

RANDOM WALK HYPOTHESIS IN FINANCIAL MARKETS

Nicolae-Marius JULA*
Nicoleta JULA**

Abstract

Random walk hypothesis states that the stock market prices do not follow a predictable trajectory, but are simply random. If you are trying to predict a random set of data, one should test for randomness, because, despite the power and complexity of the used models, the results cannot be trustworthy.

There are several methods for testing these hypotheses and the use of computational power provided by the R environment makes the work of the researcher easier and with a cost-effective approach. The increasing power of computing and the continuous development of econometric tests should give the potential investors new tools in selecting commodities and investing in efficient markets.

Keywords: *random walk, financial market, stock price, statistical test, R*

1. Introduction

When analyzing the market, there are two historical approaches. One theory is based on the randomness of the data and the other supports the predictability of the market, thus meaning a non-random hypothesis.

The random walk is an old concept, which can be traced back to 19th century, when the French broker Jules Regnault¹ published his research in 1863. In 1900, Louis Bachelier² wrote his PhD. Thesis on the same theory. The first use of random walk on stock market is attributed to Maurice Kendall³ and his paper from 1953. The history also records the test performed by the professor Burton G. Markiel⁴, when he demonstrated the randomness of the stock market using a flip coin to decide the growth of the closing price of an equity.

There are also researchers that do not agree to the random walk hypothesis in predicting the market behavior. They state that the prices are somehow predictable and one could identify some trends analyzing the historical data. Many studies are presenting tests that support the predictability of the trends in the financial markets, like professors Andrew W. Lo and Archie Craig MacKinlay's book⁵.

Returning to the randomness approach, one cannot analyze the market without defining the market efficiency theory. According to Eugene Fama, the "father" of modern empirical finance, in his research "Random Walks in Stock Market Prices" he stated that

"in an efficient market, competition among the many intelligent participants leads to a situation where, at any point in time, actual prices of individual securities already reflect the effects of information based both on events that have already occurred and on events which as of now the market expects to take place in the future. In other words, in an efficient market at any point in time the actual price of a security will be a good estimate of its intrinsic value". The idea that you cannot have a profit on an efficient market is not supported by all economists. Some stipulates that you can profit on such a market by selecting / buying leveraged ETFs (exchange trade funds). And, like almost always in trading, big risks can bring big winnings or big losses.

Depending on how much information is taken into account when forming the current price, we can distinguish weak, semi-strong and strong efficient markets. A weak efficient market is the one which forms the price based on all historical prices. Semi-strong markets bring into the formation of the price all relevant and public available information. The strong form of the efficient market takes into account not only all the relevant information, but also even insider information, usually not publicly available.

Usually, the market efficiency may be considered a by-product of market involvement of information arbitrageurs. These information arbitrageurs shall buy the assets that are supposed to be undervalued and sell the overvalued one. These agents may have some information that is scares of publicly not so accessible. By doing this, they may be considered to regulate the

* Lecturer, PhD Faculty of Economics and Business Administration, "Nicolae Titulescu" University of Bucharest (email: mariusjula@univnt.ro).

** Professor, PhD, Faculty of Economics and Business Administration, "Nicolae Titulescu" University of Bucharest (email: nicoletajula@univnt.ro).

¹ Regnault, Jules, 1863, Calcul des chances et philosophie de la bourse, Paris : Mallet-Bachelier and Castel

² Bachelier, L., 1900, Théorie de la spéculation, Gauthier-Villars.

³ Kendall, M. G.; Bradford Hill, A (1953). "The Analysis of Economic Time-Series-Part I: Prices". Journal of the Royal Statistical Society. A (General). Blackwell Publishing. 116 (1): 11–34. JSTOR 2980947.

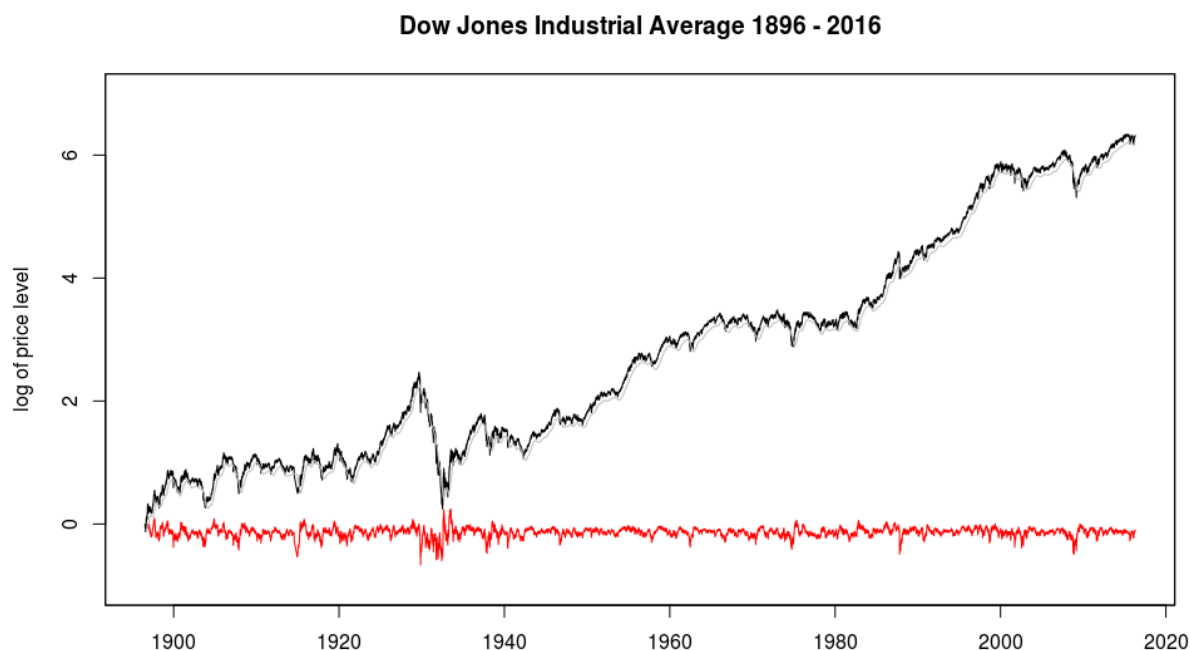
⁴ Malkiel, Burton G. (1973). A Random Walk Down Wall Street (6th ed.). W.W. Norton & Company, Inc. ISBN 0-393-06245-7.

⁵ Lo, Andrew W.; MacKinlay, Archie Craig (2002). A Non-Random Walk Down Wall Street (5th ed.). Princeton University Press. pp. 4–47. ISBN 0-691-09256-7.

price of the assets on that market and the new price should include also these sets of information known by these arbitrageurs. If the markets were perfectly efficient, the existence of the arbitrageurs would be unnecessary. Also, their actions are “boosted” by the existence of the noise traders. These so-called noise traders are the agents who trade based on other motives than new information (for example, an insurance company selling some stocks for paying a large claim).

investors who expect some gain. Stuart Ride explains the difference between a complete random market and the actual one in the Figure 1 below. In this graph, the investors are rewarded (the trend is ascending). The gray line is a compounded 126 day rolling average return, a so-called equity risk premium, the red line shows the compounded excess/residual return of the market, based on the calculations made by Ride. The gray line can be seen as the signal and the red one as a noise, or a Martingale process. Stuart Ride comes with

Figure 1



Source: <http://www.turingfinance.com/testing-the-efficient-market-hypothesis-with-r/>

2. Random Walk Hypothesis

The random walk hypothesis is a theory that stipulates the price of an equity should be described by a random walk, more specific a sub-martingale process. In probability theory, this martingale⁶ process (the term being first used by Ville, in 1939) is used to describe a fair game, in which the past events do not influence the outcome of the current winnings and only the present events matter.

A weak efficient market can be regarded as a random model, because the historical price changes are reflected in the current price. So, this is an argument to test for randomness in the markets. Stuart Ride continues this idea and states that the market is only approximately random. If the market would be completely random, no one should invest. Investors should be somehow rewarded for their risk-taken actions of holding assets. There is the so-called market risk premium and is the reason that there are long term

4 ways of testing the random hypothesis:

1. by predicting or finding statistically significant patterns in the so-called equity risk premium
2. by predicting of or finding statistically significant patterns in the residuals
3. by predicting of or finding statistically significant patterns in the sign of the residuals or in the rank
4. by using non-parametric tests of randomness

Stuard Ride⁷ suggests a different approach, meaning running all of the above tests on financial price time series. And this is only possible using dedicated software, like the R environment and the *emh* package.

2.1. The importance of randomness tests

Usually, when investing there are some already called classical approaches. First, MVO (mean variance portfolio optimisation) assumes that the prices are stationary random walks that can be completely determined by their first two moments.

⁶ Ville, Jean (1939). Étude critique de la notion de collectif. Monographies des Probabilités (in French). 3. Paris: Gauthier-Villars. Zbl 0021.14601.

⁷ <http://www.turingfinance.com/testing-the-efficient-market-hypothesis-with-r/>.

The next approach is based on Monte Carlo simulations. These stochastic processes suppose to effectively model these totally random prices.

On the market there are also high frequently traders or arbitrageurs who insist that the prices are far from random and patterns can be identified in the price formation in high frequencies. Also information scarcity can be used to anticipate the price trends.

The fundamental analysts argue that in low frequency trading there is not a random price forming, but a strong correlation to the fundamental information about the company issuing the asset/commodity. Also there are the macroeconomic analysts, who also insist that in low frequency trading the economic factors do influence the prices, thus making them not random. Factors such business cycles can be recorded in the prices and making them predictable.

The last, but not the least case is of the technical analysts. In this case, the prices and volumes are not random, but follow some significant patterns, more or less easy identifiable and exploitable by using deterministic indicators.

3. Random Walk tests in R

The *emh* package consists (in the 0.1.0 version) in 6 tests: Wald – Wolfowitz runs test, Durbin – Watson test, Ljung – Box test, Breusch - Godfrey test, Bartell Rank - Based Variance Ratio test and Lo - Mackinlay Variance Ratio test.

Test 1. The Wald – Wolfowitz runs test is one of the most used tests for randomness. It consists in indentifying the number of uninterrupted aequance of identical bits. In financial markets, we translate this series in moments of growth (1s) and moments of decrease in in price of an equity (0s).

The runs test is defined as:

H₀: the sequence was produced in a random manner

H₁: the sequence was not produced in a random manner

The statistic test is $Z = (R - \hat{R}) / s_R$

where R is the observed number of runs, \hat{R} is the expected number of runs, and s_R is the standard deviation of the number of runs. The values of \hat{R} and s_R are computed as follows:

$$\hat{R} = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$s_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

with n_1 and n_2 denoting the number of positive and negative values in the series.

Significance Level: α

Critical Region: The runs test rejects the null hypothesis if $|Z| > Z_{1-\alpha/2}$.

Test 2. The Durbin-Watson test (Durbin J., Watson G.S., 1950, 1951) is the most commonly used procedure for identifying first-order autocorrelation of errors in linear regression models.

The test is

$$dw = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2}$$

where u_t represent the residuals from the regression.

In order to test the hypothesis⁸ H₀: $\rho = 0$ (absence of error autocorrelation), the alternative H₁: $\rho \neq 0$ (the presence of the error autocorrelation phenomenon) the bilateral Durbin-Watson test is used. From the Durbin - Watson bilateral test tables, the critical values dL and dU are selected for k - the number of explanatory variables in the model and n - the size of the sample (starting from the significance level - usually 0.05 or 0.01).

- H₀ hypothesis is accepted - no first-order autocorrelation, if $dU \leq dw \leq 4 - dU$

- Reject H₀ if $dw \leq dL$ or $dw \geq 4 - dL$.

- If $dL \leq dw \leq dU$ or $4 - dU \leq dw \leq 4 - dL$, the test is inconclusive.

Test 3. Ljung – Box test, named after Greta Ljung and George Box, finds out if there is any significant autocorrelation in a time series. We should expect the autocorrelation when the financial series has a momentum or a mean-reversion. The test starts with the Ljung – Box statistic:

$$Q = n(n + 2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}, \text{ where}$$

N is the dimension of the series, $\hat{\rho}_k^2$ is the autocorrelation of the series at k lag, h is the number of lags. Under the null hypothesis, the statistics follows a chi-squared distribution.

Test 4. Breusch - Godfrey test is based on LM model and it is used to identify the autocorrelation in the errors in an econometric regression model. It is considered more powerful than Durbin – Watson tests (doesn't have the drawbacks like identifying only AR1).

The procedure consists in obtaining the residuals from the initial regression model. These residuals are forming a new regression:

$$u_t = \alpha_0 + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \dots + \alpha_p Z_{pt} + \varepsilon_t$$

The determination coefficient R² is calculated for this new regression and the nR² should follow a Chi squared distribution and the null hypothesis (there is no autocorrelation of any order) is rejected for $nR^2 > \chi^2$.

Test 5. Bartell Rank - Based Variance Ratio test is a version of the 1941's John von Neumann Ration Test. This version doesn't require anymore the property

⁸ Jula D., Jula N., 2017, MODELARE ECONOMICĂ. MODELE ECONOMETRICE ȘI DE OPTIMIZARE, Editura Mustang, Bucuresti.

of normal distribution of the series. The statistic test is calculated:

$$RVN = \frac{\sum_{i=1}^{n-1} (R_i - R_{i-1})^2}{\sum_{i=1}^n (R_i - (n+1)/2)^2}$$

Where R_i is the logarithmic return r_i and n is the length of the series. Bartel demonstrated that the $(RVN-2)/\sigma$ follows a normal distribution, where

$$\sigma^2 = \frac{4(n-2)(5n^2 - 2n - 9)}{5n(n+1)(n-1)^2}$$

Test 6. Lo - Mackinlay Variance Ratio⁹ test is used to identify the heteroskedasticity in a series. Not all financial series can be tested with this procedure. According to Reid, "the test is only valid if security price changes have finite variances".

EMH package in R

To test the random walk hypothesis on financial data can be rather time consuming. The today's technology allows the researcher to access a plethora of tools and the R environment proves once again that it is one of the leading source of utensils when dealing with statistic and econometric data. Each of the tests presented above are already implemented and can be used freely by everyone. The EMH package developed by Stuard Reid gathers these tests.

This package can be downloaded or installed in R from GitHub:

```
library(devtools)
devtools::install_github(repo="stuartgordonreid/emh")
```

The test of randomness can be performed easily, using the following syntax:

```
results<-emh::is_random(financial_data)
emh::plot_results(results)
View(results)
```

4. Conclusions

As Eugene Fama stated in his iconic article (*Random Walks in Stock Market Prices*), "If the random-walk theory is valid and if security exchanges are "efficient" markets, then stock prices at any point in time will represent good estimates of intrinsic or fundamental values. Thus, additional fundamental analysis is of value only when the analyst has new information which was not fully considered in forming current market prices, or has new insights concerning the effects of generally available information which are not already implicit in current prices. If the analyst has neither better insights nor new information, he may as well forget about fundamental analysis and choose securities by some random selection procedure.". So, one interested in investing on a market should benefit from all the information he can get. Testing a market for random walk helps decide if it is worth the risk of investing.

The future development of this research should focus on testing the model on emerging financial markets, like the Romanian one. The increasing power of computing and the continuous development of econometric tests should give the potential investors new tools in selecting and investing.

References:

- Bachelier, L., 1900, *Théorie de la spéculation*, Gauthier-Villars;
- Fama, E., 1995, *Random Walks in Stock Market Prices*. Financial Analysts Journal, 51(1), 75-80. Retrieved from <http://www.jstor.org/stable/4479810>;
- Jula D., Jula N., 2017, *Modelare Economică. Modele Econometrice și de Optimizare*, Editura Mustang, Bucuresti;
- Kendall, M. G.; Bradford Hill, A., 1953, *The Analysis of Economic Time-Series-Part I: Prices*. Journal of the Royal Statistical Society. A (General). Blackwell Publishing. 116 (1): 11–34. JSTOR 2980947;
- Lo, Andrew W., and A. Craig MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test. Review of financial studies 1.1: 41-66;
- Lo, Andrew W.; Mackinlay, Archie Craig, 2002, *A Non-Random Walk Down Wall Street (5th ed.)*. Princeton University Press. pp. 4–47. ISBN 0-691-09256-7;
- Malkiel, Burton G., 1973, *A Random Walk Down Wall Street (6th ed.)*. W.W. Norton & Company, Inc. ISBN 0-393-06245-7;
- Regnault, Jules, 1863, *Calcul des chances et philosophie de la bourse*, Paris : Mallet-Bachelier and Castel;
- Ville, Jean, 1939, *Étude critique de la notion de collectif*. Monographies des Probabilités (in French). 3. Paris: Gauthier-Villars. Zbl 0021.14601;
- <http://www.turingfinance.com/testing-the-efficient-market-hypothesis-with-r/>.

Annex

One can perform the above tests from EMH package test by test in R.

1. For The Wald – Wolfowitz runs test the syntax suggested by Stuard Reid is:

```
test_runs <- function(rets, a = 0.99) {
```

⁹ Lo, Andrew W., and A. Craig MacKinlay. "Stock market prices do not follow random walks: Evidence from a simple specification test." Review of financial studies 1.1 (1988): 41-66.

```

# Check and convert the data.
.check_data(data = rets)
binrets <- as_binary(rets)
# Get the numbers of bits.
k <- length(binrets)
k.ones <- count_ones(binrets)
k.zeros <- count_zeroes(binrets)
# Calculate the expectations.
mean <- ((2 * k.ones * k.zeros) / k) + 1
variance <- ((mean - 1) * (mean - 2)) / (k - 1)
# Calculate the number of runs.
k.runs <- test_runs_number(binrets)
# Compute the z score of k.runs.
z.score <- (k.runs - mean) / sqrt(variance)
# Compute the p-value of the z-score.
p.value <- pnorm(z.score)
# Compute the required threshold.
thresh <- abs(qnorm((1 - a) / 2))
# Return the results object.
return(c(k.runs, p.value, z.score,
abs(z.score) > thresh))
}

```

2. The Durbin-Watson test syntax suggested by Stuard Reid is:

```

test_durbinwatson <- function(rets, a = 0.99) {
# Check and convert the data.
.check_data(data = rets)
rets <- as.numeric(rets)
# Now construct the data frame.
k <- length(rets)
y.var <- tail(rets, k - 1)
x.var <- head(rets, k - 1)
data <- data.frame(y.var, x.var)
# Use lmtest to compute the p-values.
colnames(data) <- c("y", "x")
# Fit the linear model.
lmfit <- lm(formula = y ~ x,
data = data)
# Now compute the durbin-watson statistic.
dw <- lmtest::dwtest(formula = lmfit,
alternative = "two.sided")
# Get the test statistic (D) for the test.
stat <- dw$statistic
# Get the p-value for the D statistic.
p.value <- dw$p.value
# Determines the Z-score.
z.score <- qnorm(p.value)
# Compute the required threshold.
thresh <- abs(qnorm((1 - a) / 2))
# Return the results object.
return(c(stat, p.value, z.score,
abs(z.score) > thresh))
}

```

3. Ljung – Box test syntax suggested by Stuard Reid is:

```

test_ljungbox <- function(rets, a = 0.99, n.lags = 15) {
# Number of lags.
k <- length(rets)
rets <- as.numeric(rets)
# The denominator.

```

```

den <- k - seq(1, n.lags)
# Compute the lagged autocorrelations.
autocors <- c()
for (l in 1:n.lags) {
  rets.head <- head(rets, k - l)
  rets.tail <- tail(rets, k - l)
  correl <- cor(rets.head, rets.tail)
  autocors <- c(autocors, correl)
}

```

4. Breusch - Godfrey test syntax suggested by Stuard Reid is:

```

test_breuschgodfrey <- function(rets, a = 0.99) {
# Check and convert the data.
.check_data(data = rets)
rets <- as.numeric(rets)
# Now construct the data frame.
k <- length(rets)
y.var <- tail(rets, k - 1)
x.var <- head(rets, k - 1)
data <- data.frame(y.var, x.var)
# Use lmtest to compute the p-values.
colnames(data) <- c("y", "x")
# Fit the linear model.
lmfit <- lm(formula = y ~ x,
data = data)
# To compute the Beusch-Godfrey statistic.
bg <- lmtest::bgtest(formula = lmfit)
# Get the test statistic (D) for the test.
stat <- bg$statistic
# Get the p-value for the D statistic.
p.value <- bg$p.value
# Determines the Z-score.
z.score <- qnorm(p.value)
# Compute the required threshold.
thresh <- abs(qnorm((1 - a) / 2))
# Return the results object.
return(c(stat, p.value, z.score,
abs(z.score) > thresh))
}

```

5. Bartell Rank - Based Variance Ratio syntax suggested by Stuard Reid is:

```

test_bartellrank <- function(rets, a = 0.99) {
# Check and convert the data.
.check_data(data = rets)
rets <- as.numeric(rets)
# Run the bartell's rank test in randtests.
br <- randtests::bartels.rank.test(rets)
# Get the test statistic (D) for the test.
stat <- br$statistic
# Get the p-value for the D statistic.
p.value <- br$p.value
# Determines the Z-score.
z.score <- qnorm(p.value)
# Compute the required threshold.
thresh <- abs(qnorm((1 - a) / 2))
# Return the results object.
return(c(stat, p.value, z.score,
abs(z.score) > thresh))
}

```

6. *Lo - Mackinlay Variance Ratio syntax is:*

- using vrtest package

#Usage

Lo.Mac(y, kvec)

#Arguments

#y a vector of time series, typically financial return

#kvec a vector of holding periods