.

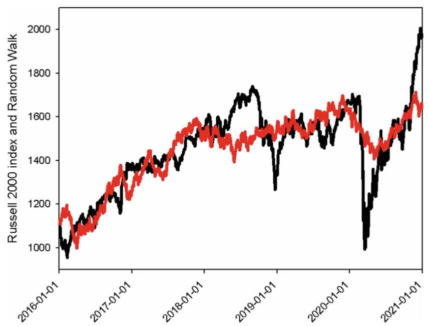

Fig. 3. The Russell 2000 index from 2016 to 2020 (black line) and its simulation (red line) generated by the random walk using any of six seeds from 6.02980 to 6.02985 with standard deviation of the Russell 2000 index from 2016 to 2020 as upper and lower ranges.

should be treated differently and separately in order to have a better fitting. Thus it is clear that there is no best seed, whose random numbers can fit the Russell 2000 index over all the time course.

Figure 4 prolonged the fitting for five years. Visibly, the fitting appears better than that in Fig. 3 because the relatively suitable fitting covered the period of time from 2011 to 2018. Actually, the Russell 2000 index quite stably increased over this period of time. Because of this stable trend, the dominant force plays a major role in determination of the upper and lower range of generated random numbers.

Figure 5 tells similar story along Figs. 2, 3 and 4, that is, the fitting becomes better because the sharp and sudden movements in the Russell 2000 index in 2018 and 2020 occupy far less fitting points. Therefore, the fitting can approximately follow the majority of the Russell 2000 index on the expense of failure of following the latest part of Russell 2000 index in 2020.

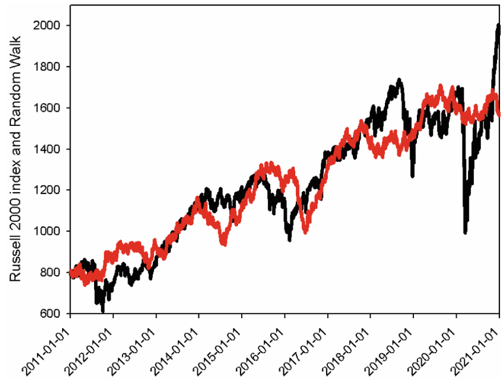

Fig. 4. The Russell 2000 index from 2011 to 2020 (black line) and its fitting (red line) generated by the random walk using any one of ten seeds ranged from 7.49042 to 7.49051 with standard deviation of the Russell 2000 index from 2011 to 2020 as upper and lower ranges.

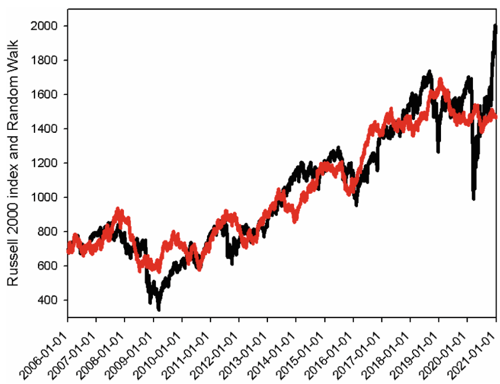

Fig. 5. The Russell 2000 index from 2006 to 2020 (black line) and its fitting (red line) generated by the random walk using any of two seeds, 3.12379 and 3.12380, with standard deviation of the Russell 2000 index from 2006 to 2020 as upper and lower ranges.

Figure 6 in reality appears no big difference from Fig. 5 although the data from 2001 to 2005 were added into the fitting. Figure 6 also gives us the impression that the fitting would look similar even if we incorporate more data into the fitting, which could be one of reasons that we did not expand our fittings into the 20th century.

As mentioned in Introduction, the weakness of random walk simulation/fitting is that it is hard to make statistical inference. An approach, which may be possible in the future, could be the bootstrap method for deriving confidence values. This should be set as an objective in our future studies.

Moreover, how to reconcile the random walk fitting with the random walk statistical tests still requires more studies in the future because the data that do not pass the statistical tests can be fitted by the random walk model and the results seem encouraging. The contradiction is that the small sample size can render a better random walk

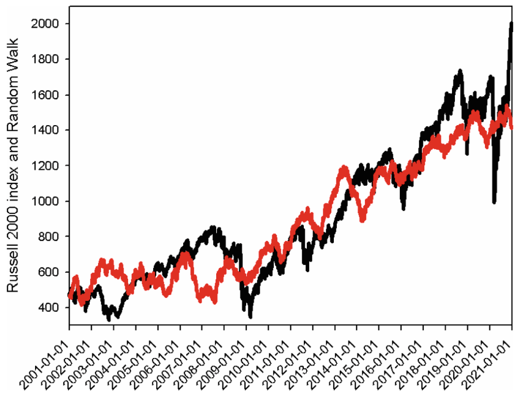

Fig. 6. The Russell 2000 index from 2001 to 2020 (black line) and its fitting (red line) generated by the random walk using one of two seeds, 3.74918 and 3.74919, with standard deviation of the Russell 2000 index from 2001 to 2020 as upper and lower ranges.

fitting/simulation but a poorer result from statistical tests, whereas the large sample size can give a worse random walk fitting/simulation but a better result from statistical tests. Clearly, more studies are needed to work along this direction.

To a broader sense, whether the movement of stock index is a random walk can be regarded as classification problem, which is currently intended to be solved by means of neural network, deep learning and artificial intelligence (AI). Therefore, the random walk model simulation/fitting should move to these directions in order to revive.

4 Conclusions

In this study, we conducted a random walk to fit the Russell 2000 index for the first 20 years in this century because the Russell 2000 index has a smaller capitalization with less manipulation. The results show that the random walk model can fit the Russell 2000 index in five different periods of time, and the fitting becomes better when more data were involved in the fitting because the sharp and sudden movements in the Russell 2000 index have less impact on the entire dataset. However, how to test the difference between the targeted data and fitted results to produce a confidence intervals, how to compare the fitting results with the statistical tests still require more studies in the future. More importantly how to apply the artificial intelligence (AI) to the random walk model to fit the stock index in order to decide whether EMH is valid demands many new studies.

Acknowledgements. The authors are thankful to the Scientific Development Fund of Guangxi Academy of Sciences (2021YFJ1203).

References

1. Abbasi W, Hájek P, Ismailova D, Yessimzhanova S, Khelifa ZB, Amonov K (2016) Kou Jump diffusion model: An application to the Standard and Poor 500, Nasdaq 100 and Russell 2000 index options. Int J Econ Fin 6:1918–1929
2. Aktan C, Iren P, Omay T (2019) Market development and market efficiency: evidence based on nonlinear panel unit root tests. Euro J Finan 25:979–993
3. Boya CM (2019) From efficient markets to adaptive markets: evidence from the French stock exchange. Res Int Bus Finan 49:156–165
4. Cai J, Houge T (2008) Long-term impact of Russell 2000 index rebalancing. Fin Anal J 64:76–91
5. Chen H-L (2006) On Russell index reconstitution. Rev Quan Fin Account 26:409–430
6. Chow KV, Denning KC (1993) A simple multiple variance ratio test. J Econometr 58:385–401
7. Deo RS, Richardson M (2003) On the asymptotic power of the variance ratio test. Economet Theor 19:231–239
8. Feller W (1968). An introduction to probability theory and its applications, 3rd edn. Wiley, New York
9. Kapadia N, Szado E (2011). 15 Years of the Russell 2000 buy-write. SSRN Electron J. https://doi.org/10.2139/ssrn.1928822
10. Keim DB (1999) An analysis of mutual fund design: the case of investing in small-cap stocks. J Fin Econ 51:173–194
11. Liu CY, He J (1991) A variance ratio tests of random walks in foreign exchange rates. J Fin 46:773–785
12. Lo AW, MacKinlay AC (1988) Stock market prices do not follow random walks: evidence from a simple specification test. Rev Fin Stud 1:41–66
13. Madhavan A (2003) The Russell reconstitution effect. Fin Anal J 59:51–64
14. SPSS Inc., SigmaPlot 2002 for Windows Version 8.02 (1986–2001)
15. Urquhart A, McGroarty F (2016) Are stock markets really efficient? Evidence of the adaptive market hypothesis. Int Rev Finan Anal 47:39–49
16. Yahoo Finance (2021). https://de.finance.yahoo.com/
17. Yan S, Wu G (2011) Fitting of SSEC index (Shanghai composite) from January 2000 to July 2010 using random walk model. Guangxi Sci 18:92–96
18. Yan S, Wu G (2020) Fitting of impact of COVID-19 pandemic on S&P 500 Index using random walk. In Conference publishing services: ICEMME, vol 4, pp 1007–1011
19. Yan S, Wu G (2021a) Fit Hang Seng index for 21st century with random walk model. In Conference publication series: CBFD, pp 66–69
20. Yan S, Wu G (2021b) Simulation of impact of COVID-19 pandemic on Dow Jones index using random walk. In Proceedings in the 2nd international conference on computing and data science (CONFCDS), pp 59–64
21. Yan S, Wu G (2021c) Simulation of NIKKEI 225 index for 21st century using random walk. In 2021 2nd international conference on big data and informatization education (ICBDIE), pp 330–333