

《研究ノート》

# An Essay on Random Walk Process: Features and Testing

Shinji Yoshioka

## Abstract

In econometric literature, a lot of kinds of stochastic processes are employed, of course, including an autoregression (AR), a moving average (MA), an ARMA, and an ARIMA processes, etc. Among those, a random walk process is very unique and has some remarkable features. One of the most possible reasons to take a certain stochastic process is the fact that no one knows the reality of the time series data to be analyzed. This brief essay focused on a random walk process that is one of the most popular stochastic processes to unveil its essence. Moreover, some non-parametric testing methodologies such as a run test, a rank test, and a mean square successive difference test are also reviewed and applied to the normal random series generated by Box-Muller transformation.

**Keywords:** Stochastic process, Random walk, Martingale, Central limit theorem, Efficient market hypothesis

**JEL Classifications:** C02, C12, C16, C46, C65, and G12

## 1. An introduction

In econometric literature, a lot of kinds of stochastic processes are employed, of course, including an autoregression (AR), a moving average (MA), an ARMA, and an ARIMA processes, etc. Among those, a random

walk process is very unique and has some remarkable features. One of the most possible reasons to take a certain stochastic process is the fact that no one knows the reality of the time series data to be analyzed. This brief essay focused on a random walk process that is one of the most popular stochastic processes to unveil its essence.

Apart from this introduction, the paper consists of some chapters as follows; next chapter summarizes basic concepts on some distributional processes, which will be helpful to go forward; the third deals with martingale and shows a random walk process and martingale are identical; the fourth is the main part of the paper and focuses on the essence of a random walk process itself that Brownian motion takes over; the fifth features some testing methodologies for a random walk process and; the final chapter briefly concludes the paper and points out some remaining issues.

In the ensuing part of the paper, additionally, following two points are assumed: The time series data adopted hereafter are discrete and; the suffixes like  $t$  or  $s$  that indicate periods are thus natural numbers. These assumptions are based on practical economic data and enable us to avoid differentiability problems and also to utilize  $\sum$  to sum up data instead of integrations.

## 2. An overview of basic distributional processes

Random walk process is only defined to *iid*, which is an abbreviation of independently and identically distributed but other related distributional processes would interest most economists. This paper picks up four processes, i.e., *NID*, *iid*, white noise and stationarity. These four processes are stronger in this order, i.e., *NID* *iid* *WN* *stationarity* holds.

1 ) NID

$X_t \sim NID(\mu, \sigma^2)$  indicates that a series  $X_t$  is subject to a normal distribution with its mean of  $\mu$  and variance of  $\sigma^2$  and independent from each other. *NID* indicates normally and independently distributed. A mathematical expression is given as follows:

- a )  $E(X_t) = \mu$
- b )  $\text{var}(X_t) = \sigma^2$
- c )  $\text{cov}(X_t, X_s) = 0$  when  $t \neq s$
- d )  $X_t \sim N(\mu, \sigma^2)$

Among above four features, the third condition of c) that indicates zero covariance does not mean non-correlation as usual but independence. More precisely expressed employing probability function  $p$ ,  $p(X_t, X_s) = p(X_t)p(X_s)$  will hold under  $t \neq s$ , which means that the coincidental probability distribution is identical to their multiplied product. The concept of independence is a broader than non-correlation. Hence, when independent, non-correlation will always hold but the inverse relation does not.

2 ) iid

*iid* is main concept for random walk process and  $X_t \sim iid(\mu, \sigma^2)$  indicates that a series  $X_t$  distributes with its mean of  $\mu$  and variance of  $\sigma^2$  and independent from each other. The fourth condition of *NID* that is subject to a normal distribution therefore drops while other three hold as follows:

- a )  $E(X_t) = \mu$
- b )  $\text{var}(X_t) = \sigma^2$
- c )  $\text{cov}(X_t, X_s) = 0$  when  $t \neq s$

## 3 ) white noise (WN)

$X_t \sim WN(\mu, \sigma^2)$  indicates that a series  $X_t$  distributes with its mean of  $\mu$  and variance of  $\sigma^2$  and without auto-correlation. The mathematically expressed conditions are same as *iid* but the third does not depict independence but non-correlation as usual:

$$a) E(X_t) = \mu$$

$$b) \text{var}(X_t) = \sigma^2$$

$$c) \text{cov}(X_t, X_s) = 0 \quad \text{when } t \neq s$$

## 4 ) stationarity

Stationarity holds when a series  $X_t$  distributes with its mean of  $\mu$  and variance of  $\sigma^2$  and its covariance is subject to a function of period  $|t - s|$ . It is not required that covariance is identical to zero but subject to the difference of period. Stationarity here is given as its weak concept, which is reported as follows:

$$a) E(X_t) = \mu$$

$$b) \text{var}(X_t) = \sigma^2$$

$$c) \text{cov}(X_t, X_s) = \gamma(|t-s|) \quad \text{when } t \neq s$$

### 3. Martingale

'Martingale' originates from a French gambling game. In the original game, the bet will be doubled after losing a trick and the game ends when you win. Any type of game will do but it will be helpful to imagine a coin toss or something. The most important element of martingale is that the bet will be doubled after losing a trick. Martingale was firstly introduced to probability and statistics literature by early works of Lévy and later, Ville (1939) and

Doob (1950) widened its applicable fields.

Martingale is defined for a filtration, which receives information cumulatively. In the case of a financial asset price of  $X_t$ , e.g., stock price, at the time point of  $t$ , when certain information set of  $I_t$  includes all available information relating to its price and  $I_t$  determines  $X_t$ , this series of  $X_t$  is called to be 'adopted to  $I_t$ .' Since a filtration does not lose any information and accumulates all available information, the relation of  $I_1 I_2 I_3 \dots$  will hold. The definition of Martingale of a series  $X_t$  to a filtration  $I_t$  is as follows:

- a) expectation of the series is finite, i.e.,  $E(|X_t|) < \infty$
- b)  $X_t$  is adopted to  $I_t$
- c) expectation of  $X_{t+1}$  under  $I_t$  at the time point  $t$  is equal to contemporaneous value of  $X_t$ , i.e.,  $E(X_{t+1}|I_t) = X_t$

Moreover, it is well-known that the conditional expectation is the best expectancy to minimize squared mean of prediction error. Martingale therefore contains a certain level of expectation. Of course, the relation of  $E(X_{t+1}|I_t) = E(X_{t+2}|I_t) = E(X_{t+3}|I_t) = \dots = X_t$  will hold, in other words  $E(X_t - X_{t+s}|I_t) = 0$  when  $s = 1, 2, 3, \dots$ . This implies that mean prediction for future  $X_t$  variation is equal to zero, which indicates that the future prediction for martingale is impossible and totally useless. The prediction of series  $X_t$  is therefore equal to the contemporaneous value when  $X_t$  is subject to martingale.

#### 4. Random walk

In econometric literature, two types of random walk are employed, i.e., without and with drift. It is also plausible that random walk without drift is a special case that drift is equal to zero. Under the condition that given  $X_0$  as initial value is set to be zero and  $\epsilon_t \sim iid(0, \sigma^2)$  holds, they are defined as

follows:

a) random walk without drift

$$X_t = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \dots + \varepsilon_{t-1} + \varepsilon_t \quad \text{or} \quad X_t = X_{t-1} + \varepsilon_t$$

b) random walk with drift

$$X_t = \mu t + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \dots + \varepsilon_{t-1} + \varepsilon_t \quad \text{or} \quad X_t = \mu + X_{t-1} + \varepsilon_t$$

Some statistical values are calculated as follows:

a) expectation

$$\text{random walk without drift} \quad E(X_t) = 0$$

$$\text{random walk with drift} \quad E(X_t) = \mu t$$

b) variance

$$\text{random walk without drift} \quad \text{var}(X_t) = t\sigma^2$$

$$\text{random walk with drift} \quad \text{var}(X_t) = t\sigma^2$$

c) covariance (when  $t > s$ ) for both random walk processes with and without drift

$$\text{cov}(X_t, X_s) = (t - s)\sigma^2$$

This random walk contains some useful and interesting features as follows:<sup>1</sup>

- a) Random walk is equivalent to martingale;
- b) Symmetric random walk is recursive;
- c) For any constant  $c$ , probability that random walk is located under  $c$  is asymptotic to zero when  $t \rightarrow \infty$ ;
- d) Shock is permanent;
- e) First-differenced random walk is stationary;
- f) Variation of random walk increases proportionally to square root of time  $t$ ;
- g) Increment of random walk is independent; and
- h) Correlation of random walk elements  $X_t$  and  $X_s$  is equal to square root of their proportion.

---

<sup>1</sup> Brownian motion also takes over these eight features of random walk process.

Related to above a) , since  $E(\epsilon_t) = 0$  holds according to the definition, random walk is equivalent to martingale as follows:

random walk without drift

$$E(X_{t+1}|I_t) = E(X_t|I_t) + E(\epsilon_{t+1}|I_t) = E(X_t|I_t) = X_t$$

random walk with drift

$$E(X_{t+1}|I_t) = \mu + E(X_t|I_t) = \mu + X_t$$

Here, some kinds of particular random walk processes should be given. At first, for confirming above b) , simple random walk is defined as follows:

$$X_t = \epsilon_1 + \epsilon_2 + \epsilon_3 + \dots + \epsilon_{t-1} + \epsilon_t \quad \text{or} \quad X_t = X_{t-1} + \epsilon_t \quad X_0 = 0$$

and

$$P(\epsilon_t = 1) = p \quad \text{and} \quad P(\epsilon_t = -1) = 1 - p^2$$

Among simple random walk, those under the condition that  $p = 1 - p = 1/2$  are defined as symmetric random walk. For simple random walk, starting from initial value of zero, the probability to return to zero is calculated as follows:

$$P = 1 - |2p - 1|$$

Apparently,  $P = 1$  holds when  $p = 1/2$  , which implies that symmetric random walk is recursive or returns to its initial value. The first passage time that is defined as time from initial to returning to initial value will be infinite since it is calculated as follows:

$$\text{first - passage - time} = \frac{4p(1 - p)}{\sqrt{1 - 4p(1 - p)}}$$

The third point of above c) is rather interesting. For random walk series  $X_t$  with its initial value of zero and without drift, central limit theorem derives a standardized normal variable  $Z_t$  from  $X_t$  as follows:

$$Z_t = \frac{X_t}{\sqrt{t}} = \frac{1}{\sqrt{t}} \sum_{i=1}^t \epsilon_i \sim N(0, 1)$$

<sup>2</sup> Even in this case, of course,  $\epsilon_i$  is independent since  $\epsilon_t \sim iid(0, \sigma^2)$ .

Probability that random walk does not exceed a certain constant  $c$  is calculated employing distribution function of above  $Z_t$  as follows:

$$P(|X_t| \leq c) = P(|Z_t \sqrt{t} \leq c) = P\left(|Z_t| \leq \frac{c}{\sqrt{t}}\right) \approx 2 \left(\frac{c}{\sqrt{t}}\right) - 1$$

where

$$(u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

Of course,  $\left(\frac{c}{\sqrt{t}}\right)_t \rightarrow 0$  holds since  $\frac{c}{\sqrt{t}} \rightarrow 0$ . Therefore, probability that random walk series  $X_t$  does not exceed  $c$  is asymptotic to zero, which is identical to  $P(|X_t| \leq c) \rightarrow 0$ . Hence, for any constant  $c$ , probability that random walk is located under  $c$  is asymptotic to zero when  $t \rightarrow \infty$ .

The fourth and fifth points of above d) and e) seem almost self-evident. Since  $\epsilon_t \sim iid(0, \sigma^2)$ ,  $\sum_{i=s}^t \epsilon_i = 0$ ,  $s, t$  holds. Therefore, when a shock is given at an arbitrary point  $s$ , i.e.,  $\epsilon_s = shock$ , following relation will hold:

$$X_t = \sum_{i=1}^{s-1} \epsilon_i + shock + \sum_{j=s+1}^t \epsilon_j = 0 + shock + 0 = shock \quad \text{when } s < t$$

Above relation indicates that a shock will continue permanently.

Continuously, above point of e) is equivalent to  $X_t \sim I(1)$  or  $X_t \sim I(0)$ , which appears self-evident. Of course,  $X_t = \sum_{i=1}^t \epsilon_i \sim I(0)$  since  $\epsilon_t \sim iid(0, \sigma^2)$ . That's all.

Relating to above point f), variance of a random walk series is calculated as follows:

$$\begin{aligned} \text{var}(X_t) &= E(\epsilon_1 + \epsilon_2 + \dots + \epsilon_t)^2 = E(\epsilon_1^2) + \text{var}(\epsilon_2) + \dots + \text{var}(\epsilon_t) = \sigma^2 t \\ &\text{since } E(\epsilon_i \epsilon_j) = 0 \quad \text{when } i \neq j \end{aligned}$$

Therefore, standard deviation of series  $X_t$  is calculated as  $stdev(X_t) = \sqrt{\text{var}(X_t)} = \sigma \sqrt{t}$ , which implies that variation measured by standard



deviation increases proportionally to square root of time  $t$ .

Similar to above d) and e) and employing variance calculation of f), since increment of random walk series  $X_t$  can be expressed as sum of  $\epsilon_t$ , variance of increment is also written by  $\epsilon_t$  as follows:<sup>3</sup>

$$X_s - X_r = \sum_{i=r+1}^s \epsilon_i \quad \text{and} \quad X_t - X_s = \sum_{i=s+1}^t \epsilon_i \quad \text{when} \quad r < s < t$$

$$\begin{aligned} \text{var}(X_s - X_r) &= \text{var}(\epsilon_{r+1} + \epsilon_{r+2} + \dots + \epsilon_s) = \text{var}(\epsilon_{r+1}) + \text{var}(\epsilon_{r+2}) + \dots + \text{var}(\epsilon_s) \\ &= (s - r) \quad \text{and} \quad \text{var}(X_t - X_s) = \text{var}(\epsilon_{s+1} + \epsilon_{s+2} + \dots + \epsilon_t) = \text{var}(\epsilon_{s+1}) + \text{var}(\epsilon_{s+2}) + \dots + \text{var}(\epsilon_t) = (t - s) \quad \text{when} \quad r < s < t \end{aligned}$$

Above relation confirms the point of g) that increment of random walk is independent

Relating to the final point of h), for random walk process  $X_t$ , we already confirms that  $\text{var}(X_t) = t$  hold. Moreover, when  $s < t$ , covariance is calculated as follows:

$$\begin{aligned} \text{cov}(X_s - X_t) &= E(X_s - X_t) = E((\epsilon_1 + \epsilon_2 + \dots + \epsilon_s) - (\epsilon_1 + \epsilon_2 + \dots + \epsilon_t)) = E(\epsilon_{s+1} + \epsilon_{s+2} + \dots + \epsilon_t) \\ &= 0 \quad \text{since} \quad E(\epsilon_i - \epsilon_j) = 0 \quad \text{when} \quad i \neq j \end{aligned}$$

Therefore, correlation coefficient, i.e.,  $\text{corr}(X_s, X_t)$  when  $s < t$ , is equal to square root of their proportion, calculated as follows:

$$\text{corr}(X_s, X_t) = \frac{\text{cov}(X_s, X_t)}{\sqrt{\text{var}(X_s)\text{var}(X_t)}} = \frac{0}{\sqrt{s * t}} = 0$$

Before proceeding to next chapter, examples of random walk process are reported at Figure 1 above. The series of error term, i.e.,  $\epsilon_t$  are generated from uniform random number  $[0, 1)$  by Box-Muller transformation indicated at Box and Muller (1958). According to this methodology,  $U_1$  and  $U_2$  subject to uniform distribution are converted to  $Y_1$  and  $Y_2$  subject to normal

<sup>3</sup> Here,  $r$  is also assumed a natural number besides  $t$  and  $s$ .

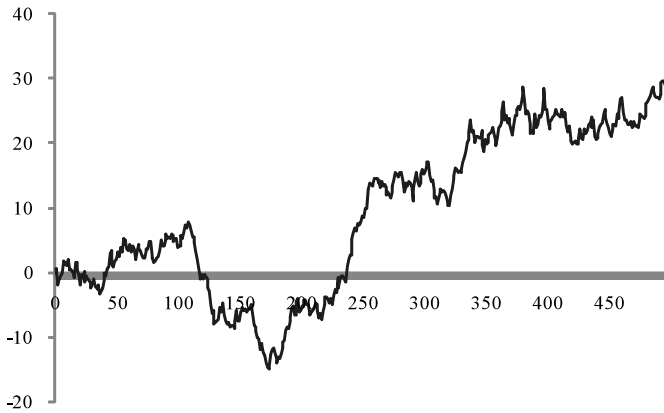
random distribution as follows:

$$Y_1 = \sqrt{-2 \ln U_1} \cos 2\pi U_2$$

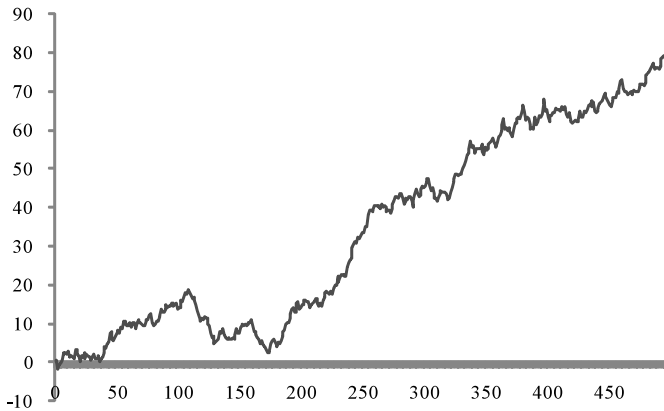
$$Y_2 = \sqrt{-2 \ln U_1} \sin 2\pi U_2$$

Figure 1: Example of Random Walk

(1) random walk without drift:  $X_t = X_{t-1} + \epsilon_t$



(2) random walk with drift:  $X_t = \mu + X_{t-1} + \epsilon_t$      $\mu = 0.1$



Note: Both (1) and (2) contain zero of initial value and identical error term that follows  $\epsilon_t \sim NID(0, 1)$ , which is generated by Box-Muller transformation.

Source: Author.

## 5. Tests for random walk

After confirming some remarkable and interesting features of random walk process at previous chapter, now, testing methodologies should be presented. At first, we have to test following three points:

- a ) Randomness or independence, of course;
- b ) Homoscedasticity; and
- c ) Rank of integration, which must be first.

Among above three points, the second test for homoscedasticity and the third for rank of integration are usually completed by various tests, including augmented Dickey-Fuller test for the latter. The first test for randomness or independence, however, seems distinctive for random walk. Therefore, this paper focuses on randomness or independence test and saves further discussion on the second and the third tests for other literature. This paper picks up three types of non-parametric testing for randomness or independence of series, which are a run test,<sup>4</sup> a rank test<sup>5</sup> and a mean square successive difference test ( MSSD ). Among these three, the first two tests are for randomness and the third is for independence. These three tests are applied for the first differenced series of random walk process reported at Figure 1.

At first, run test is focused on, in which the number is calculated and compared against its sampling distribution under the random walk hypothesis. A run is a sequence of consecutive positive or negative deviations. A positive deviation is represented by a “ + ” sign and a negative drop is represented by a “ - ” sign. A run is defined as a deviation sequence of the same sign.

---

<sup>4</sup> Kamal and Kashif-ur-Rehman ( 2006 ) employs a run test for security prices and Hashimoto et al . ( 2008 ) does for the dollar-yen and euro-dollar exchange rates.

<sup>5</sup> Luger ( 2001 ) proposes a class of linear signed rank statistics to test for a random walk with unknown drift in the presence of arbitrary forms of conditional heteroscedasticity.

Using the laws of probability, it is possible to estimate the number of runs in each of the two categories and given the sample size. Too many or too few runs in a series seems to be caused by autocorrelation. By comparing the total number of runs in the data with the expected number of runs under random walk hypothesis, the test of the random walk hypothesis may be constructed. It has been shown that the distribution of the number of runs converges to a normal distribution<sup>6</sup> asymptotically using numbers of runs , “ + ” and “ - ” signs as follows:

$$Z = \frac{R - \frac{2mn}{m+n} + 1}{\sqrt{\frac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)}}} \approx N(0, 1)$$

where R number of runs  
 m number of “ + ” signs  
 n number of “ - ” signs

Here, error terms of  $\epsilon_t$  calculated according to Box-Muller transformation for above Figure 1 are employed for examples during  $1 \leq t \leq 50$  . Table 1 reports tick-by-tick data and runs of  $\epsilon_t$ .

**Table 1: Sample data generated by Box-Muller transformation for run and rank tests**

$t$	$\epsilon_t$	dev.	run	rank	$t$	$\epsilon_t$	dev.	run	rank
1	-2.42118	-	1	1	26	-0.25671	-	17	20
2	0.74111	+	2	38	27	-0.24707	-		21
3	0.46276	+		31	28	-0.91784	-		9
4	0.60398	+		34	29	0.69227	+	18	35
5	1.80444	+		48	30	0.72917	+		37

<sup>6</sup> More precisely, this asymptotic approximation holds when  $20 < m, n$ .

6	-0.55085	-	3	14	31	-1.15014	-	19	7	
7	0.40488	+	4	29	32	-0.10238	-		23	
8	-0.49905	-	5	15	33	-0.31368	-		18	
9	0.77875	+	6	39	34	0.72231	+	20	36	
10	-1.59876	-	7	5	35	-1.50848	-	21	6	
11	0.22819	+	9	8	27	36	0.55589	+	22	33
12	-0.21130	-		22	37	0.43263	+	30		
13	-0.29153	-		19	38	1.17692	+	44		
14	-0.81117	-		10	39	1.27169	+	46		
15	2.35330	+	10	50	40	-0.62908	-	23	12	
16	-0.04015	-	11	24	41	1.22818	+	24	45	
17	-1.09004	-		8	42	0.21897	+			
18	-0.74658	-		11	43	2.13348	+			
19	-1.71899	-		3	44	0.53229	+			
20	1.62863	+	12	47	45	-2.14295	-	25	2	
21	-0.45344	-	13	17	46	-0.45987	-		16	
22	0.96854	+	14	41	47	0.78319	+	26	40	
23	-1.69407	-	15	4	48	0.39365	+		28	
24	1.03155	+	16	43	49	0.98925	+		42	
25	-0.56090	-	17	13	50	0.04053	+		25	

Note: The column of ' run ' as well as that of ' dev. ' is employed for a run test while that of ' rank ' is utilized for a rank test, later developed. The number of ' rank ' indicates the order of each  $t_i$  when sorted from small to large.

Source: Author.

According to above Table 1 , substituting  $R = 26$  ,  $m = 24$  and  $n = 24$  to above equation of run test, the calculation result of  $Z$  is  $Z = 4.77E-06$  . On the other hand, the critical value of the standardized normal distribution at 5 percent significance is 1.96 that does not reject the value of  $Z$ . Above series of  $t_i$  should be therefore regarded as random.

The second non-parametric methodology to test randomness is a rank test

that examines whether a series contains a linear trend or not. The procedure of a rank test is developed hereafter. For given series of  $X_1, X_2, \dots, X_n$ , we count the number  $Q$ , which is the number of pair  $(i, j)$  for  $X_i > X_j$  when  $j > i$ ,  $i = 1, 2, 3, \dots, n - 1$ , and  $j = 2, 3, \dots, n$ . According to Table 1,  $Q = 559$ . Of course, the scope of  $Q$  is between zero and  ${}_2C_n = \frac{n(n - 1)}{2}$ .  $Q = 0$  denotes  $X_1 < X_2 < \dots < X_n$  while  $Q = \frac{n(n - 1)}{2}$  reflects  $X_n < X_{n-1} < \dots < X_1$ . Therefore, a linear increasing trend is observed when  $Q$  is small and a linear decreasing trend is supposed if  $Q$  is large, vice a versa. Expectation and variance of this  $Q$  is given as follows:

$$E(Q) = \frac{n(n - 1)}{4}$$

$$\text{var}(Q) = \frac{B_n}{12} n(n - 1)(2n + 5)$$

where  $B_n$  is Bernoulli number defined as  $B_0 = 1$  and

$$B_n = -\frac{1}{n + 1} \sum_{k=0}^{n-1} C_k B_k = -\frac{1}{n + 1} \sum_{k=0}^{n-1} \frac{(n + 1)!}{(n - k + 1)! k!} B_k$$

Therefore, standardized statistics of  $Q$ , i.e., statistics  $Z$  given at ensuing equation, converges to a normal distribution asymptotically as follows:<sup>8</sup>

$$Z = \frac{Q - E(Q)}{\sqrt{\text{var}(Q)}} = \frac{Q - \frac{n(n - 1)}{4}}{\sqrt{\frac{B_n}{12} n(n - 1)(2n + 5)}} \quad \text{and} \quad |Z| \approx N(0, 1)$$

According to Table 1, substituting  $Q = 559$ ,  $n = 50$  and  $B_{50} =$

$7.500866746076964366856\text{E}+24$  or  $495057205241079648212477525/66^9$  to

<sup>7</sup> More precisely, this is a recurrence equation for Bernoulli number. The original function for its coefficients that is  $f(x) = \frac{x}{e^x - 1}$  is not stable when differentiated. This recurrence equation is calculated based on a concept that the product of the original function and its Maclaurin-expanded inverse function is equal to 1.

<sup>8</sup> More precisely, this asymptotic approximation holds when  $10 < n$ .

above equation of rank test, the calculation result of  $Z$  is  $|Z| = 4.72E - 12$  . On the other hand, the critical value of the standardized normal distribution at 5 percent significance is 1.96 that does not reject the value of  $|Z|$  . An example series of  $t$  reported at Table 1 should be therefore regarded as random.

The third and final test employed in the paper is a mean square successive difference test (MSSD) ,which is introduced by von Neumann et al ( 1941 ) and Young ( 1941 ) . This employs a ratio between a mean square and a sample variance as follows:

$$MS = \frac{\sum_{i=1}^{n-1} (X_{i+1} - X_i)^2}{n - 1}$$

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

where  $MS$  mean square  
 $s^2$  sample variance  
 $\bar{X}$  sample mean

When observations, i.e.,  $X_1, X_2, X_3, \dots, X_n$  , are independent and derived from a population with expectation of  $M$  and variance of  $\sigma^2$  , above  $s^2$  is an unbiased and coincident estimator to  $\sigma^2$  . Moreover, following estimator  $ss^2$  is also coincident to  $\sigma^2$ :

$$ss^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{2(n - 1)}$$

---

<sup>9</sup>  $B_{50}$  is derived by pari/GP, which is a software package developed and maintained at Bordeaux I University and distributed under the GPL license for mathematical education. More detailed information is provided at its web site at Bordeaux I University of <http://www.math.u-bordeaux.fr/~belabas/pari/>.

Additionally, following ratio is introduced:

$$C = 1 - \frac{SS^2}{s^2}$$

When observations are independent, above  $C$  takes close to zero and large  $C$  in absolute value implies that they are not independent. According to Young (1941), expectation and variance of  $C$  are derived as follows:

$$\begin{aligned} E(C) &= 0 \\ \text{var}(C) &= E(C^2) = \frac{2n - 3 - b_2}{2n(n - 1)} \\ \text{where } b_2 &\quad \text{sample kurtosis} \end{aligned}$$

When observations are subject to  $NID(0, \sigma^2)$ , expectation of sample kurtosis is equal to  $\frac{3(n - 1)}{n + 1}$ . Therefore, variance of  $C$  is nearly equal to  $\frac{n - 2}{(n + 1)(n - 1)}$ . When  $20 \leq n$ , it is possible to approximate  $C$  as standardized, hence, statistics  $Z$  given at ensuing equation, converges to a normal distribution asymptotically as follows:

$$Z = \frac{C}{\sqrt{\frac{n - 2}{(n + 1)(n - 1)}}} \approx N(0, 1)$$

Based on data reported at Table 1, the calculation result is that  $|Z| = 1.03$  which does not reach rejection region of 1.96 at 5 percent significance. An example series of  $\dots$ , reported at Table 1 should be therefore regarded as independent.

## 6. Conclusion

This paper summarizes some remarkable features of random walk according to a mathematical statistics methodology and applies some tests to the



normal random series generated by Box-Muller transformation. Among those features, one of the most important is unpredictability. If a certain financial asset price follows random walk, it's nonsense and impossible to predict it, i.e., as Malkiel ( 2004 ) insists that " a blindfolded monkey throwing darts at a newspaper's financial pages could select a portfolio that would do just as well as one carefully selected by experts, " which strongly suggests efficient market hypothesis holds.

Here are two interesting points are remained: one is practical application of random walk process to an individual financial asset price, e.g., a stock price or exchange rate, etc., and; the other is to proceed Brownian motion, which takes over a lot of remarkable features from random walk. The latter way should go straight forward to Ito integration or stochastic differential equation and to a Black-Scholes model. The author is very satisfied if the paper builds a base for further study.

## References

- Box, G. E. P. and Mervin E. Muller ( 1958 ) " A Note on the Generation of Random Normal Deviates, " *Annals of Mathematical Statistics* 29(2) , pp.610-611 , 1958
- Doob, Josef Leo ( 1950 ) *Stochastic Processes*, reprinted by Wiley Classics Library , 1990
- Hashimoto, Yuko, Takatoshi Ito, Takaaki Ohnishi, Misako Takayasu, Hideki Takayasu, Tsutomu Watanabe ( 2008 ) " Random Walk or A Run: Market Microstructure Analysis of the Foreign Exchange Rate Movements based on Conditional Probability, " *Working Paper Series No . 25* , Research Center for Price Dynamics, Institute of Economic Research, Hitotsubashi University, August 2008
- Kamal, Yasir and Kashif-ur-Rehman ( 2006 ) " Random Walk of Security Prices: Empirical Evidence from KSE, LSE, and ISE, " *Journal of Independent Studies and Research* 4(1) , 2006 , pp.17-23
- Luger, Richard ( 2001 ) " Exact Non-Parametric Tests for a Random Walk with Unknown Drift under Conditional Heteroscedasticity, " *Bank of Canada Working Paper 2001-2* , February 2001

- Malkiel, Burton G , ( 2004 ) *A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing*, W. W. Norton & Co. , 2004
- Ville, Jean ( 1939 ) *Etude Critique de la Concept du Collectif*, Gauthier-Villars , 1939
- von Neumann, J., R. H. Kent, H. R. Bellinson and B. I. Hart ( 1941 ) " The mean square successive difference, " *Annals of Mathematical Statistics* 12(2) , 1941 , pp.153-162
- Young, L. C . ( 1941 ) " On randomness in ordered sequences, " *Annals of Mathematical Statistics* 12(3) , 1941 , pp.293-300