A Time-Series Analysis on the S&P 500 Stock Index

Chunmei Gao

500887771

December 18, 2018

Drivers and Objectives

- **Drivers**
  - Work Needs
  - Research & Learn

- **Objectives**
  - Time Series Analysis Techniques
  - Practical Procedures
  - Applications

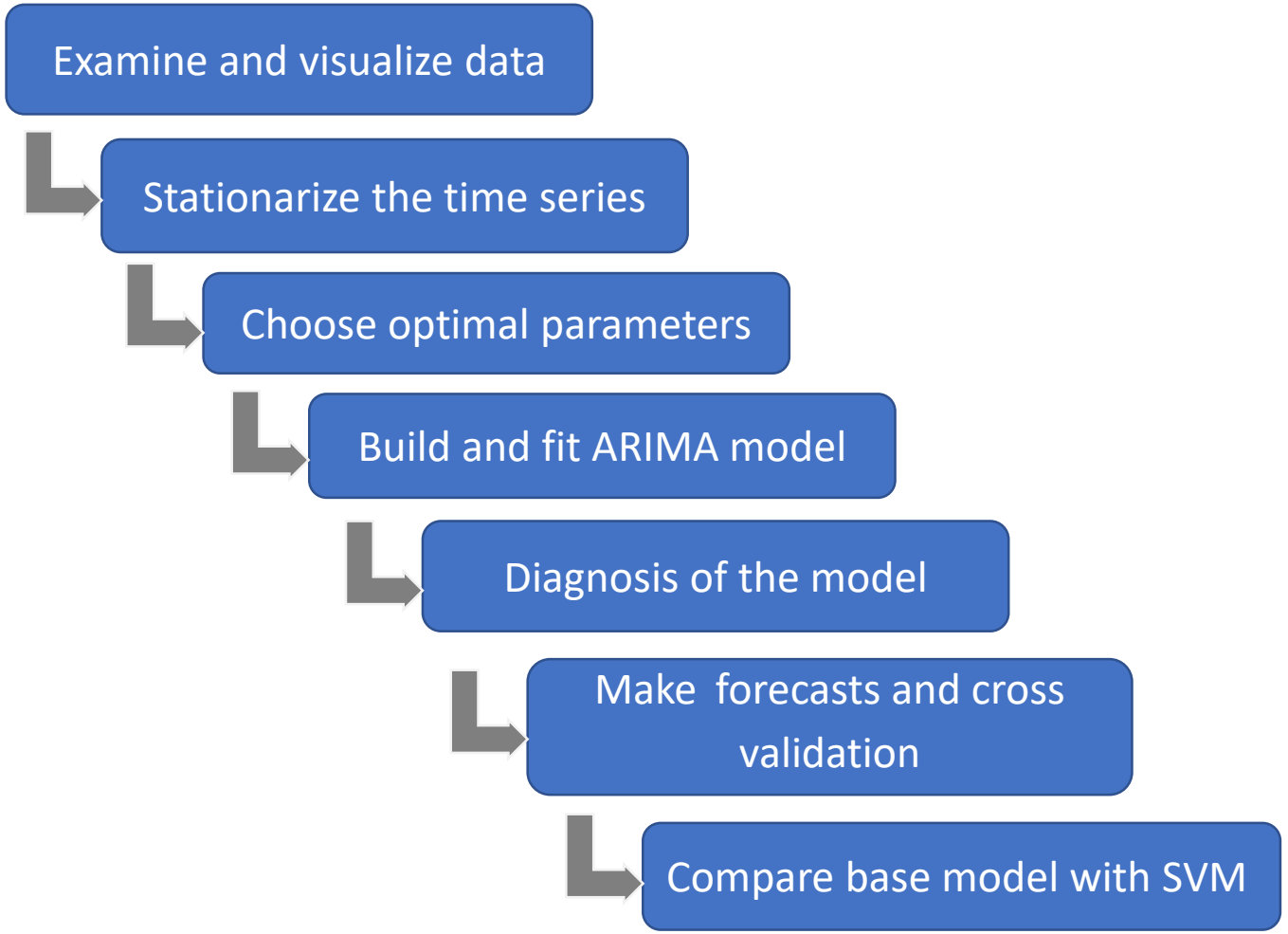
Table of Contents

Time Series Analysis - ARIMA Model and Approach

- **Time Series**
  - A sequence of numbers
  - Constant Intervals
  - Univariate & Multivariate
  - Width & Depth

- **ARIMA** (p,d,q)
  - **AR** - Auto Regressive (p)
  - **MA** – Moving Average (q)
  - **Integrated** component (d)

- **Approach**

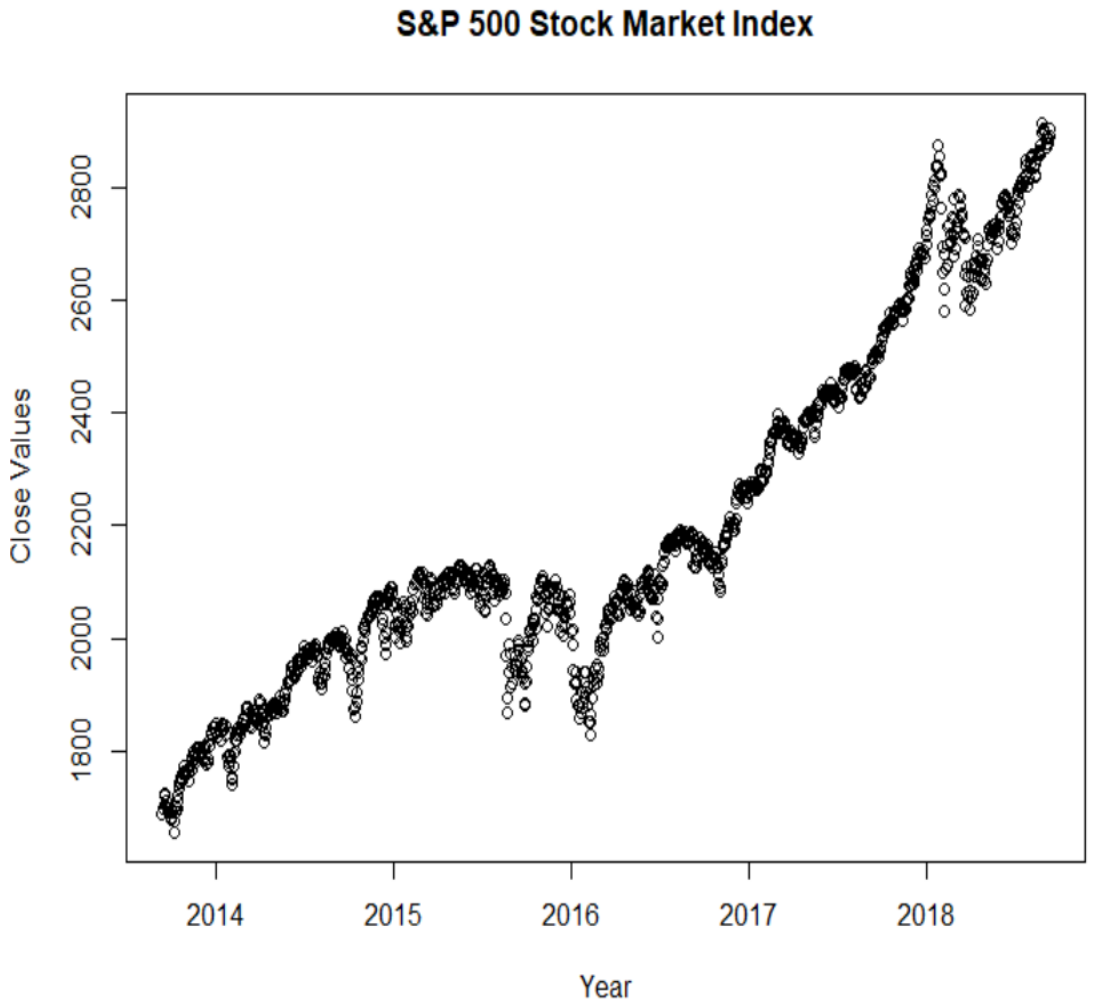


Dataset – S&P 500 Daily Stock Index

- **Daily S&P 500 Stock Index Dataset**
  - 5 years historical data:
    September 2013 – September 2018
  - 30 days forecasts:
    October 1st – November 9th, 2018

Date	Open	High	Low	Close*	Adj Close**	Volume
Sep 13, 2018	2,896.85	2,906.76	2,896.39	2,904.18	2,904.18	3,254,930,000
Sep 12, 2018	2,888.29	2,894.65	2,879.20	2,888.92	2,888.92	3,264,930,000
Sep 11, 2018	2,871.57	2,892.52	2,866.78	2,887.89	2,887.89	2,899,660,000
Sep 10, 2018	2,881.39	2,886.93	2,875.94	2,877.13	2,877.13	2,731,400,000

- **Characteristics of the Data**
  - Trend: uptrend
  - Seasonality: non-seasonal
  - Stationarity: not stationary



Data Preprocessing and Preparation

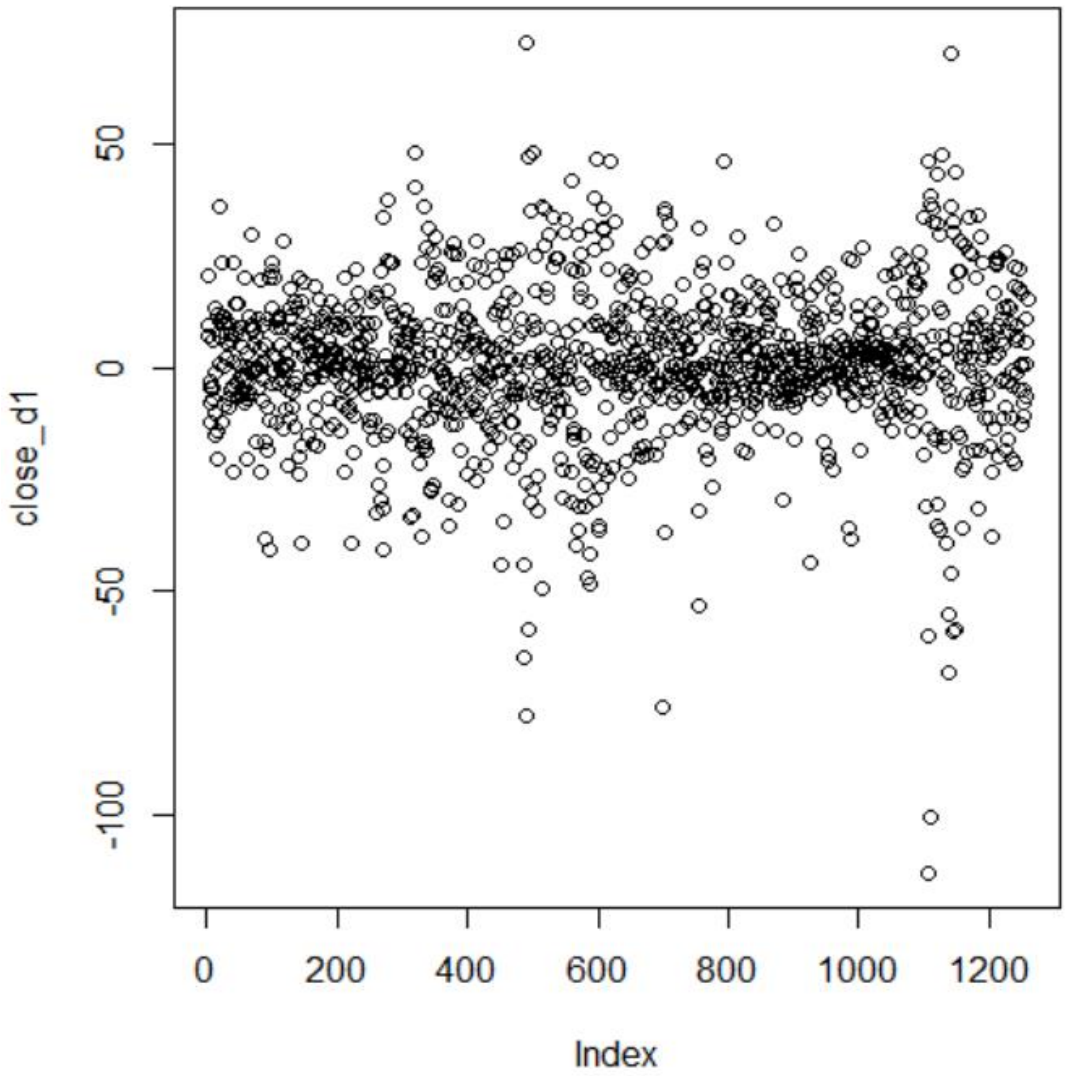
- **Convert data to time series object:** ts()
- **Clean the data:** tsclean()
- **Stationarize the data**

  - ADF: test for stationarity

```
adf.test(ts_close, alternative = "stationary")
Augmented Dickey-Fuller Test data:  ts_closeDickey
-Fuller = -1.6706, Lag order = 10, p-value = 0.717
8 alternative hypothesis: stationary
```

  - Differencing

```
close_d1 <- diff(ts_close, differences = 1)
plot(close_d1)
```



ARIMA Model Parameters Selection and Fitting

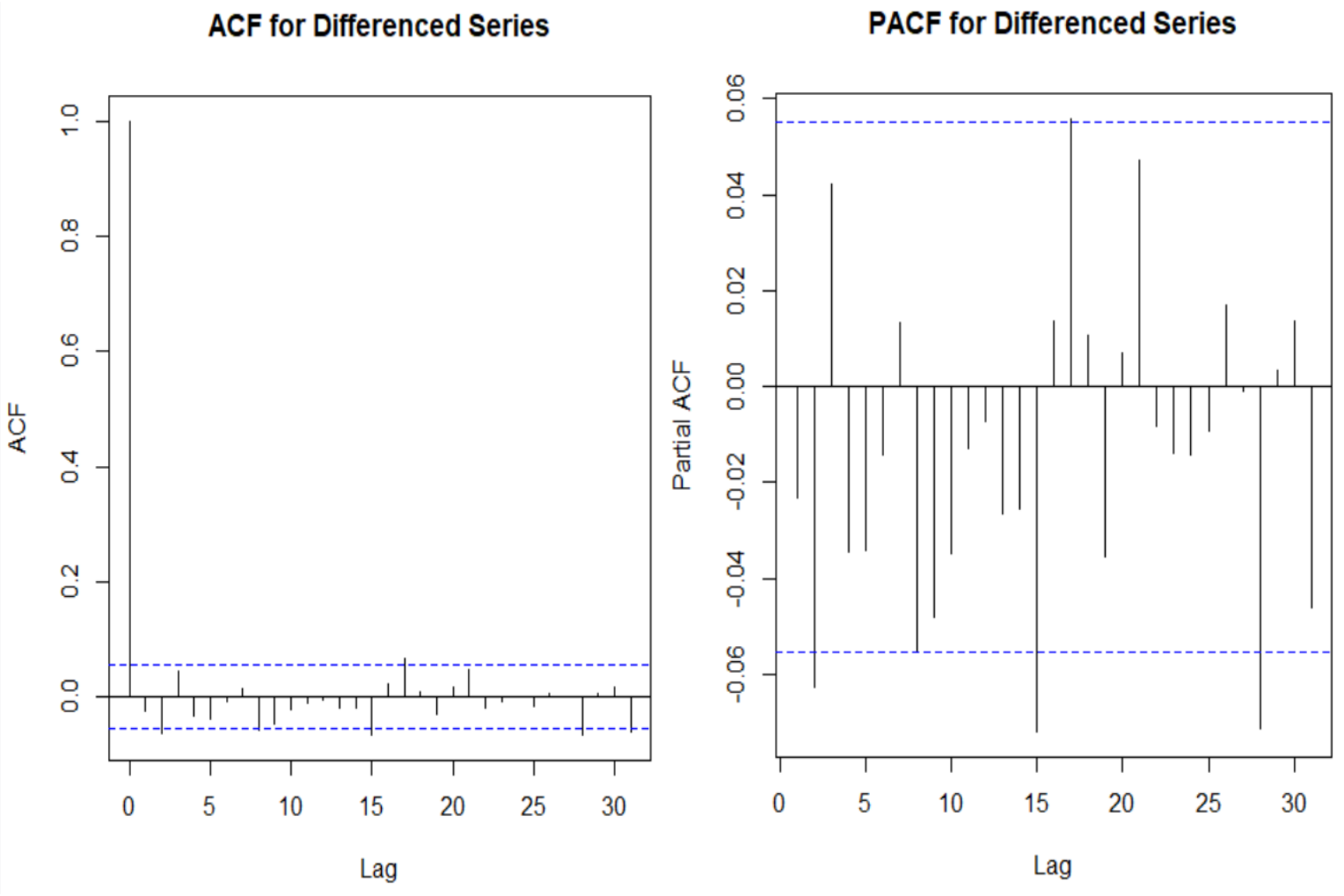
- **Determine Parameters**
  - ACF: MA(**q**)
  - PACF: AR(**p**)
  - Times of Differencing: **d**

ARIMA(0, 1, 0)

- **Identify Optimal Parameters**
  - AIC: the lowest

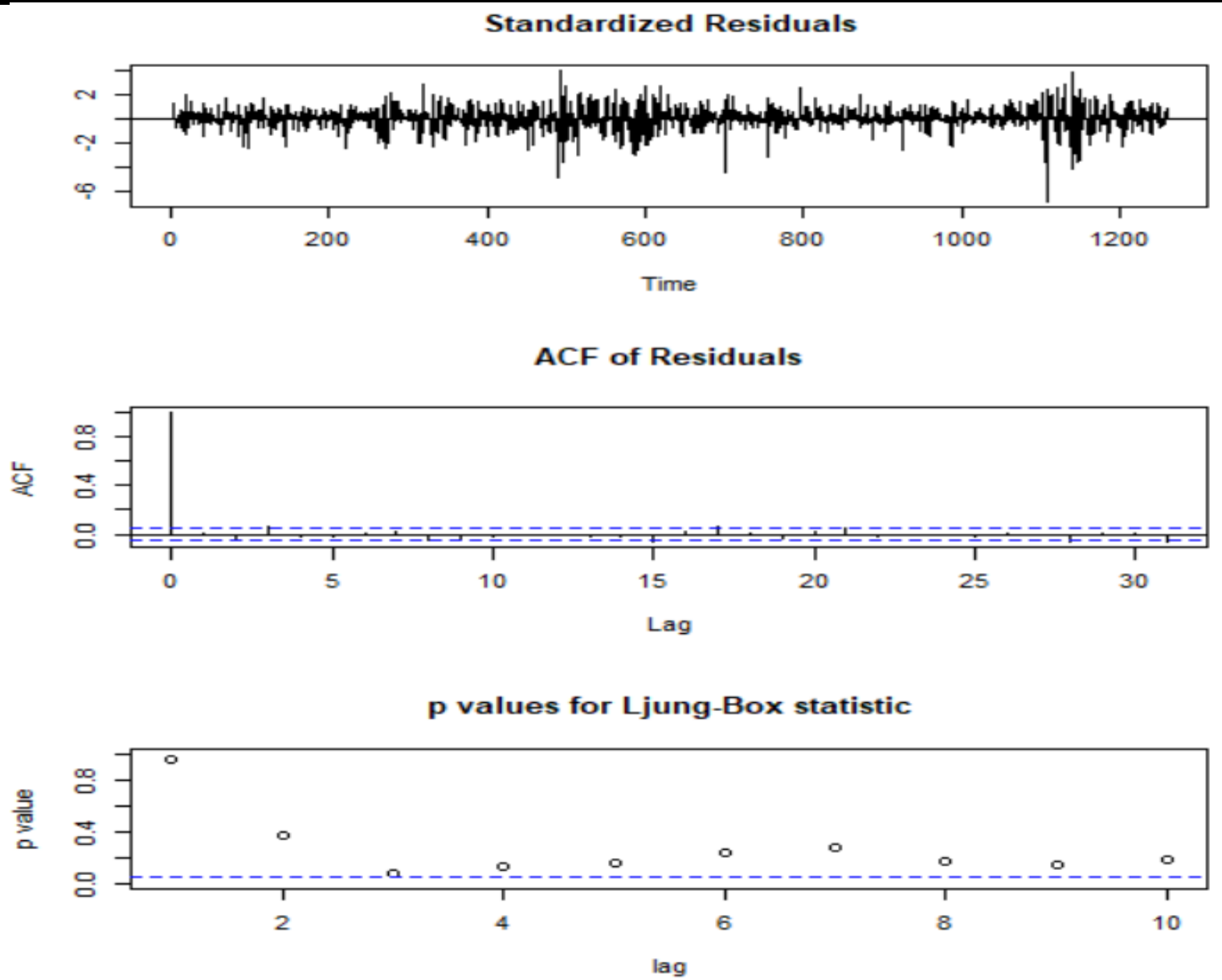
arima(close_d1, order=c(0,0,0)) #ARIMA(0,1,0) aic = 10685.41
arima(close_d1, order=c(1,0,0)) #ARIMA(1,1,0) aic = 10686.73
arima(close_d1, order=c(1,0,1)) #ARIMA(1,1,1) aic = 10679.06, lowest
arima(close_d1, order=c(0,0,1)) #ARIMA(0,1,1) aic = 10686.63
arima(close_d1, order=c(0,0,2)) #ARIMA(0,1,2) aic = 10683.67
arima(close_d1, order=c(2,0,0)) #ARIMA(2,1,0) aic = 10683.75

ARIMA(1, 1, 1)

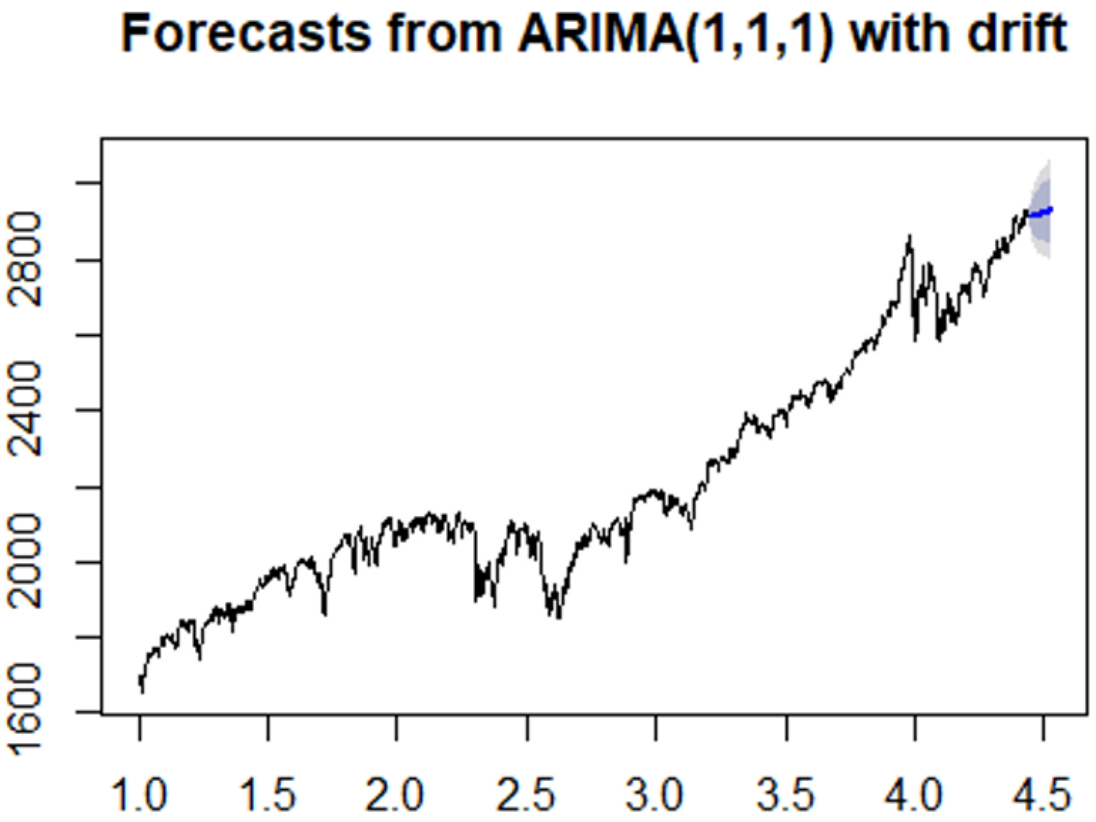


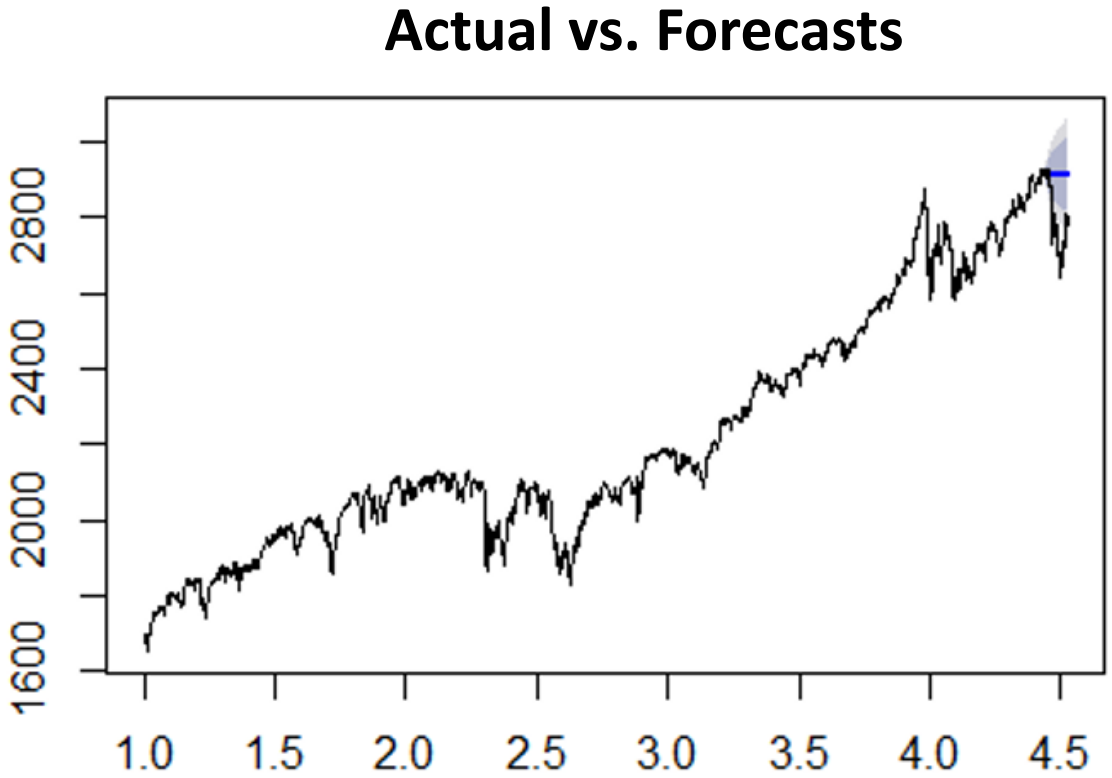
Diagnosis of the model

- White Noise
- Residuals: *tsdiag()*



Forecasts and Cross Validation





Techniques Comparison – SVM vs. ARIMA

ARIMA Model Errors

```
accuracy(fcast, sp500_Oct$Close)

##                      ME      RMSE      MAE        MPE      MAPE
## Training set  0.02250337  16.58955 11.55024 -0.006219626 0.5332682
## Test set   -139.01892711 162.13242 141.11746 -5.089648731 5.1614038
##                    MASE       ACF1
## Training set  0.9897036 -0.01070866
## Test set     12.0919119         NA
```

SVM Model Errors

```
accuracy(svm_fcast, sp500_Oct$Close)

##                    ME    RMSE      MAE      MPE     MAPE
## Test set 1060.035 1065.328 1060.035 38.04468 38.04468
```

Conclusion

- Time series analysis has clear characteristics.
- ARIMA model is well studied for time series analysis with well documented procedures to follow and tools to use.
- ARIMA has limitations.
- This study and project implementation enabled me to perform actual time series data analysis at work place.

Q & A