Predictive Analytics

Week 13: ARIMA models (optional)

Semester 2, 2018

Discipline of Business Analytics, The University of Sydney Business School

Week 13: ARIMA models (optional)

1. Stationarity and Box-Jenkins methodology
2. Differencing
3. Models for stationary series
4. ARIMA models
5. Seasonal ARIMA models

This module provides a discussion of **ARIMA models**, the most widely used methods for univariate time series forecasting.

ARIMA models aim to describe the serial dependence in the data, rather than to directly describe the time series components as in exponential smoothing. The two approaches are complementary.

Stationarity and Box-Jenkins methodology

Stationarity (key concept)

Intuitively, a **stationary** time series is one whose properties do not depend on the time at which we observe it.

Time series with trend and seasonality are not stationary, since these patterns affect the change the mean of the series over time.

Strict stationarity (key concept)

Formally, a time series process is **strictly stationary** when the joint distribution of $Y_t, Y_{t-1}, \ldots, Y_{t-k}$ does not depend on t. That is, the joint density

$$p(y_t, y_{t-1}, \ldots, y_{t-k})$$

does not depend on t.

Weak stationarity (key concept)

A process is **weakly stationary** or **covariance stationary** if its mean, variance and autocovariances do not change over time. That is,

$$E(Y_t) = \mu,$$

$$\text{Var}(Y_t) = \sigma^2,$$

$$\text{Cov}(Y_t, Y_{t-k}) = \text{Cov}(Y_t, Y_{t-k}) = \gamma_k,$$

for all t and k.

ARIMA models

- Introduced in the seminal book "Time Series Analysis Forecasting and Control" (1970) by Box and Jenkins.

- The Box-Jenkins approach relies on (a) finding a stationary transformation of the data (b) modeling the autocorrelations in the transformed data.

- This approach contrast with exponential smoothing, where we explicitly model the different time series components through additive or multiplicative specifications.

Box Jenkins methodology

- We consider log or Box-Cox transformations to stabilise the variance of series.

- Differencing (next section) leads to stationarity in the mean by removing changes in the level of the series (due for example to trend and seasonality).

- Autocorrelation (ACF) and partial autocorrelation (PACF) plots help us to assess stationarity and to identify suitable specification for the stationary transformation of the series.

Partial autocorrelation function (PACF)

- The partial autocorrelation of order k (labelled ρ_{kk}) is the correlation between Y_t and Y_{t-k} net of effects at times $t-1, t-2, \ldots, t-k+1$.

- r_{kk} estimates ρ_{kk}.

PACF: interpretation via autoregressions

$$Y_t = \rho_{10} + \rho_{11} Y_{t-1} + a_t$$

$$Y_t = \rho_{20} + \rho_{21} Y_{t-1} + \rho_{22} Y_{t-2} + a_t$$

$$Y_t = \rho_{k0} + \rho_{k1} Y_{t-1} + \rho_{k2} Y_{t-2} + \ldots + \rho_{kk} Y_{t-k} + a_t$$

Differencing

Box and Jenkins advocate difference transforms to achieve stationarity. The **fist difference** of a time series is

$$\Delta Y_t = Y_t - Y_{t-1}$$

Example: random walk

In the random walk model

$$Y_t = Y_{t-1} + \varepsilon_t,$$

the first difference leads to stationary white noise series

$$\Delta Y_t = Y_t - Y_{t-1} = \varepsilon_t.$$

Second order differencing

In rare cases, it may be necessary to difference the series a second time to obtain stationarity:

$$\Delta^2 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

Differencing

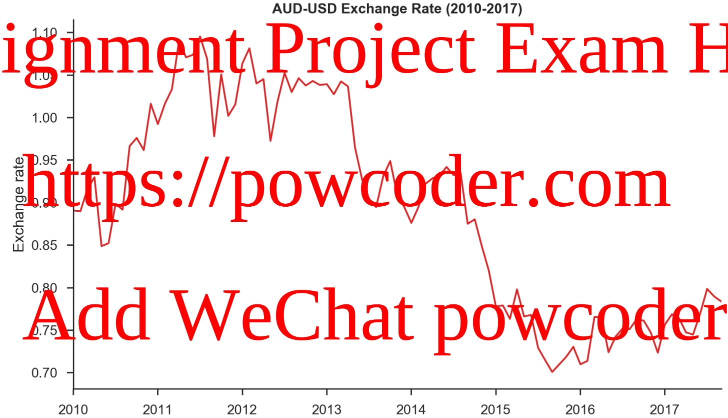
The ACF helps us to to determine whether the time series needs differencing (or further differencing).

- The ACF of a non-stationary series will decrease slowly.

- The ACF of a stationary series should drop to zero relatively quickly.

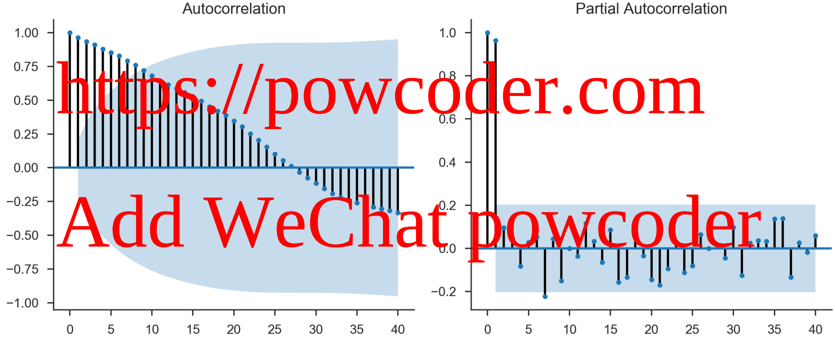
Unit root tests are also common for determining the need for differencing, but sensitive to assumptions. When in doubt, use model selection for model selection, not hypothesis testing.

Example: AUD/USD exchange rate

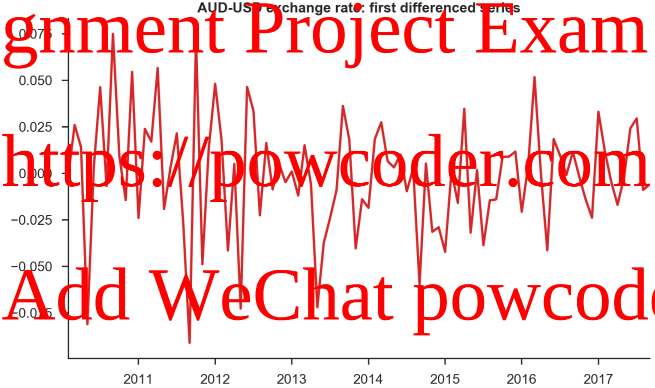


Example: AUD/USD exchange rate

ACF and PACF for the time series

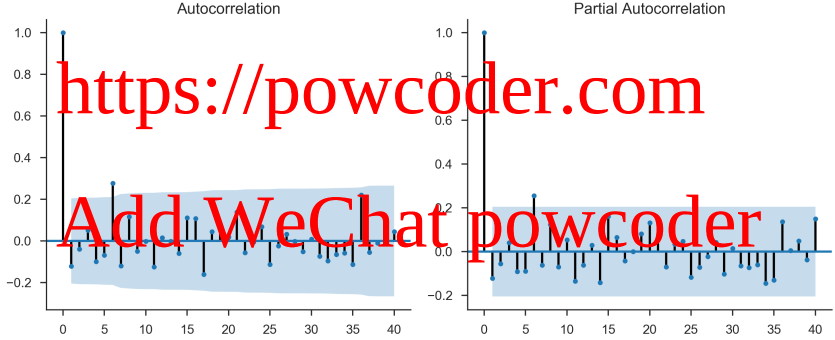


Example: AUD/USD exchange rate



Example: AUD/USD exchange rate

ACF and PACF for the first differenced series



Seasonal differencing

We can use seasonal differencing to address non-stationarity caused by seasonality:

$$\Delta_m Y_t = Y_t - Y_{t-m},$$

where m is the number of seasons.

The ACF of a series that needs seasonal differencing will decrease slowly at the seasonal lags m, $2m$, $3m$, etc.

First and seasonal differencing

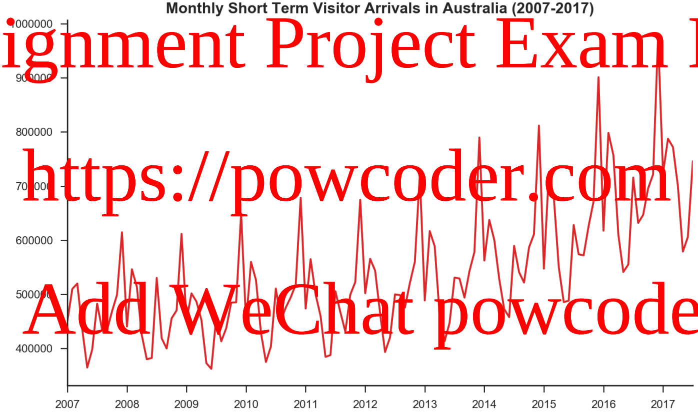
Time series that have a changing level and a seasonal pattern may require both first and seasonal differencing for stationarity.

The first and seasonally differenced series is

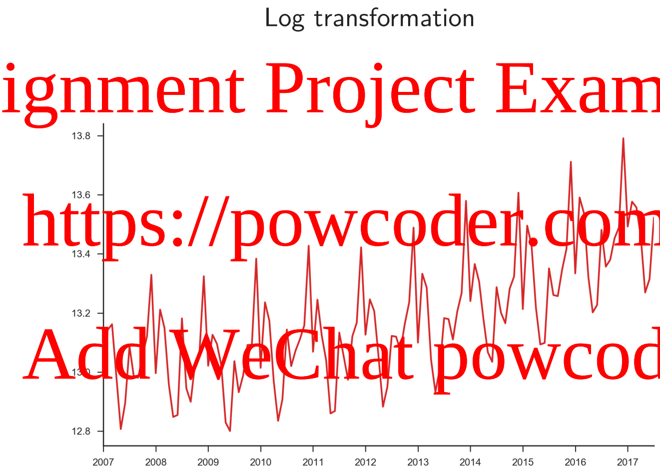
$$\Delta_m(\Delta Y_t) = (Y_t - Y_{t-1}) - (Y_{t-m} - Y_{t-m-1}),$$

noting that the order of differencing does not matter.

Example: Visitor Arrivals in Australia



Example: Visitor Arrivals in Australia

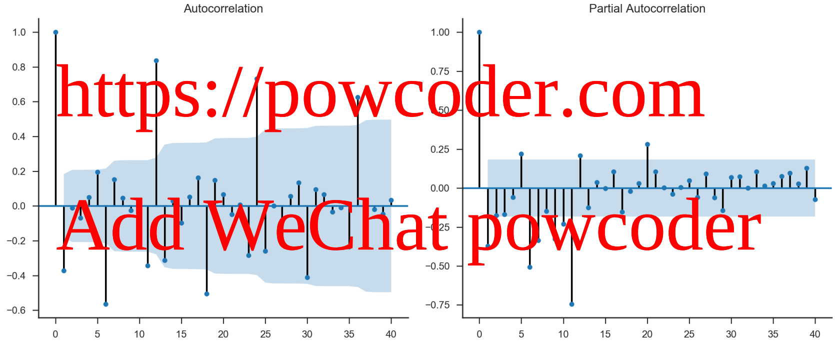


Example: Visitor Arrivals in Australia

First differenced log series

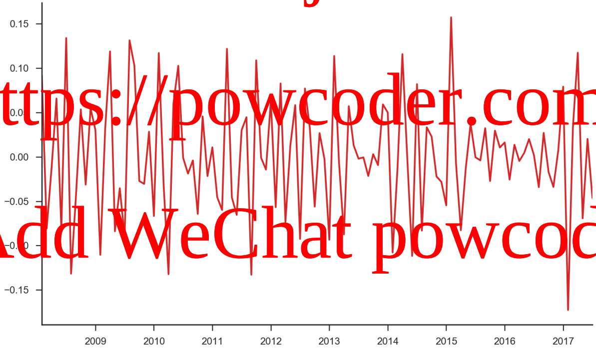
Example: Visitor Arrivals in Australia

ACF and PACF for the first differenced log series



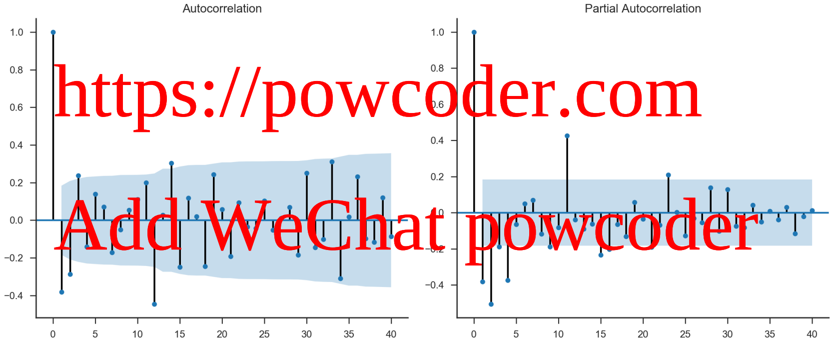
Example: Visitor Arrivals in Australia

First and seasonally differenced series



Example: Visitor Arrivals in Australia

ACF and PACF for the first and seasonally differenced log series



Backshift notation

The **backshift operator** is a useful notational device for ARIMA models.

$$BY_t = Y_{t-1}$$

We can manipulate the backshift operator with standard algebra, for example

$$B^2Y_t = B(BY_t) = BY_{t-1} = Y_{t-2}.$$

Therefore,

$$B^kY_t = Y_{t-k}.$$

Differencing in backshift notation

First differenced series:

$$(1 - B)Y_t = Y_t - BY_t = Y_t - Y_{t-1}$$

Seasonally differenced series:

$$(1 - B^m)Y_t = Y_t - B^m Y_t = Y_t - Y_{t-m}$$

First and seasonally differenced series:

$$\begin{aligned}(1 - B)(1 - B^m)Y_t &= (1 - B - B^m + B^{m+1}) \\ &= (Y_t - Y_{t-1}) - (Y_{t-m} - Y_{t-m-1})\end{aligned}$$

Models for stationary series

Autoregressive (AR) model (key concept)

The **autoregressive model** of order p, or **AR(p)** model, is

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t,$$

where ε_t is a white noise series.

Example: AR(1) model

AR(1) model:

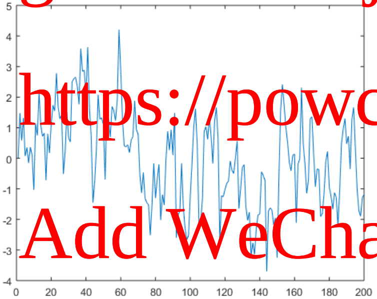
$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t,$$

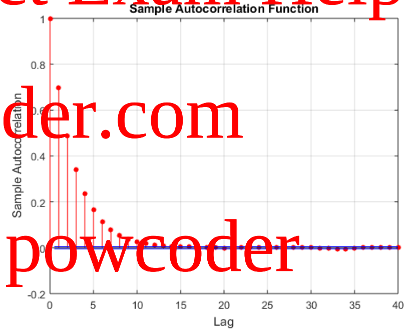
where ε_t is i.i.d. with mean zero and variance σ^2.

$$E(Y_t|y_1, \ldots, y_{t-1}) = E(Y_t|y_{t-1}) = c + \phi_1 y_{t-1}.$$

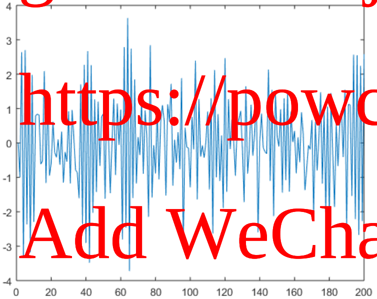
$$\text{Var}(Y_t|y_1, \ldots, y_{t-1}) = \text{Var}(Y_t|y_{t-1}) = \sigma^2.$$

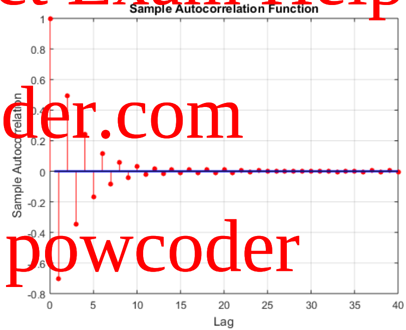
AR(1) illustration: $\phi = 0.7$





AR(1) illustration: $\phi = -0.7$





AR model: ACF and PACF identification (key concept)

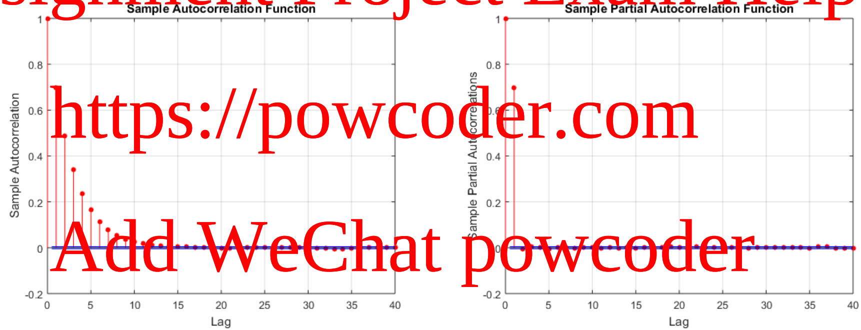
For an AR(p) process, we can show that:

- The theoretical autocorrelations ρ_k decrease exponentially.

- The theoretical partial autocorrelation ρ_{kk} cuts off to zero after lag p.

- The pth partial autocorrelation ρ_{pp} is ϕ_p.

AR(1) with $\phi = 0.7$: ACF (left) and Partial ACF (right)



AR(p) model forecasts

From the linearity of expectations,

$$E(Y_{t+h}|y_{1:t}) = c + \phi_1 E(Y_{t+h-1}|y_{1:t}) + \ldots + \phi_p E(Y_{t+h-p}|y_{1:t}),$$

where

$$E(Y_{t+h-i}|y_{1:t}) = \begin{cases} \widehat{y}_{t+h-i} & \text{if } h > 1 \\ y_{t+h-i} & \text{if } h \leq i. \end{cases}$$

Example: AR(1) model

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t$$

For $t + 1$,

$$\begin{aligned}\widehat{y}_{t+1} &= E(Y_{t+1}|y_{1:t}) \\ &= E(c + \phi_1 Y_t + \varepsilon_{t+1}|y_{1:t}) \qquad = c + \phi_1 y_t\end{aligned}$$

$$\text{Var}(Y_{t+1}|y_{1:t}) = \sigma^2.$$

Example: AR(1) model

For $t+2$,

$$
\begin{aligned}
\widehat{y}_{t+2} &= c + \phi_1 \widehat{y}_{t+1} \\
&= c(1+\phi_1) + \phi_1^2 y_t.
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{Var}(Y_{t+2}|y_{1:t}) &= \mathsf{Var}(\phi_1 Y_{t+1} + \varepsilon_{t+2}|y_{1:t}) \\
&= \phi_1^2 \mathsf{Var}(Y_{t+1}|y_{1:t}) + \sigma^2 \\
&= (1+\phi_1^2)\sigma^2
\end{aligned}
$$

Example: AR(1) model

$$
\begin{aligned}
\widehat{y}_{t+h} &= c + \phi_1 y_{t+h-1} \\
&= c(1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^{h-1}) + \phi_1^h y_t
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{Var}(Y_{t+h}|y_{1:t}) &= \phi_1^2 \mathsf{Var}(Y_{t+h-1}|y_{1:t}) + \sigma^2 \\
&= \sigma^2(1 + \phi_1^2 + \ldots + \phi_1^{2(h-1)}).
\end{aligned}
$$

As h gets larger, both the point forecast and the conditional variance converge exponentially to a constant.

Illustration: AR(1) forecast

Stationarity conditions

AR(p) model:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t.$$

We need to impose restrictions on the AR coefficients such that the model is stationary.

AR(1): $-1 < \phi_1 < 1$.

AR(2): $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.

AR(p) with $p > 2$: more technical.

Moving average (MA) model (key concept)

The **moving average** model of order q, or **MA(q)** model, is

$$Y_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q},$$

where ε_t is a white noise series.

Example: MA(1) process

The

$$Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

$$E(Y_t|y_{t-1}) = c + \theta_1 \varepsilon_{t-1}$$

$$\text{Var}(Y_t|y_{t-1}) = \sigma^2$$

MA model: ACF and PACF identification

For an MA(q) process, we can show that:

- The theoretical autocorrelation ρ_k cuts off after lag q.
- The theoretical partial autocorrelations ρ_{kk} decrease exponentially.

- An MA(q) process is **invertible** when we can write it as a linear combination of its past values (an AR(∞) process) plus the contemporaneous error term.

- Estimation and forecasting methods for MA models rely on invertibility. We therefore impose restrictions on the MA coefficients such that invertibility holds.

ARMA(p, q) model (key concept)

The **ARMA(p, q)** model is

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

where ε_t is a white noise series

In backshift notation,

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) Y_t = c + \left(1 + \sum_{i=1}^{p} \theta_i B^i\right) \varepsilon_t.$$

The autocorrelations and partial autocorrelations decrease exponentially for ARMA processes.

Example: ARMA(1, 1)

The ARMA(1,1) model is

$$Y_t = c + \phi_1 Y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t.$$

In backshift notation,

$$(1 - \phi_1 B) Y_t = c + (1 + \theta_1 B) \varepsilon_t.$$

ARIMA models

ARIMA(p,d,q) model (key concept)

The **ARIMA(p,d,q)** model is

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) (1 - B)^d Y_t = c + \left(1 + \sum_{i=1}^{p} \theta_i B^i\right) \varepsilon_t,$$

p : autoregressive order.

d : degree of first differencing (nearly always $d = 0$ or $d = 1$).

q : moving average order.

ARIMA(p,d,q) model

ARIMA(p,d,q) model:

$$\underbrace{\left(1-\sum_{i=1}^{p}\phi_i B^i\right)}_{\text{AR }(p)\text{ component}} \underbrace{(1-B)^d}_{\text{Differencing}} \; Y_t = c + \underbrace{\left(1+\sum_{i=1}^{p}\theta_i B^i\right)}_{\text{MA}(q)\text{ component}} \varepsilon_t.$$

The ARIMA(p,d,q) model specifies a stationary ARMA(p,q) model for the differenced series.

Example: ARIMA(0,1,1) model

The ARIMA(0,1,1) model is an MA(1) model for the first differenced series,

$$Y_t - Y_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

In backshift notation,

$$(1 - B)Y_t = (1 + \theta_1 B)\varepsilon_t.$$

With an intercept:

$$Y_t - Y_{t-1} = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

ARIMA(0,1,1): relation to exponential smoothing

$$\text{ARIMA(0,1,1): } Y_t = Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$\begin{aligned} E(Y_t|y_{1:t-1}) &= y_{t-1} + \theta_1 \varepsilon_{t-1} \\ &= y_{t-1} + \theta_1 (y_{t-1} - y_{t-2} - \theta_1 \varepsilon_{t-2}) \\ &= (1+\theta_1) y_{t-1} - \theta_1 (y_{t-2} + \theta_1 \varepsilon_{t-2}) \end{aligned}$$

Now, label $\ell_{t-1} = y_{t-1} + \theta_1 \varepsilon_{t-1}$ and $\alpha = (1+\theta_1)$. We get:

$$\ell_{t-1} = \alpha y_{t-1} + (1-\alpha)\ell_{t-2}$$

The simple exponential smoothing model.

Intercept in a first differenced series

The inclusion of an intercept induces a linear trend in an ARIMA(p,1,q) model.

For example, in the **random walk plus drift** model

$$Y_t = c + Y_{t-1} + \varepsilon_t,$$

we can derive

$$Y_{t+h} = Y_t + \sum_{i=1}^{h} (c + \varepsilon_{t+i}),$$

$$\widehat{y}_{t+h} = y_t + c \times h,$$

$$\mathrm{Var}(Y_{t+h}|y_{1:t}) = h\sigma^2.$$

ARIMA modelling

- Estimation: maximum likelihood.

- Order selection (p, q): visual identification, AIC, and model validation.

- Intercept terms induce permanent trends. Use model selection.

Seasonal ARIMA models

Seasonal ARIMA: ACF and PACF identification (key concept)

We refer to a seasonal ARIMA model as

$$\text{ARIMA} \underbrace{(p, d, q)}_{\text{Non-seasonal}} \underbrace{(P, D, Q)_m}_{\text{Seasonal}},$$

where D is the order of seasonal differencing, P and Q are the orders of the seasonal AR and MA components, and m is the number of seasons.

Seasonal ARIMA: ACF and PACF identification (key concept)

ARIMA(0,0,0)(P,0,0)

- Sample autocorrelations decrease exponentially for lags m, $2m$, $3m$, etc.
- Sample partial autocorrelations cuts off at lag Pm.

ARIMA(0,0,0)(0,0,Q)

- Sample autocorrelations cuts off at lag Qm.
- Sample partial autocorrelations decrease exponentially for lags m, $2m$, $3m$, etc.

Seasonal AR(1) or ARIMA$(0,0,0)(1,1,0)_{12}$:

$$Y_t - Y_{t-12} = c + \phi_1(Y_{t-12} - Y_{t-24}) + \varepsilon_t$$

Seasonal MA(1) or ARIMA$(0,0,0)(0,1,1)_{12}$:

$$Y_t - Y_{t-12} = c + \theta_1 \varepsilon_{t-12} + \varepsilon_t$$

Seasonal ARIMA models

ARIMA(1,0,0)(0,1,1)$_{12}$ model:

$$(1-\phi_1 B)(1-B^{12})Y_t = c + (1+\theta_1 B^{12})\varepsilon_t$$

$$Y_t - Y_{12} = c + \phi_1(Y_{t-1} - Y_{13}) + \varepsilon_t + \theta_1 \varepsilon_{t-12}$$

Seasonal ARIMA models

ARIMA(1,1,1)(1,1,0)$_{12}$ model:

$$(1-\phi_1 B)(1-\phi_2 B^{12})(1-B)(1-B^{12})Y_t = c + (1+\theta_1 B)\varepsilon_t$$

$$\begin{aligned}(Y_t - Y_{t-1}) - (Y_{t-12} - Y_{t-13}) = c &+ \phi_1\left[(Y_{t-1} - Y_{t-2}) - (Y_{t-13} - Y_{t-14})\right] \\ &+ \phi_2\left[(Y_{t-12} - Y_{t-13}) - (Y_{t-24} - Y_{t-25})\right] \\ &+ \phi_1\phi_2\left[(Y_{t-13} - Y_{t-14}) - (Y_{t-25} - Y_{t-26})\right] \\ &+ \varepsilon_t + \theta_1\varepsilon_{t-1}\end{aligned}$$

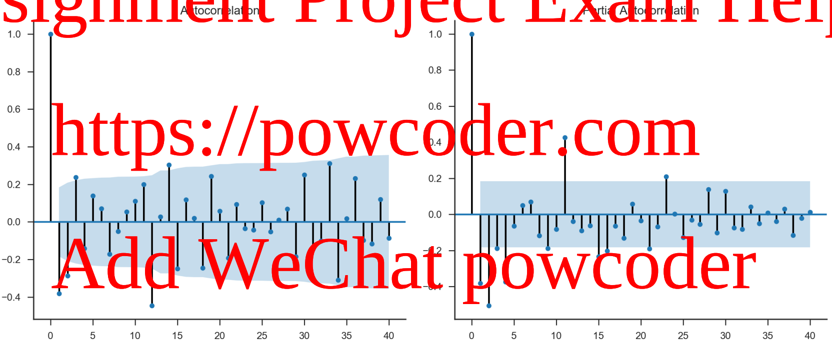
- Estimation: maximum likelihood.

- Order selection (p, q, P, Q): visual identification, AIC, and model validation.

- Usually only one seasonal AR or MA term is needed.

Example: Visitor Arrivals in Australia

Recall that we obtained the following ACF and PACF plots the first and seasonally differenced log series:

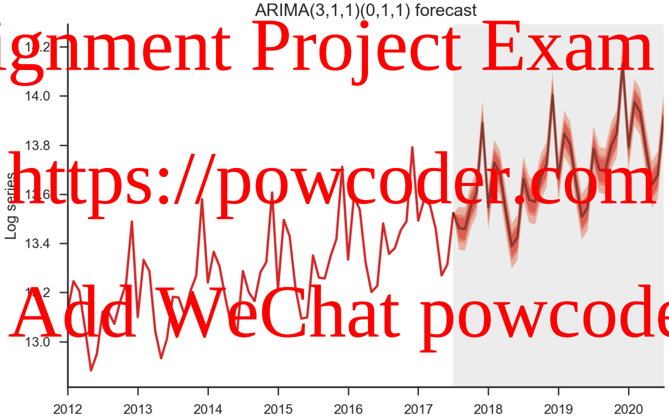


We select an ARIMA(3,1,0)(0,1,1)$_{1}2$ specification based on the AIC.

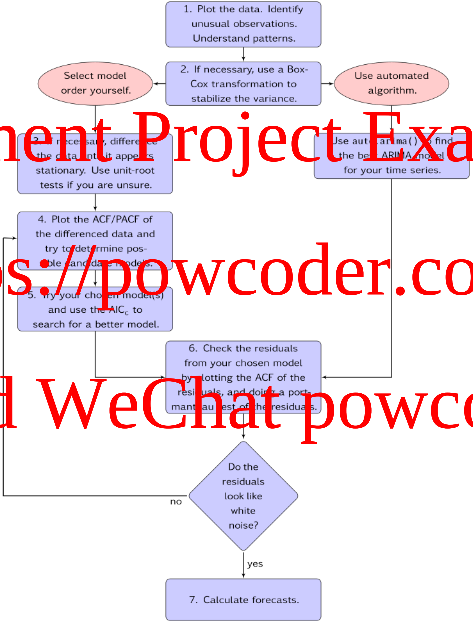
Example: Visitor Arrivals in Australia

```
                                 Statespace Model Results
==========================================================================================
Dep. Variable:                             Arrivals   No. Observations:                127
Model:             SARIMAX(3, 1, 1)x(0, 1, 1, 12)   Log Likelihood                 212.670
Date:                                [illegible]   AIC                           -411.[illegible]39
Time:                                [illegible]   BIC                           -391.[illegible]3[illegible]
Sample:                                01-31-2007   HQIC                          -403.250
                                     - 07-31-2017
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0007      0.000      2.957      0.003       0.000       0.001
ar.L1          0.0532      0.118      0.450      0.652      -0.178       0.285
ar.L2         -0.0454      0.112     -0.403      0.687      -0.266       0.175
ar.L3          0.2426      0.112      2.166      0.030       0.023       0.462
ma.L1         -0.9726      0.166     -5.873      0.000      -1.297      -0.648
ma.S.L12      -0.9976      7.7[illegible]8     -0.[illegible]28      0.89[illegible]     -16.2[illegible]2      14.247
sigma2         0.0010      0.008      0.12[illegible]      0.89[illegible]      -0.015       0.017
===================================================================================
Ljung-Box (Q):                       65.43   Jarque-Bera (JB):                  0.72
Prob(Q):                              0.01   Prob(JB):                          0.70
Heteroskedasticity (H):               0.49   Skew:                              0.10
Prob(H) (two-sided):                  0.03   Kurtosis:                          2.66
===================================================================================
```

Example: Visitor Arrivals in Australia



Summary of modelling process (FPP)



Review questions

- What is stationarity and why is it a fundamental concept in ARIMA modelling?
- What transformation do we apply to a time series to make it stationary?
- How do we identify AR vs MA processes from ACF and PACF plots?
- What is an ARIMA model?
- Write the equation for a seasonal ARIMA model using backshift notation.