Informatics and Control Problems 42 Issue 1 (2022)
journal homepage: www.icp.az

The use of autoregressive process AR(1) in the analysis of some finance indices

H.A. Jafarova[1*], R.T. Aliyev[2,3]

[1]*Azerbaijan State University of Economics, Baku, Azerbaijan*
[2]*Baku State University, Baku, Azerbaijan*
[3]*Institute of Control Systems of Azerbaijan National Academy of Sciences, Baku, Azerbaijan*

A R T I C L E   I N F O

Article history:
Received 16.03.2022
Received in revised form 30.03.2022
Accepted 11.04.2022
Available online 25.05.2022

Keywords:
Autoregressive process AR(1)
Markov walk
Finance
Price of stock Indices
R programming language
Economic efficiency

A B S T R A C T

The article considers an autoregressive process AR(1) with finance application. The solution of many control problems comes down to autoregressive process. We study the first-order autoregressive process AR(1) with application to the S&P500 and NASDAQ indices for 01/01/2017-01/08/2021. The main purpose of the study is to investigate the effects of the Covid-19 pandemic on stock indices. For statistical computing and visualization, R language and environment are used.

1. Introduction

Frequently, the values of a time series are highly correlated with the values that precede and succeed them. This type of correlation is called autocorrelation. Autoregressive modeling is a technique used to forecast time series with autocorrelation. First-order autocorrelation refers to the association between consecutive values in a time series. A $p > 2$ order autocorrelation refers to the correlation between values in a time series that are p periods apart.

In [1]-[3] different types of random walks were considered. But this article considers the application in the finances of random walks caused by AR(1).

In [4], a family of moments of the first passage time of a parabolic boundary by a perturbed Markov random walk described by an autoregressive process AR(1) is considered.

In [5], the tail behavior of the stopping time $T_0 = min\{n \geq 1 : X_n \leq 0\}$ is considered for a class of one-dimensional autoregressive sequences(X_n), and existing general analytical approaches to this and related problems discussed, proposing a new one, on the basis of a renewal-type decomposition for the moment generating function of T_0 and on the analytical Fredholm alternative.

In [6]-[7], limit theorems for a family of the first passage times of a parabola by the sums of the squares and the generalization autoregression process of order one AR (1) are studied.

In [8], estimation of a missing value for a stationary AR(1) with exponential innovations considered, and two methods of estimation of the missing value, with respect to Pitman's measure of closeness, are compare.

*Corresponding author.
E-mail addresses: hilala-jafarova@unec.edu.az (H.A. Jafarova), rovshanaliyev@bsu.edu.az (R.T. Aliyev).

www.icp.az/2022/1-03.pdf https://doi.org/10.54381/icp.2022.1.03

[9] answers the requests of teachers of financial mathematics and engineering by making a bias towards probabilistic and statistical ideas and the methods of stochastic calculus in the analysis of market risks.

In [10], random coefficient AR(1) processes is investigated, where the random coefficient satisfies some suitable conditions. Conditional least squares estimator is shown to be consistent and to be asymptotically normality distributed. This extends the limit theory for stationary and near-stationary cases.

In [11], the focus is on the relationship between U.S. Dollar Index (USDX) and Gold prices from 2010.01.01 to 2019.01.01. The aim of this study is to analyze and determine the character of the comovement between price levels. For best visualization, graphics of USDX and Gold price at monthly intervals with indicators Exponential Moving Average (EMA) and Relative Strength Index (RSI) are given. Main results of Regression Model (MSE, RMSE, MAE, RMSLE, Mean Residual Deviance, R-squared, Null Deviance and etc.) for reported on validation and training data are compared.

Unlike the above studies, first-order autoregressive process $AR(1)$ with application to the S&P500 and NASDAQ indices for 01/01/2017-01/08/2021 is considered.

Regression analysis refers to classical statistical methods. With their wide range of possibilities, different regression procedures have long and successfully been used in engineering practice to identify processes. Another approach to the description of the main trend of the time series and forecasting is the autoregressive model. Its construction is preceded by an assessment of the availability of autocorrelation in the studied series. Autocorrelation is an interdependence between successive values of the time series. First-order autocorrelation assesses the degree of interdependence between adjacent values of the time series.

Let $\xi_n, n \geq 1$ be a sequence of independent and identically distributed random variables defined on some probability space $(\Omega, \mathcal{F}, P)$. It is well known that the first-order autoregressive process ($AR(1)$) is defined using a recurrent relation of the form

$$X_n = \beta X_{n-1} + \xi_n, \quad n \geq 1, \tag{1}$$

where $\beta \in R = (-\infty, +\infty)$ is some fixed number and it is assumed that the initial value X_0 does not depend on the innovation $\{\xi_n\}$. Some aspects of Markov renewal and renewal-reward processes also consider in the papers Rahimov et al (2019), (2020), Aliyev and Khaniyev (2013), Aliyev et al (2015), Aliyev and Bayramov (2017).

2. Problem statement

The simplest autoregressive model uses only the most recent outcome of the time series observed to predict future values. For a time series Y_t such a model is called a first-order autoregressive model, often abbreviated *AR(1)*, where the 1 indicates that the order of autoregression is one:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

is the *AR(1)* population model of a time series Y_t.

Main purpose of this study is application of first-order autoregressive process $AR(1)$ for the prediction of a stock's future prices based on its past performance. As it is known, S&P500 and NASDAQ, the world's leading indices, also influenced the change in the prices of other currency pairs. This paper presents these indices based on empirical data describing the *AR(1)* first-order autoregressive process.

The Standard and Poor's 500, or simply the S&P 500, is a stock market index tracking the performance of 500 large companies listed on stock exchanges in the United States. It is one of the most commonly followed equity indices. As of December 31, 2020, more than $5.4 trillion was invested in assets tied to the performance of the index.

The S&P 500 index is a free-float weighted/capitalization-weighted index. As of September 30, 2021, the 9 largest companies on the list of S&P 500 companies accounted for 28.1% of the market capitalization of the index and were, in order of weighting, Apple Inc., Microsoft, Alphabet Inc. (including both class A&C shares Amazon.com, Facebook, Tesla, Inc., Nvidia, Berkshire Hathaway and JPMorgan Chase&Co.

The components that have increased their dividends in 25 consecutive years are known as the S&P 500 Dividend Aristocrats and used to forecast the direction of the economy.

3. Mathematical model and solution algorithm of the optimization problem

Initially, descriptive analysis was performed for S&P500 and NASDAQ.

Table 1.

Descriptive analysis of S&P500 indices price

	Highest:	Lowest:	Difference:	Average:	Change %:
S&P500	4,544	2,191.6	2,352.72	3,047	92.73

As a result of the descriptive analysis, it was determined that the average price of the S&P500 indices was 3,047 USD, minimum price was 2,191.6 USD in January 2017, and the maximum price was 4,544 USD in August 2021. Percentage of change was 92.73% for 01/01/2017-01/08/2021 time interval.

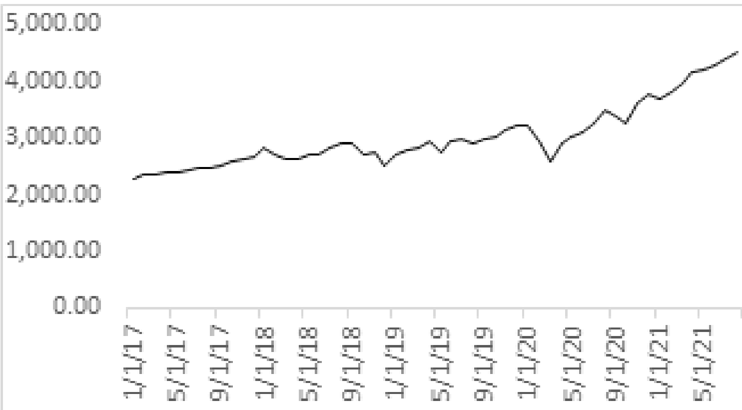

Fig. 1. Price of S&P500 indices beetween 01/01/2017-01/08/2021

```
> plot.ts(sp500_ts, main="SP500 Stock Prices", xlab="Data", ylab="Prices")
```

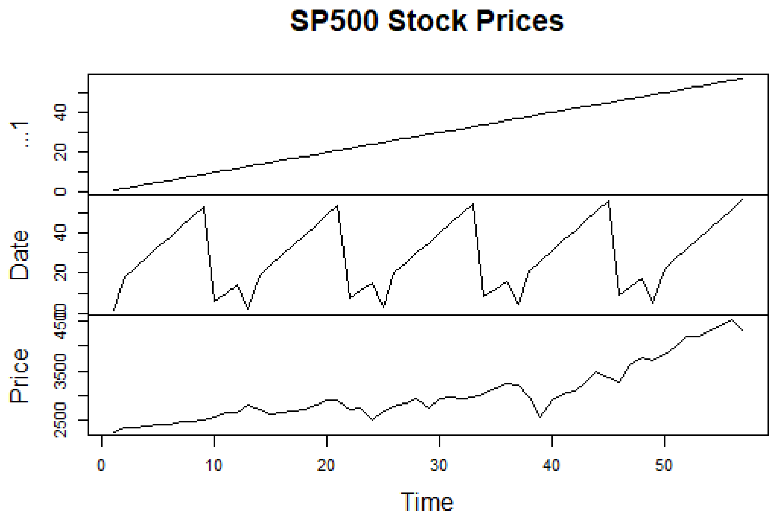

Fig. 2. Price of S&P500 index beetween 01/01/2017-01/08/2021

Let us consider the S&P500 stock prices from 01/01/2017 to 01/09/2021, which we have as a time series object sp500_ts. Below we have plotted the stock prices in the left chart and a scatter plot of the stock prices with a lag of 1 on the right-hand side. We can clearly see a strong positive correlation between the two.

The **sp500_ar** object also contains the residuals (ε). Using the **summary()** function, you can see that the object contains a time series of residuals.

Results of summary() function construct the equation of the model *AR (1)* as follows between S&P500 and Laq1 for 01/01/2017-01/01/2021:

$$y= - 0.02133*Laq1+3534$$

The fact that the indicators indicating the adequacy (*Multiple R-squared=0.00045*, *Adjusted R-squared =-0.01806)*, of the *AR (1)* model are very small and the p-values=0.87 are very large points to a violation of the autogressive process. *Cor(y, Laq1)=-0.021* autocorrelation points to a weak inverse relationship between the indicators.

Statistical significance is another way of saying that the p-value of a statistical test is small enough to reject the null hypothesis of the test. The most common threshold is $p < 0.05$; that is, when you would expect to find a test statistic as extreme as the one calculated by your test only 5% of the time.

Autocorrelation is an important part of time series analysis. It helps us understand how each observation in a time series is related to its recent past observations. When autocorrelation is high in a time series, it becomes easy to predict their future observations.

Results of summary() function *AR(1)* model between S&P500 and Laq1 for 01/01/2017-01/01/2019 before COVID-19 pandemic period:

```
Call:
lm(formula = Laq1 ~ y)

Residuals:
     Min       1Q   Median       3Q      Max 
-232.525  -33.314    8.919   56.797  154.181 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 502.45100  256.95214   1.955   0.0634 .  
y             0.81053    0.09853   8.226 3.71e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 85.46 on 22 degrees of freedom
Multiple R-squared:  0.7547,	Adjusted R-squared:  0.7435 
F-statistic: 67.67 on 1 and 22 DF,  p-value: 3.705e-08

> cor(y,Laq1)
[1] 0.8687077
```

Results of summary() function *AR(1)* model between SP500 and Laq1 for 01/01/2019-01/08/2021 after COVID-19 pandemic period:

```
Call:
lm(formula = Laq1 ~ y)

Residuals:
    Min      1Q  Median      3Q     Max 
-426.40  -64.23   27.53   90.74  300.34 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 111.05768  180.93003   0.614    0.544    
y             0.98184    0.05345  18.370   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 162.1 on 30 degrees of freedom
Multiple R-squared:  0.9184,	Adjusted R-squared:  0.9156 
F-statistic: 337.5 on 1 and 30 DF,  p-value: < 2.2e-16

> cor(y,Laq1)
[1] 0.9583096
```

The next economic indices to which we apply the autoregressive model are the NASDAQ indices. The NASDAQ stock market ranks second on the list of stock exchanges by market capitalization of shares traded, after the New York Stock Exchange. The exchange platform is owned by NASDAQ, Inc., which also owns the NASDAQ Nordic stock market network and several US stock and options exchanges.

The NASDAQ stock market has three different market levels: Capital Market, Global Market, Global Select Market.

Table 2.

Descriptive analysis of NASDAQ indices price

	Highest:	Lowest:	Difference:	Average:	Change %:
NASDAQ	15,288	5,398	9,890.10	8,852	183.5

As a result of the descriptive analysis, it was determined that the average price of the NASDAQ indices was 8,852 USD, minimum price was 5,398 USD in January 2017, and the maximum price was 15,288 USD in August 2021. Percentage of change was 183.5% for 01/01/2017-01/08/2021 time interval.

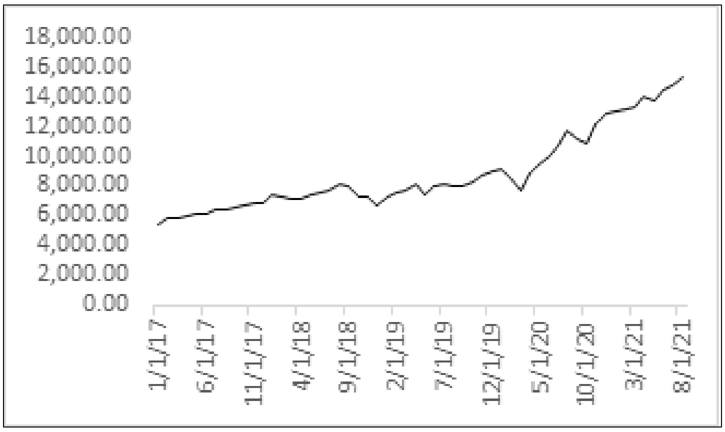

Fig. 3. Price of NASDAQ indices between 01/01/2017-01/08/2021

Results of summary() function *AR(1)* model between NASDAQ and Laq1 for 01/01/2017-01/08/2021:

```
Call:
lm(formula = y ~ Laq1)

Residuals:
     Min       1Q   Median       3Q      Max
-1015.87  -246.50   -56.41   224.27   974.17

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 326.12816  205.78091   1.585    0.119
Laq1          0.94373    0.02216  42.580   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 428.2 on 53 degrees of freedom
Multiple R-squared:  0.9716,    Adjusted R-squared: 0.9711
F-statistic: 1813 on 1 and 53 DF, p-value: < 2.2e-16
>cor(y~Laq1)
[1] 0.9856966
```

Compared to the S&P500 indices, the adequacy ratios of the *AR (1)* model built for the Nasdaq indices, correlation coefficient are high, with a *p*-value of less than 0.05. Another difference is that the percentage change in price is about twice as high.

4. Conclusion

In this study, we investigate the first-order autoregressive process $AR(1)$ with application to the S&P500 and NASDAQ indices for 01/01/2017-01/08/2021. The results of the study showed that the negative effects of the Covid-19 pandemic did not go unnoticed by stock indices. Adverse effects are expressed in the low adequacy of the *AR (1)* model to S&P500 indices established over the time interval. Evidence of this can be explained by the construction of a highly adequate model of the results of the study over two time intervals (before (01/01/2017-01/01/2019) and after (01/01/2019-01/08/2021) the Covid-19 pandemic). Percentage of change of NASDAQ indices was 183.5% for 01/01/2017-01/08/2021 time interval. This is about twice as much as the S&P500 indices. In the descriptive analysis, visualization, construction of the autoregressive model *AR (1)* of the stock indices R statistical programming language was used.

Acknowledgment

This work was supported by the Science Development Foundation under the President of the Republic of Azerbaijan – Grant No EIF-ETL-2020-2(36)-16/05/1-M-05.

References

[1] R. Aliyev, T. Khaniyev, On the moments of a semi-Markovian random walk with gaussian distribution of summands, Communications in statistics – Theory and methods. 43 No.1 (2013) pp.90-104.

[2] R. Aliyev, O. Ardic, T. Khaniyev, Asymptotic approach for a renewal-reward process with a general interference of chance, Communications in statistics – Theory and methods. 45 No.14 (2015) pp.4237-4248.

[3] R. Aliyev, V. Bayramov, On the asymptotic behaviour of the covariance function of the rewards of a multivariate renewal-reward process, Statistics and probability letters. 127 (2017) pp.138-149.

[4] R.T. Aliyev, F. Rahimov, A. Farhadova, On the first passage time of the parabolic boundary by the Markov random walk, Communications in statistics – Theory and methods. No.1 (2022) pp.1-10 https://doi.org/10.1080/03610926.2021.2024852

[5] V.F. Melfi, Nonlinear Markov renewal theory with statistical applications, The annals of probability. 20 No.2 (1992) pp.753-771.

[6] F.H. Rahimov, I.A. Ibadova, V.S. Khalilov, Limit theorems for a family of the first passage times of a parabola by the sums of the squares autoregression process of order one (AR (1)), Uzbek mathematical journal. No.2 (2019) pp.81-88.

[7] F.H. Rahimov, I.A. Ibadova, T.E. Hashimova, Limit theorem for first passage times in the random walk described by the generalization of the autoregression process, Uzbek mathematical journal. No.4 (2020) pp.102-110.

[8] A. Saadatmand, A.R. Nematollahi, S.M. Sadooghi-Alvandi, On the estimation of missing values in *AR(1)* model with exponential innovations, Communications in statistics – Theory and methods. 46 No.7 (2016) pp.3393-3400.

[9] A.N. Shiryaev, Essentials of stochastic finance: facts, models, theory, Word Scientific, (2003) 843 p.

[10] Y. Zhang, X. Yang, Limit theory for random coefficient first-order autoregressive process, Communications in statistics – Theory and methods. 39 No.11 (2010) pp.1922-1931.

[11] H.A. Jafarova, R.T. Aliyeva, T.N. Abdullayev, Analysis of the relationship between USDX and Gold Prices in R, Informatics and control problems. 39 No.2 (2019) pp.24-34.