Prediction of Tesla's Stock Price Using R

VE406

Cao Kaiyan Fu Tianchi Zhang Junjie

2019.12.6

Stock prediction is important

- Many big companies spend millions of dollars to hire expert analysts just for stock prediction!

- It helps investors understand the stock market trend,
- make better decisions and EARN MONEY!

Tesla's Stock Price at Close

ARIMA Model

- ARIMA model (A.K.A. Box-Jenkins model) is *one of the most used and popular* **model** in forecasting financial *time series* over short time periods.

- It is in the form ARIMA(*p,d,q*) where *p* is the order of autoregressive terms, *d* is the order of differences needed for stationarity, and *q* is the order of moving average terms.

Tesla's Close Price from 2016.11.27-2019.11.25

- We will use the data from 2019.11.26 to 2019.12.3 (altogether 5 days) as our test set

Training Set

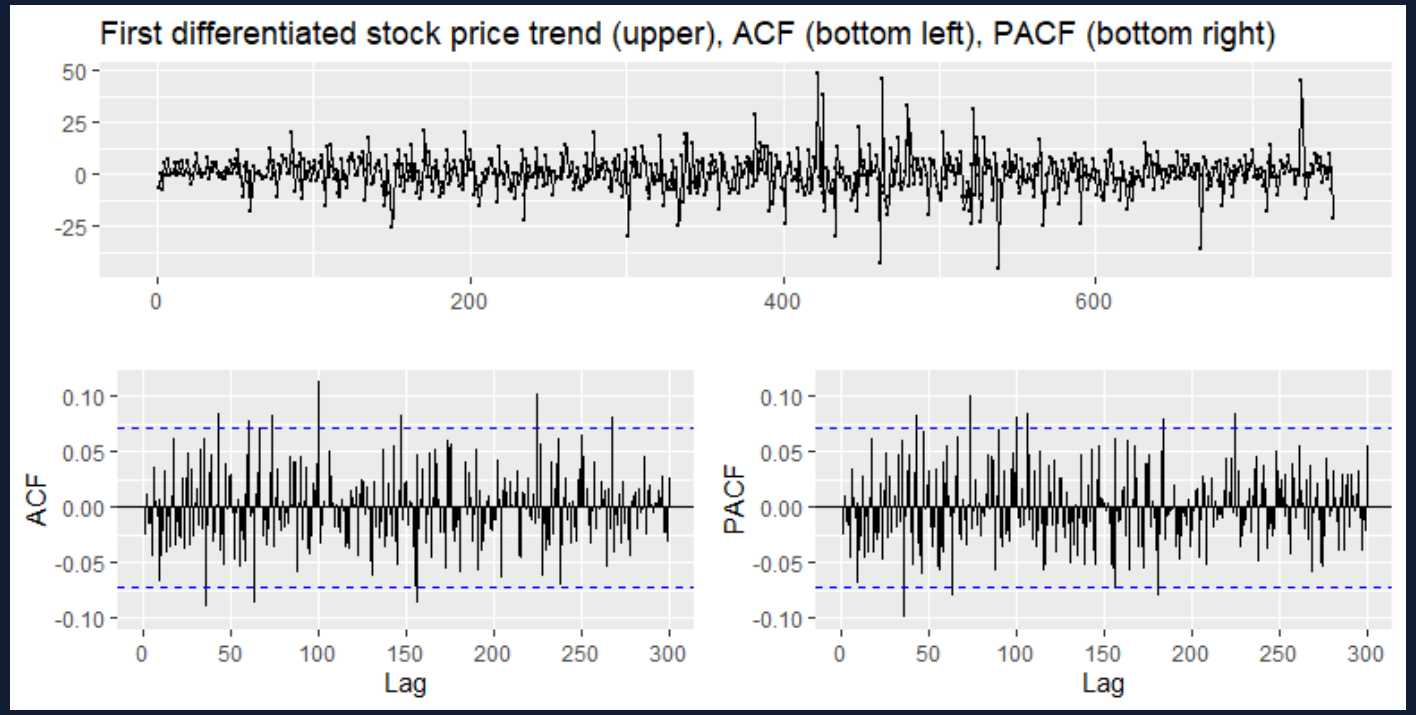
- But this time series seems obviously NOT stationary!

```
        Augmented Dickey-Fuller Test

data:  close.ts
Dickey-Fuller = -3.1644, Lag order = 9, p-value = 0.09406
alternative hypothesis: stationary
```

Differentiate the data

- We take the first order difference and it looks much better.

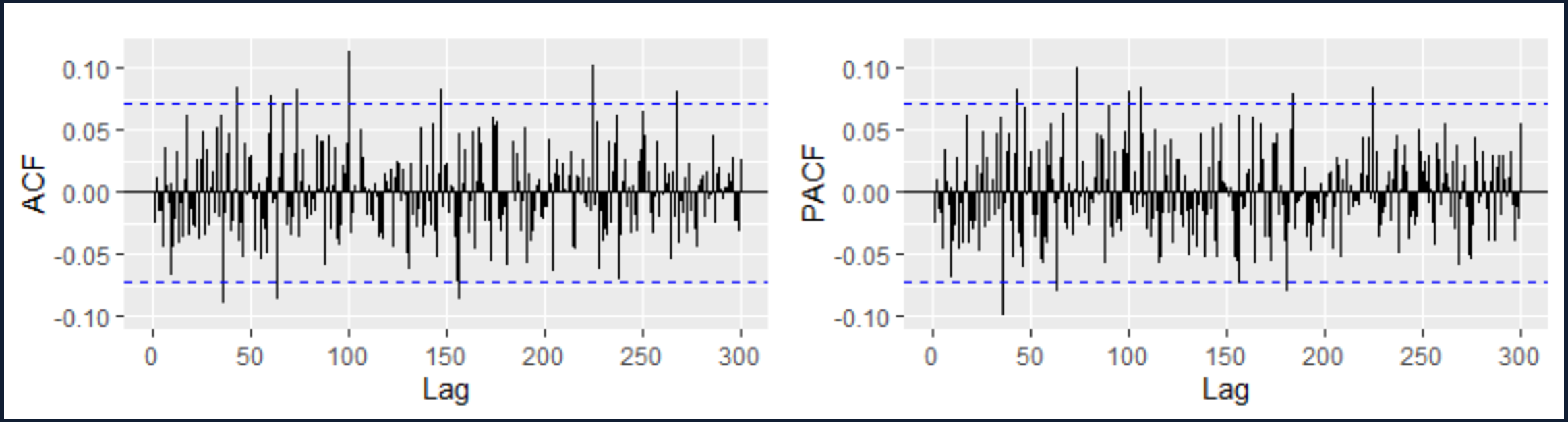


- First order difference is enough to make our data stationary.
  - *d=1*

```
        Augmented Dickey-Fuller Test

data:  diff_close.ts
Dickey-Fuller = -9.2007, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary
```

Parameter *p and q*



We can tell *q=0*

We can tell *p=0*

Final ARIMA model

- Now that we have the model ARIMA (0,1,0), let us check using auto.arima

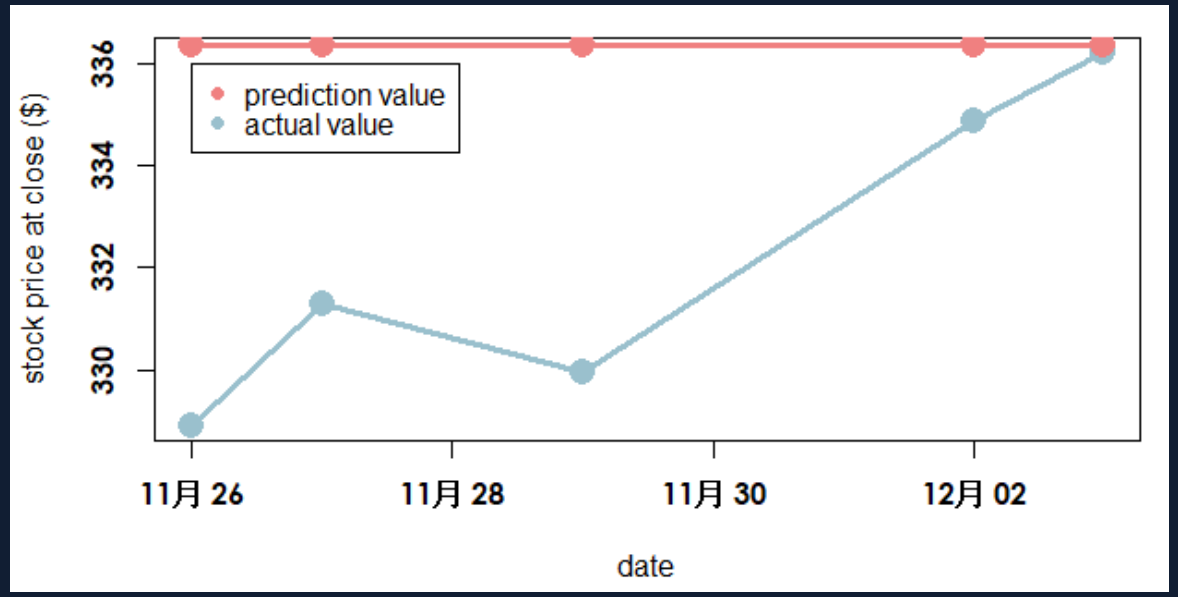
```
Series: close.ts
ARIMA(0,1,0)

sigma^2 estimated as 81.63:  log likelihood=-2722.25
AIC=5446.51   AICc=5446.51   BIC=5451.13
```

- Now we can proceed to plot the data!

Final ARIMA model

- Bad result
  - It is just a straight line!
  - It is not surprising…

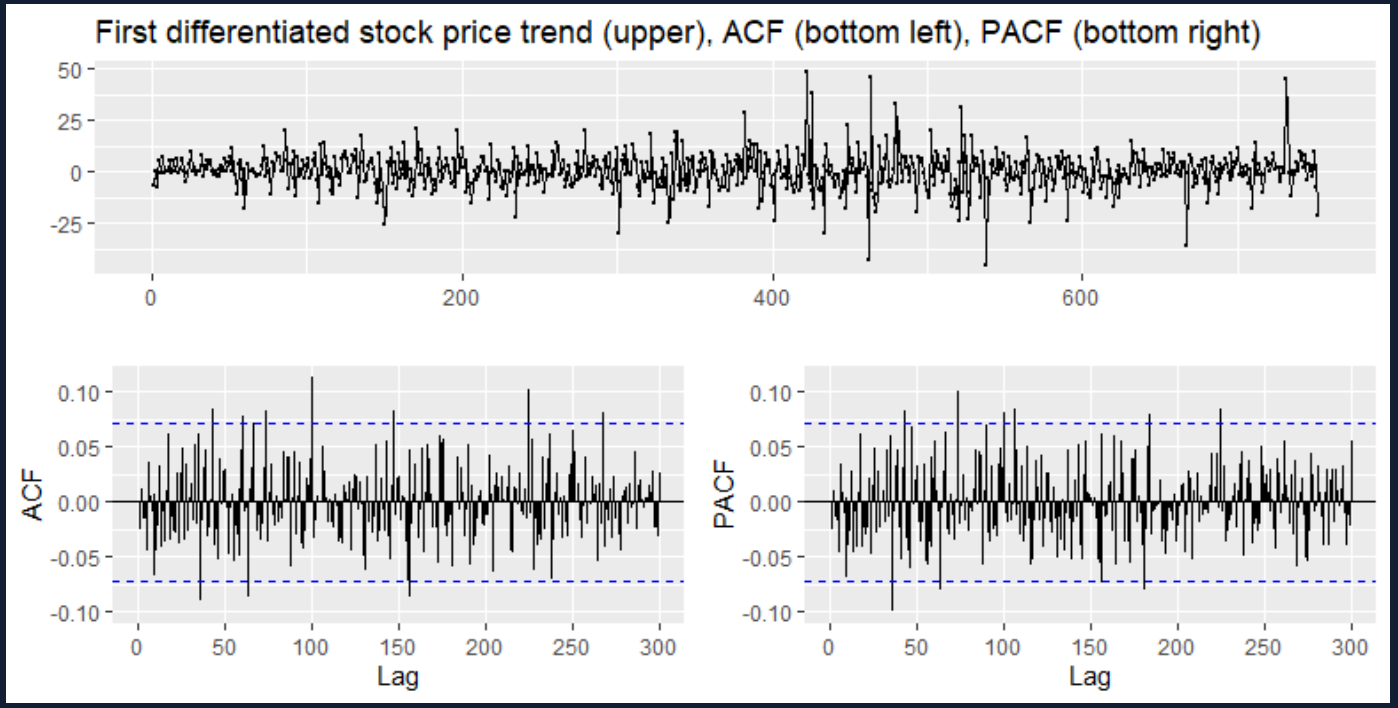


ARIMA model prediction values vs. actual values

- How to add some trend into our prediction instead of having a straight line?
- Is there a better model?

Motivation of Seasonality

- Look at the differenced data from **2016.11.27-2019.11.25**
  - Is there seasonality?
  - Can T possibly be 100?



- Use seasonal dummy test to verify

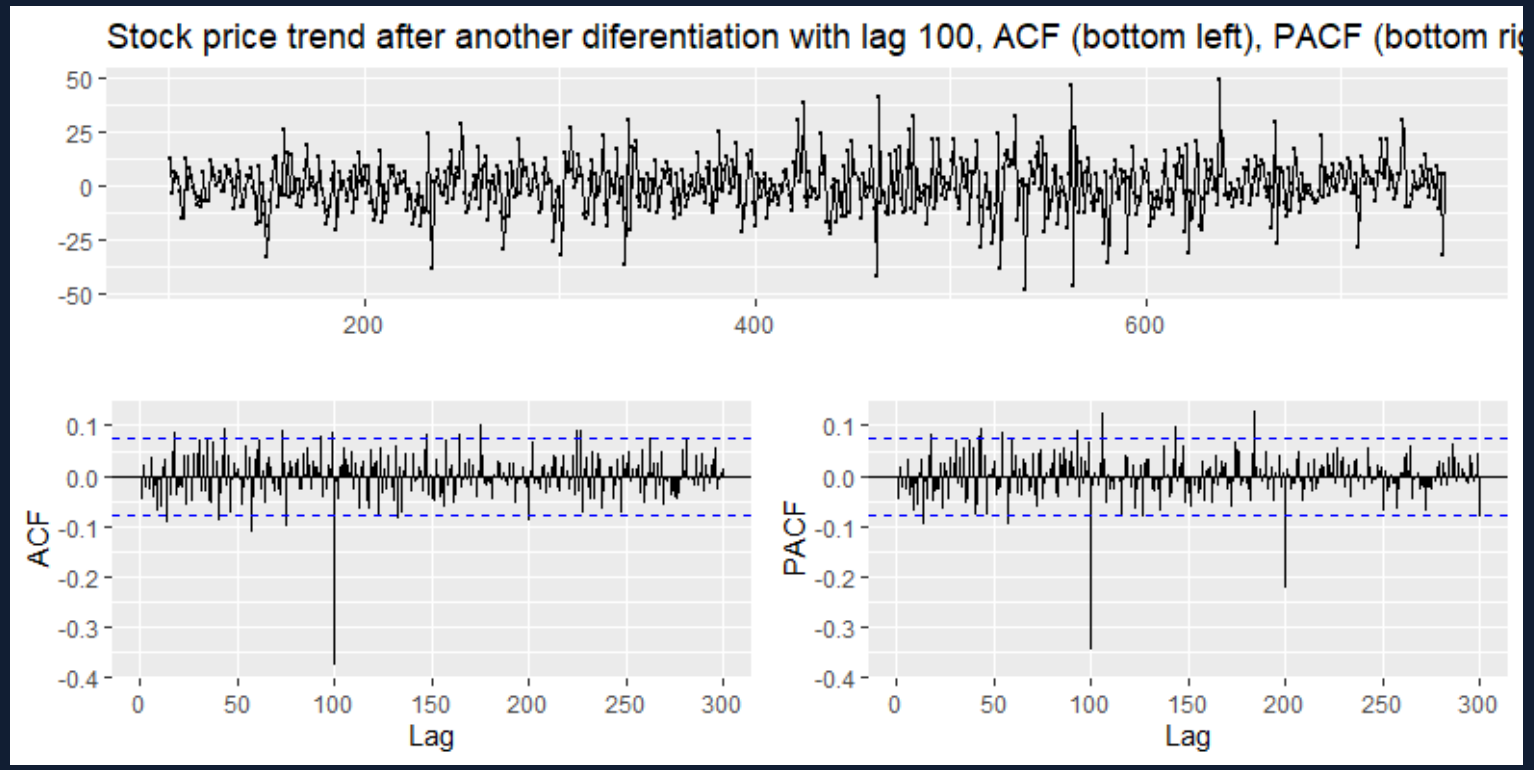
```
> isSeasonal(diff_close.ts, freq = 100, test = 'seasdum')
[1] TRUE
```

ARIMA to SARIMA

- SARIMA or Seasonal ARIMA, is an *extension* of ARIMA that takes *seasonality* into account.
- It takes the form SARIMA$(p, d, q)(P, D, Q)_m$, with *P*, Seasonal autoregressive order; *D*, Seasonal difference order; *Q*, Seasonal moving average order and *m*, the number of time steps for a single seasonal period.

SARIMA

- Taking another difference with lag 100
- We can tell
  P=2
  Q=1
  D=1

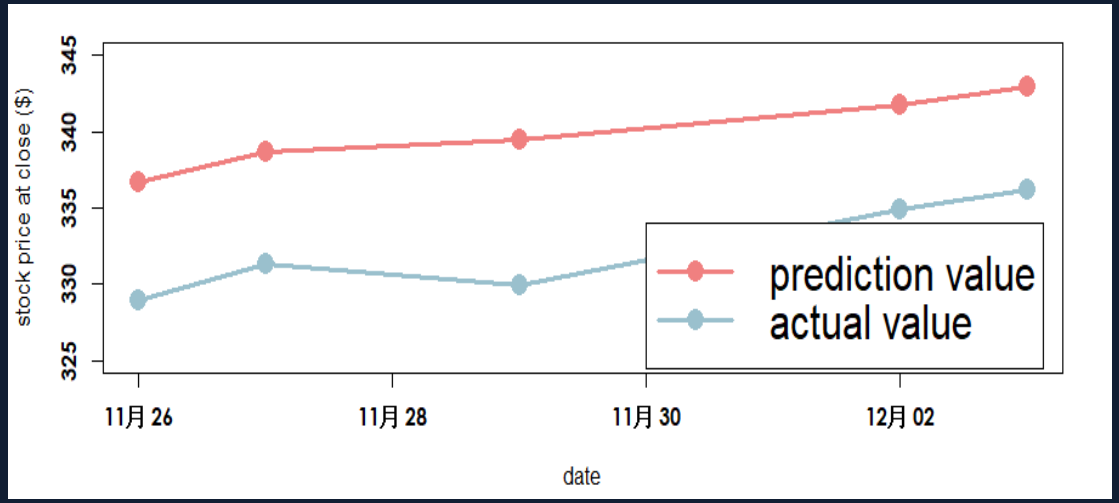


The result of taking difference with lag 100

- Now we have the model $SARIMA(0,1,0)(2,1,1)_{100}$

SARIMA Prediction

- Result
  - Error: 0.02302

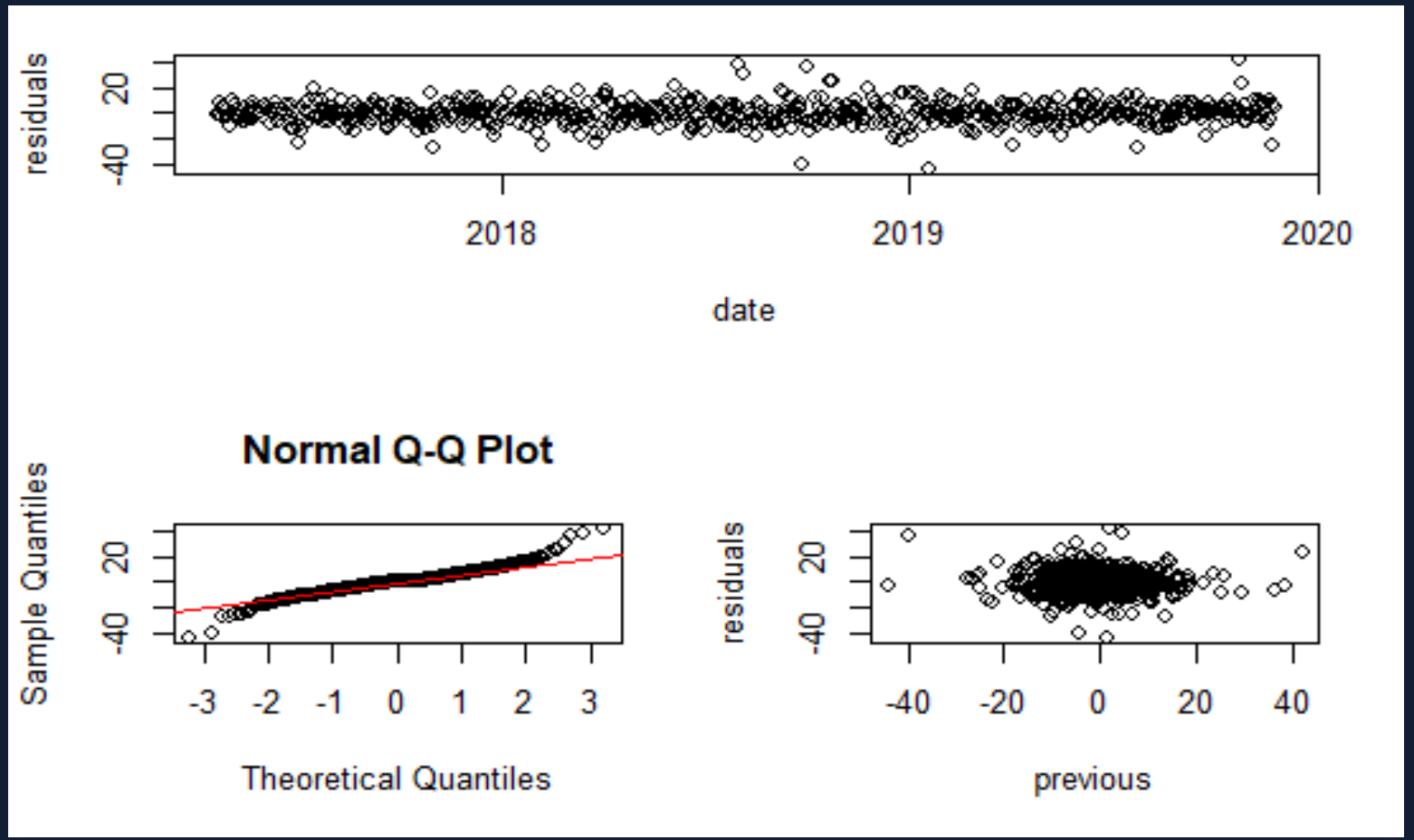


SARIMA model prediction values vs. actual values

- In terms of trend, it seems better than ARIMA model
- But we have to check the diagnostic tests first!

Diagnostic

- It seems fine.



Autocorrelation Problem

- Ljung-Box test for residuals

P-Value
0.2874
0.4209
0.5394
0.7024
0.7816
0.7965

- Ljung-Box test for residuals squared

P-Value
0.0001
0.0005
0.0003
0.0010
0.00175
0.0033

- There can possibly be autocorrelations between the residuals!

Search method

- Use Rugarch Package in R
  - No straight-forward method for Sarima-Garch
- Go back to Arima model

Abandon Original Arima

- Lack explanation of information
- Find New Arima with the largest AIC
  - Get Parameter (4,1,5)
- Use the Protmanteau Test
  - Check the Arch Effect

P-Value

0.0001402

0.0006291

3.665e-05

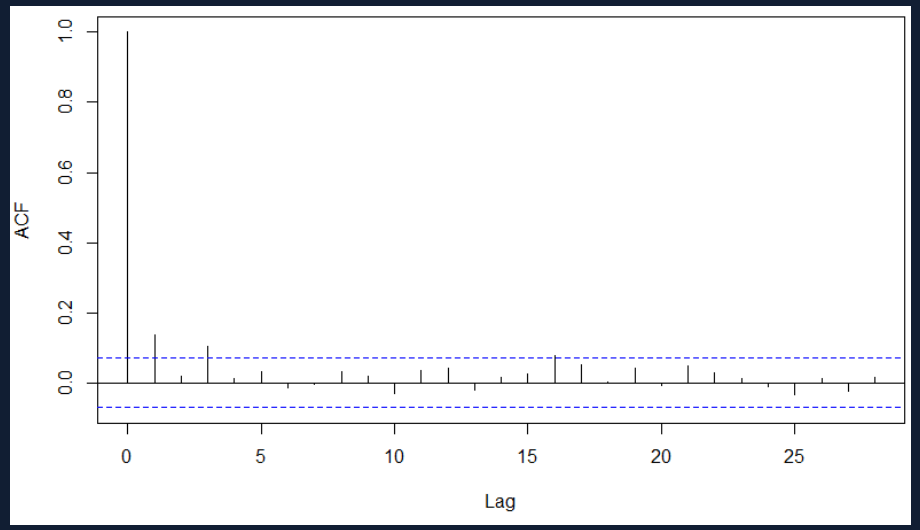
0.000108

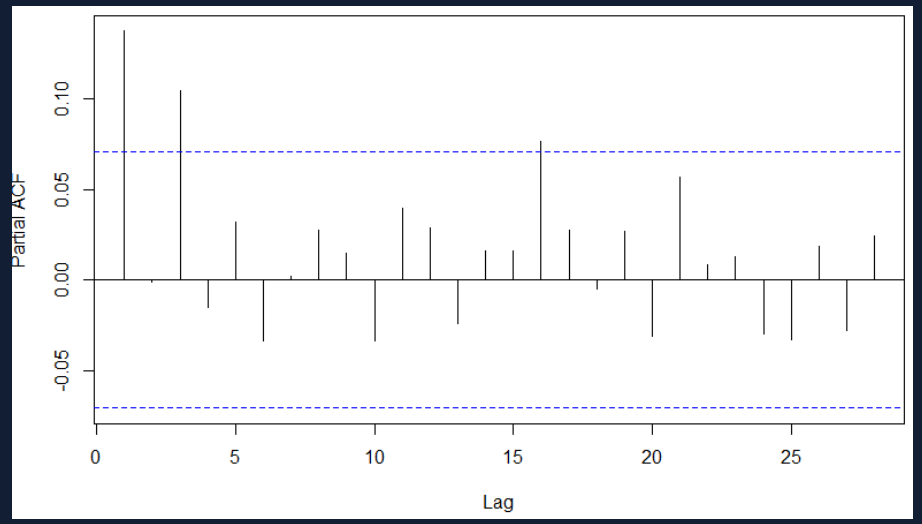
0.0002091

0.0004744

Choose the Garch specification

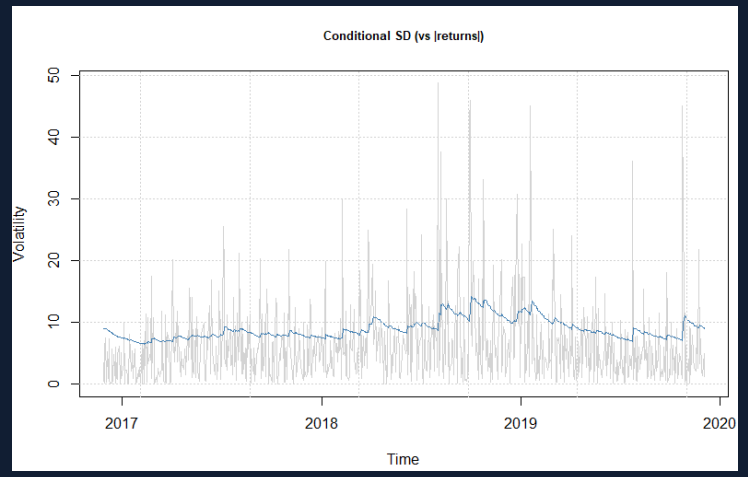
- Variance Order
  - Check ACF and PACF plot of the squared residuals

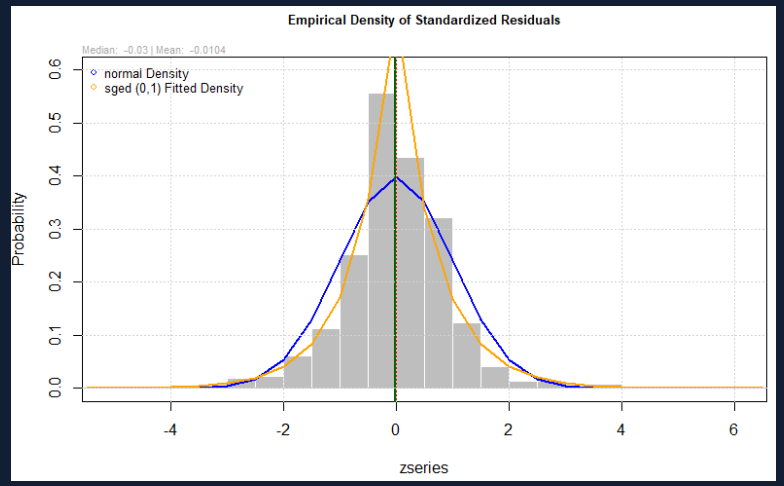




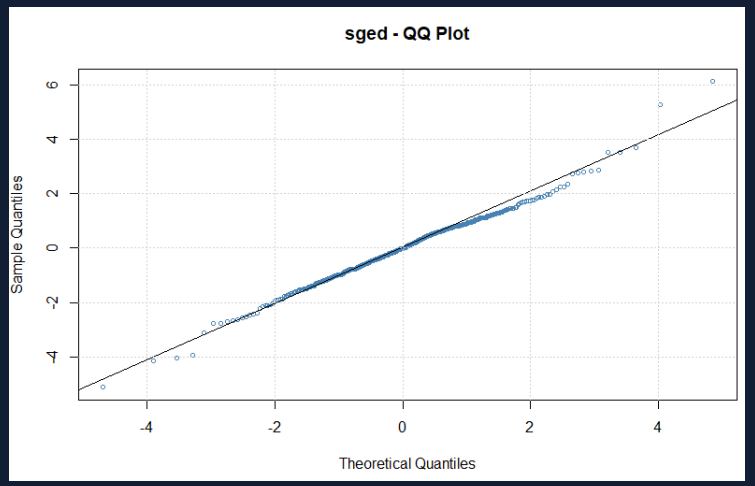
Diagnostic

- Explanation
  - Conditional SD
  - Original Standardized Residual



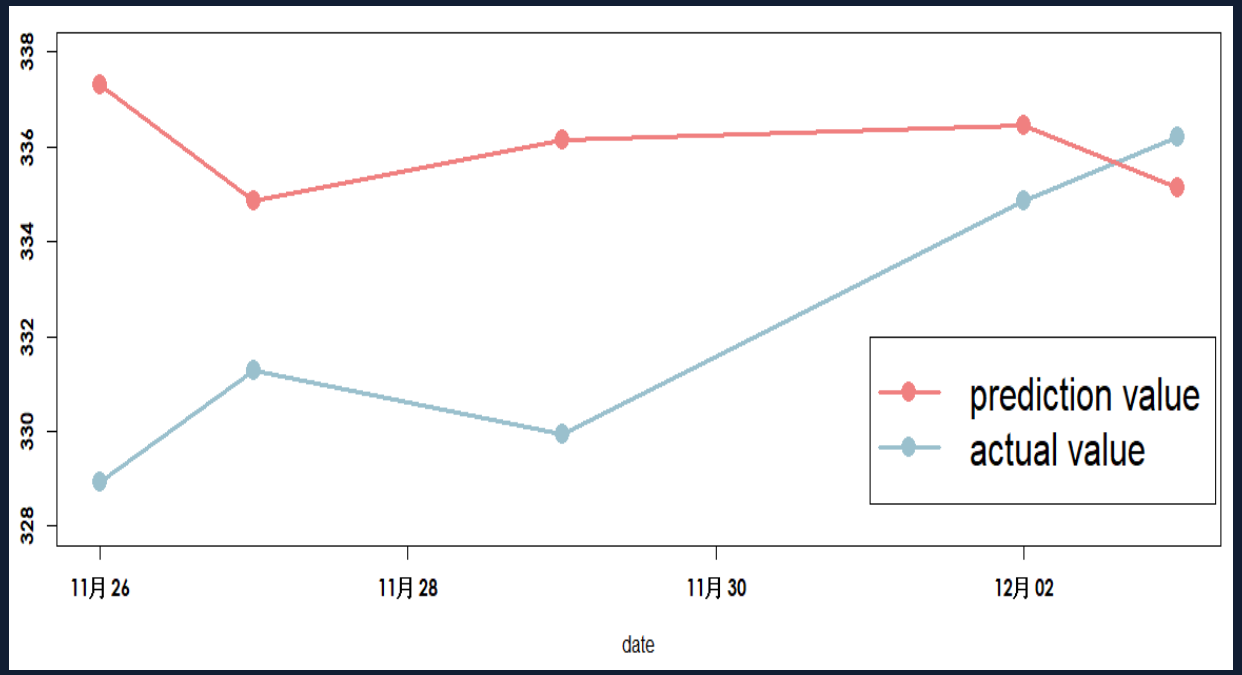


- Normality



Validation

- Plot
  - Prediction and Actual Value
- Error
  - 0.01258502



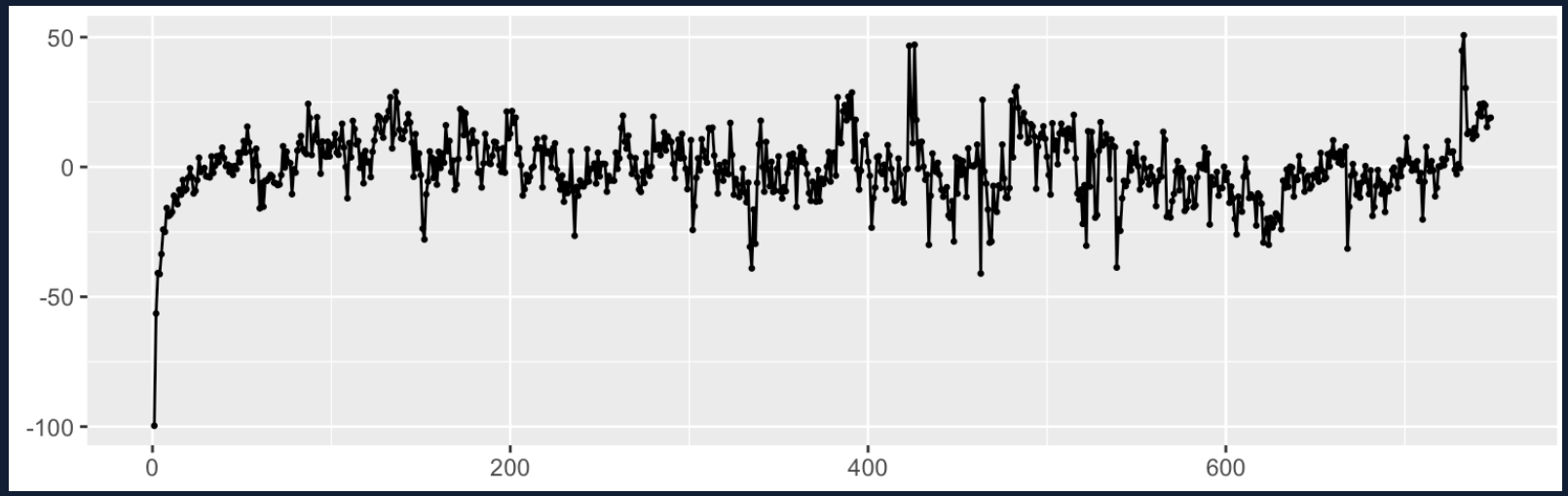
ARIMA-GARCH model prediction values vs. actual values

Motivation of Fractional Integrated Model

- Stock price is series that long-term trend is important
- Integer difference may not be the best
- fractional difference (d<0.5) can obtain long term memories

Have a try!

- Differ by 0.4995839 (given by program under p=q=0)

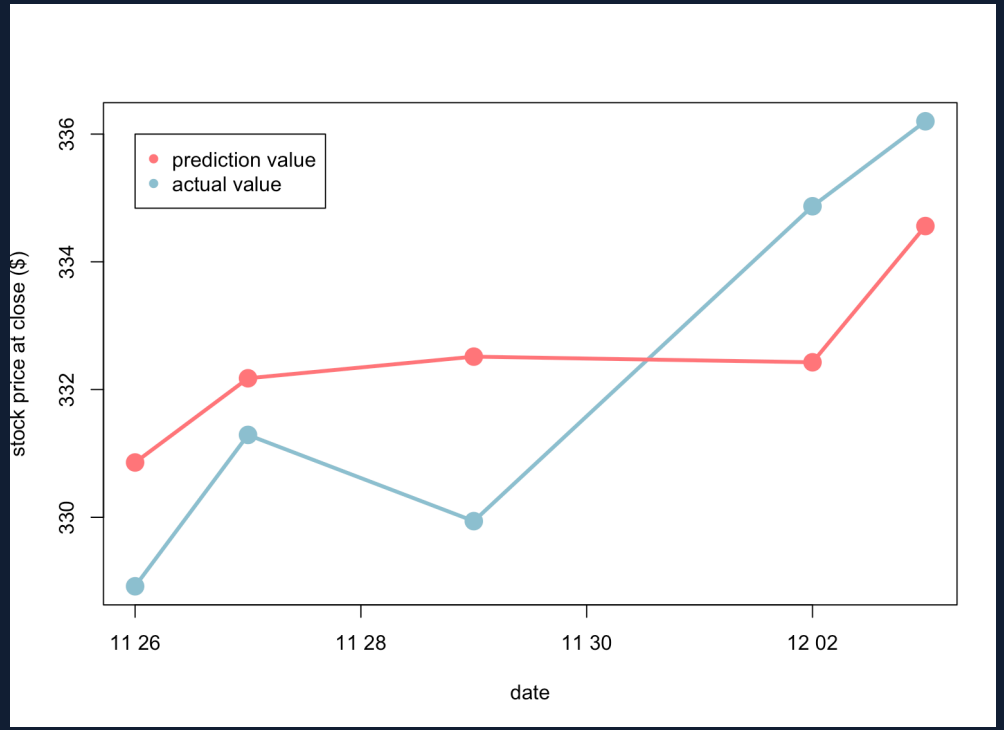


- ADF test

```
Warning message:
In adf.test(close.res.fd) : p-value smaller than printed p-value
```

Move further

- Parameter of SARIMA as the starting point
- Get the optimal solution with numerical method provided in R
- Result
  - Error 0.005709598
  - Coefficient



SARFIMA model prediction values vs. actual values

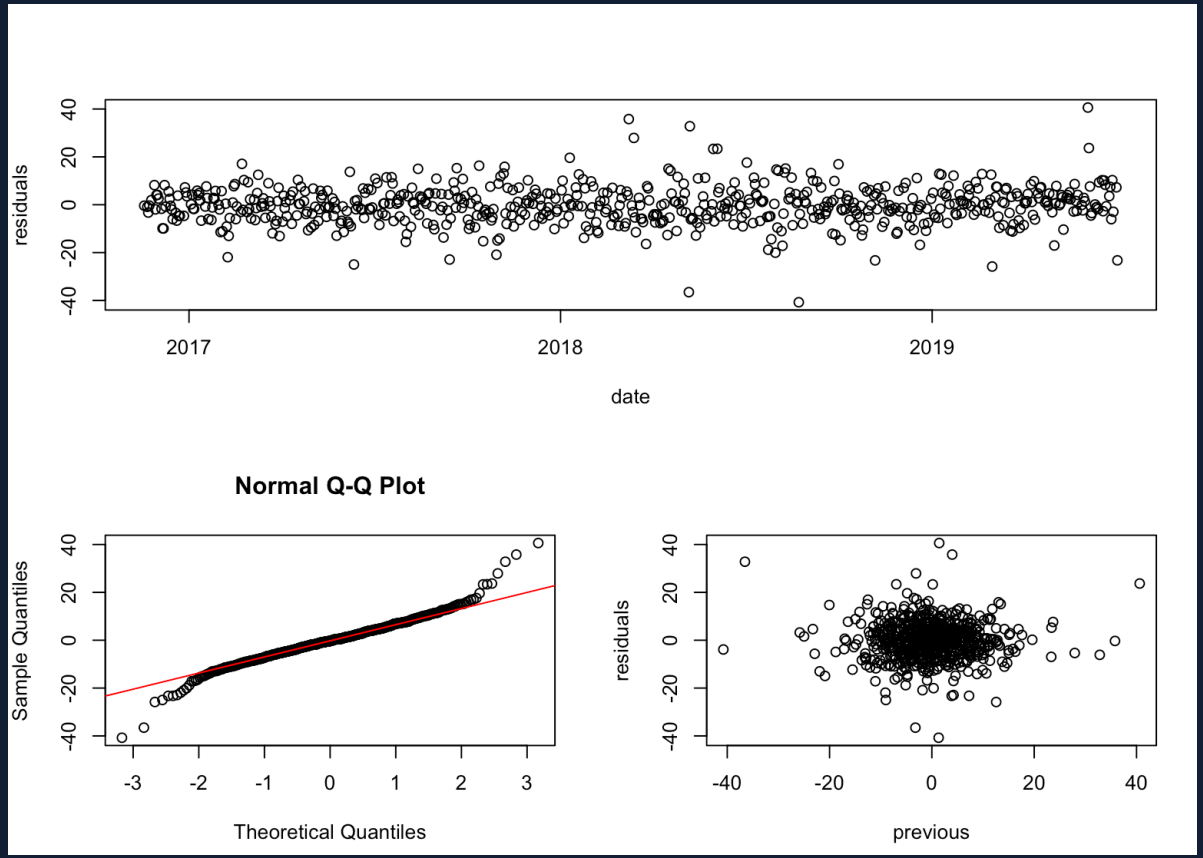
phi.100(1)	phi.100(2)	theta.100(1)	d.f	d.f.100	Fitted mean
0.6327175	-0.1922292	0.7491553	-0.03524325	-0.9991771	-0.0820306

Validation

- Ljung-Box test of squared term

P-Value
0.001809
0.007654
0.011
0.02252
0.04295
0.04361

- Diagnostic Plot



Conclusion

- Final model is SAFIMA. why?
  - best error performance
  - best trend
  - relative fine arch effect
- Can we do better? may be SAFIMA-GARCH
- Issue regarding massive data set
  - may suffer from overfit (flexibility of our model)
  - computational efficiency
  - do not contains other regressor

Thanks