Simulation of Gold Price for the First 20 Years in the 21st Century with Random Walk Model

Shaomin Yan[1†], Guang Wu[1*]

*Corresponding author: e-mail: †shaominyan@gxas.cn, *hongguanglishibahao@gxas.cn

[1]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

Abstract—Whether or not a stock index can be interpreted by the random walk model is important but debatable because it is closely linked to the weak form of the efficient market hypothesis (EMH). As an important commodity, a precious metal, and the reserve in central banks, the gold price is the objective of numerous studies. The issue of whether the gold price is subject to the weak form of EMH is interesting and important. Although the gold price is the oldest recorded price in human history, we select the first 20 years in the 21st century for this report to be consistent with our previous studies. In this report, the random walk model was used to simulate the gold daily close price because this task was previously approached using statistical tests. Our focus concentrates on the simulation in the decimal configuration, and the results show that the simulation becomes worse with increase of the number of years involved in simulation.

Keywords-Gold price; random walk; simulation; finance and trade; stock market

1. INTRODUCTION

Whether or not a stock index can be interpreted by the random walk model is important but debatable because it is closely linked to the efficient market hypothesis (EMH), which classifies: (i) the strong form relying on private information, (ii) the semi-strong form relying on publicly available information, and (iii) the weak form relying on historical prices, with which the random walk model is concerned [1].

Actually not only the stock indices and their derivates are subject to the EMH studies [2], the foreign currency exchange rates are the objective of EMH studies [3]. As an important commodity, a precious metal, and the reserve in central banks, the gold price is also the aim of many studies including the random walk studies [4, 5].

In the works to test the random walk hypothesis, the statistical tests are widely applied to historical data, for example, the root test [6], run test [7], variance ratio test [8, 9], multiple variance ratio test [10]. From a purely statistical viewpoint, these tests can be divided into the parametric such as autocorrelation test, and non-parametric such as run test.

ICIDC 2022, June 17-19, Qingdao, People's Republic of China

DOI 10.4108/eai.17-6-2022.2322799

Over recent years, we conducted several studies to use the random work model directly simulate/fit the stock indices [11-15]. This approach, though the computational time is huge, does reveal the randomness in stock indices from a viewpoint different from statistical tests. Clearly, we should not limit our effort to stock indices.

In this report, we expand the random walk simulation to the gold price in order to add more pieces of evidence to the discussion on the weak form of EMH.

2. Historical Data and Random Walk Model

2.1 Historical data

The gold price is the oldest recorded price in human history. Due to the heavy computational time and being consistent with our studies on stock indices [11-15], only the data for the first 20 years in the 21st century are used in this report. The daily gold price was downloaded from Yahoo Finance [16]. Indeed, another source of the gold price, which has a longer history than Yahoo Finance does, is the London Bullion Market Association [17], but we prefer to the data in Yahoo Finance in order to be consistent with our studies on stock indices [11-15], whose data were also downloaded from Yahoo Finance.

The data are composed of daily open, high, low, close, adjusted close, and volume. For our purpose, we use only the daily close. The period from January 2, 2001 to December 31, 2020 has 4984 trading days. The period from January 3, 2006 to December 31, 2020 has 3741 trading days. The period from January 3, 2011 to December 31, 2020 has 2491 trading days. The period from January 4, 2016 to December 31, 2020 has 1246 trading days. The period from January 2 to December 31, 2020 has 251 trading days.

2.2 Random Walk Model

Random walk practically is a path generated by tossing a fair coin continuously over time [18]. As one side of coin and the other side are defined as 1 and –1, we can record either 1 or –1 following each tossing. The addition of the record results in a random walk along the time course. As the turnout of tossing of a coin is a random event, a series of 1 or –1 is also a series of random events.

2.3 Gold Price in the Configuration of A Random Walk

The gold price was set into a configuration, which is similar to the abovementioned random walk. For this purpose, we compare a gold price with that in its previous trading day, if it is higher or lower, then we record 1 or –1 correspondingly, and then we add this series of 1 or –1 along the time course. This addition is exactly a random walk.

2.4 Random Walk in the Decimal Configuration

Clearly, the random walk in the 1 or –1 configuration is no longer suitable for the current research demand. Therefore, we design a random walk in the decimal configuration to simulate the gold price, which is the decimal configuration. This is easily done with the algorithm, which generates random numbers in the decimal configuration.

2.5 Simulation

The simulations were performed by using a computer to generate multiple series of random numbers with different seeds, constructing the random walks based on these series of random numbers, and comparing each random walk with the gold price close in both 1/–1 and decimal configurations. Eventually, we can find the seed whose random walk is most similar to the gold price close.

TABLE 1. PROCEDURE TO CONSTRUCT A RANDOM WALK SIMULATION

Date	Gold Price Close	Compare Preceding Close	Random Walk in 1 or–1 Configuration	Generated Random Number	Compare Preceding Random Number	Random Walk in 1 or–1 Configuration	Generated Random Number	Random Walk in Decimal Configuration
Jan 2, 2020	1524.5		0	–0.56759		0	17.7517	1524.5
Jan 3, 2020	1549.2	1	1	0.10146	1	1	–18.9686	1505.5
Jan 6, 2020	1566.2	1	2	0.12706	1	2	–23.1902	1482.4
Jan 7, 2020	1571.8	1	3	0.81422	1	3	8.37399	1490.8
Jan 8, 2020	1557.4	–1	2	–0.32088	–1	2	20.99567	1511.8
Jan 9, 2020	1551.7	–1	1	0.03117	1	3	13.54696	1525.3
Jan 10, 2020	1557.5	1	2	0.66874	1	4	23.02406	1548.4
Jan 13, 2020	1548.4	–1	1	–0.90877	–1	3	0.1063	1548.5
Jan 14, 2020	1542.4	–1	0	–0.33833	1	4	–5.15531	1543.3
Jan 15, 2020	1552.1	1	1	–0.59799	–1	3	–1.34943	1542.0

3. RESULTS AND DISCUSSION

Table 1 explains how we conduct the random walk simulation in both 1/–1 and decimal configurations. Columns 1and 2 include the date and gold price close for the beginning of 2020. Column 3 records whether the gold price is higher or lower than that in its preceding day in the 1/–1 configuration. For instance, $ 1549.2, the gold price close on January 3, 2020 is higher than $ 1524.5, the gold price close on January 2, 2020, so 1 was recorded in cell 2, column 3. Column 4 is the addition of each datum in column 3, and creates the gold price in the 1/–1 configuration. Column 5 is the generated random numbers by a computer. Column 6 records whether the generated random number is larger or smaller than its preceding random number in the 1/–1 configuration. Column 7 is the addition of each cell in column 6, and creates a random walk simulation for comparison with column 4. The last two columns manifest how we construct a random walk simulation in the decimal configuration. Column 8 is the random numbers generated with the elaboration of the gold price close in 2020, because the command for generation of random numbers usually includes 4 terms, i.e. a seed, the number to be generated, and upper and lower ranges, of which the best seed is what we are looking for, the upper and lower ranges are difficult to determine. According to our previous studies [11-15], we found that the standard deviation is suitable, so we use it to generate the random numbers for the simulations in the decimal configuration. Column 9 is the random walk simulation by adding each random number in column 8 to the corresponding gold price close value in column 9.

Technically, Figure 1 follows the procedure in column 2 (black line) and column 7 (red line). Figures 2 to 5 follows the procedure in column 2 (black line) and the last two columns (red line).

The black line in Figure 1 is the gold price in the 1/–1 configuration. Because the black line is going up continuously, there are more trading days whose gold price is higher than that in its previous day. The random walk simulation can meet such an almost monotonic uptrend.

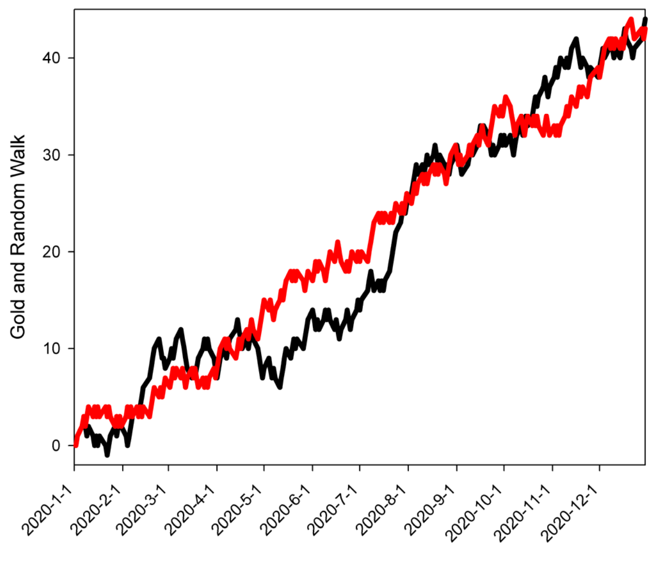

Figure 1. The gold price in 2020 in the 1/ –1 configuration (black line) and its random walk simulationin in the 1/ –1 configuration (red line) using the seed of 0.78654.

In Figure 2, the random walk simulation turns to the decimal configuration, because theoretically the random walk simulation in the 1/–1 configuration always has a chance for a perfect fit. As seen in Figure 2, the simulation is relatively good although the Covid-19 impact in mid March cannot be fully simulated.

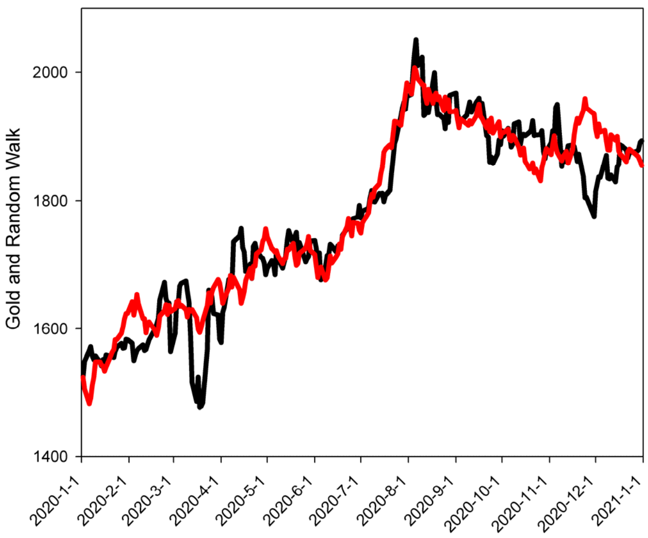

Figure 2. The gold price in 2020 (black line) and its random walk simulation (red line) in the decimal configuration using any of five seeds from 5.50369 to 5.30373 with an increment of 0.00001.

For Figure 3, the simulation appears good for the general trend although two segments of the simulation from 2016 to 2017 and from mid-2018 to mid-2019 are on the opposite sides. Comparison with the simulations on stock indices, it seems that the impact of Covid-19 is not as strong as that on stock indices. This perception should be verified in our future studies.

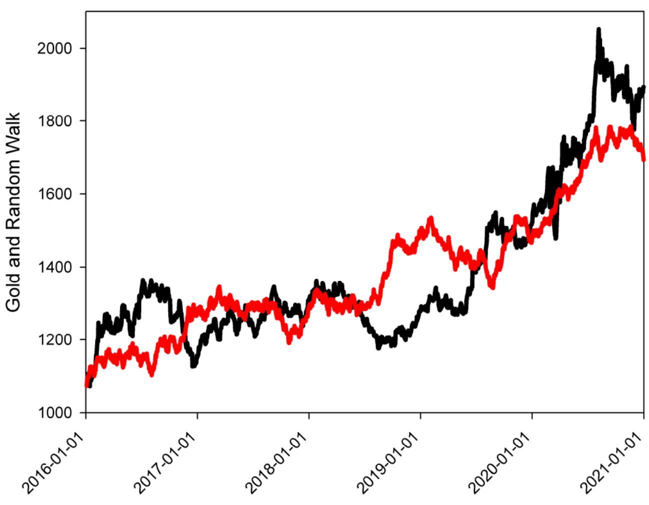

Figure 3. The gold price from 2016 to 2020 (black line) and its random walk simulation (red line) generated in the decimal configuration using any of eight seeds from 6.95463 to 6.95470 with an increment of 0.00001.

It is hard to say whether the simulation in Figure 4 is better or worse than that in Figure 3, because there is only one segment from 2013 to 2015 that the simulation is on the opposite side, and the most part of simulation follows the gold price quite closely. But the simulation fails to catch up with the peak in 2020. As our experience increases, we feel that it may need to use different numbers of simulations for different periods of the gold prices. In our cases, we use initially one million seeds to generate one million simulations for all periods of gold prices. After finding that the interval between sequential seeds is too small to distinguish two sequential simulations, we apply 100 000 seeds to generate 100 000 simulations for all periods of gold prices. But the effect should be different when using 100 000 seeds for 2020, which has 251 trading days and the period from 2001 to 2020, which has 4984 trading days. More likely, we need to choose a proportional number of seeds with respect to the length of time.

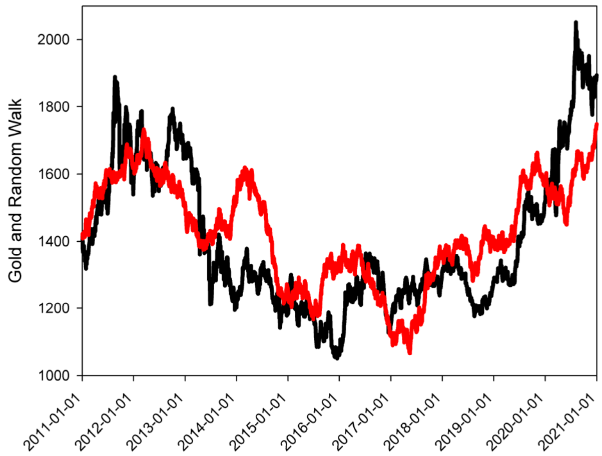

Figure 4. The gold price from 2011 to 2020 (black line) and its random walk simulation (red line) in the decimal configuration using any of two seeds of 3.07944 and 3.07945.

Figure 5 shows that the simulation reaches to a plateau from as early as 2010. Indeed, there is a plateau in gold prices from 2013 to 2020, but the simulation does not pick up for the peak in 2020. Highly likely, the first uphill in 2010 is averaged for its following years. In the view of Figure 5, the simulation is a failure although it could be better if we had taken more computational time to simulate it with more seeds as mentioned in the above paragraph.

Figure 6 actually is the prolonged simulation in Figure 5 because it follows the uptrend from 2001 to 2015, and then enters a plateau to the end of simulation. As the simulation did in Figure 5, the simulation in Figure 6 completely ignores two peaks from 2009 to 2013 and in 2020. This is very disappointing for finding a reasonable seed in limited and acceptable computational time. On the other hand, this also shows the difficulty for large-scale random walk simulations in terms of seeds, which can generate really different profiles.

It is not very clear how many samples, i.e. the sample size, are used in statistical tests to determine whether a stock index or a commodity price can be classified as a random walk. Although no statistical test is sample-size dependent, there could be a period of time whose gold price is either more suitable or less suitable due to the nature of simulation

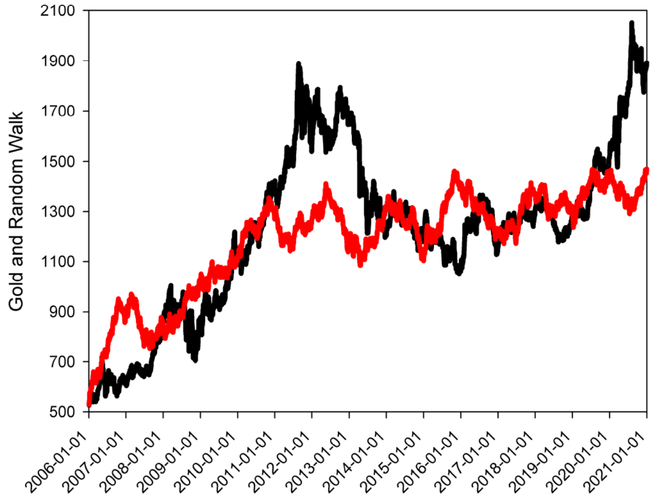

Figure 5. The gold price from 2006 to 2020 (black line) and its random walk simulation (red line) in the decimal configuration using any of six seeds from 5.45096 to 5.45101 with an increment of 0.00001.

In this report, we use the term, configuration, instead of form, format, and type. This modification makes our studies inconsistent and bizarre. But this is due to the fact that the wide application of anti-plagiarism software packages to check the manuscripts to make the authors impossible to use the same scientific terms to describe the same thing to work on different subjects, also make the English wording in manuscripts exhausted.

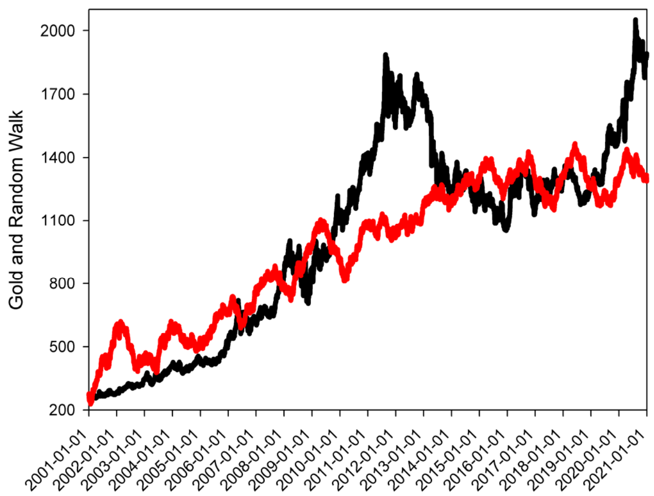

Figure 6. The gold price from 2001 to 2020 (black line) and its random walk simulation (red line) in the decimal configuration using a seed of 1.81432.

In our previous studies [11-15], the classical random walk is referred to the 1/–1 configuration, which should be corresponded to the statistical tests used for the weak form of EMH. Perhaps these tests need to expand to the decimal configuration.

4. CONCLUSION

In this report, we continue our efforts to use a random walk model to simulate a market dataset to answer the question of whether an important commodity, the gold price, follows the weak form of EMH. The results show that the random walk model can satisfactorily simulate the gold price for a short period of time, especially for the random walk in the 1/–1 configuration. For the long-term simulation, the issue of whether the increase of the number of simulations can be helpful is yet to know. This should be the topic in our future studies.

Acknowledgment: The authors gratefully acknowledge the Scientific Development Fund of Guangxi Academy of Sciences (2021YFJ1203).

REFERENCES

[1] A. Abraham, F. J. Seyyed, and S. A. Alsakran, “Testing the random walk behavior and efficiency of the Gulf stock markets,” Finan. Rev., vol. 37, pp. 469-480, 2002.

[2] H. Claessen, and S. Mittnik, “Forecasting stock market volatility and the informational efficiency of the DAX-index options market,” Euro. J. Finan, vol. 8, pp. 302–321, 2002.

[3] R. Ahmad, S. G. Rhee, and Y. M. Wong, “Foreign exchange market efficiency under recent crises: Asia Pacific focus,” J. Int. Money Finan., vol. 31, pp. 1-26, 2012.

[4] J. P. Botha, “The random walk model and the behaviour of gold prices: a note,” Invest. Analy. J., vol. 9, pp. 9-17, 1980.

[5] G. Smith, “Tests of the random walk hypothesis for London gold prices,” Appl. Econ. Lett., vol. 9, pp. 671-674, 2002.

[6] C. Aktan, P. Iren, and T. Omay. “Market development and market efficiency: evidence based on nonlinear panel unit root tests,” Euro J Finan, vol. 25, pp. 979-993, 2019.

[7] J. Mahmoud, and Z. Hussien, “Testing the weak form efficiency of Amman stock exchange,” Int. Res. J. Finan. Econ., vol. 65, pp. 93-97, 2011.

[8] K. V. Chow, and K. C. Denning, “A simple multiple variance ratio test.” J. Econometrics, vol. 58, pp. 385–401, 1993.

[9] R. S. Deo, and M. Richardson, “On the asymptotic power of the variance ratio test,” Econometric Theory, vol. 19, pp. 231–239, 2003.

[10] K. V. Chow, and K. C. Denning, “A simple multiple variance ratio test.” J. Econometrics, vol. 58, pp. 385–401, 1993.

[11] S. Yan, and G. Wu, “Fitting of SSEC index (Shanghai composite) from January 2000 to July 2010 using random walk model,” Guangxi Sci vol. 18, pp. 92–96, 2011.

[12] S. Yan, and G. Wu, “Fitting of impact of COVID-19 pandemic on S&P 500 Index using random walk,” Conference Publishing Services: ICEMME vol. 4, pp. 1007–1011, 2020.

[13] S. Yan, and G. Wu, "Simulation of NIKKEI 225 index for 21st century using random walk," 2nd International Conference on Big Data and Informatization Education (ICBDIE), pp. 330-333, 2020.
[14] S. Yan, and G. Wu, "Fit Hang Seng index for 21st century with random walk model," Conference Publication Series: CBFD pp. 66–69, 2021.
[15] S. Yan, and G. Wu, "Simulation of NASDAQ Composite Index for the first 20 years of the 21st century by means of random walk model," Conference Publishing Services: EIMSS, pp. 454-457, 2021.
[16] https://de.finance.yahoo.com/
[17] https://www.lbma.org.uk/
[18] W. Feller, An Introduction to Probability Theory and its Applications. 3rd ed., New York: John Wiley, 1968.