A Prediction Approach for Stock Market Volatility Based on Time Series Data

Higher School of Economics
Faculty of Computer science

The forecasting Process

Problem identification: Who needs these predictions and how they will be used.

Data Gathering: Nifty and Sensex index prices - two main indicators of the Indian economy.

Data Analysis: Identifying important data features and separating redundant information.

Model Comparison: Different models are analysed to select the best for forecasting.

Implementation: Cyclical evaluation and tuning, the model is put into practice.

Problem

This research tries to analyze the time series data of the Indian stock market and build a statistical model that could efficiently predict the future stocks.

Actuality

Forecasting applications in multiple different areas of the world.

Financial forecasting is an essential tool that predicts future business conditions

India has become an attractive market for investors

Growing attention towards the research of time series data

Forecasting Models

AR

- takes previous values of the time series

MA

- uses the errors of previous values

ARMA

- AR + MA

ARIMA

- ARMA + differencing

ARIMA

The ARIMA(p, d, q) model is an AutoRegressive Integrated Moving Average model, where $\hat{y}_t$ is:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}$$

Arima combines three fundamental parts:

AR, Differencing, MA

Methodology

01 Stationarity

Test the stationarity of data -> Series needs to be differenced

02 Plot ACF and PCF

These graphs helps to define parameters for “AR and MA” terms

03 Method

Decomposition of time series to study seasonality and check stationarity by DF test

04 Model selection

Forecast the results using auto.arima() function to find model parameters

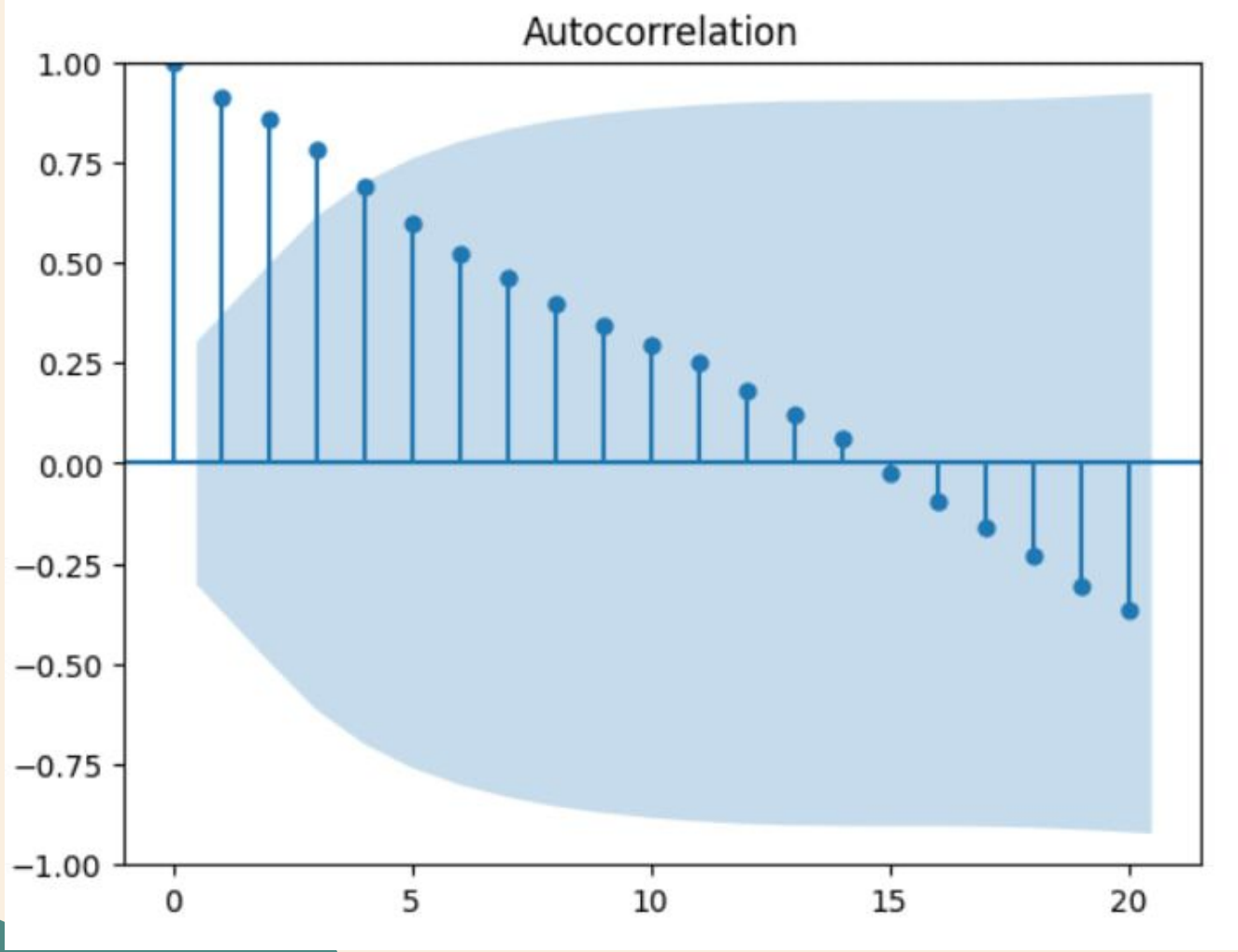
05 Results

Check the residuals using the L-Jung-Box test to ensure their randomness and therefore the effectiveness of the model

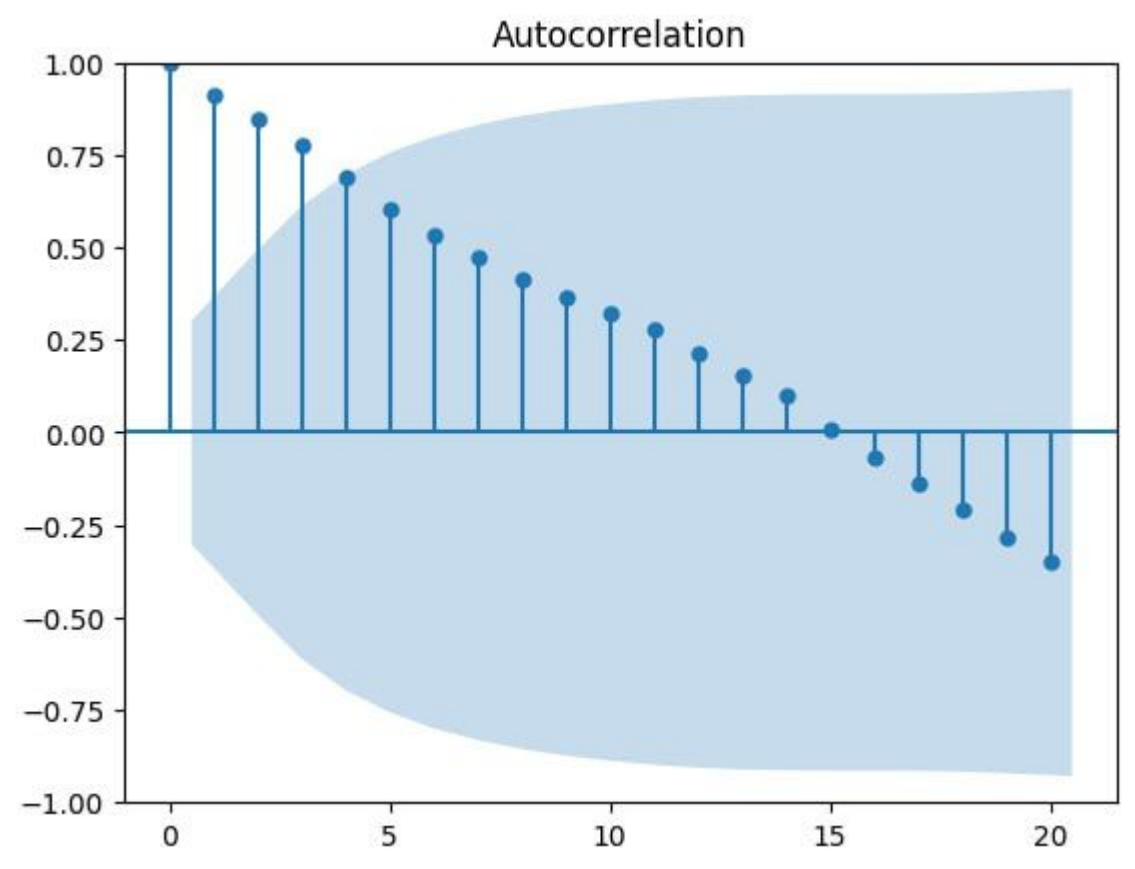
Stationarity, ACF and PCF

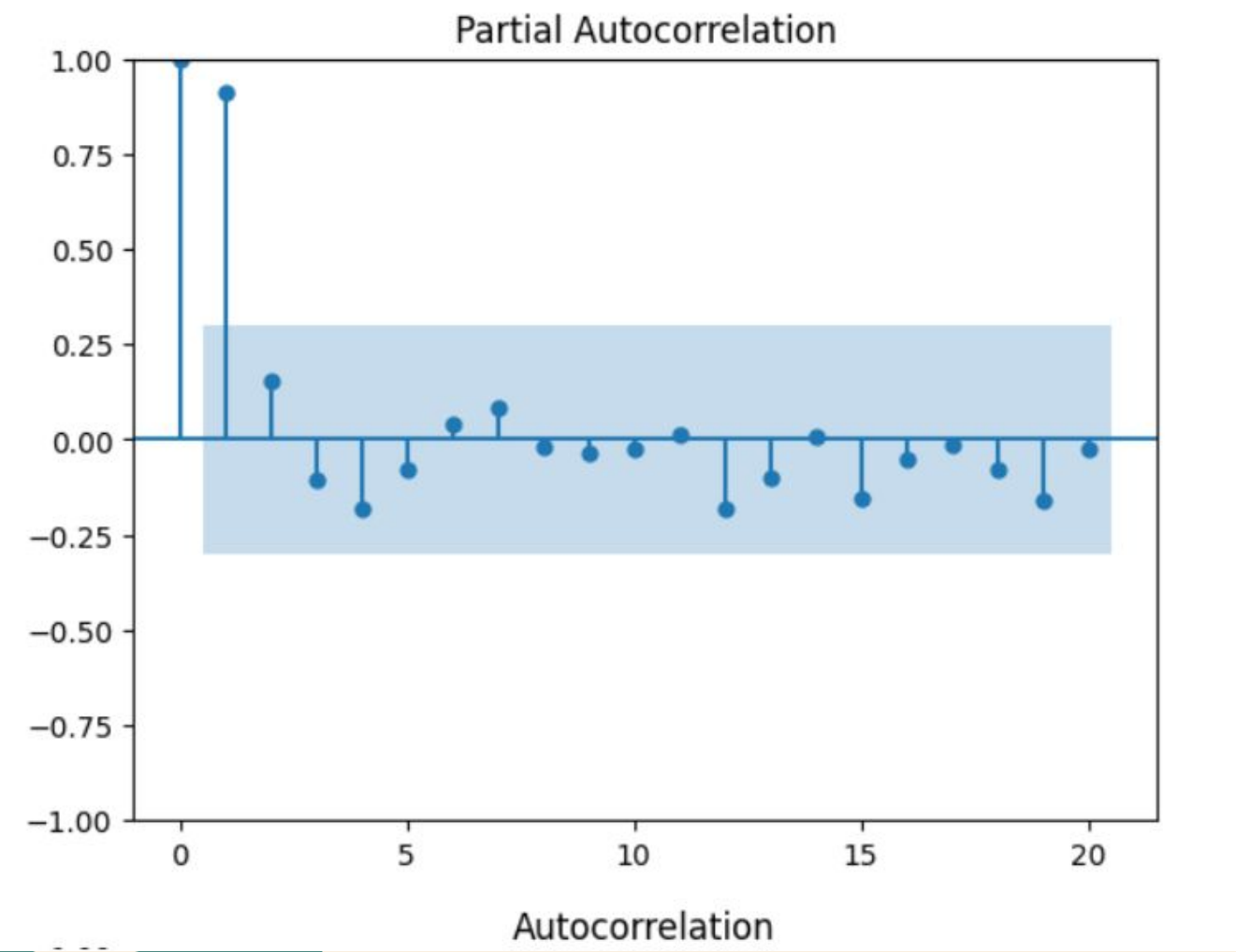
- Time series needs to be lacking trend and seasonality, in order to be stationary.
- Augmented Dickey-Fuller (ADF test) was used to check the stationarity of the model.
- Initially data was unstationary, so the difference between consecutive observations was used.
- ACF is used to find the order of p, while PCF is used to find the order of q. ACF shrinks in of “AR” model, while PCF shrinks in case of “MA” model, this is why both models are needed

Sensex


Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20

Nifty


Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20


Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
5
10
15
20
Autocorrelation

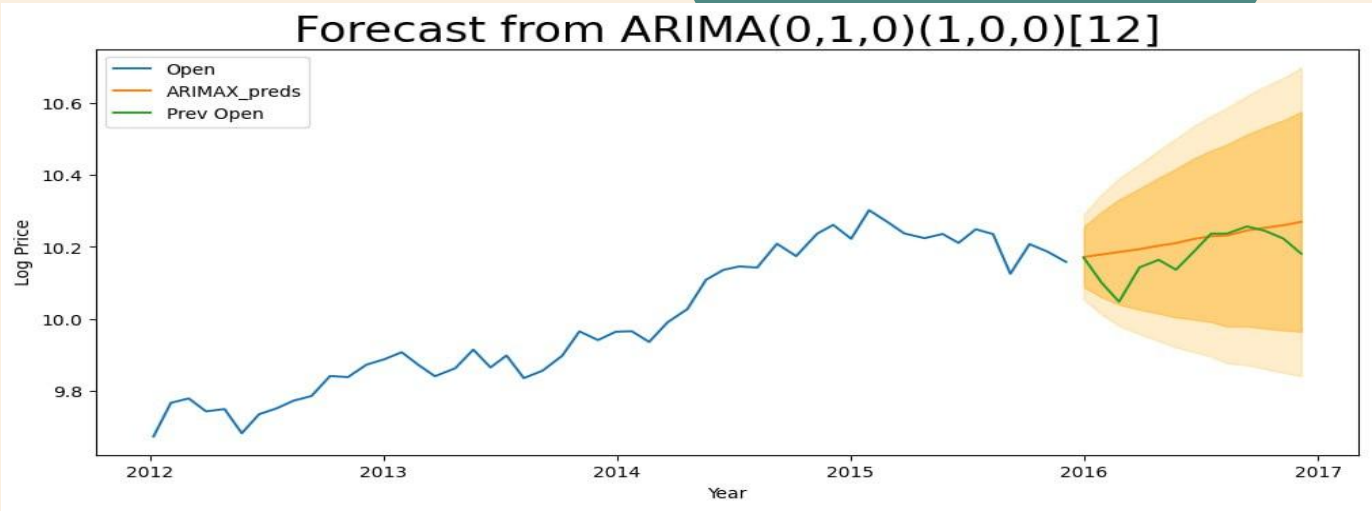
Model selection and results

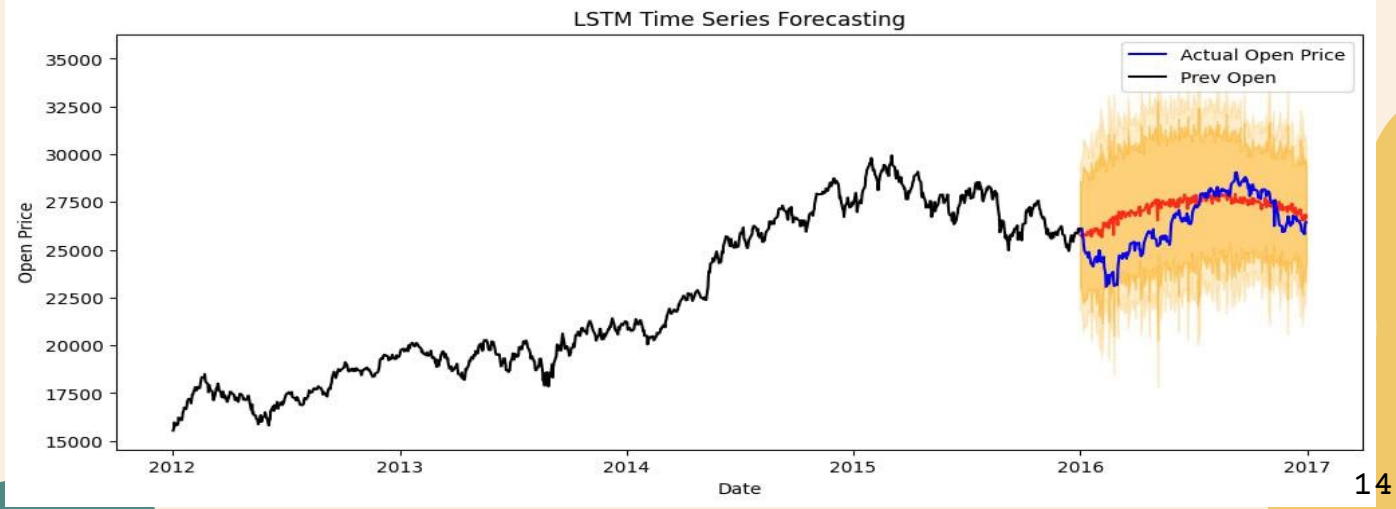
- The auto.arima() function uses a ‘trace’ that justifies why the parameters (p,d,q) chosen are best suited for the ‘‘ARIMA(p,d,q) model’’.
- Predicted results and actual data will always differentiate due to irregular component. To verify that residuals are in fact random and therefore there is no significant correlation between variables and the model is valid the L-Jung-Box test is used.

LSTM

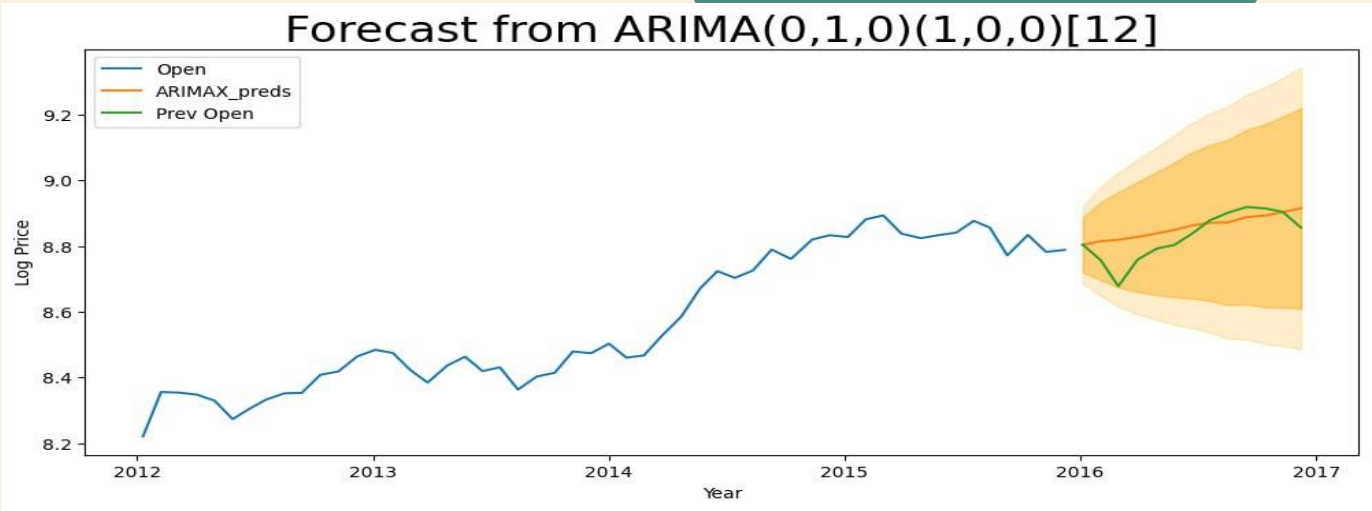
- LSTM is a RNN model which is able to capture long-term dependencies and complex pattern in data.
- It has a “Memory” and “Forget Gate” mechanism which allows to retain crucial information while discarding irrelevant noise.

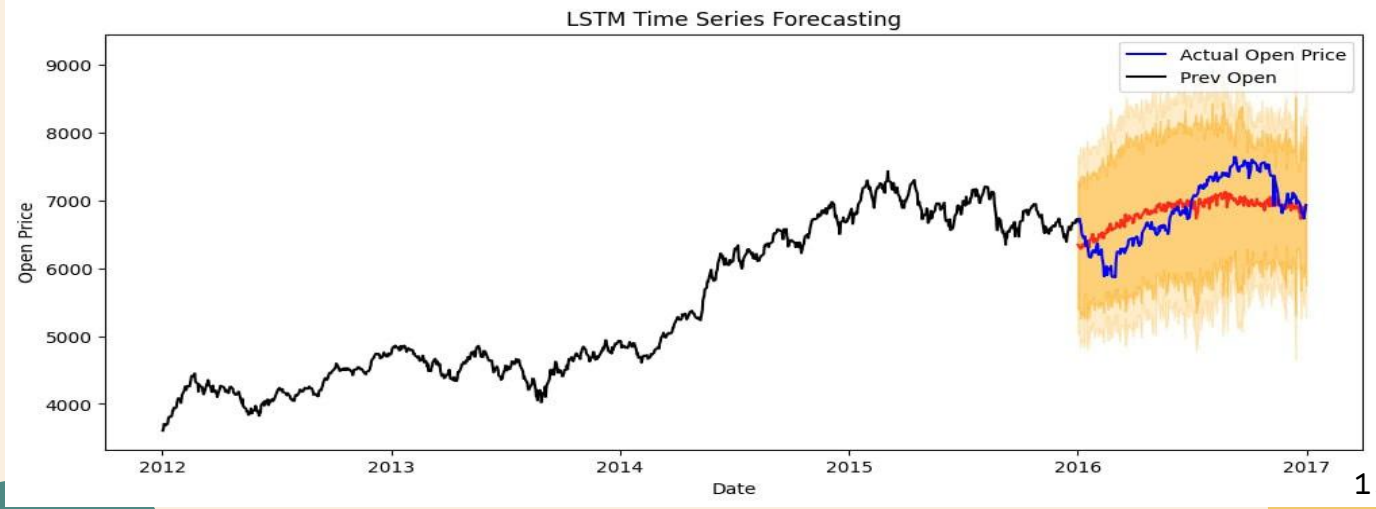
Sensex Forecasting





Nifty Forecasting





Results

	ARIMA MAE	LSTM MAE	ARIMA MPE	LSTM MPE
NIFTY	278	514	4.2%	7.7%
SENSEX	1166	1885	4.6%	7.4%

Observation: ARIMA performed better than the LSTM by both metrics

Conclusion: ARIMA is more appropriate than LSTM for making forecasts on the Indian financial market

Thank you for your attention!

Methodology

01 Stationarity

Test the stationarity of data -> Differentiate series if needed

02 Plot ACF and PCF

These graphs helps to define parameters for “AR and MA” terms

03 Method

Decomposition of time series to study seasonality and check stationarity by DF test

04 Model selection

Forecast results using auto.arima() function to find model parameters

05 Results

Check the residuals using L-Jung-Box to ensure their randomness and therefore the

06 Final analysis

You can describe the topic of the section here