"Ovidius" University Annals, Economic Sciences Series
Volume XIX, Issue 2 /2019

Statistical Analysis of the Dija Index in the Period 2009-2019

Emilia Anuța Corovei
"1st of December 1918" University of Alba Iulia, Doctoral School in Accounting, Romania
emilia_lud@yahoo.com

Abstract

In this paper we analyzed the Dow Jones Industrial Average (DJIA) in the period 2009-2019. The Dow Jones Industrial Average (DJIA) is one of the most popular indexes; it indicates the value of 30 large, publicly owned companies from the United States. We used a lot of relevant statistic tests; we applied the ADF and PP tests both on the series of returns and on the series of the index analyzed. While the ADF test corrects the serial correlation by adding differentiated terms (lags), the PP test corrects the coefficient of the autoregressive process of the order. There are three versions of random walk: RW1, RW2 and RW3. In our paper we analyzed RW3 model for DIJA index. In the RW3 model, the errors are serially uncorrelated which allows movement dependencies.

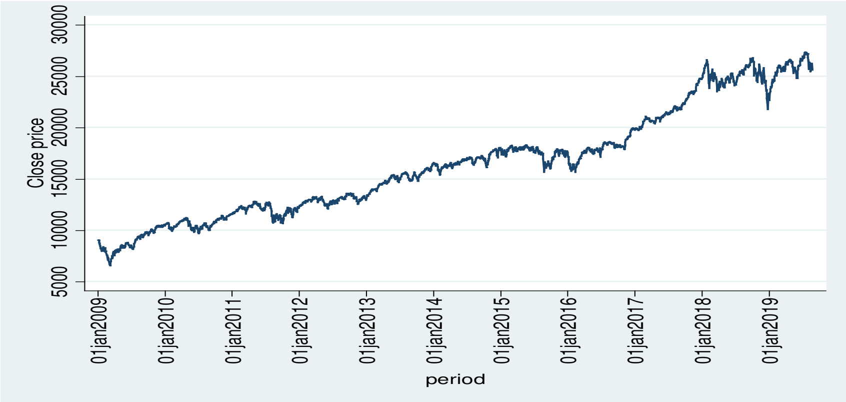
Key words: DIJA, statistical analysis, stationarity tests
J.E.L. classification: C23, F43, O47

1. Introduction

The Dow Jones Industrial Average (DJIA) is one of the most popular indexes; it indicates the value of 30 large, publicly owned companies from the United States. All these companies are also included in the S&P 500 Index. The Index is traded on New York Stock Exchange (NYSE) and the NASDAQ since 1896. The companies included in the DJIA as of March 18, 2019 are: the "3M Company, the Amercian Express Company, Apple Inc., The Boeing Company, Caterpillar Inc., Chevron Corporation, Cisco Systems, the Coca-Cola Company, DowDuPont Company, Exxon Mobil Corporation, The Goldman Sachs Group, The Home Depot Inc., International Business Machines Corporation, Intel Corporation, Johnson &Johnson, JP Morgan Chase &Co., McDonald's Corporation, Merck & Company Inc, Microsoft Corporation, Nike Inc., Pfizer Inc., Proctor and Gamble Co., The Travelers Companies Inc., UnitedHealth Group, United Technologies Corporation, Verizon Communications Inc., Visa Inc., Walmart Inc., Walgreens Boots Alliance and the Walt Disney Company" (*Investopedia, https://www.investopedia.com/terms/d/djia.asp*).

We considered the daily closing price of the DIJA Index; analyzed in the period 01.01.2009-23.08.2019. The figure below shows the evolution of the DIJA Index. It can be observed that in this process ($x_t, t \in Z$) starting with the date t = 12mar2009 a rupture is observed; so it will be a non-stationary variable (it does not fulfill the first condition, E(x_t)= constant). The evolution of the index fluctuated during the analyzed period. We observe a continuous increase but marked by continuous decreases at some points.

Figure no.1 - Evolution of the DIJA Index in the period 2009-2019



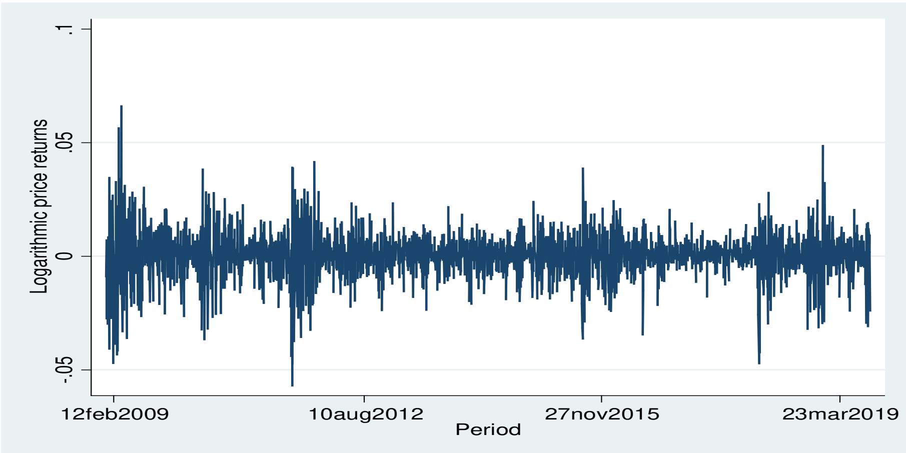
Source: Author's own calculation

The following formula was used to determine the price returns:

$$R_t^L = \ln(\frac{C_t + D_t}{C_{t-1}}) \text{ (equation 1)}$$

Based on the graph below, we can observe periods of high volatility (large fluctuations are followed by large fluctuations of the opposite direction). Volatility is due to new information coming on the market, information that influences the market as a whole. This volatility can be determined by the financial post- crisis (eg. in 2009 high values in 2012 small values).

Figure no. 2 - Logarithmic price returns for DIJA Index in the period 2009-2019



Source: Author's own calculation

The distribution of daily returns is shown in the table below.

Table no. 1 - The distribution of daily returns for DIJA index in the period 2009-2019

Skewness/Kurtosis tests for Normality			
Variable	**Obs Pr(Skewness)**	**Pr(Kurtosis) adj chi2(2)**	**Prob>chi2**
Logarithmic price return	2.7e+03 0.0000	0.0000	0.0000

Source: Author's own calculation

The Jarque-Bera statistic tests the hypothesis of normality of returns. According to the obtained result the distribution of returns deviates from a normal law or the probability of the Jarque-Bera test = 0, 00 <5%, therefore the returns do not follow the normal law.

2. Theoretical background

We found a lot of papers emphasizing the statistical analysis of an index. We presented below a random selection of these papers. Balan, S., Tejas and C. Sohong, C., (2018) made a time series analysis of the IT stock market during the 2007 – 2009 recession. Wang, J., and Wang, T., (2008) made a statistical analysis and data analysis of stock market by interacting particle models, Singh, P., & Thakral, A.,(2017) presented the statistical analysis of the indexes and its constituents. Bentian, L., and Dechang, P., (2018) analyzed the global stock index data during crisis period via complex network approach. Quentin, C.C, Wen-liang G., HsiehbYiuman, T., (1999) made a analysis of the price discovery on the S&P 500 index markets. In the last years, we have seen an increase in the number of studies that take into account the impact of the stock market volatility on the economy, on the economic growth and on finding the most efficient investment strategies.

We continue to consider a range of random processes, called stationary processes. We assume a random process Y_t, where $t \in Z$. For the observation related to moment t, the random variable Y_t, is defined:

- The mean: $E(Y_t) = \mu_t$,
- The variance: $Var(Y_t) = \sigma_t^2$,
- The covariance between two variables Y_t and Y_s: $\gamma_{ts} = cov\,(Y_t\,, Y_s\,)$.

Because we have only one observation for each variable Y_t , it is impossible to estimate these elements. The estimation becomes possible for a particular class of random processes, called stationary processes. A process is stationary when the probability distribution is stable regardless of time. In the case of stationarity in the strong sense, all the moments of the variables are constant, and in the case of stationarity in the weak sense, only the mean, the variance and the covariance are constant over time. By definition, a series whose mean, dispersion and covariance are constant over time is stationary. A stationary series tends to return to the value of the mean and to fluctuate around it (it has a finite variance). A nonstationary series has a different average at different times. The question of the stationarity of a series depends on the existence of a root.

3. Research methodology and Findings

3.1. Testing the stationarity hypothesis

In this section we will apply the ADF and PP tests both on the series of returns and on the series of the index analyzed. While the ADF test corrects the serial correlation by adding differentiated terms (lags), the PP test corrects the coefficient of the autoregressive process of the order, AR (1).

Augmented Dickey-Fuller (ADF) test detects non-stationarity. It is used to test whether a series is stationary (relative to the average or relative to the deterministic tendency), respectively to determine the order of integration. The decision on the null hypothesis is as follows:

- the null hypothesis is rejected, if the series of returns has no unit root (random walk), is relative stationary to the average or deterministic.
- the null hypothesis is accepted, if the series of returns has a unit root, is nonstationary, with stochastic tendency.

If the null hypothesis is not rejected, then the ADF test will continue to be applied to determine the unit root within the first order differences. In order to determine the integration order, the test is applied for the initial data, differentiated by order 1 and 2 respectively.

3.2. Application of the ADF test on the index series

We consider three variations of the ADF test for the index data series (Please see: http://www.econ.uiuc.edu/~econ508/Stata/e-ta8_Stata.html):

- models with intercept and trend;
- models with intercept, but without trend;
- models without both intercept and trend.

Table no. 2 - ADF models for DIJA index series in the period 2009-2019

Results	(a) models with intercept and trend		(b) models with intercept, but without trend;		(c) models without both intercept and trend	
	t-Statistic	Prob.*	t-Statistic	Prob.*	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.437	0.3599	-0.049	0.9543	1.575	Is not in the confidence interval
Index(-1)	-2.44	0.015	-0.05	0.961	1.58	0.115
Constant	2.49	0.013	0.56	0.578		
Trend	2.45	0.014				

Source: Author's own calculation

Based on the above results the series of index has a unit root, is nonstationary with a stochastic tendency (of a random walk type), but not deterministic (the constant and the trend have no significant coefficients). In order to determine the degree of integration we will apply the test for first differences data. The resluts are presented below:

Table no. 3 - ADF models for the first differences in DIJA index series in the period 2009-2019

	(b) models with intercept, but without trend;		(c) models without both intercept and trend	
Results	**t-Statistic**	**Prob.***	**t-Statistic**	**Prob.***
Augmented Dickey-Fuller test statistic	-21.262	0.0000	-21.205	Is not in the confidence interval
D.Index(-1)	-21.26	0.000	-21.20	0.000
Constant	1.42	0.157		

Source: Author's own calculation

Based on the above results, in both analyzed cases the first difference of the index series data is stationary, the process will be integrated in order 1 (I (1)).

3.3. Application of the ADF test on the return index series.

From what we have seen the ADF test supports the first order integration hypothesis of the series of the index, which leads to the conclusion that the returns are stationary, because they are determined as the difference between the logarithm of two consecutive exchange rates. The results obtained by applying the test are the following:

Table no. 4 - ADF models for DIJA return index series in the period 2009-2019

Results	(a) models with intercept and trend		(b) models with intercept, but without trend;		(c) models without both intercept and trend	
	t-Statistic	Prob.*	t-Statistic	Prob.*	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-26.297	0.000	-26.306	0.000	-26.253	Is not in the confidence interval
Index(-1)	-26.30	0.000	-26.31	0.038	-26.25	0.000
Constant	0.01	0.994	1.55	0.121		
Trend	0.77	0.439				

Source: Author's own calculation

If we apply the ADF test on the differences of order 1 and 2 we obtained the same results.

In conclusion, the series presenting the index is non-stationary and has a single unit root, ie. it is integrated by the first order, while the series of returns is stationary. In conclusion the first random walk condition is satisfied.

3.4. Martingale and the random walk model

The random walk model is very useful in studying stock exchange rates and indexes characterized by periods of stock market growth and decline, in which the stock market often changes its position. The random walk hypothesis was based on two categories of researchers - theorists and practitioners - who reached the same conclusions. In their studies Moore (1962), Granger & Morgenstern (1963) and Fama (1965) empirically demonstrate that stock price developments are random. At the same time, Samuelson (1965) and Mandelbrot (1966) demonstrate that if information is gathered at low costs, the trading of shares is performed without costs, and all participants in the capital market interpret the new information identically, then the evolution of stock prices behaves like a random walk.

In the theory of probability a martingale model is a model of a fair game where no information from past events can help predict future winnings. In particular, a martingale is a sequence of random variables (for example, a stochastic process) for which, the best forecast for the next course is today's course. It is observed that the "random walk" model coincides with the martingale model, if it is reasoned only in terms of hope. The martingale model does not make any assumptions about moments greater than one, while the "random walk" model requires that all moments are independent.

There are three versions of random walk: RW1, RW2 and RW3. RW1 is the most restrictive version and takes into account the fact that the errors are independent and identically distributed, that is, they have an average equal to zero and the variance, which indicates that the returns are uncorrelated in series so that future prices cannot be predicted based on past prices. The RW2 model has independent errors but they are not identically distributed, which allows unconditional heteroscedasticity. In the RW3 model, the errors are serially uncorrelated which allows movement dependencies.

Testing random walk model - RW3. In order to investigate the random walk hypothesis that the residues show poor white noise, ie they are dependent, but not correlated, we applied the Ljung-Box test (Q Statistic).

The Ljung-Box test (Q-statistic) for lag k has the null hypothesis: there is no autocorrelation until lag k. The Q-statistic test is used to test whether a series is a white noise. However, a problem arises in determining the number of lags. If it is too small it might not detect the serial correlations for a larger order of the lags. If the test is too high, it may be weaker, the significant correlations for one lag may be diminished with the insignificant ones for other lags.

Decision on the null hypothesis: If the calculated value of the test is greater than the value existing at a fixed significance level (1%, 5%, 10%), then the null hypothesis is rejected. When Q does not differ significantly from 0, the first M self-correlations are insignificant. In practice M is considered between 12 and 20, depending on the length of the series.

Table no. 5 - Ljung-Box Q-Statistic Test

LAG	AC	PAC	Q	Prob>Q
1	-0.060	-0.060	9.3564	0.002
2	0.017	0.013	10.093	0.006
3	-0.017	-0.015	10.843	0.013
4	0.018	0.016	11.722	0.020
5	-0.055	-0.053	19.502	0.002
6	0.002	-0.005	19.508	0.003
7	0.026	0.028	21.296	0.003
8	-0.013	-0.012	21.751	0.005

9	-0.014	-0.015	22.286	0.008
10	0.028	0.024	24.250	0.007
11	0.007	0.009	24.375	0.011
12	-0.016	-0.013	25.012	0.015
13	0.018	0.016	25.860	0.018
14	-0.054	-0.055	33.448	0.002
15	-0.041	-0.045	37.783	0.001
16	0.016	0.015	38.415	0.001
17	0.034	0.032	41.501	0.001
18	-0.033	-0.028	44.252	0.001
19	-0.030	-0.038	46.626	0.000
20	0.009	0.000	46.827	0.001

Source: Author's own calculation

For correlation (AC) and partial correlation coefficients (PAC) the null hypothesis is rejected. According to the Ljung-Box Q-Statistic test, the return series have autocorrelations, which means that the residuals have linear dependencies, the return series do not have a random walk as well.

The model for which AIC and BIC will be minimal is chosen and we estimated the parameters of the chosen model. The most attractive model is ARMA (5,1). This model is used to remove the linear structure.

Table no. 6 - AIB and BIC results

AIC							
	p						
q		**0**	**1**	**2**	**3**	**4**	**5**
	0		-6.010278	-6.126809	-6.202760	-6.227478	-6.258950
	1	-6.421654	-6.424373	-6.426416	-6.425666	-6.425893	**-6.428763**
BIC							
	p						
q		**0**	**1**	**2**	**3**	**4**	**5**
	0		-6.005733	-6.119989	-6.193664	-6.216104	-6.245298
	1	**-6.417556**	-6.417323	-6.414296	-6.412245	-6.412835	-6.407729

Source: Author's own calculation

For the chosen model, we will build the residuals correlogram to see if there are any autocorrelations after the linear component filtering. If the model is a valid one, the residuals should be uncorrelated, a condition that is verified in the table below. The probabilities are much higher than the acceptance threshold, which means that the null hypothesis is accepted and the residuals do not have linear dependencies.

Table no. 7 - Ljung-Box Q-Statistic Test for the residual series

LAG	**AC**	**PAC**	**Q**	**Prob>Q**
1	-0.001	-0.001	0.0010	
2	-0.001	-0.001	0.0042	
3	-0.005	-0.005	0.0596	
4	-0.002	-0.002	0.0670	
5	-0.005	-0.005	0.1380	

6	-0.004	-0.004	0.1758	
7	0.027	0.027	2.1060	0.147
8	-0.019	-0.019	3.0787	0.215
9	-0.011	-0.011	3.4054	0.333
10	0.020	0.021	4.4883	0.344
11	0.011	0.011	4.7940	0.442
12	-0.012	-0.011	5.1408	0.526
13	0.010	0.010	5.3866	0.613

Source: Author’s own calculation

4. Conclusions

In this paper we analyzed the DIJA index for a period between 2009-2019. The logitmatic returns for this index were determined. In a first phase I studied the hypotheses of stationarity and unitary root through ADF and PP tests. While the ADF test corrects the serial correlation by adding differentiated terms (lags), the PP test corrects the coefficient of the autoregressive process of the order. There are three versions of random walk: RW1, RW2 and RW3. RW1 is the most restrictive version and takes into account the fact that the errors are independent and identically distributed, that is, they have an average equal to zero and the variance, which indicates that the returns are uncorrelated in series so that future prices cannot be predicted based on past prices. The RW2 model has independent errors but they are not identically distributed, which allows unconditional heteroscedasticity. In the RW3 model, the errors are serially uncorrelated which allows movement dependencies. In our paper we analyzed RW3 model for DIJA index. The rejection of the random walk hypothesis is due to the linear and nonlinear correlations we identified in the analyzed series.

5. Acknowledgment

This paper was supported by the project "Dezvoltarea învățământului terțiar universitar în sprijinul creșterii economice – PROGRESSIO - Contract no. POCU/380/6/13/125040", project co-funded from European Social Fund through Human Capital Operational Programme POCU 2014-2020.

6. References

- Balan, S., Tejas, C. Sohong, C., 2018. *A Time Series Analysis of the IT Stock Market during the 2007 – 2009 Recession,* Big Data (Big Data) 2018 IEEE International Conference on, pp. 5295-5297, 2018.
- Bentian, L., Dechang, P., 2018. Analysis of global stock index data during crisis period via complex network approach. *PLoS ONE* 13(7): e0200600. https://doi.org/10.1371/journal.pone.0200600
- Quentin, C.C, Wen-liang G., HsiehbYiuman, T., 1999. Price discovery on the S&P 500 index markets: An analysis of spot index, index futures, and SPDRs, *International Review of Financial Analysis*, Vol. 8, Issue 1, 1999, Pages 21-34,
- Singh, P., & Thakral, A., 2017. *Stock market: Statistical analysis of its indexes and its constituents.* 962-966. 10.1109/SmartTechCon.2017.8358514.
- Wang, J., Wang, T., 2008. Statistical Analysis and Data Analysis of Stock Market by Interacting Particle Models, *Journal of Computers*, Vol. 3, No. 12, december 2008;
- www.finance. yahoo.com
- http://www.econ.uiuc.edu/~econ508/Stata/e-ta8_Stata.html