Anthony Anh Quoc Doan

Time Series & ARIMA Forecasting with R

For LA East R Users Group
Meetup - Jan 14, 2019

About Me

- A statistician/data scientist.
- A BS in Computer Science (UCR) and working on my Master of Applied Statistic (CSULB).
- 10+ years of professional experiences as software developer.
- 2+ years as a statistician doing mostly modeling

Acknowledgement

- This presentation is what I've learned from the book, Forecasting : Principles and Practice by Rob J Hyndman and George Athanasopoulos
- Most of the examples and pictures is from this free online book

Motivation

- Time series data forecasting models are used in many industries
- Stock price prediction (this is not a financial advice)
- Retail industry for inventory planning
- Uber uses to predict how many driver needed certain time of the day

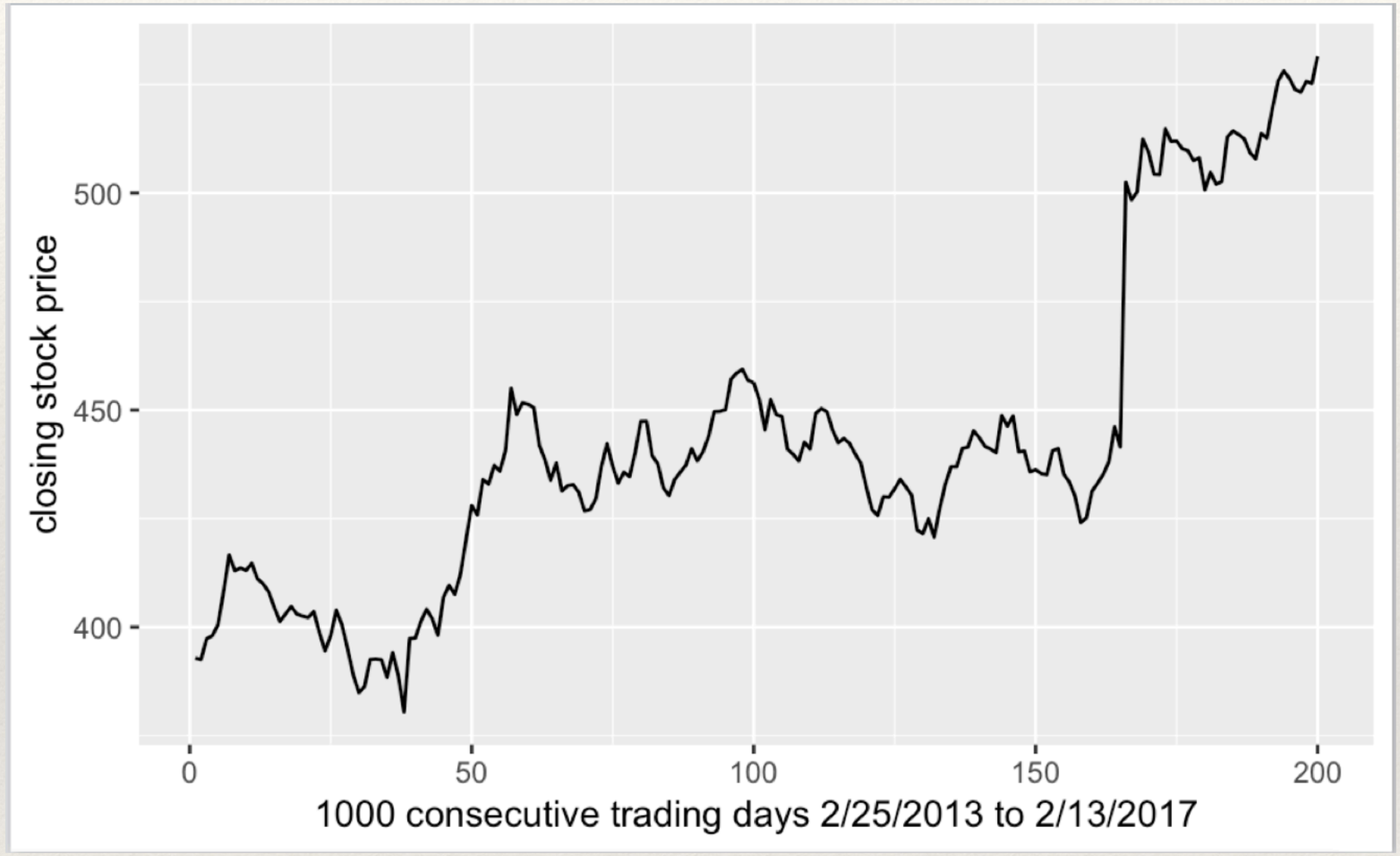
Why do we need a set of different models?

- In statistic, once you identify what type of data you have you have a whole set of tools for that type of data.
- Why?
- Take the linear regression model, it is for inference first and foremost and it doesn't take into account all information from the data (such as correlation among the observations).
- Time Series forecasting model is created for prediction and take into account all the information within the data

What is Time Series Data?

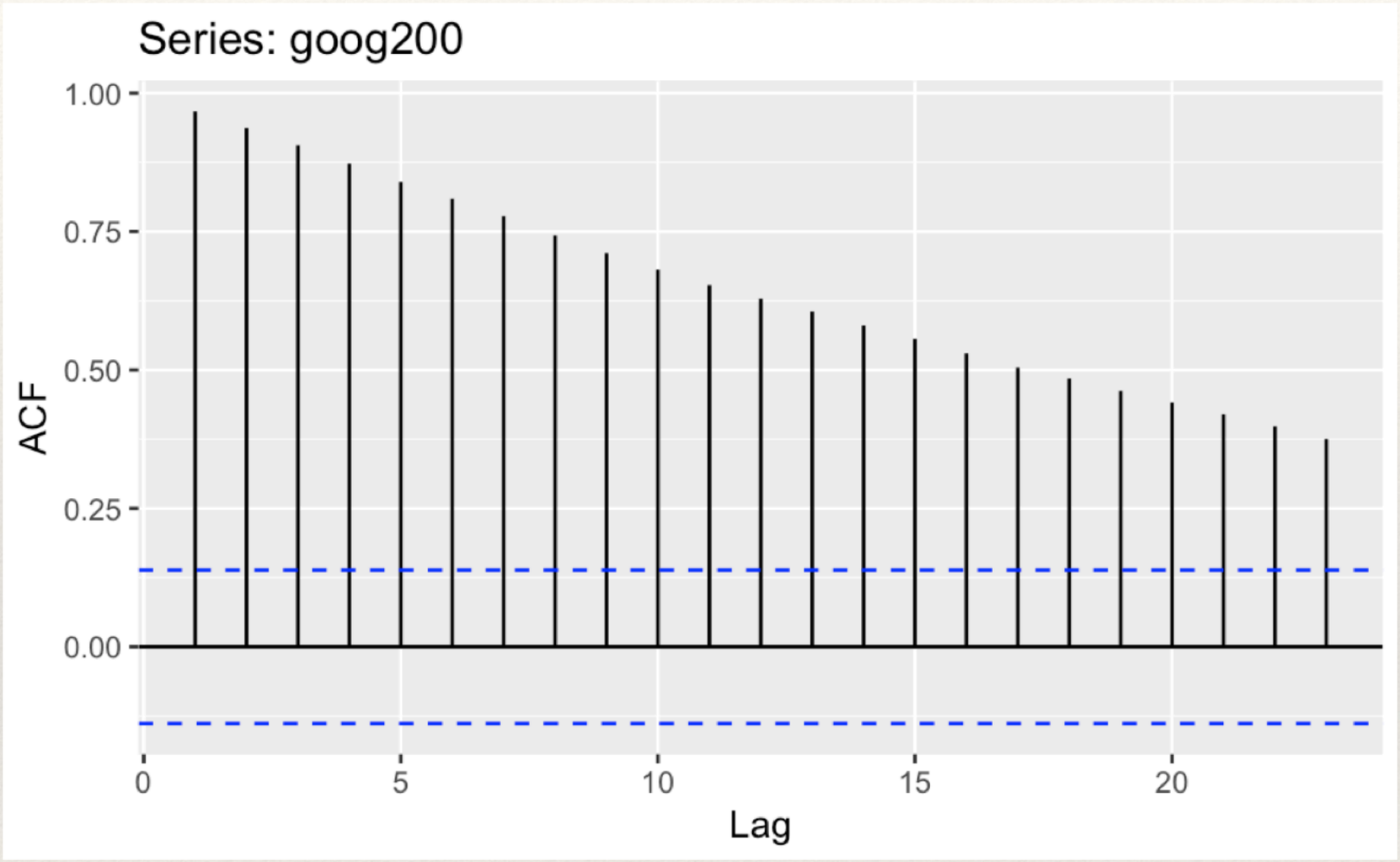
- We're going to talk about just univariate time series data.
- Time series data is a sequence of data points that measure the same thing over time
- The data is time dependent
- Every observation/values/responses are affected by past values
- There is a natural order in time (past, present, future)

Plotting Time Series Data




```
require('fpp2')
autoplot(goog200) +
  xlab("1000 consecutive trading days 2/25/2013 to 2/13/2017") +
  ylab("closing stock price")
```

ACF Plot




```
> ggAcf(goog200)
```

Now that we figure out our data is a time series data. Let's create a forecast model.

Why ARIMA forecast Model?

- There is a renown Time Series forecasting competition, Makridakis Competitions.
- It is currently on its 4th competition.
- The top algorithm is two statistical models (Exponential Smoothing & ARIMA).
- If you know those two models you're good with all univariate time series data out there.

Fitting a Model with ARIMA using R

```
21  (fit1 <- auto.arima(goog200))
```

```
> (fit1 <- auto.arima(goog200))
Series: goog200
ARIMA(0,1,0) with drift

Coefficients:
      drift
      0.6967
s.e.  0.4373

sigma^2 estimated as 38.25:  log likelihood=-644.45
AIC=1292.91   AICc=1292.97   BIC=1299.5
```

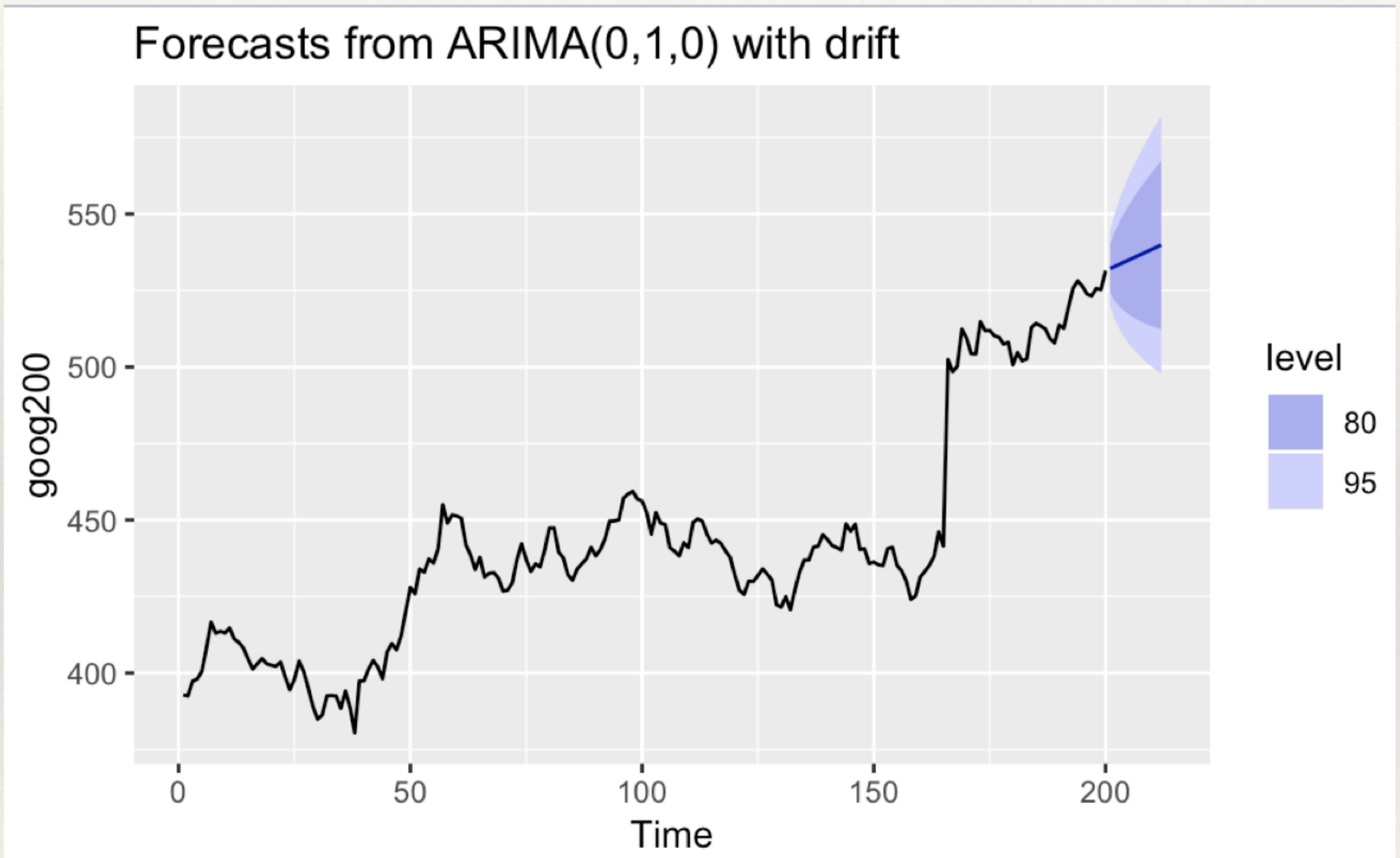
Forecasting

```
22  fit1 %>% forecast(h=1)
```

```
> fit1 %>% forecast(h=1)
    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
201        532.175 524.2492 540.1008 520.0535 544.2965
```

Graphing the forecasting

```
fit1 %>% forecast(h=12) %>% autoplot()
```



Auto ARIMA without optimization

```
(fit2 <- auto.arima(goog200, stepwise = FALSE,
                    approximation = FALSE))
```

```
> (fit2 <- auto.arima(goog200, stepwise = FALSE,
+                     approximation = FALSE))
Series: goog200
ARIMA(2,1,2)

Coefficients:
         ar1      ar2      ma1     ma2
      0.5815  -0.7574  -0.6222  0.9077
s.e.  0.0863   0.1976   0.0531  0.1504

sigma^2 estimated as 37.32:  log likelihood=-640.74
AIC=1291.49   AICc=1291.8   BIC=1307.96
```

Comparison between the two

```
> (fit1 <- auto.arima(goog200))
Series: goog200
ARIMA(0,1,0) with drift

Coefficients:
       drift
      0.6967
s.e.  0.4373

sigma^2 estimated as 38.25:  log likelihood=-644.45
AIC=1292.91   AICc=1292.97   BIC=1299.5
```

```
> (fit2 <- auto.arima(goog200, stepwise = FALSE,
+                     approximation = FALSE))
Series: goog200
ARIMA(2,1,2)

Coefficients:
         ar1      ar2      ma1     ma2
      0.5815  -0.7574  -0.6222  0.9077
s.e.  0.0863   0.1976   0.0531  0.1504

sigma^2 estimated as 37.32:  log likelihood=-640.74
AIC=1291.49   AICc=1291.8   BIC=1307.96
```

Accuracy of the two models

```
34  accuracy(fit1)
35  accuracy(fit2)
```

```
> accuracy(fit1)
                     ME      RMSE       MAE           MPE
Training set 0.001960665 6.153549 3.807244 -0.01512911
                    MAPE     MASE         ACF1
Training set 0.8591933 1.01779 -0.06043606
> accuracy(fit2)
                     ME      RMSE       MAE        MPE       MAPE
Training set 0.6392715 6.031873 3.695635 0.1313824 0.8384858
                    MASE         ACF1
Training set 0.9879536 -0.03037047
```