ARMA

Sean Hellingman ©

Regression for Applied Data Science (ADSC2020)

shellingman@tru.ca

Winter 2025

THOMPSON RIVERS UNIVERSITY

Topics

2. Introduction
3. Moving Averages Process
4. Autoregressive Process
5. ARMA
6. Stationarity
7. Exercises and References

Introduction

- So far, we have looked at some theoretical models and visualizations of AR(p) and MA(q) processes.
- It is often difficult to tell the process just through visualizations of the series.
- We will now look to construct models.

Recall: Autocorrelation

- The **autocorrelation function** is a measure of the linear correlation between values of the process at different times.

$$Corr(Y_t, Y_s) = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t)Var(Y_s)}} \quad \text{for } t, s = 0 \pm 1, \pm 2, ... \tag{1}$$

- In other words, how linearly related values are across time.

Moving Averages AR(q)

Recall: Moving Average Process

- A **moving average** process is a random variable indexed in time that partially depends on the current and previous white noise terms.

$$Y_t = \mu + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + e_t \tag{2}$$

- Where e_t follows a white noise process.
- MA(3) Example: $Y_t = \mu + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3} + e_t$

ACF Plot

- The ACF plot will help us Identify correlation structures in the time series.
- It can help us determine the order of the MA(q) portion of the process.
- In a *stationary* process there should be an exponential decay in the correlation structure.

ACF Plot in R

- Using the *TSA* package in R you can use the `acf(series)` function (other functions exist).
- The confidence bands give information about the significance of the autocorrelation at each lag.
- The value at lag $= 0$ is the correlation of the process with itself ($Corr(Y_t, Y_t)$).

Example 1

1. Import the *Ex1A.RData* and *Ex1B.RData* files into R.
2. Use base R to plot the time series.
3. Try to determine the order of the MA(q) processes

Model Estimation

- We can use `model <- arima(data, order = c(p,d,q))` in R (*TSA* package).
- Estimates are calculated through MLE.
- You can use the `coeftest(model)` function from the *lmtest* package to obtain *p*-values.

Example 2

1. Estimate models for the two time series from Example 1 using the q values you identified.
2. Are these estimates significant.
3. What do these results imply?
4. Write out the model formula for one of the models.

Autoregressive Process AR(p)

Recall: Autoregressive Process

- An **autoregressive** process is a random variable indexed in time that partially depends on the previous values and a white noise terms.

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + e_t \tag{3}$$

- Where e_t follows a white noise process.

- AR(3) Example: $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + e_t$

PACF Plot

- The PACF plot will help us Identify correlation structures in the time series.
- It can help us determine the order of the AR(p) portion of the process.
- In a *stationary* process there should be an exponential decay in the correlation structure.

PACF Plot in R

- Using the *TSA* package in R you can use the `pacf(series)` function (other functions exist).
- The confidence bands give information about the significance of the partial autocorrelation at each lag.
- The value at lag $= 0$ is the correlation of the process with itself ($Corr(Y_t, Y_t)$).

Example 3

1. Import the *Ex3A.RData* and *Ex3B.RData* files into R.
2. Use base R to plot the time series.
3. Try to determine the order of the AR(p) processes

Model Estimation

- We can use `model <- arima(data, order = c(p,d,q))` in R (*TSA* package).
- Estimates are calculated through MLE.
- You can use the `coeftest(model)` function from the *lmtest* package to obtain *p*-values.

Example 4

1. Estimate models for the two time series from Example 3 using the p values you identified.
2. Are these estimates significant.
3. What do these results imply?
4. Write out the model formula for one of the models.

Auto Regressive Moving Averages ARMA(p,q)

Autoregressive Moving Average Process

- An **autoregressive moving average** process is partially autoregressive and partially moving averages.

$$\begin{aligned} Y_t =& \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} \\ &+ \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} + e_t \end{aligned} \tag{4}$$

- Where e_t follows a white noise process.

- ARMA(1,1) Example: $Y_t = \phi_1 Y_{t-1} + \theta_1 e_{t-1} + e_t$

ACF and PACF Plot in R

- We may use the ACF and PACF plots to determine the values of p and q.
- The confidence bands give information about the significance of the partial autocorrelation at each lag.
- The value at lag $= 0$ is the correlation of the process with itself ($Corr(Y_t, Y_t)$).

Example 5

1. Import the *Ex5A.RData* and *Ex5B.RData* files into R.
2. Use base R to plot the time series.
3. Try to determine the order of the ARMA(p,q) processes.

Model Estimation

- We can use `model <- arima(data, order = c(p,d,q))` in R (*TSA* package).
- Estimates are calculated through MLE.
- You can use the `coeftest(model)` function from the *lmtest* package to obtain p-values.

Example 6

1. Estimate models for the two time series from Example 5 using the p and q values you identified.
2. Are these estimates significant.
3. What do these results imply?
4. Write out the model formula for one of the models.

`auto.arima()`

- Because we are using MLE to estimate the models, AIC and BIC may be used to pick the *best* model.
- The `auto.arima(data)` function from the forecast package can be used.
- *The function may take many additional arguments.*
- We can still use the `coeftest()` function to test for significance.

Example 7

1. Use the `auto.arima()` function to select models from the data used in Example 6.

2. How did we do from ACF and PACF the plots?

Recall: Stationarity

- A process is said to be **strictly stationary** if the joint distribution of Y_{t_1}, Y_{t_2}, ..., Y_{t_n} is the same as Y_{t_1-k}, Y_{t_2-k}, ..., Y_{t_n-k}.
  - For all time points t_1, t_2, ..., t_n and all time lags k.
- In other words, the nature of the process does not change over time.
- If a process is strictly stationary and has finite variance, then the variance function must depend only on the time lag.

Recall: Weak Stationarity

- A stochastic process is said to be **weakly stationary** if:
  1. The mean function is constant over time.
  2. $\gamma_{t,t-k} = \gamma_{0,k}$ for all t and k.
- *We will refer to weak stationarity during this course.*

Weak Stationarity and ARMA

- **For a general ARMA(p,q) model, we require both stationarity and invertibility.**
- We will discuss solutions in later classes.
- We have ways to test for stationarity.

Augmented Dickey-Fuller (ADF) Test

- The null hypothesis of the ADF test is that there is a unit root (not stationary).
- **A rejection of the null hypothesis suggests a stationary process.**
- In R we can use the `adf.test(data)` function from the *tseries* package.

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

- The null hypothesis of the KPSS test is that the process is stationary.
- **A failure to reject the null hypothesis suggests a stationary process.**
- In R we can use the `kpss.test(data)` function from the *tseries* package.

Example 8

1. Import the *Ex8A.RData* and *Ex8B.RData* files into R.
2. Visualize the processes and their correlations structures.
3. Are they stationary?
4. Can you estimate appropriate models for these datasets?

Exercise 1

- Take some time to examine time series data (real and simulated).
- Do any of these processes seem to be stationary (use formal tests)?
- Can you use the ACF and PACF plots to guess the ARMA(p,q) models?
- Estimate some models for your data.

References & Resources

1. Shumway, R. H., & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications: With R Examples*. Springer Texts in Statistics. Link

2. Jonathan, D. C., & Kung-Sik, C. (2008). *Time series analysis with applications in R*.

- arima.sim()
- TSA
- zoo
- arima()
- auto.arima()
- adf.test()
- kpss.test()