ARIMA processes

Practical Time Series Analysis

Thistleton and Sadigov

Objectives

- Describe autoregressive, integrated, moving average models
- Rewrite autoregressive, integrated, moving average models using backshift and difference operators

ARMA processes

Remember ARMA(p, q) process

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots \beta_q Z_{t-q}$$

can be written as

$$\phi(B) X_t = \beta(B) Z_t$$

where

$$\beta(B) = \beta_0 + \beta_1 B + \cdots + \beta_q B^q$$

$$\phi\,(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

- z – complex variable

- Roots of the polynomials $\beta(z)$ and $\phi(z)$ lie outside of the unit circle

- ARMA(p, q) process will be stationary and invertible

Non-stationary data

- Real life datasets are non stationary
- They might have a systematic change in trend
- We need to remove trend
- Difference operator $\nabla = 1 - B$

Difference operator

Remember

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

So, the random walk model

$$X_t = X_{t-1} + Z_t$$

can be written

$$\nabla X_t = Z_t$$

ARIMA(p, d, q) process

A process X_t is Autoregressive INTEGRATED Moving Average of order (p, d, q) if

$$Y_t := \nabla^d X_t = (1 - B)^d X_t$$

is ARMA(p, q).

$$Y_t \sim ARMA(p, q)$$

$$X_t \sim ARIMA(p, d, q)$$

$$\phi(B)\nabla^d X_t = \beta(B)Z_t$$

or

$$\phi(B)(1 - B)^d X_t = \beta(B)Z_t$$

d – order of differencing

- $d = 1$ or $d = 2$
- Over differencing may introduce dependence
- ACF might also suggest differencing is needed
- $\phi(z)(1 - z)^d$ has unit root with multiplicity of d
- ACF will decay very slowly

Modeling

- Trend suggests differencing
- Variation in variance suggests transformation
- Common transformation: log, then differencing
- It is also known as log-return
- ACF suggests order of moving average process (q)
- PACF suggests order of autoregressive process (p)
- Akaike Information Criterion (AIC)
- Sum of squared errors (SSE)
- Ljung-Box Q-statistics (Next lecture)
- Estimation!

What We've Learned

- Describe autoregressive, integrated, moving average models

- Rewrite autoregressive, integrated, moving average models using backshift and difference operators

Ljung-Box Q-statistic

Practical Time Series Analysis

Thistleton and Sadigov

Objectives

- Define Ljung-Box Q-statistic
- Learn the decision rule to test the null hypothesis that several autocorrelation coefficients are zero
- Test the null hypothesis that several autocorrelation coefficients are zero using R

Portmanteau statistic

Box and Pierce (1970) proposed Portmanteau statistic

$$Q^*(m) = T \sum_{l=1}^{m} r_l^2$$

as a test statistic for the null hypothesis

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0$$

against the alternative hypothesis

$$H_a: \rho_i \neq 0$$

for some $i \in \{1,2, \dots, m\}$.

Under i.i.d condition of $\{r_t\}$,

$$Q^*(m) \sim \chi^2(df = m)$$

asymptotically.

Ljung and Box (1978) modified statistic to increase the power of the test in finite samples

$$Q(m) = T(T+2)\sum_{l=1}^{m} \frac{r_l^2}{T-l}$$

Decision rule

We reject the null hypothesis if $\mathrm{Q}(m)$ is large enough i.e.,

$$Q(m) > \chi^2_\alpha$$

where χ^2_α is $100(1-\alpha)$-th quantile of Chi-Squared distribution with m degrees of freedom.

Most packages will actually calculate p-value. We will reject the null hypothesis if the p-value is sufficiently small, i.e.

$$p < \alpha$$

where α is the significance level.

Choice of m and R routine

Usually we take

$$m \approx \ln(T)$$

R routine

$$Box.test(data, lag = \log(T))$$

What We've Learned

- Define Ljung-Box Q-statistic
- Learn the decision rule to test the null hypothesis that several autocorrelation coefficients are zero
- Test the null hypothesis that several autocorrelation coefficients are zero using R

ARIMA fitting: Daily female births in California in 1959

Practical Time Series Analysis

Thistleton and Sadigov

Objectives

- Fit an ARIMA model to a real world data set
- Judge various fitting tools such as ACF, PACF and AIC
- Examine Ljung-Box test for testing autocorrelation in a time series

Modeling

- Trend suggests differencing
- Variation in variance suggests transformation
- Common transformation: log, then differencing
- It is also known as log-return
- ACF suggests order of moving average process (q)
- PACF suggests order of autoregressive process (p)
- Akaike Information Criterion (AIC)
- Sum of squared errors (SSE)
- Ljung-Box Q-statistics
- Estimation!

Daily female births in CA, 1959

- Time Series Data Library (TSDL)
- Created by Rob Hyndman, Professor of Statistics at Monash University, Australia.
- Link: https://datamarket.com/data/list/?q=provider%3Atsdl
- Category: Demography
- Name: Daily total female births in California, 1959
- 01 January 1959 – 31 December 1959
- Daily time series

Obtaining the data

- Click on the link: https://datamarket.com/data/set/235k/daily-total-female-births-in-california-1959#!ds=235k&display=line
- Export as CSV file
- Open the file, clean up the bottom row.
- Put the file into your working directory and read it to R
- OR read it directly from its path to R

ARIMA(0,1,2)

Then we have,

$$(1-B)X_t = 0.015_{0.015} + Z_t - 0.8511_{0.0496}\, Z_{t-1} - 0.1113_{0.0502}\, Z_{t-2}$$

where moving average coefficients are significant in the level of 0.05, and indices are standard errors.

Thus, the fitted model is

$$X_t = X_{t-1} + 0.015 + Z_t - 0.8511\, Z_{t-1} - 0.1113\, Z_{t-2}$$

where

$$Z_t \sim Normal\ (0, 49.08)$$

What We’ve Learned

- How to fit an ARIMA model to a real world data set using various fitting tools such as ACF, PACF and AIC
- Examine Ljung-Box test for testing correlation in a time series