ISSN: 2278-3369
International Journal of Advances in Management and Economics
Available online at www.managementjournal.info

RESEARCH ARTICLE

The Evolution of S & P 500 Index, Forecasted Using an Autoregressive Integrated Model

Florin Dan Pieleanu

Academy of Economic Studies Bucharest, School of Cybernetics, Statistics and Economic Informatics.

Abstract

The indexes Standard&Poor's and Dow Jones Industrial Average were each, along time, the benchmarks of the American stock exchange. But lately, the first one gained the advantage, because it is composed of the stocks from multiple companies, carefully chosen on the base of strict criteria. Hence, its evolution is interesting, due to many reasons. The present article makes a forecast for S&P 500's value on a 30-day period, with input data covering 1 year. The forecast is constructed on an autoregressive integrated model, because this type of model proved to be adequate for short- and medium-term periods. The offered results are compaired in the end with the real values of the index, in order to see how accurate is the estimation. The conclusions will either confirm once again, or will refute the ability of the autoregressive model to forecast future values, but they will be 100% true only for the studied context.

Keywords: *Index, S&P 500, Forecast, Benchmark, Criteria.*

Introduction

The Standard&Poor's 500 index, or shorter S&P 500 is an American index, based on the market value of 500 companies from NYSE (New York Stock Exchange) or NASDAQ (National Association of Securities Dealers Automated Quotations). Many are considering it as the best picture of the American stock exchange, due to its special weighting scheme and to its different composition. The first S&P 500 index was introduced in 1923, and its current form exists from March 1957. Its evolution from 1950 to 2013 looks like this:

Figure 1: The evolution of S&P 500 in 1950-2013 period

The composition of the index is chosen by a committee, using eight main criteria for this selection: stock value, liquidity, the public held percent, financial viability, time since public listing, domicile, activity area, the action exchange.

Weighting of S&P 500 is made by the stock value of the included companies: the evolution of prices for companies with a higher value has a bigger impact then the one for companies with a lower value.

For calculating the index's value, the sum of the stock value for all 500 included companies is divided by a parameter, called divisor:

$S \& P\ 500 = (\Sigma P_i{*}Q_i)/Div.$
Where:

P_i = the price of the shares for company "i", included in the index
Q_i = number of available public shares for company "i"
Div. = divisor with an established value

S&P 500 is considered by many as the benchmark of the American stock exchange, or even the definition of it. This position was also held by Dow Jones Industrial Average index, but because it is composed of only 30 companies, the first gained the upper-hand.

The Autoregressive Model. Related Studies

The ARIMA model (or the Box-Jenkins model) was quite vastly used in various forecasting attempts, being considered one of the most successful. It is based on the past values of the series, and on its residual values, and it does not assume any basic relation, unlike other models. ARIMA proves to be a strong model, with good results, mainly in short- and medium-term periods.

The model can be written as ARIMA (p,d,q), where p,d and q are parameters, with the following meanings:

p = the order of the autoregressive (AR) model
d = the integration order
q = the order of the moving average (MA) model

The future value of the analyzed variable is, according to the model, a linear combination between its past values and its residual terms, as:

$$Y_t = \Phi_0+\Phi_1*Y_{t-1}+\Phi_2*Y_{t-2}+\ldots\ldots+\Phi_p*Y_{t-p}+\xi_t-\theta_1*\xi_{t-1}-\theta_2*\xi_{t-2}-\ldots\ldots-\theta_q*\xi_{t-q}$$

Where:

Y_t is the time series for the analyzed variable
Φ_i *and* θ_j are the coefficients' series
ξ_t is the residual terms' series

The ARIMA model (or its more complex forms) was widely used for estimating future value attempts, [1,5,7,11,14] being only some examples in a very comprehensive list.

Methodology

Used Data

Excel and Eviews softwares were used. Data were daily recordings for S&P 500 index, for 1 year (stock exchange), from 05.01.2015 to 31.12.2015, with a total of 258 observations. The goal is to estimate the future value of the index for 30 days, using an autoregressive model (more about this in a previous article by the same author [10]), then to compaire the forecasted values with the real ones, in order to see how accurate or inaccurate the estimation was.

Data Characteristics

In the specified time horizon, the series' graphic looks like this:

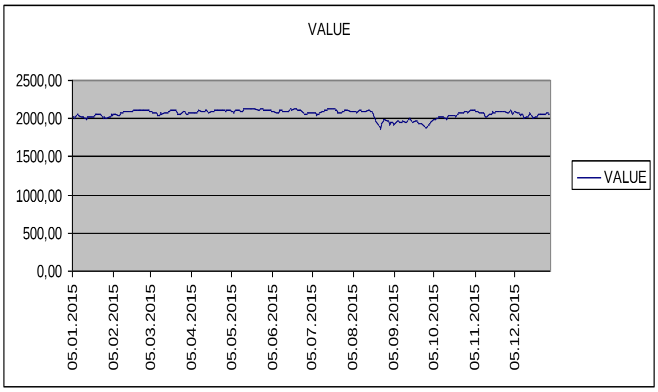


Fig.2: The index' evolution in year 2015

Showing a certain stability, without important variations. The next analysis is aiming to check for the series' stationarity. For this, 3 elements will be considered: the graph, the correlogram and the Dickey-Fuller test. These are offering the below situation:

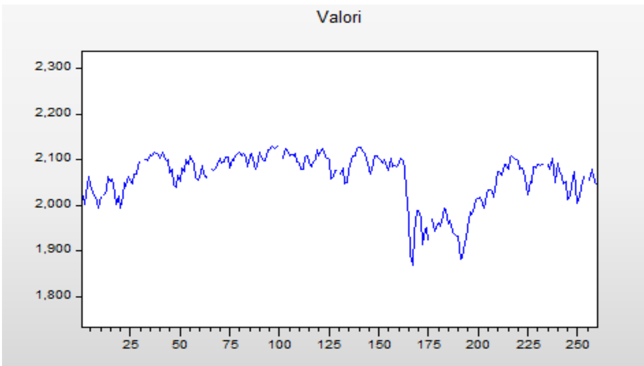


Fig.3: The original series' graph in the studied time period

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.909	0.909	210.03	0.000
		2	0.835	0.048	387.92	0.000
		3	0.786	0.112	546.13	0.000
		4	0.737	-0.001	685.78	0.000
		5	0.713	0.140	816.91	0.000
		6	0.689	0.023	940.07	0.000
		7	0.660	-0.002	1053.3	0.000
		8	0.622	-0.054	1154.4	0.000
		9	0.575	-0.066	1241.1	0.000
		10	0.551	0.091	1321.1	0.000
		11	0.539	0.061	1397.9	0.000
		12	0.526	0.021	1471.4	0.000
		13	0.511	-0.012	1541.0	0.000
		14	0.483	-0.048	1603.4	0.000
		15	0.445	-0.058	1656.8	0.000
		16	0.423	0.048	1705.1	0.000
		17	0.406	0.009	1749.8	0.000
		18	0.381	-0.068	1789.5	0.000

Fig.4: The original series' correlogram

Null Hypothesis: VALORI has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.967288	0.0395
Test critical values:	1% level	-3.458347	
	5% level	-2.873755	
	10% level	-2.573355	

*MacKinnon (1996) one-sided p-values.

Fig.5: The Dickey-Fuller test - original series

Which prove the non-stationarity of the series. Before the actual forecast process, it is necessary to transform the series into a stationary one. The first difference will be implied, which leads to the following changes, regarding the before-mentioned criteria:

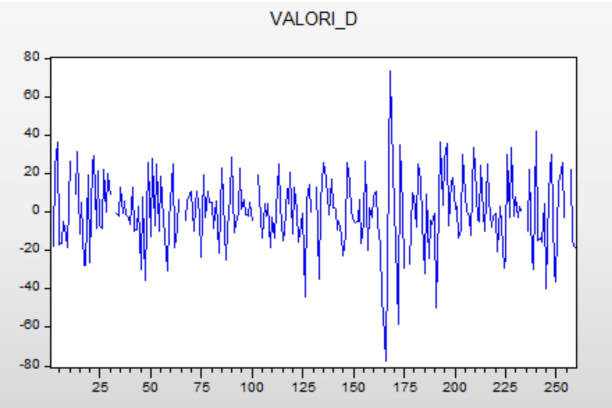


Fig.6: The integrated series' graph in the studied time period

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.079	0.079	1.5202	0.218
		2	-0.125	-0.132	5.3895	0.068
		3	-0.076	-0.056	6.8303	0.078
		4	-0.099	-0.108	9.2785	0.055
		5	-0.045	-0.048	9.7907	0.081
		6	0.037	0.013	10.138	0.119
		7	-0.028	-0.060	10.336	0.170
		8	-0.052	-0.058	11.028	0.200
		9	0.030	0.020	11.253	0.259
		10	-0.029	-0.054	11.472	0.322
		11	0.020	0.020	11.576	0.396
		12	0.054	0.029	12.329	0.420
		13	0.069	0.068	13.572	0.405
		14	0.004	0.005	13.577	0.482
		15	-0.015	0.003	13.638	0.553
		16	-0.054	-0.032	14.409	0.568
		17	0.003	0.027	14.412	0.638

Fig.7: The integrated series' correlogram

Null Hypothesis: VALORI_D has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-14.02958	0.0000
Test critical values:	1% level	-3.458347	
	5% level	-2.873755	
	10% level	-2.573355	

*MacKinnon (1996) one-sided p-values.

Fig.8: The Dickey-Fuller test-integrated series

This time, the graph, correlogram and Dickey-Fuller test are showing the stationarity of the series. Hence, this being obtained, we can continue with our goal.

Estimating the Model's Parameters

The series proves to be 1 level integrated. The next step is to identify the p and q parameters. The criteria for this goal are represented by: lowest values for the Akaike, Schwarz and S.E (standard error), highest value for adjusted R^2. By studying the autocorrelation and the partial autocorrelation, one cannot get a precise clue regarding the model, so one must try several ARIMA models, in order to see which one is the best for estimating the future S&P 500 values. The summary for the tests with different models can be found in the below table:

Table 1: comparison of chosen criteria for the studied models

ARIMA	Akaike	Schwarz	Adjusted R^2	S.E
(1,0,0)	8,796	8,825	0,0023	19,593
(1,0,1)	8,795	8,838	0,0080	19,537
(0,0,1)	8,794	8,823	0,0046	19,570
(2,0,0)	8,786	8,829	0,0182	19,436
(0,0,2)	8,785	8,828	0,0195	19,424
(3,0,0)	8,791	8,849	0,0166	19,452
(0,0,3)	8,787	8,845	0,0216	19,402
(2,0,1)	8,781	8,839	0,0278	19,341
(1,0,2)	8,780	8,838	0,0286	19,333

The ARIMA (1,0,2) model seems to be the fittest for our goal, because it checks the most criteria, having the lowest values for Akaike and Standard Error, and the highest value for adjusted R^2. As a consequence, it will be used in the present paper.

Its testing form is: $S\&P_t = \Phi_1 * S\&P_{t-1} + \xi_t - \theta_1 * \xi_{t-1} - \theta_2 * \xi_{t-2}$

Meaning that the value for the index in moment t depends mainly on its value in t-1 moment and on the residual terms in t, t-1 and t-2 moments.

Testing adequacy is also required, that is checking that the residual values of the autocorrelation function are „white noise", and that the comparison of the theoretical and empirical values for the autocorrelation and partial autocorrelation is valid:

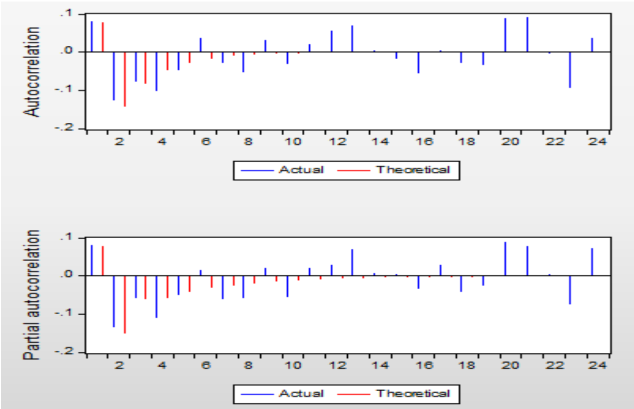
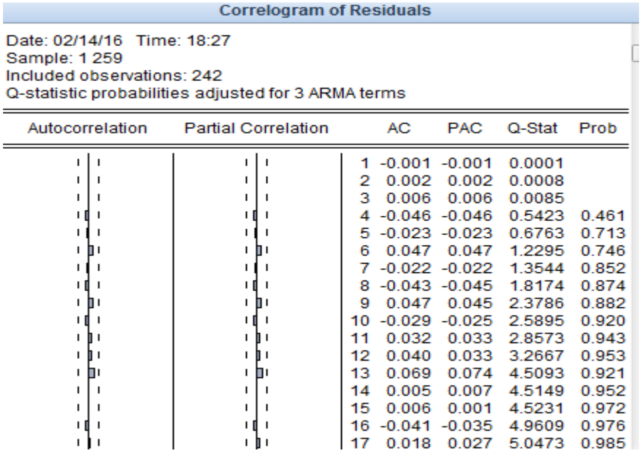


Fig.9: Autocorrelation and partial autocorrelation – theoretical vs. empirical values



Correlogram of Residuals

Date: 02/14/16 Time: 18:27
Sample: 1 259
Included observations: 242
Q-statistic probabilities adjusted for 3 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.001	-0.001	0.0001	
		2	0.002	0.002	0.0008	
		3	0.006	0.006	0.0085	
		4	-0.046	-0.046	0.5423	0.461
		5	-0.023	-0.023	0.6763	0.713
		6	0.047	0.047	1.2295	0.746
		7	-0.022	-0.022	1.3544	0.852
		8	-0.043	-0.045	1.8174	0.874
		9	0.047	0.045	2.3786	0.882
		10	-0.029	-0.025	2.5895	0.920
		11	0.032	0.033	2.8573	0.943
		12	0.040	0.033	3.2667	0.953
		13	0.069	0.074	4.5093	0.921
		14	0.005	0.007	4.5149	0.952
		15	0.006	0.001	4.5231	0.972
		16	-0.041	-0.035	4.9609	0.976
		17	0.018	0.027	5.0473	0.985

Fig.10: The correlogram of the residual terms for the chosen model

It is obvious that the model is adequate, and its residual terms are „white noise". As a consequence, one can continue with the actual forecast.

Results and Conclusions

The below table highlights the estimated values for S&P 500, next to the real ones, for a better comparison:

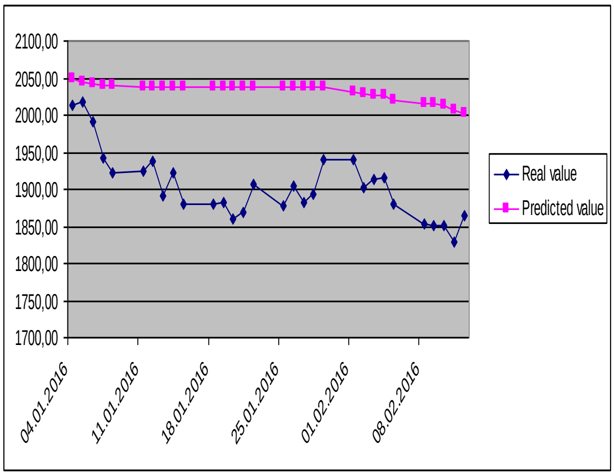


Fig.11: The common graph for estimated and real S&P 500 index values

Table 2: Estimated vs. real index values

Date	Estimated value	Real value
2016-01-04	2047,91	2012,66
2016-01-05	2043,61	2016,71
2016-01-06	2041,16	1990,26
2016-01-07	2039,76	1943,09
2016-01-08	2038,97	1922,03
2016-01-11	2038,52	1923,67
2016-01-12	2038,26	1938,68
2016-01-13	2038,12	1890,28
2016-01-14	2038,04	1921,84
2016-01-15	2037,99	1880,33
2016-01-18	2037,97	1880,67
2016-01-19	2037,95	1881,33
2016-01-20	2037,94	1859,33
2016-01-21	2037,93	1868,99
2016-01-22	2037,92	1906,90
2016-01-25	2037,91	1877,08
2016-01-26	2037,89	1903,63
2016-01-27	2037,87	1882,95
2016-01-28	2037,86	1893,36
2016-01-29	2037,85	1940,24
2016-02-01	2032,13	1939,38
2016-02-02	2028,87	1903,03
2016-02-03	2027,01	1912,53
2016-02-04	2025,95	1915,45
2016-02-05	2019,89	1880,05
2016-02-08	2016,44	1853,44
2016-02-09	2014,47	1852,21
2016-02-10	2013,34	1851,86
2016-02-11	2006,93	1829,08
2016-02-12	2003,27	1864,78

And the common graph for the two series (the estimated and the real ones) looks like this:

The autoregressive model offers a quite linear trend for Standard and Poor's value, like a slightly and steady decrease. This result confirms other conclusions, even one obtained by the same author. Hence, ARIMA tends to show mellow variations, but without having a very good ability to show major changes. The real values of the index also picture some decrease, but a more deep one, if one compaires it with the one offered by ARIMA. Because the chances for bigger changes increase as time goes by, on longer periods other types of models (ANN, fuzzy, vector machines etc.) are performing a little better than autoregressive models [1-17]. These latter ones have the upper-hand in shorter time periods, being more exact and accurate, due to the fact that only discrete changes are usually happening. Future research that use special forms of ARIMA- that is ARIMAX or ARFIMA – are targeted.

References

1. Adebiyi A. Adewumi, Ayo CK (2014) Stock Price Prediction Using the ARIMA Model", UKSim-AMSS 16th International Conference on Computer Modelling and Simulation.

2. Cao L, Tay F (2001) Financial forecasting using vector machines, Neural Computer Appl., 184-192,

3. Demuth H, Beale M (1998) Neural network toolbox: For use with MATLAB, Natick, MA: The Math Works, Inc.,

4. Hosseini H, Luo D, Reynolds K (2006) The comparison of different feed forward neural network architectures for ECG signal diagnosis, Medical Eng. Phys., 372-378.

5. Javier E. Rosario, Francisco N, Antonio J (2003) ARIMA Models to Predict Next Electricity Price, IEEE Transactions on Power Systems, 18:1014-1020.

6. Lam M (2004) Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis, Decis Support Syst., 567-581.

7. Mondal P, Shit L, Goswami S (2014) Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices", International Journal of Computer Science, Engineering and Applications (IJCSEA), vol.4,

8. O'Connor N, Madden M (2006) A neural network approach to predicting stock exchange movements using external factors, Knowl. Base Syst., 371-378.

9. Pieleanu F (2016) Comparative study in estimating Volkswagen's price: ARIMA versus ANN", sent for publishing.

10. Pieleanu F (2016) Predicting the evolution of BET index, using an ARIMA model, sent for publishing, 2016

11. Saxena P, Merh N, Pardasani R (2010) A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting", Journal of Business Intelligence, 3:23-43.

12. Vanstone G Finnie (2009) An empirical methodology for developing stock market trading systems using artificial neural networks", Expert System Appl., 6668-6680.

13. Yao J, Tan L, Poh H (1992) Neural networks for technical analysis: a case study on KLCI", International Journal of Theoretical and Applied Finance, 221-241.

14. Zhang G, Patuwo B, Hu Y (1998) Forecasting with artificial neural networks: the state of the art", International Journal of Forecasting, 35-62.

15. https://research.stlouisfed.org/fred2/series/SP500/downloaddata

16. http://www.investopedia.com/terms/s/sp500.asp

17. https://en.wikipedia.org/wiki/S%26P_500_Index