QUANTITATIVE METHODS IN ECONOMICS
Volume XIX, No. 4, 2018, pp. 339 – 346

Received: 19.07.2018
Accepted: 21.12.2018

SOME PROPOSAL OF THE TEST FOR A RANDOM WALK DETECTION AND ITS APPLICATION IN THE STOCK MARKET DATA ANALYSIS

Marcin Dudziński https://orcid.org/0000-0003-4242-8411
Konrad Furmańczyk https://orcid.org/0000-0002-7683-4787
Arkadiusz Orłowski https://orcid.org/0000-0002-6755-1830
Faculty of Applied Informatics and Mathematics
Warsaw University of Life Sciences – SGGW, Poland
e-mail: marcin_dudzinski@sggw.pl; konrad_furmanczyk@sggw.pl; arkadiusz_orlowski@sggw.pl

Abstract: According to the numerous groups of theoreticians and practitioners, who act in the area of financial markets, changes in the stock prices are random and it is almost infeasible to predict them correctly using historical data. This approach is based on the random walk theory, which states that the price of financial instrument in the subsequent time point is the sum of its price in the previous time point and some random variable with a finite variance, i.e. it is modeled with the use of a stochastic process called a random walk. The random walk hypothesis stands in contradiction to the beliefs of the ordinary technical analysis followers, where the prediction is carried out on the grounds of existing trends, and furthermore, this hypothesis regards such a modeling of financial markets as incorrect. In our work, we construct statistical test for a random walk detection, which is based on the first arcsine law. We also present simulation results that allow to check the quality of the proposed test, as well as we show the application of the introduced test in the stock exchange data analysis.

Keywords: random walk, arcsine law, test for a random walk detection, stock market data analysis

JEL classification: C10, C12, C15, C19

https://doi.org/10.22630/MIBE.2018.19.4.32

INTRODUCTION

Random walk theory states that the price of financial instrument in the subsequent time point is the sum of its price in the previous time point and some random variable with a finite variance, i.e. it is modeled with the help of a stochastic process called a random walk. We say that a stochastic process $(Y_0, Y_1, \dots, Y_n, \dots)$ is a random walk (RW), if the following relations hold (see [Żak 2012]):

$$Y_0 = y_0,$$
$$Y_1 = y_0 + \varepsilon_1,$$
$$Y_2 = y_0 + \varepsilon_1 + \varepsilon_2,$$
$$\vdots$$
$$Y_n = y_0 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n,$$

where $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, \dots$ form a sequence of independent and identically distributed (iid) random variables with a finite variance.

In our considerations, we deal with a symmetric Gaussian RW. Thus, we assume that $\varepsilon_i's$ have a standard normal distribution ($\varepsilon_i \sim N(0;1)$), and additionally that $y_0 = 0$. Thus, our RW process may be written as follows:

Thus, our RW process may be written as follows:

$$Y_t = \sum_{i=1}^{t} \varepsilon_i, \;\; t = 1,2,\dots, \text{ where } \varepsilon_i - \text{iid} \sim N(0;1).$$

Consequently, we may write that:

$$Y_t = Y_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t - \text{iid} \sim N(0;1),\; t = 1,2,\dots \tag{1}$$

Obviously, the relation in (1) is equivalent to:

$$\varepsilon_t = Y_t - Y_{t-1}, \text{ where } \varepsilon_t - \text{iid} \sim N(0;1),\; t = 1,2,\dots \tag{2}$$

From the economical point of view the RW process (Y_t), $t = 1,2,\dots,$ may be interpreted as follows. Namely, if Y_t stands for the stock price in time t, then the difference between the stock prices in the periods t and $t-1$ – given by the relation in (2) – determines an increase or a decrease in the stock price in the period between $t-1$ and t.

As far as we are concerned, there are not many statistical tests devoted to a random walk identification. To the best of our knowledge, the main tools that have been applied in this context so far are the two celebrated tests – an augmented Dickey-Fuller test (see [Dickey, Fuller 1979]) and the Runs test, and through our work, we attempt to fill in a gap related to this field of research.

Our primary objective is to construct a new test for a random walk detection, the idea of which is based on the so-called first arsine theorem. We also check the quality of the proposed test by conducting some simulation studies. For more detailed information regarding the theory of RW processes, we refer to [Maddala 2001] and [Montanari 1997].

Our paper is organized as follows. In Section 1, we present a general idea leading to the construction of our test for a random walk identification. In Section 2, we describe the construction of our test. Additionally, in Section 3 we check the efficiency of the introduced test, while in Section 4 we apply our test procedure to a dataset containing the daily returns of the WIG index at the Warsaw Stock Exchange. Finally, Section 5 summarizes our study.

A GENERAL IDEA OF THE PROPOSED TEST

Let (Y_t) denote a stochastic process given by:

$$Y_0 = 0,\ Y_t = Y_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t - \text{iid} \sim N(0;1),\ t = 1,2,\ldots$$

We define a sequence $(\tilde{\varepsilon}_t)$ as follows:

$$\tilde{\varepsilon}_t = \begin{cases} +1, & \text{if } \quad \varepsilon_t = Y_t - Y_{t-1} > 0, \\ -1, & \text{if } \quad \varepsilon_t = Y_t - Y_{t-1} \le 0. \end{cases}$$

Since $\varepsilon_t \sim N(0;1)$, we have that $\varepsilon_t{}'$s are symmetric and so are $\tilde{\varepsilon}_t{}'$s. Consequently, a sequence $(\tilde{\varepsilon}_t)$ is distributed in the way as below:

$$P(\tilde{\varepsilon}_t = +1) = P(\tilde{\varepsilon}_t = -1) = 1/2,\ t = 1,2,\ldots$$

Let moreover, a sequence $(\tilde{Y}_t)$ be such that:

$$\tilde{Y}_t = \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2 + \cdots + \tilde{\varepsilon}_{t-1} + \tilde{\varepsilon}_t = \tilde{Y}_{t-1} + \tilde{\varepsilon}_t,\ t = 1,2,\ldots$$

The construction of our test is strictly connected with *the first arcsine law*, which states that if we denote by T_n^+ the proportion of those among $(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_n)$ (i.e. among $(\tilde{\varepsilon}_1, \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2 + \cdots + \tilde{\varepsilon}_n)$), which are non-negative, then the following property is satisfied (see [Qiang, Jiajin 2018] and [Feller 1968, 79-82], for further details):

$$P(T_n^+ \le x) \to \frac{2}{\pi} arc\, sin\left(\sqrt{x}\right) \text{ for any } x \epsilon \langle 0;1 \rangle,\ as\ n \to \infty. \tag{3}$$

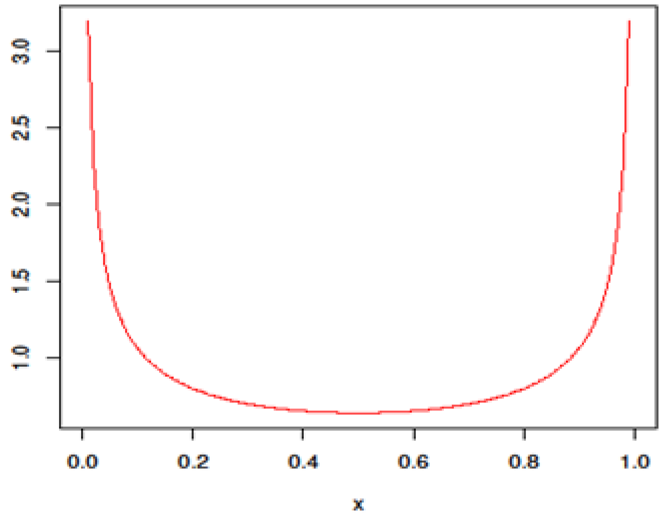
Thus, we may say that a sequence of proportions (T_n^+) converges in distribution to the arcsine distribution. The cumulative distribution function (cdf) F and the density function (df) f, of the arcsine distribution, are given by the formulas, respectively:

$$F(x) = \begin{cases} 0, & \text{if } \quad x < 0, \\ \frac{2}{\pi} arc\, sin\left(\sqrt{x}\right), & \text{if } \quad x \epsilon \langle 0;1 \rangle, \\ 1, & \text{if } \quad x > 1; \end{cases} \tag{4}$$

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}} \mathbf{I}_{(0;1)} = \begin{cases} \frac{1}{\pi\sqrt{x(1-x)}}, & \text{if } \quad x \epsilon (0;1), \\ 0, & \text{if } \quad x \notin (0;1). \end{cases} \tag{5}$$

The density f in the interval $(0;1)$ is depicted in the figure below:

Figure 1. The density function of the arcsine distribution in $(0; 1)$



Source: own elaboration

Conclusion (3), of the first arcsine law, may practically be used for $n \geq 20$, which means that:

$$P(T_n^+ \leq x) \approx \frac{2}{\pi} arc\, sin(\sqrt{x}) \text{ for } x \epsilon \langle 0; 1 \rangle, \text{ if } n \geq 20. \quad (6)$$

It intuitively seems that the probability of reaching the value close to 0.5 by T_n^+ is the largest. However, the truth is quite the opposite, namely: the values close to 0.5 are the least probable and the most probable values for T_n^+ are close to 0 or 1 (see [Feller 1966, 81-82]). It may be seen from the graph of the density f in the figure above. Thus, if we denote by α the significance level of the test:

$$H_0\colon (\tilde{Y}_t) \text{ and consequently also } (Y_t) \text{ form the random walk processes,} \quad (7)$$

we look for a critical area (or otherwise, a set of rejections) of the form $K_{c(\alpha)} = (0.5 - c(\alpha);\ 0.5 + c(\alpha))$, where $c(\alpha)$ is a certain number from the interval $(0;\ 1)$. Therefore, we may write that:

$$P(T_n^+ \in K_{c(\alpha)}) \approx \int_{0.5-c(\alpha)}^{0.5+c(\alpha)} f(x)\, dx = \alpha, \text{ if } n \geq 20,$$

which, due to the formula for a density f in (5) and the fact that $(0.5 - c(\alpha);\ 0.5 + c(\alpha)) \subset (0; 1)$, is equivalent to:

$$P(T_n^+ \in K_{c(\alpha)}) \approx \int_{0.5-c(\alpha)}^{0.5+c(\alpha)} \frac{1}{\pi\sqrt{x(1-x)}} dx = \alpha, \text{ if } n \geq 20. \quad (8)$$

It follows from (8) that:

$$\int_{0.5-c(\alpha)}^{0.5+c(\alpha)} \frac{1}{\pi\sqrt{x(1-x)}} dx = \frac{2}{\pi} arc\, sin(\sqrt{x})\Big|_{0.5-c(\alpha)}^{0.5+c(\alpha)} = \alpha. \quad (9)$$

In view of (9), we have the following condition on $c(\alpha)$:

$$arc\ sin\sqrt{0.5 + c(\alpha)} - arc\ sin\sqrt{0.5 - c(\alpha)} = \frac{\pi\alpha}{2}. \tag{10}$$

The values of $c(\alpha)$, calculated numerically according to (10) for the chosen significance levels, are collected in the following table:

Table 1. Values of $c(\alpha)$ for some selected α

α	0.01	0.05	0.1
$c(\alpha)$	0.008	0.039	0.078

Source: own elaboration

Thus, for the significance level $\alpha = 0.05$, we obtain the following critical area of the test in (7):

$$K_{c(0.05)} = (0.5 - 0.039;\ 0.5 + 0.039) = (0.461;\ 0.539). \tag{11}$$

The presented idea plays a key role in the construction of our test.

BRIEF DESCRIPTION OF THE PROPOSED TEST

Below, we present the idea of our test for a random walk detection. The notations for $\tilde{\varepsilon}_t$ and $\tilde{Y}_t$ from Section 1 will also be used throughout the current section of our paper.

As has already been mentioned in Section 1, the first arcsine law states that, for sufficiently large n, a sequence of proportions (T_n^+), where – for recollection:

$$T_n^+ = \#\{1 \le t \le n\colon \tilde{Y}_t \ge 0\}/n = \#\{1 \le t \le n\colon \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2 + \cdots + \tilde{\varepsilon}_t \ge 0\}/n,$$

has asymptotically arcsine distribution.

It follows from (8) and (11) that

$$P\left(T_n^+ \in K_{c(0.05)}\right) = P\left(T_n^+ \in (0.461;\ 0.539)\right) \approx \int_{0.461}^{0.539} \frac{1}{\pi\sqrt{x(1-x)}}\,dx = 0.05, \text{ if } n \ge 20.$$

Thus, provided $n \ge 20$, T_n^+ is a test statistic of our test and therefore, if the calculated value of T_n^+ belongs to the interval $K_{c(0.05)} = (0.461;\ 0.539)$, we reject the hypothesis H_0 from (6). Obviously, if this value does not belong to $(0.461;\ 0.539)$, we accept the hypothesis H_0 stating that $(\tilde{Y}_t)$ and (Y_t) are the random walk processes.

QUALITY OF THE PROPOSED TEST

In order to check the quality of the proposed test, we carried out the following empirical studies. Firstly, we generated 1000 samples $\boldsymbol{y}^{(i)}$, each of the size 1000, according to the model of the investigated process $(Y) = (Y_t)$:

$$y^{(1)} = \left(y_1^{(1)}, \dots, y_{1000}^{(1)}\right), y^{(2)} = \left(y_1^{(2)}, \dots, y_{1000}^{(2)}\right), \dots,$$
$$y^{(1000)} = \left(y_1^{(1000)}, \dots, y_{1000}^{(1000)}\right).$$

Secondly, based on the given samples $y^{(i)}$, we obtained 1000 samples $\varepsilon^{(i)}$:

$$\varepsilon^{(1)} = \left(\varepsilon_1^{(1)}, \dots, \varepsilon_{1000}^{(1)}\right),\ \varepsilon^{(2)} = \left(\varepsilon_1^{(2)}, \dots, \varepsilon_{1000}^{(2)}\right), \dots,$$
$$\varepsilon^{(1000)} = \left(\varepsilon_1^{(1000)}, \dots, \varepsilon_{1000}^{(1000)}\right),$$

where, for $i = 1, \dots, 1000$:

$$\varepsilon_t^{(i)} = \begin{cases} y_t^{(i)} - y_{t-1}^{(i)}, \text{ if } t = 2, \dots, 1000, \\ y_t^{(i)}, \text{ if } t = 1. \end{cases}$$

Subsequently, based on the given samples $\varepsilon^{(i)}$, we obtained 1000 samples $\tilde{\varepsilon}^{(i)}$:

$$\tilde{\varepsilon}^{(1)} = \left(\tilde{\varepsilon}_1^{(1)}, \dots, \tilde{\varepsilon}_{1000}^{(1)}\right),\ \tilde{\varepsilon}^{(2)} = \left(\tilde{\varepsilon}_1^{(2)}, \dots, \tilde{\varepsilon}_{1000}^{(2)}\right), \dots,$$
$$\tilde{\varepsilon}^{(1000)} = \left(\tilde{\varepsilon}_1^{(1000)}, \dots, \tilde{\varepsilon}_{1000}^{(1000)}\right),$$

where, for $i = 1, \dots, 1000$:

$$\tilde{\varepsilon}_t^{(i)} = \begin{cases} +1, & \text{if } \varepsilon_t^{(i)} > 0, \\ -1, & \text{if } \varepsilon_t^{(i)} \le 0. \end{cases}$$

Next, we calculated the fractions:

$$T_{1000}^{+(1)} = \#\left\{1 \le t \le 1000: \tilde{\varepsilon}_1^{(1)} + \tilde{\varepsilon}_2^{(1)} + \dots + \tilde{\varepsilon}_t^{(1)} \ge 0\right\}/1000,$$
$$T_{1000}^{+(2)} = \#\left\{1 \le t \le 1000: \tilde{\varepsilon}_1^{(2)} + \tilde{\varepsilon}_2^{(2)} + \dots + \tilde{\varepsilon}_t^{(2)} \ge 0\right\}/1000,$$
$$\vdots$$
$$T_{1000}^{+(1000)} = \#\left\{1 \le t \le 1000: \tilde{\varepsilon}_1^{(1000)} + \tilde{\varepsilon}_2^{(1000)} + \dots + \tilde{\varepsilon}_t^{(1000)} \ge 0\right\}/1000,$$

and finally, we checked the number of those among the values $T_{1000}^{+(i)}$, $i = 1, \dots, 1000$, satisfying the condition $T_{1000}^{+(i)} \,\epsilon\, K_{c(0.05)} = (0.461;\ 0.539)$, which is a number of rejections among 1000 iterations of our test procedure. It enabled us to investigate the quality of the proposed test. We analyzed it thoroughly in Section 3 of our work.

We verify the quality of our test for a random walk identification with the use of the following models: M1 – a Gaussian random walk, M2 – ARIMA(1,0,0) with: $ar = 0.99$, $\sigma = 3$, M3 – ARIMA(1,0,0) with: $ar = 0.999$, $\sigma = 3$.

Applying the R package, we carried out 1000 iterations of our test procedure and then, we computed the numbers of rejections of H_0 among those iterations, i.e. the number of those among $T_{1000}^{+(i)}$, $i = 1, \dots, 1000$, which belong to $K_{c(0.05)} = (0.461;\ 0.539)$. Below, we present these numbers for the selected models.

Table 2. The numbers of rejections of H_0 in (7) among 1000 iterations

Model	The proposed RW test	ADF test	Runs test
M1	50	45	46
M2*	49	163	50
M3*	55	41	49

* Models close to random walks

Source: own elaboration

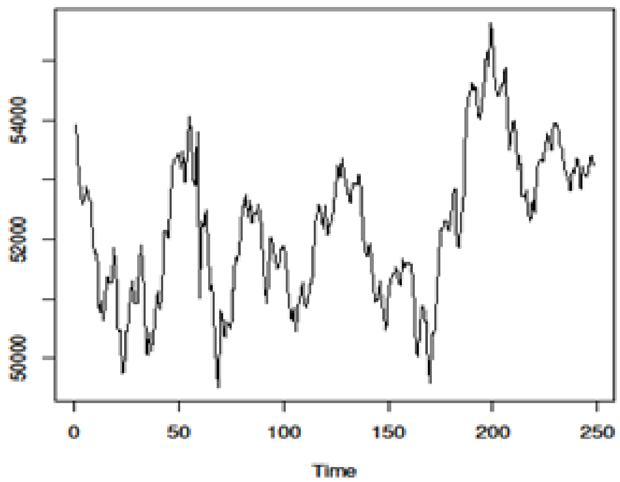
Comparing the numbers of rejections of our RW test with the numbers of rejections of the two earlier existing tests for a RW detection, we may see that the obtained fractions of rejections are close to a nominal significance level 0.05 for all of the considered tests and selected models except for the case of the ADF test and the model denoted by M2.

From the above depicted table, we may also observe that our test did not reject the hypothesis H_0: the processes $(\tilde{Y}_t)$ and (Y_t) form the random walks.

SOME APPLICATION IN THE STOCK EXCHANGE DATA ANALYSIS

In order to give an outlook on possible applications of our proposed test, we considered the daily stock exchange quotations of the WIG index from the period between December 2013 and November 2014. They are depicted in the figure below.

Figure 2. Daily stock returns for WIG from the period December 2013-November 2014



Source: own elaboration

It is supposed by a large number of academic people, primarily economists and statisticians, that the daily quotations of stock prices form a random walk process (see [Fama 2018]). Based on the given data, relating to this index in the chosen space of time, we have carried out the proposed test procedure. As a result, we obtain a value of the test statistic 0.31 and, assuming a standard significance level of 0.05, we do not reject the null hypothesis that the daily returns of the WIG index come from a random walk process. This is in accordance with the real fact that the daily quotations of WIG form a random walk. It is also worthwhile to mention that the p-values of the earlier existing augmented Dickey-Fuller test and the Runs test (which are equal to 0.08 and 0.34, respectively) confirm that the considered sequence forms a random walk.

SUMMARY

The principal objective of our work was to construct a new test devoted to a random walk detection. The main idea applied in the construction of our test procedure was based on the conclusion from the first arcsine law. Except of introduction of our test, we also checked its quality by computing the numbers of rejections of the null hypothesis that the given process forms a random walk, for three selected models, after 1000 iterations of our test procedure. The conducted research shows that the proposed approach provides an effective tool leading to reasonable conclusions regarding the subject of a random walk identification. The results of our test, which has been designed by implementation of the arcsine principle, seem very optimistic and encourage us to further investigations in the direction of a random walk detection.

REFERENCES

Dickey D. A., Fuller W. A. (1979) Distributions of the Estimators for Autoregressive Time Series with a Unit Root. Jour. Amer. Stat. Assoc., 74, 427-431.

Fama E. F. (2018) Random Walks in Stock-Market Prices. Selected Papers, 16, Graduate School of Business, University of Chicago, https://www.chicagobooth.edu/ ~/media/ 34F68FFD9CC04EF1A76901F6C61C0A76.PDF [access date: 25.08.2018].

Feller W. (1966) Wstęp do rachunku prawdopodobieństwa. PWN, Warszawa (in Polish).

Feller W. (1968) An Introduction to Probability Theory and its Applications. Vol.1, 3rd Edition. Wiley.

Maddala G. S. (2001) Introduction to Econometrics. Wiley.

Montanari A., Rosso R., Taqqu M. S. (1997) Fractionally Differenced ARIMA Models Applied to Hydrologic Time Series: Identification, Estimation, and Simulation. Water Resources Research, 33(5), 1035-1044.

Qiang L., Jiajin L. (2018) Arcsine Laws and its Simulation and Application. Research Paper Available from: http://individual.utoronto.ca/normand/Documents/MATH5501/ Project-3/Arcsine_laws_and_simu.pdf [access date: 20.03.2018].

Żak T. (2012) Błądzenie losowe i cztery najważniejsze twierdzenia rachunku prawdopodobieństwa. Educational material available from: http://prac.im.pwr.edu.pl/ ~zak/ Spacery_losowe_Kolo_Studenckie_26_04_2012.pdf [access date: 26.04.2012].