Time Series Analysis & Forecasting

Class 4

Arnab Bose, Ph.D.

MSc Analytics

University of Chicago

auto.arima: Hyndman-Kkandakar Algorithm

Refer to https://otexts.com/fpp2/arima-r.html

Time Series with Drift

```
set.seed(1)         # so you can reproduce the results
v = rnorm(100,1,1)  # v contains 100 iid N(1,1) variates
x = cumsum(v)       # x is a random walk with drift = 1
plot(x)
model1 <- arima(x, order = c(1,1,0))
plot(forecast(model1, h=12))
kpss.test(x) # not stationary due to drift
kpss.test(x, null = "Trend") # cannot detect drift, only deterministic trend

# same effect as it does not consider drift term
model2 <- Arima(x, order = c(1,1,0))
plot(forecast(model2, h=12))

# correct results with drift included
model3 <- Arima(x, order = c(1,1,0), include.drift = T)
plot(forecast(model3, h=12))
# Refer to   https://www.stat.pitt.edu/stoffer/tsa4/Rissues.htm
```

Model Estimation

- TS $\{y_t\}$
- Maximum Likelihood (ML) is estimating the unknown parameters such that the probability of observing the given $\{y_t\}$ is as high as possible
- Use differential calculus to solve for the parameters and get the maximum value
- t-value given by

$$t = \frac{\hat{\beta} - \beta}{std\_error(\beta)} = \frac{\hat{\beta} - \beta}{\frac{\hat{\sigma}}{\sqrt{\sum x_i^2}}}$$

- ARIMA models include any parameter whose |t-value| > 2

Model Diagnostic – Residual Analysis

- Define residual = actual – estimate

- Autocorrelation – use Durbin Watson

$$d = 2(1 - r)$$

r => 1st lag sample autocorrelation of the residual

d = 2 => no autocorrelation

- Normality – QQ plot

- Normality – Shapiro-Wilk test where the null hypothesis is that the dataset is normally distributed. If *p*-value is lower than threshold, then reject

Ljung-Box Test

- Let sample autocorrelation function of the residuals be denoted as $\widehat{r_k}$

- Box-Pierce statistic

$$Q = n \sum_k \widehat{r_i^2}$$

- For ARMA(p,q) for large n, Q has approximate χ^2 (chi-squared) distribution

- Ljung-Box noted that even with large n, n=100, the approximation to χ^2 distribution is not satisfactory

- Ljung-Box modified Q

$$Q = n(n+2) \sum_k \widehat{r_i^2}$$

Trend and Seasonal Decompositions

- Trend adjustments
  - Fit a regression model and subtract from TS to get residuals with no trend
    - Assumes constant trend historically and continuing into (immediate) future
  - Difference the TS
    - No parameter estimation (simpler)

- Seasonal adjustments
  - To eliminate seasonality say with lag = D

$$\nabla^D Y_t = (1 - B^D)Y_t = Y_t - Y_{t-D}$$

- Most TS require a combination of trend and seasonal adjustments

R code – Trends and Drift

- *ts.plot(cumsum(rnorm(1000, mean = 2, sd=sqrt(9))), col="purple", ylab="yt" , ylim=range(-10:2000)) # RW + stochastic trend with beta = 2 and white noise variance = 9*
- *x <- 1:1000*
- *yt <- 2*x + rnorm(1000, sd = sqrt(9)) # RW + deterministic trend with beta = 2*
- *lines(x, yt, col="blue")*
- *lines(cumsum(rnorm(1000)), col="green") # RW with no trend*
- *legend(1,1999, legend=c("drift/stochastic trend", "deterministic trend", "no trend"), col=c("purple", "blue", "green"), lty=1:3, cex=0.8)*

Seasonal ARIMA

- Mathematically using the backward shift operator

$$B^s z^t = z_{t-s}, s \Rightarrow seasonal\ period$$

- TS represented as

$$\Phi(B^s)Y_t = \Theta(B^s)e_t$$

- Also, since errors are autocorrelated as in regular ARIMA, we also have

$$\Phi(B)Y_t = \Theta(B)e_t$$

- So multiplicative seasonal ARIMA is represented as

$$\Phi(B)\Phi(B^s)\nabla^d \nabla_s^D Y_t = \Theta(B)\Theta(B^s)e^t$$

R code – seasonal and trend difference

```
ts.plot(AirPassengers)
auto.arima(AirPassengers) # model_1 with d = 1 (ie trend diff) and D = 1 (ie seasonal diff)

# try to do both the diff manually
# trend diff
trendDiff <- diff(AirPassengers)
ts.plot(trendDiff) # the plot has seasonality

# seasonal diff
seasonalDiff <- diff(AirPassengers, 12)
ts.plot(seasonalDiff) # the plot has trend

# both trend and seasonal diff – order does not matter
comboDiff <- diff(seasonalDiff)
ts.plot(comboDiff)
auto.arima(comboDiff) # the model should be identical to mode_1
```


Textbook Chapters

- Materials covered available in books:
  - FPP: Chapter 8, MJK: Chapter 5, TSA: Chapters 7, 8, 10