QBUS6840 Lecture 8

ARIMA models (II)

Discipline of Business Analytics

The University of Sydney Business School

Last week: ACF and PACF for non-seasonal time series

- if the ACF either cuts off fairly quickly or dies down fairly quickly, then the time series should be considered stationary
- if the ACF dies down extremely slowly, then it should be considered nonstationary

Last week: Autoregressive, $AR(p)$ Processes

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

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

Data characteristics

- The ACF dies down
- The PACF has spikes at lags $1, 2, ..., p$ and cuts off after lag p

Table of contents

Outline

Moving average (MA) processes
- MA(q) process
- Backshift operators
- Invertibility

$ARMA(p, q)$ and $ARIMA(p, d, q)$ processes
- $ARMA(p, q)$ processes
- $ARIMA(p, d, q)$ processes

Moving average MA(q) processes

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

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

- a weighted moving average of the past few forecast errors.
- appropriate to model quantities y_t, such as economic indicators, which are affected by random events that have both immediate and persistent effect on y_t
- sometimes, the ε_t are called random shocks: shocks caused by unpredictable events

See example `Lecture08_Example01.py`

MA(1) process

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

Unconditional mean:

$$\mathbb{E}[Y_t] = \mathbb{E}[c + \varepsilon_t + \theta_1 \varepsilon_{t-1}] = c + 0 + \theta_1 \times 0 = c$$

Unconditional variance:

$$\begin{aligned}\mathbb{V}(Y_t) &= \mathbb{V}(c) + \mathbb{V}(\varepsilon_t) + \mathbb{V}(\theta_1 \varepsilon_{t-1}) \\ &= 0 + \sigma^2 + \sigma^2 \theta_1^2 = \sigma^2(1 + \theta_1^2)\end{aligned}$$

$MA(1)$ process: Properties

Covariance:

$$\begin{aligned}
\text{Cov}(Y_t, Y_{t-1}) =& \text{Cov}(c + \varepsilon_t + \theta_1 \varepsilon_{t-1}, c + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2}) \\
=& \text{Cov}(c, c) + \text{Cov}(c, \varepsilon_{t-1}) + \text{Cov}(c, \theta_1 \varepsilon_{t-2}) + \text{Cov}(\varepsilon_t, c) \\
& + \text{Cov}(\varepsilon_t, \varepsilon_{t-1}) + \text{Cov}(\varepsilon_t, \theta_1 \varepsilon_{t-2}) + \text{Cov}(\theta_1 \varepsilon_{t-1}, c) \\
& + \text{Cov}(\theta_1 \varepsilon_{t-1}, \varepsilon_{t-1}) + \text{Cov}(\theta_1 \varepsilon_{t-1}, \theta_1 \varepsilon_{t-2}) \\
=& \theta_1 \text{Cov}(\varepsilon_{t-1}, \varepsilon_{t-1}) = \theta_1 \mathbb{V}(\varepsilon_{t-1}) = \theta_1 \sigma^2
\end{aligned}$$

Autocorrelation:

$$\rho_1 := \frac{\text{Cov}(Y_t, Y_{t-1})}{\mathbb{V}(Y_t)} = \frac{\theta_1 \sigma^2}{\mathbb{V}(Y_t)} = \frac{\theta_1}{1 + \theta_1^2}$$

$MA(1)$ process: Properties

$$\mathrm{Cov}(Y_t, Y_{t-2}) = 0(\text{Why?}),$$

$$\rho_2 = 0.$$

We have

$$\rho_k = 0 \qquad \text{for } k > 1.$$

Conclusion: MA(1) process is stationary for every θ_1, and its ACF plot cuts off after lag 1

Partial ACF:

$$\rho_{kk} = -\frac{\theta_1^k(1-\theta_1^2)}{1-\theta_1^{2(k+1)}}, \quad k \geq 1.$$

Partial ACF plot dies down exponentially when $|\theta_1| < 1$.

$MA(1)$ process: Forecasting

$$\mathbb{E}(Y_{t+1}|y_{1:t}) = c + \theta_1 \varepsilon_t$$

- We use the forecast errors $\widehat{\epsilon}_1, ..., \widehat{\epsilon}_t$ from the previous periods to construct the next forecast at time $t+1$
- Let the forecast at time t is $\widehat{y}_t$, and forecast error

$$\widehat{\varepsilon}_t = y_t - \widehat{y}_t = y_t - (c + \theta_1 \widehat{\varepsilon}_{t-1})$$

- Forecast of Y_{t+1} is

$$\widehat{y}_{t+1} = c + \theta_1 \widehat{\varepsilon}_t$$

  and forecast error

$$\widehat{\varepsilon}_{t+1} = y_{t+1} - \widehat{y}_{t+1} = y_{t+1} - (c + \theta_1 \widehat{\varepsilon}_t).$$

The variance of the forecast is

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

$MA(1)$ process: Forecasting

$$\mathbb{E}(Y_{t+2}|y_{1:t}) = c + \mathbb{E}(\varepsilon_{t+2}|y_{1:t}) + \theta_1 \mathbb{E}(\varepsilon_{t+1}|y_{1:t}) = c,$$

so

$$\widehat{y}_{t+2|t} = c.$$

$$\mathbb{V}(Y_{t+2}|y_{1:t}) = \sigma^2(1 + \theta_1^2)$$

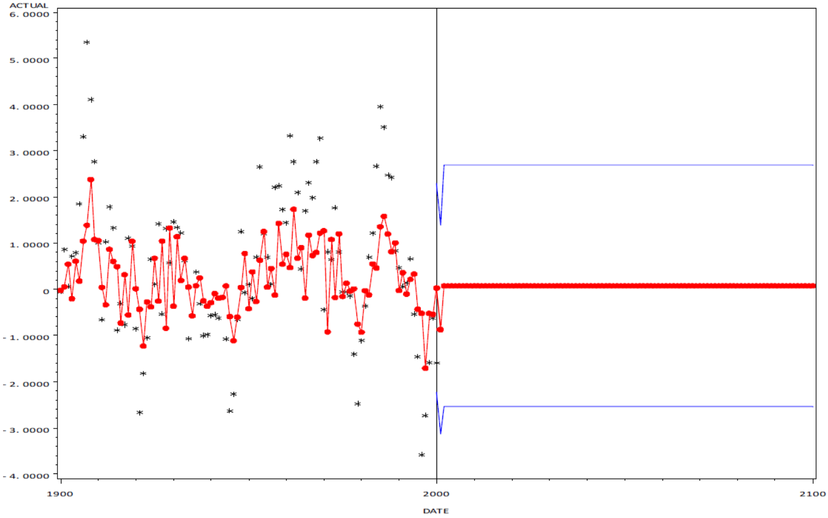
In general, it's easy to see that

$$\widehat{y}_{t+h|t} = c \qquad \text{for } h > 1$$

and

$$\mathbb{V}(Y_{t+h}|y_{1:t}) = \sigma^2(1 + \theta_1^2) \qquad \text{for } h > 1$$

$MA(1)$ process: Forecasting



$MA(q)$ processes: Properties

Consider the unconditional variance:

$$
\begin{aligned}
\mathbb{V}(Y_t) =& \mathbb{V}(c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}) \\
=& \sigma^2 (1 + \theta_1^2 + \cdots + \theta_q^2).
\end{aligned}
$$

$MA(q)$ processes: Properties

Consider the unconditional variance:

$$\begin{aligned}\mathbb{V}(Y_t) =& \mathbb{V}(c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}) \\ =& \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2).\end{aligned}$$

Covariance at lag 1:

$$\begin{aligned}\mathrm{Cov}(Y_t, Y_{t-1}) =& \mathrm{Cov}(c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}, \\ & c + \varepsilon_{t-1} + \theta_1\varepsilon_{t-2} + \theta_2\varepsilon_{t-3} + \ldots + \theta_q\varepsilon_{t-q-1}) \\ =& \sigma^2(\theta_1 + \theta_1\theta_2 + \theta_2\theta_3 + \ldots + \theta_{q-1}\theta_q).\end{aligned}$$

Hence

$$\rho_1 = \frac{\theta_1 + \theta_1\theta_2 + \theta_2\theta_3 + \ldots + \theta_{q-1}\theta_q}{1 + \theta_1^2 + \ldots + \theta_q^2}$$

$MA(q)$ processes: Properties

Covariance at lag q:

$$\begin{aligned}\mathrm{Cov}(Y_t, Y_{t-q}) =& \mathrm{Cov}(c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q},\\ & c + \varepsilon_{t-q} + \theta_1\varepsilon_{t-q-1} + \theta_2\varepsilon_{t-q-2} + \ldots + \theta_q\varepsilon_{t-2q})\\ =& \sigma^2\theta_q.\end{aligned}$$

Hence

$$\rho_q = \frac{\theta_q}{1 + \theta_1^2 + \ldots + \theta_q^2}$$

We can also find that (take-home exercise):

$$\rho_k = 0 \qquad \text{for } k > q$$

Question: What about ρ_k if $2 \leq k < q$?

$MA(q)$ processes: Properties

Covariance at lag q:

$$\begin{aligned}\text{Cov}(Y_t, Y_{t-q}) =& \text{Cov}(c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}, \\ & c + \varepsilon_{t-q} + \theta_1\varepsilon_{t-q-1} + \theta_2\varepsilon_{t-q-2} + \ldots + \theta_q\varepsilon_{t-2q}) \\ =& \sigma^2\theta_q.\end{aligned}$$

Hence

$$\rho_q = \frac{\theta_q}{1 + \theta_1^2 + \ldots + \theta_q^2}$$

We can also find that (take-home exercise):

$$\rho_k = 0 \qquad \text{for } k > q$$

Question: What about ρ_k if $2 \leq k < q$?

$$\rho_k = \frac{\theta_k + \theta_{k+1}\theta_1 + \cdots + \theta_q\theta_{q-k}}{1 + \theta_1^2 + \ldots + \theta_q^2}$$

Take-home exercise: derive this!

$MA(q)$ processes: Properties

- ρ_k (ACF) cuts off after lag q.
- ρ_{kk} (PACF) dies down exponentially.

$MA(q)$ processes: Forecasting

$$\begin{aligned}\widehat{y}_{t+h|t} &= \mathbb{E}(Y_{t+h}|y_{1:t}) \\ &= c + \theta_1\mathbb{E}(\varepsilon_{t+h-1}|y_{1:t}) + \ldots + \theta_q\mathbb{E}(\varepsilon_{t+h-q}|y_{1:t}),\end{aligned}$$

where

$$E(\varepsilon_{t+h-i}|y_{1:t}) = \begin{cases} 0 & \text{if } h > i \\ \hat{\varepsilon}_{t+h-i} & \text{if } h \leq i \end{cases}$$

$$\mathbb{V}(Y_{t+h}|y_{1:t}) = \sigma^2\left(1 + \sum_{i=1}^{\min(q,h-1)} \theta_i^2\right)$$

Example: $MA(3)$ Forecasting

$$E(Y_{t+h}|y_{1:t}) = c + \theta_1 E(\varepsilon_{t+h-1}|y_{1:t}) + \theta_2 E(\varepsilon_{t+h-2}|y_{1:t}) + \theta_3 E(\varepsilon_{t+h-3}|y_1$$

Hence, given previous forecast errors $\widehat{\epsilon}_1, ..., \widehat{\epsilon}_t$,

$$\begin{aligned}
\widehat{y}_{t+1} =& c + \theta_1 E(\varepsilon_t|y_{1:t}) + \theta_2 E(\varepsilon_{t-1}|y_{1:t}) + \theta_3 E(\varepsilon_{t-2}|y_{1:t}) \\
=& c + \theta_1 \widehat{\varepsilon}_t + \theta_2 \widehat{\varepsilon}_{t-1} + \theta_3 \widehat{\varepsilon}_{t-2} \\
\widehat{y}_{t+2} =& c + \theta_1 E(\varepsilon_{t+1}|y_{1:t}) + \theta_2 E(\varepsilon_t|y_{1:t}) + \theta_3 E(\varepsilon_{t-1}|y_{1:t}) \\
=& c + \theta_1 \times 0 + \theta_2 \widehat{\varepsilon}_t + \theta_3 \widehat{\varepsilon}_{t-1} = c + \theta_2 \widehat{\varepsilon}_t + \theta_3 \widehat{\varepsilon}_{t-1} \\
\widehat{y}_{t+3} =& c + \theta_1 E(\varepsilon_{t+2}|y_{1:t}) + \theta_2 E(\varepsilon_{t+1}|y_{1:t}) + \theta_3 E(\varepsilon_t|y_{1:t}) \\
=& c + \theta_1 \times 0 + \theta_2 \times 0 + \theta_3 \widehat{\varepsilon}_t = c + \theta_3 \widehat{\varepsilon}_t \\
\widehat{y}_{t+3} =& c
\end{aligned}$$

Outline

Moving average (MA) processes
- MA(q) process
- Backshift operators
- Invertibility

$ARMA(p, q)$ and $ARIMA(p, d, q)$ processes
- $ARMA(p, q)$ processes
- $ARIMA(p, d, q)$ processes

Backshift operators

We now introduce the Backshift operator, which is very useful for describing time series models

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

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

$$B^k Y_t = Y_{t-k}$$

Particularly for a constant series $\{d\}$, we define

$$Bd = d$$

Backshift operators

In context: $AR(1)$

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

We derive in the last lecture, $\mu = E(Y_t) = E(Y_{t-1}) = c/(1-\phi_1)$

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

$$(1 - \phi_1 B)(Y_t - \mu) = \varepsilon_t$$

which comes from the fact $c = (1-\phi_1)\mu = (1-\phi_1 B)\mu$, which is from $Bd = d$ for any constant d.
Denote $Z_t = Y_t - \mu$, then

$$(1 - \phi_1 B) Z_t = \varepsilon_t \Longrightarrow Z_t = \phi_1 Z_{t-1} + \varepsilon_t$$

Backshift operators

In context: $MA(1)$

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

which gives $\mu = E(Y_t) = c$.

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

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

Backshift operators

In context: $MA(1)$

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

which gives $\mu = E(Y_t) = c$.

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

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

Denote $Z_t = Y_t - \mu$, then

$$Z_t = (1 + \theta_1 B)\varepsilon_t \Longrightarrow Z_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Backshift operators

In context: $AR(p)$

$$\begin{aligned} Y_t &= c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \\ &= c + \phi_1 B(Y_t) + \ldots + \phi_p B^p(Y_t) + \varepsilon_t \end{aligned}$$

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)(Y_t - \mu) = \varepsilon_t$$

where $\mu = c/(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$,

$$(1 - \sum_{i=1}^{p} \phi_i B^i)(Y_t - \mu) = \varepsilon_t$$

Backshift operators

In context: $MA(q)$

$$
\begin{aligned}
Y_t &= c + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \\
&= c + \theta_1 B(\varepsilon_t) + \ldots + \theta_q B^q(\varepsilon_t) + \varepsilon_t
\end{aligned}
$$

$$(Y_t - \mu) = (1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q)\varepsilon_t$$

$$(Y_t - \mu) = (1 + \sum_{i=1}^{q} \theta_i B^i)\varepsilon_t$$

Outline

Moving average (MA) processes

- MA(q) process
- Backshift operators
- Invertibility

$ARMA(p, q)$ and $ARIMA(p, d, q)$ processes

- $ARMA(p, q)$ processes
- $ARIMA(p, d, q)$ processes

Invertibility

Definition

An $MA(q)$ process is invertible when we can rewrite it as a linear combination of its past values (an $AR(\infty)$) plus the contemporaneous error term.

Invertibility: $MA(1)$

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

Note: For MA processes $c = \mu$

$$(Y_t - \mu) = (1 + \theta_1 B)\varepsilon_t \Rightarrow \varepsilon_t = \frac{Y_t - \mu}{(1 + \theta_1 B)}$$

[Note $\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots$ for $|x| < 1$] Under the condition $|\theta_1| < 1$, we have

$$\varepsilon_t = (1 - \theta_1 B + \theta_1^2 B^2 - \theta_1^3 B^3 + \ldots)(Y_t - \mu)$$

$$\varepsilon_t = -\mu(1 - \theta_1 + \theta_1^2 - \theta_1^3 + \ldots) + Y_t - \theta_1 B Y_t + \theta_1^2 B^2 Y_t - \cdots$$

$$\therefore \quad Y_t = c^* - \sum_{i=1}^{\infty} (-1)^i \theta_1^i Y_{t-i} + \varepsilon_t$$

Invertibility: $MA(1)$ (alternative route)

The MA(1) gives

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

hence

$$\begin{aligned}
Y_t &= c + \theta_1 \varepsilon_{t-1} + \varepsilon_t = c + \theta_1 (y_{t-1} - c - \theta_1 \varepsilon_{t-2}) + \varepsilon_t \\
&= c(1-\theta_1) + \theta_1 y_{t-1} - \theta_1^2 \varepsilon_{t-2} + \varepsilon_t \\
&= c(1-\theta_1 + \theta_1^2) + \theta_1 y_{t-1} - \theta_1^2 y_{t-2} + \theta_1^3 \varepsilon_{t-3} + \varepsilon_t \\
&\quad \vdots \\
&= c(1-\theta_1+\theta_1^2-\theta_1^3+\ldots) - \sum_{i=1}^{\infty} (-1)^i \theta_1^i Y_{t-i} + \varepsilon_t
\end{aligned}$$

$$\therefore \quad Y_t = c^* - \sum_{i=1}^{\infty} (-1)^i \theta_1^i Y_{t-i} + \varepsilon_t \text{ or}$$

$$\varepsilon_t = Y_t - c^* + \sum_{i=1}^{\infty} (-1)^i \theta_1^i Y_{t-i}$$

Invertibility: Why it matters

- If we want to find the value ε_t at a certain period and the process is invertible, we need to know the current and past values of Y. For a noninvertible representation we would need to use all future values of Y!

- Convenient algorithms for estimating parameters and forecasting are only valid if we use an invertible representation.

Notes

- Every invertible $MA(q)$ model can be written as an AR model of infinite order.
- Every stationary $AR(p)$ model can be written as an MA model of infinite order.

Example: $AR(1)$ as $MA(\infty)$

$$
\begin{aligned}
Y_t &= c + \phi_1 Y_{t-1} + \varepsilon_t \\
&= c(1+\phi_1) + \phi_1^2 Y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&= c(1+\phi_1+\phi_1^2) + \phi_1^2 Y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&\quad \vdots \\
&= c(1+\phi_1+\ldots+\phi_1^{t-1}) + \phi_1^t y_0 + \sum_{i=1}^{t-1} \phi_1^i \varepsilon_{t-i} + \varepsilon_t
\end{aligned}
$$

$$
Y_t = \frac{c}{1-\phi_1} + \sum_{i=1}^{\infty} \phi_1^i \varepsilon_{t-i} + \varepsilon_t
$$

Outline

Moving average (MA) processes

- MA(q) process
- Backshift operators
- Invertibility

$ARMA(p, q)$ and $ARIMA(p, d, q)$ processes

- $ARMA(p, q)$ processes
- $ARIMA(p, d, q)$ processes

$ARMA(p, q)$ processes

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

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

Example: $ARMA(0, 0)$ [constant + white noise]

$$Y_t = c + \varepsilon_t,$$

Example: $ARMA(1, 1)$

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

$ARMA(p, q)$ processes: Properties

$$E(Y_t) = \frac{c}{1 - \phi_1 - \ldots - \phi_p}$$

- ρ_k dies down.
- ρ_{kk} dies down.
- See Examples `Lecture08_Example02.py`

$ARMA(1,1)$: Forecasting

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

$$\widehat{y}_{t+1} = E(Y_{t+1} | y_1, \ldots, y_t) = c + \phi_1 y_t + \theta_1 \varepsilon_t$$

$$\mathrm{Var}(Y_{t+1} | y_1, \ldots, y_t) = \sigma^2.$$

$ARMA(1,1)$: Forecasting

$$
\begin{aligned}
Y_{t+2} &= c + \phi_1 Y_{t+1} + \theta_1 \varepsilon_{t+1} + \varepsilon_{t+2} \\
&= c + \phi_1 (c + \phi_1 Y_t + \theta_1 \varepsilon_t + \varepsilon_{t+1}) + \theta_1 \varepsilon_{t+1} + \varepsilon_{t+2} \\
&= c(1 + \phi_1) + \phi_1^2 Y_t + \phi_1 \theta_1 \varepsilon_t + (\phi_1 + \theta_1) \varepsilon_{t+1} + \varepsilon_{t+2}
\end{aligned}
$$

$$
\widehat{y}_{t+2} = c(1 + \phi_1) + \phi_1^2 y_t + \phi_1 \theta_1 \varepsilon_t
$$

$$
\mathrm{Var}(Y_{t+2} | y_1, \ldots, y_t) = \sigma^2 (1 + (\phi_1 + \theta_1)^2).
$$

Stationary transforms

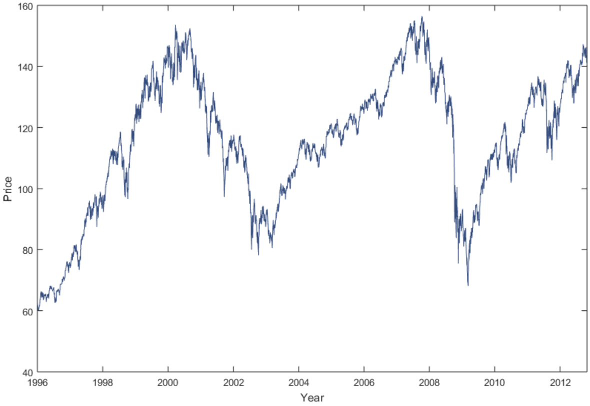
Box and Jenkins advocate difference transforms to achieve stationarity, e.g

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

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

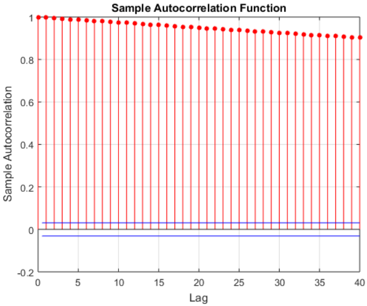
Stationary transforms

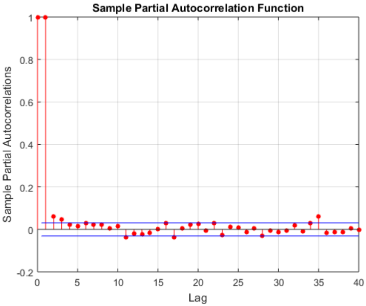
Example: S&P 500 index



Stationary transforms

Example: S&P 500 index

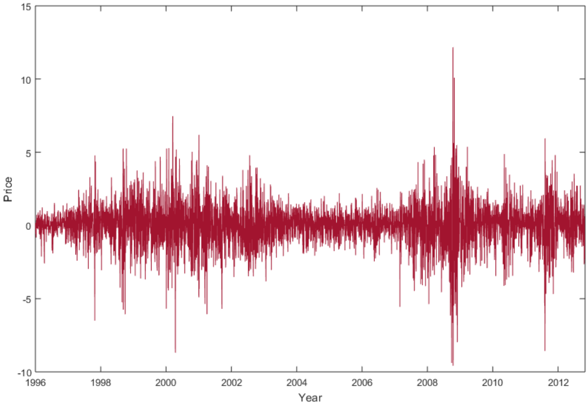




Stationary transforms

Example: S&P 500 index

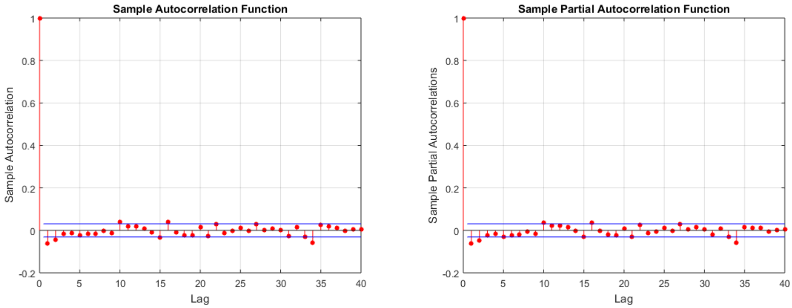
Taking the first difference:



Stationary transforms

Example: S&P 500 index

Autocorrelations for the differenced series:



Outline

Moving average (MA) processes

- MA(q) process
- Backshift operators
- Invertibility

$ARMA(p, q)$ and $ARIMA(p, d, q)$ processes

- $ARMA(p, q)$ processes
- $ARIMA(p, d, q)$ processes

Autoregressive Integrated Moving Average Models: ARIMA(p, d, q)

- Suppose we consider the d-order difference of the original time series $\{Y_t\}$. Denote $Z_t = \Delta^d Y_t$
- An ARMA(p, q) model on $\{Z_t\}$ is called an ARIMA(p, d, q) model on $\{Y_t\}$
- Examples `Lecture08_Example03.py`

ARIMA(0, 1, 0)

Random walk plus drift model

$$Z_t = c + \varepsilon_t \text{ or } Y_t = c + Y_{t-1} + \varepsilon_t$$

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

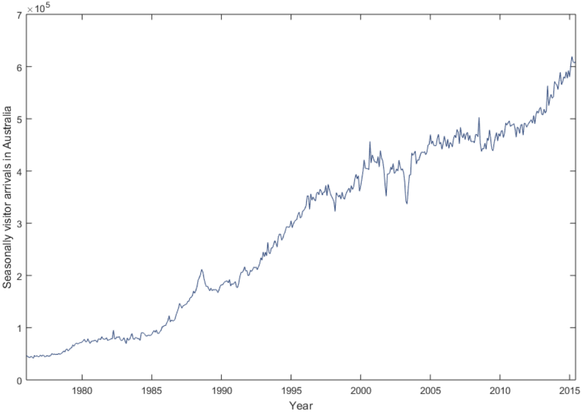
$$\widehat{y}_{t+h} = y_t + c \times h$$

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

It is the formal statistical model for the drift forecasting method mentioned early in the course.

Seasonally adjusted visitor arrivals in Australia

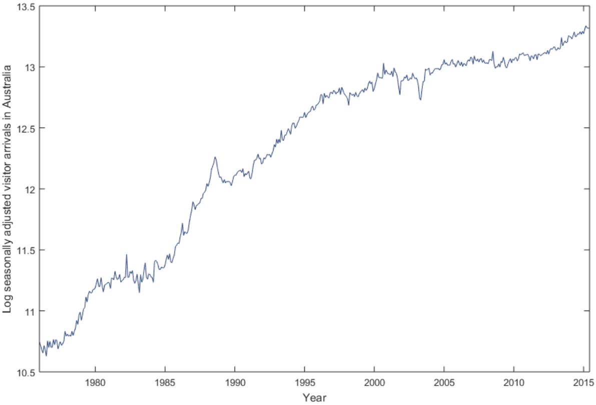
Example of modelling process



Seasonally adjusted visitor arrivals in Australia

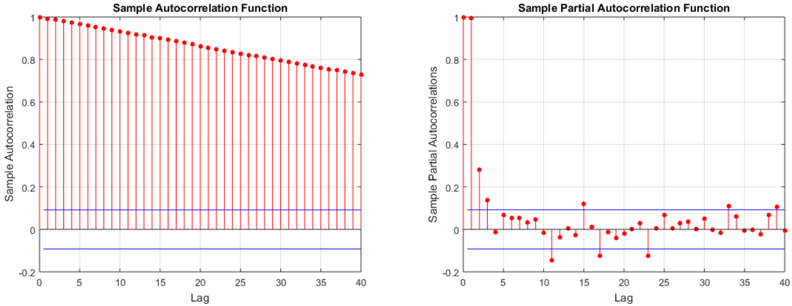
Variance stabilising transform

We first take the log of the series as a variance stabilising transformation:



Log seasonally adjusted visitor arrivals in Australia

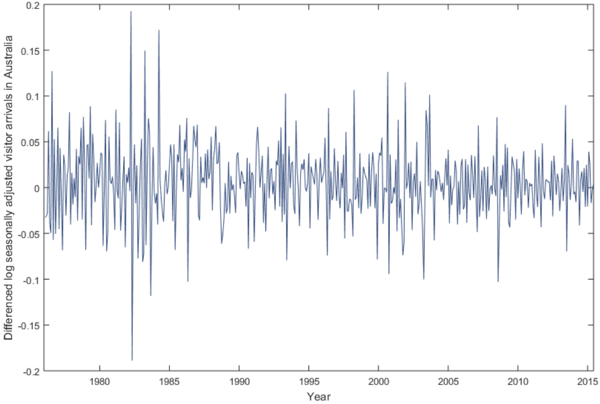
ACF and PACF for the log series



Log seasonally adjusted visitor arrivals in Australia

Stationary transform

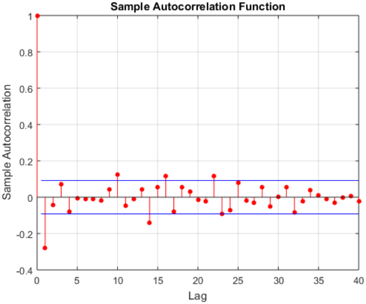
We then take the first difference:

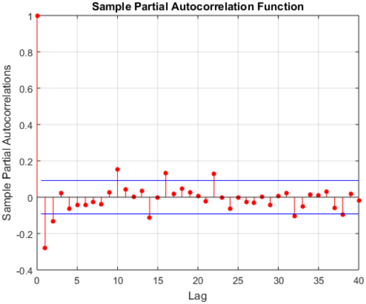


Log seasonally adjusted visitor arrivals in Australia

Differenced series

Autocorrelations for the differenced series:





Log seasonally adjusted visitor arrivals in Australia

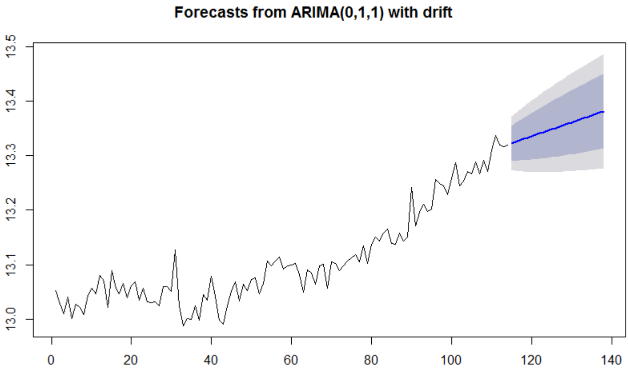
Tentative model identification

- The ACF of the differenced series cuts off after lag one.
- The PACF seems to die down.
- This suggests that the differenced series may be an MA(1) process.
- The original log series would then be an $ARIMA(0,1,1)$ process.

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

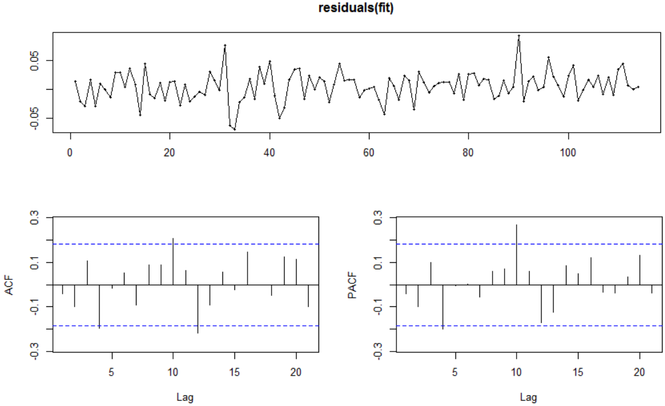
Log seasonally adjusted visitor arrivals in Australia

Forecasting



Log seasonally adjusted visitor arrivals in Australia

Residual analysis



ARIMA(0, 1, 1) model

Reinterpreting the model: SES is ARIMA(0,1,1)

Consider the ARIMA(0, 1, 1) model with the intercept $c = 0$:

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

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

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

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

This is the simple exponential smoothing model.

$ARMA(p, q)$ processes: Formulation with backshift operators

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

$ARIMA(p, d, q)$ processes: Formulation with backshift operators

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

Procedure to Estimate $ARMA(p, q)/ARIMA(p, d, q)$ processes: `Lecture08_Example04.py`

1. For the given time series $\{Y_t\}$, check its stationarity by looking at its Sample ACF and Sample PACF.
2. If ACF does not die down quickly, which means the given time series $\{Y_t\}$ is nonstationary, we seek for a transformation, e.g., log transformation $\{Z_t = log(Y_t)\}$, or the first order difference $\{Z_t = Y_t - Y_{t-1}\}$, or even the difference of log time series, or the difference of the first order difference, so that the transformed time series is stationary by checking its Sample ACF
3. When both Sample ACF and Sample PACF die down quickly, check the orders at which ACF or PACF die down to indentify tentatively the lags p and q of the ARIMA, and the order of difference will be d.
4. Estimate the identified $ARIMA(p, d, q)$, or $ARMA(p, q)$ (if we did not do any difference transformation)
5. Make forecast with estimated $ARIMA(p, d, q)$, or $ARMA(p, q)$ model

$ARIMA(p, d, q)$ processes: Order selection

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

How to choose p (the order of AR) and q (the order of MA) when the ACF and PACF do not give us a straightforward answer?

ARIMA order selection: AIC

- We define Akaike's Information Criterion as

$$\text{AIC} = -2\log(L) + 2(p + q + k + 1),$$

  where L is the likelihood of the data and $k = 1$ if the model has an intercept.

- The model with the minimum value of the AIC is often the best model for forecasting.

ARIMA order selection: Corrected AIC

- The corrected Akaike's Information Criterion is

$$\text{AICc} = AIC + \frac{2(p+q+k+1)(p+q+k+2)}{n-p-q-k-2},$$

  where n is the number of observations.

- The corrected AIC penalises extra parameters more heavily, often has better performance in small samples.

- The AICc is the foremost criterion used by researchers in selecting the orders of ARIMA models.

- The AICc is based on the assumption of normally distributed residuals.

ARIMA order selection: BIC

- A related measure is Schwarz's Bayesian Information Criterion (known as SBIC, BIC or SC):

$$\begin{aligned} \text{BIC} &= -2\log(L) + log(n)(p+q+k+1) \\ &= \text{AIC} + (\log(n)-2)(p+q+k+1). \end{aligned}$$

- As with the AIC, minimizing the BIC is intended to give the best model. The model chosen by BIC is either the same as that chosen by AIC, or one with fewer parameters. This is because BIC penalises the model complexity more heavily than the AIC.

- Under some mathematical assumptions, BIC can select the true model with enough data (if the true model exists!)

Recap

We have looked at

- Moving average processes, MA(q)
- ARMA and ARIMA processes

Next lecture: Seasonal ARIMA Models

Recap

We have looked at

- Moving average processes, MA(q)
- ARMA and ARIMA processes

Next lecture: Seasonal ARIMA Models

Thank you and see you next week!