adani

Institute of Digital Technology Management

PGDM in Big Data Analytics

Term II (Batch 2023-25)

Course name = Time series and Forecasting Techniques & Analysis

Auto-Regressive 2 model

Submitted to: Dr. Dhyani Mehta

Submitted by:

Name	Enrollment no.
Aakanksha Bapat	20231001
Aakanxa Yadav	20231002
Shraddha Gosavi	20231051
Sanidhya Jain	20231049
Dheeraj Saini	20231016
Deepanshu	20231018

Table of content:

INTRODUCTION:

Because of the complexity of the stock market, stock price prediction is thought to be one of the hardest financial forecasting tasks to do. A lot of people are eager to learn any forecasting method that lowers investment risk and guarantees easy earnings from the stock market. This is still the main motivation for academics to create new prediction models.

This report addresses the prediction of stock prices with the Autoregressive model of order 2, a kind of time series model widely applied in econometrics and statistics. The AR (2) model is a linear regression of a time series' current value based on its two most recent historical values, and it is used in time series analysis and stock price prediction.

It can mathematically be expressed as follows:

$$Yt = c + \phi 1\ Yt - 1 + \phi 2\ Yt - 2 + \epsilon t$$

Here:

Yt = The time series value at time t is denoted by Yt

$c =$ constant/ intercept

$\phi 1$ and $\phi 2$ = this represents lag (1) and lag (2)

$\epsilon t =$ error

Day-to-day closing prices have been studied for fifty financial companies for the last three years. On assuming that the present returns are based on the previous returns for stock we can say that today' return is a function of the returns of the previous day (Lag 1) and the returns of the day before yesterday (Lag 2). Mathematically we can define this relationship as:

$$\boldsymbol{Return} = \boldsymbol{f}\ (\boldsymbol{return}(-\mathbf{1}), \boldsymbol{return}(-\mathbf{2}))$$

Defining Variables:

1. Dependant variable: return

2. Independent variable: return (-1), return (-2)

Variables	measurements	code	Description
Return	(previous day closing price – present-day closing price)/ Previous day closing price X 100)	RETURN	This variable represents the current return in the dataset. It represents the rate of return or % change for a specific financial instrument, recording changes in price or performance over a predetermined time frame.
Return (-1)	Lag 1	RETURN (-1)	It represents the return value from the previous one period, and it symbolizes the return value from the prior period, enabling the comparison between the recent return and the recent past performance.
Return (-2)	Lag 2	RETURN (-2)	It represents the return value from the previous two periods, giving information about the performance from further back in time and how it might have affected the return today.

DESCRIPTIVE STATISTICS:

Mean	0.000520
Median	0.000261
Maximum	0.076004
Minimum	-0.051478
Standard deviation	0.011705
Skewness	0.064726
Kurtosis	7.122490
Jarque-Bera	527.3622
Probability	0.000000

INTERPRETATIONS:

1. The standard deviation of return is more than the mean of the return, so we can say that there is high volatility which means that there will be more fluctuations in the prices.

2. As we can see the mean of return is greater than the median, so there is a positive skewness.

3. The kurtosis of the return is more than 3, which means the curve of the return will be leptokurtic, which indicates that there are more extreme values in the data as compared to the normal distribution.

4. The positive skewness indicates that the distribution is skewed to the right, and the high kurtosis suggests heavy tails and a peaky distribution.

5. The low p-value in the Jarque-Bera test indicates that the data is not normally distributed.

6. The wide range between the maximum and minimum values, along with the positive skewness, suggests that the data may have a long left tail with some extreme values pulling the mean to the left.

7. The small standard deviation suggests that the values are relatively close to the mean on average.

COVARIANCE ANALYSIS

	RETURN	RETURN (-1)	RETURN (-2)
RETURN	1	0.081657	-0.061771
RETURN (-1)	0.081657	1	0.081790
RETURN (-2)	-0.061771	0.081790	1

INTERPRETATIONS:

1. In the above table, we came to know that the correlation between RETURN and it is the first difference (RETURN (-1)) is 0.081657 which signifies that there is a very weak positive correlation.

2. The correlation between RETURN and its second difference (RETURN (-2)) is -0.061771 which shows that there is a very weak negative correlation.

3. The correlation between RETURN (-1) and RETURN (-2) is 0.003499 which shows there is a very weak positive correlation.

AUTO-REGRESSION FUNCTION

VARIABLE	CO-EFFICIENT	PROBABILITY
C	0.000507	0.2383
RETURN (-1)	0.087293	0.0176
RETURN (-2)	-0.068957	0.0608

R^2	0.011385
ADJUSTED R^2	0.008709
F STAT	4.255155
PROBABILITY (F STAT)	0.014540
DW STAT	2.007149

INTERPRETATIONS:

1. The P value of lag (1) of returns is more than 0.05, which means that it is insignificant.

2. The P value of lag (2) of returns is more than 0.05, which means that it is insignificant.

3. The value of R-squared is 0.011385, which means that it explains only 1.13% of our model.

4. The P value of the F-statistic is 0.014, which is less than 0.05, which makes it significant. F-stats are explained by the Explained sum of squares and Residual sum of squares. This means that the combined effect of the lagged returns in predicting current returns is significant.

NORMALITY TEST FOR VARIABLES

VARIABLES	JARQUE-BERA	CORRELOGRAM	ADF TEST
RETURN	0.0000	0.110	0.00
RESULT	Not Normal	Normal	Normal

1. JARQUE-BERA

H0: *DATA IS NORMAL*

H1: *DATA IS NOT NORMAL*

INTERPRETATIONS:

1. The P. value of Jarque-Bera is 0.000 which is less than 0.05, meaning that we will accept H1, which says that the data is not normal.

2. CORRELOGRAM:

H0: *THERE IS NO AUTO-CORRELATION/ PARTIAL AUTO-CORRELATION*

H1: *THERE IS AUTO-CORRELATION/ PARTIAL AUTO-CORRELATION*

INTERPRETATIONS:

1. As the P value is more than 0.05, it means that we will accept H0, which means there is no autocorrelation or partial autocorrelation.

2. It indicates that the data is normal.

3. ADF UNIT ROOT TEST:

H0: *THERE IS UNIT ROOT*

H1: *THERE IS NO UNIT ROOT*

INTERPRETATIONS:

1. As we can see the P. value of the ADF test is 0, which is less than 0.05, meaning that we will accept H1. It says that there is no unit root in the data. It means that the data is normal.

NORMALITY TEST FOR VARIABLES:

Sr. no	Tests	Probability value	result
1	Jarque-Bera	0.00	Not Normal
2	Correlogram	0.878	Normal
3	ADF Test	0.00	Not Normal
4	BGLM Test	0.0535	Normal
5	DW Test	2.007149	Normal

1. JARQUE-BARA

H0: *ERROR IS NORMAL*

H1: *ERROR IS NOT NORMAL*

INTERPRETATIONS:

1. The P. value of Jarque-Bera is 0.000 which is less than 0.05, meaning that we will accept H1, which says that the error is not normal.

2. CORRELOGRAM

H0: *THERE IS NO AUTO-CORRELATION/ PARTIAL AUTO-CORRELATION*

H1: *THERE IS AUTO-CORRELATION/ PARTIAL AUTO-CORRELATION*

INTERPRETATIONS:

1. As most of the P values are more than 0.05, it means that we will accept H0, which means there is no autocorrelation or partial autocorrelation.
2. It indicates that the error is normal.

3. ADF UNIT ROOT TEST

H0: *THERE IS UNIT ROOT*

H1: *THERE IS NO UNIT ROOT*

INTERPRETATIONS:

1. As we can see the P value of the ADF test is 0, which is less than 0.05, meaning that we will accept H1. It says that there is no unit root in the error. It means that the error is normal.

4. BGLM TEST

H0: *There is no autocorrelation in the residuals as lag =0*

H1: *There is autocorrelation in the residuals because of lag =0*

INTERPRETATIONS:

1. The P value of F-stats which is more than 0.05, which means we will accept H0 which says there is no autocorrelation. It indicates that the error is normally distributed.

5. DW TEST

DL = 1.75

DU = 1.79

INTERPRETATIONS:

1. As we can see the value of the DW test is 2.007149 which is slightly greater than 2 which makes it fall between 2 and 4-DU, which means there is no autocorrelation.

2. It means that the error is normal.

6. HETROSCADASTICITY TESTING

Sr. no	Tests	Probability value	Chi-square
1	White test	0.0002	0.0003
2	Breusch Pagan Godfrey	0.0223	0.0224
3	Glejser	0.0092	0.0020

a.) WHITE TEST

H0: *THERE IS NO HETEROSCEDASTICITY*

H1: *THERE IS HETEROSCEDASTICITY*

INTERPRETATION:

1. Here we can see that the P-chi-value of the Observed R-squared is less than 0.05, then we will accept H1, which says that there is heteroscedasticity.

b) BREUSH-PAGAN-GODGREY

H0: *THERE IS NO HETEROSCEDASTICITY*

H1: *THERE IS HETEROSCEDASTICITY*

INTERPRETATION:

1. Here we can see that the P-chi-value of the Observed R-squared is less than 0.05, then we will accept H1, which says that there is heteroscedasticity.

c) GLEJSER TEST

H0: *THERE IS NO HETEROSCEDASTICITY*

H1: *THERE IS HETEROSCEDASTICITY*

INTERPRETATION:

1. The P value of F stat is less than 0.05, then we will accept H0, which says that there is no heteroscedasticity.

7. MULTICOLLINEARITY

	COEFFICIENT VARIANCE	**CENTERED VIF**
C	**1.84E**	**NA**
RETURN (-1)	**0.001347**	**1.006735**
RETURN (-2)	**0.001349**	**1.006735**

INTERPRETATION:

1. The value of Centered VIF of the lag value of the returns is less than 5, which means there is no multi-collinearity.

CONCLUSION:

The AR (2) model for predicting returns of NIFTY FINANCIAL SERVICE stock does not appear to be statistically significant, with low explanatory power (low R2 and adjusted R2) and non-significant coefficients. The model does not provide a meaningful improvement in explaining the variability in returns compared to a simple mean. The findings suggest that the lagged returns in the AR (2) model are not effective predictors for the current returns of NIFTY FINANCIAL SERVICE stock based on the provided features. It might be required to explore more features, improve the model even more, or consider different modelling strategies. Furthermore, it might be crucial to look for outliers or significant observations to improve the model.