QBUS 6840 Lecture 8

ARIMA models (II)

Professor Junbin Gao

The University of Sydney Business School

Reschedule

- Week 8: MA Process and ARIMA: ARIMA Part II and B-Operator
- Week 9: Seasonal ARIMA and Forecast Combination
- Week 10: Neural Network and Recurrent Neural Networks (For Group Project Purpose)
- Week 11: State Space Models
- Week 12: Hierarchical and Group Time Series
- Week 13: Into the Future

Review of ACF and PACF

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

Review of $AR(p)$ Processes

Data characteristics

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

Model characteristics

- For an AR(1) model $Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t$ to be statioanry:

$$-1 < \phi_1 < 1$$

- For an AR(2) model $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$ to be stationary:

$$-1 < \phi_1 < 1, \phi_1 + \phi_2 < 1, \phi_2 - \phi_1 < 1.$$

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

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

See example `Lecture08_Example01.py`

MA(1) process

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

Unconditional Expectation

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

Unconditional:

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

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

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

$MA(1)$ process
Properties

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

(Why?) hence

$$\rho_2 = 0.$$

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

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

A MA(1) is stationary for every θ_1

$MA(1)$ process
Forecasting

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

$$\begin{aligned} E(Y_{t+1}|y_{1:t}) &= \widehat{y}_{t+1} = c + \theta_1 \widehat{\varepsilon}_t, \\ \mathrm{Var}(Y_{t+1}|y_{1:t}) &= \sigma^2. \end{aligned}$$

where $\widehat{\varepsilon}_t = y_t - \widehat{y}_t$ if we know the predict $\widehat{y}_t$ at time t; otherwise we can set $\widehat{\varepsilon}_t = 0$.

$MA(1)$ process

Forecasting

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

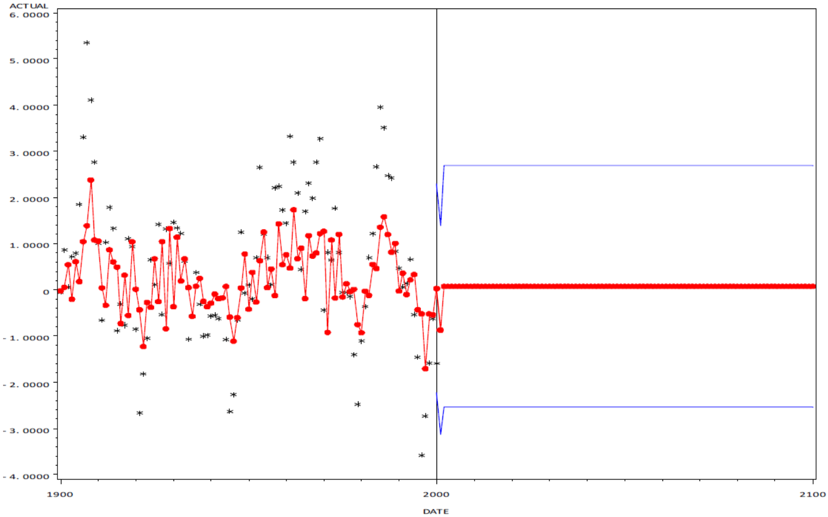
$$\text{Var}(Y_{t+2}|y_{1:t}) = \sigma^2(1+\theta_1^2)$$

$$E(Y_{t+h}|y_{1:t}) = c \qquad \text{for } h > 1$$

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

$MA(1)$ process

Forecasting



Unconditional:

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

and

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

Hence

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

Specially

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

Hence

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

And

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

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

Specially

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

Hence

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

And

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

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

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

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

where

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

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

$$\begin{aligned}\widehat{y}_{t+h} =& E(Y_{t+h}|y_{1:t})\\ =& c+\theta_1 E(\varepsilon_{t+h-1}|y_{1:t})+\theta_2 E(\varepsilon_{t+h-2}|y_{1:t})+\theta_3 E(\varepsilon_{t+h-3}|y_{1:t}),\end{aligned}$$

Hence

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

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

Backshift operators

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

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

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

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

$$Bd = d$$

Backshift operators

In context: $AR(1)$

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

where gives $\mu = E(Y_t) = E(Y_{t-1}) = c/(1 - \phi_1)$

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

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

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

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

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

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: $AR(p)$

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

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

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

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

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

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

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

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

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.

用于估计参数和预测的便捷算法仅在我们使用可逆表示时才有效。

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

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

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

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

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

The MA(1) gives

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

hence

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

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

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

We wish to find θ_1 such that

$$\min \sum_{t=2}^{T} \varepsilon_t^2$$

From previous formula, we know, given $Y_1, Y_2, ..., Y_T$,

$$\begin{aligned}
\varepsilon_2 &= Y_2 - c^* - \theta_1 Y_1 \\
\varepsilon_3 &= Y_3 - c^* - \theta_1 Y_2 + \theta_1^2 Y_1 \\
\varepsilon_4 &= Y_4 - c^* - \theta_1 Y_3 + \theta_1^2 Y_2 - \theta_1^3 Y_1 \\
&\dots \\
\varepsilon_T &= Y_T - c^* - \theta_1 Y_{T-1} + \theta_1^2 Y_{T-2} - \cdots + (-1)^{T-1} \theta_1^{T-1} Y_1
\end{aligned}$$

This results in a nonlinear least square problem. Harder than AR estimate.

Invertibility

What about $MA(1)$ in the case of $\theta_1 > 1$?

Note that for any t

$$\varepsilon_t = \frac{1}{\theta_1}(Y_{t+1} - c - \varepsilon_{t+1})$$

Hence

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

We have to use future Ys to express any ε_t, so there is no way to minimise $\sum_{t=2}^{T}\varepsilon_t^2$ for estimating θ_1.

Notes

- Every invertible $MA(q)$ model can be written as an AR model of infinite order.
- If the coefficient terms on y_{t-k} in the AR representation decline with k then the MA model is invertible. So is AR(p) invertible?
- An $MA(1)$ requires that $|\theta_1| < 1$ for invertibility.
- Every stationary $AR(p)$ model can be written as an MA model of infinite order.

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

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

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

Checking Stationarity of AR(p)

- Consider the AR(p) process

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

- Accordingly, define the characteristic equation

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

  whose roots are called the characteristic roots. There are p such roots, although some of them may be equal.

- Conclusion: The AR(p) is stationary if all the roots satisfy $|z| > 1$.
- For example, the AR(1) is $Y_t = \phi_1 Y_{t-1} + \varepsilon_t$. The characteristic equation is $1 - \phi_1 z =$ and its only root is $z^* = 1/\phi_1$. $|z^*| > 1$ implies the AR(1) stationarity. This means $|\phi_1| < 1$.

Checking Invertibility of MA(q)

- Consider the MA(q) process

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

- Accordingly, define the characteristic equation

$$1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q = 0$$

  whose roots are called the characteristic roots. There are q such roots, although some of them may be equal.

- Conclusion: The MA(q) is invertible if all the roots satisfy $|z| > 1$.

- For example, the MA(1) is $Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$. The characteristic equation is $1 + \theta_1 z =$ and its only root is $z^* = -1/\theta_1$. $|z^*| > 1$ implies the MA(1) is invertible. This means $|\theta_1| < 1$.

$ARMA(p, q)$ 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: $ARMA(0, 0)$: (White Noise)

$$Y_t = c + \varepsilon_t,$$

Example: $ARMA(1, 1)$:

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

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

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

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

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

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

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

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

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

Stationary transforms

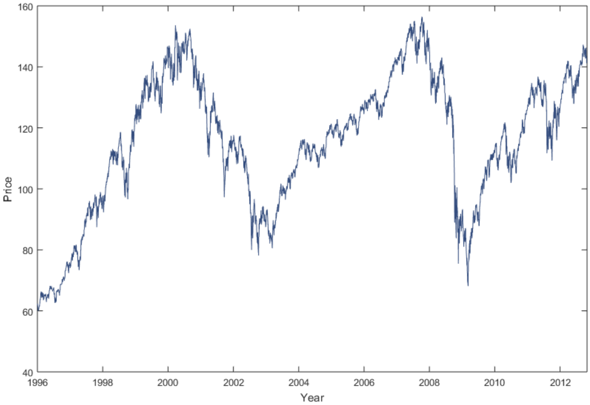
Box and Jenkins advocate difference transforms to achieve stationarity, e.g

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

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

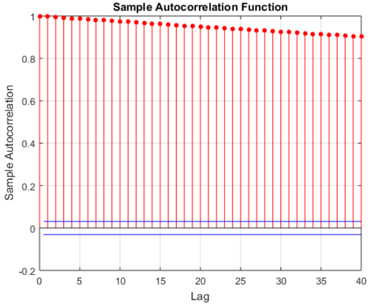
Stationary transforms

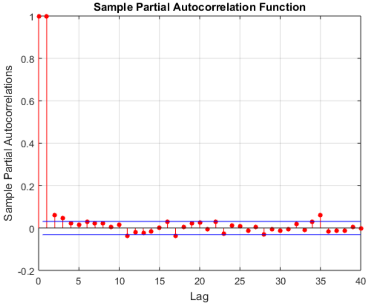
Example: S&P 500 index



Stationary transforms

Example: S&P 500 index

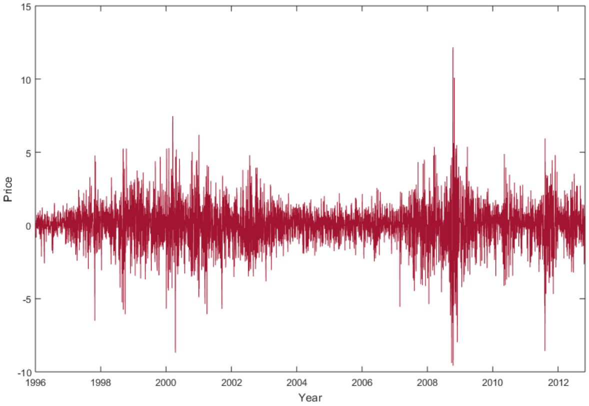




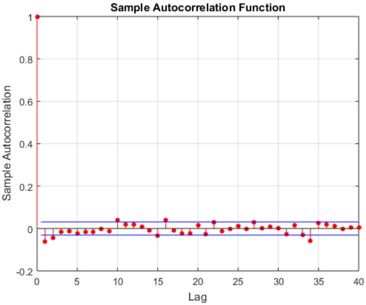
Stationary transforms

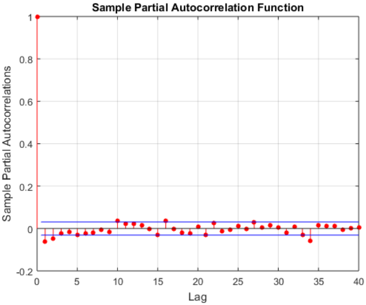
Example: S&P 500 index

Taking the first difference:



Autocorrelations for the differenced series:





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

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

ARIMA$(0, 1, 0)$ model: the random walk model

After taking the first difference, the series ΔY_t is white noise, i.e., $\Delta Y_t = \varepsilon_t$. We can therefore write:

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

The random walk model is:

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

Adding an intercept, we obtain the random walk plus drift model:

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

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

ARIMA(0, 1, 0)

Random walk model

$$
\begin{aligned}
Y_t &= Y_{t-1} + \varepsilon_t \\
&= Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\
&= Y_{t-3} + \varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
&= Y_1 + \sum_{i=2}^{t} \varepsilon_i .
\end{aligned}
$$

ARIMA(0, 1, 0)

Random walk model

Model equation: $Y_t = Y_{t-1} + \varepsilon_t$

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

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

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

ARIMA(0, 1, 0)

Random walk plus drift model

Model equation: $Y_t = c + Y_{t-1} + \varepsilon_t$

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

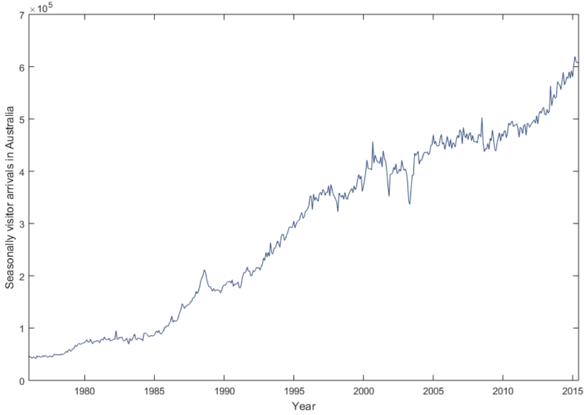
$$\widehat{y}_{t+h} = y_t + c \times h$$

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

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

Seasonally adjusted visitor arrivals in Australia

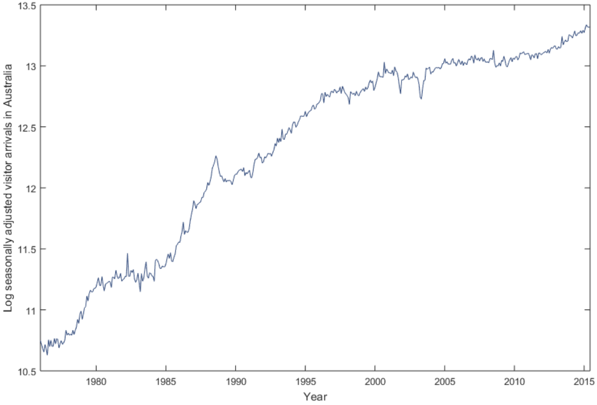
Example of modelling process



Seasonally adjusted visitor arrivals in Australia

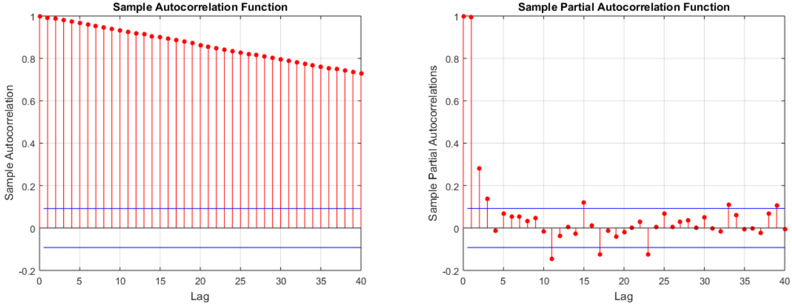
Variance stabilising transform

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

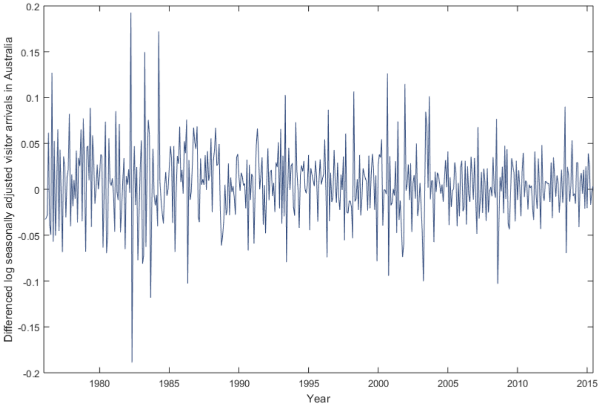


Log seasonally adjusted visitor arrivals in Australia

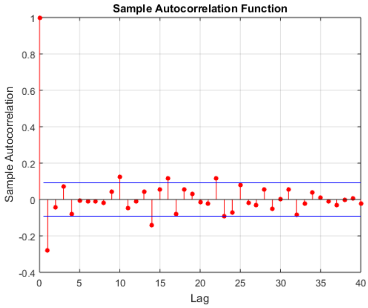
ACF and PACF for the log series

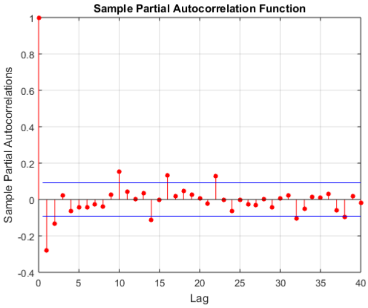


We then take the first difference:



Autocorrelations for the differenced series:





- 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.

Log seasonally adjusted visitor arrivals in Australia

ARIMA$(0,1,1)$ model

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

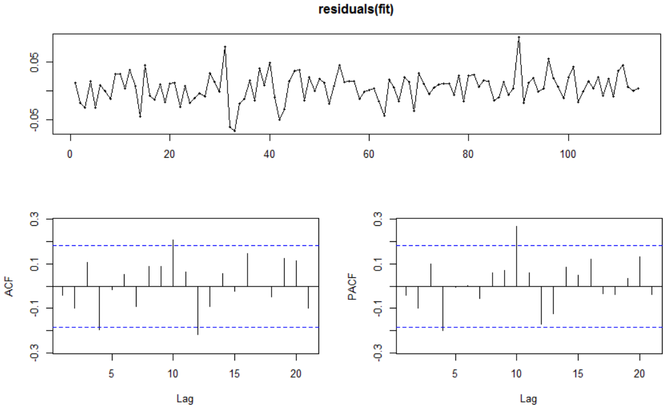
$$(1 - B)Y_t = (1 + \theta_1 B)\varepsilon_t$$

With an intercept:

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

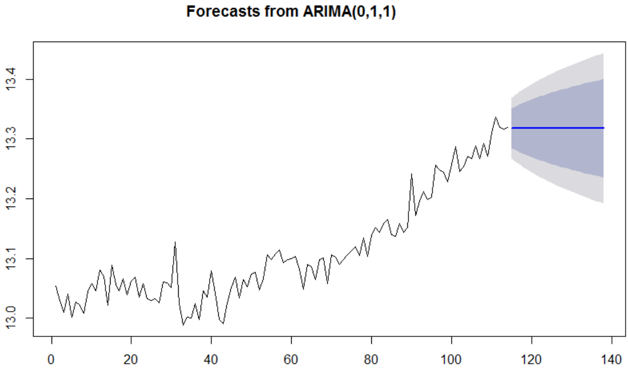
Log seasonally adjusted visitor arrivals in Australia

Residual analysis



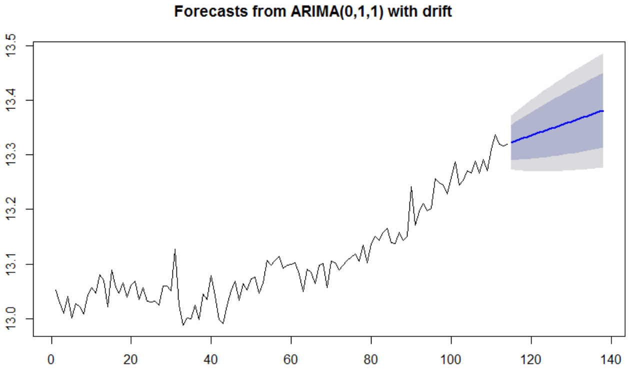
Log seasonally adjusted visitor arrivals in Australia

Forecasting



Log seasonally adjusted visitor arrivals in Australia

Forecasting by adding an intercept to the model



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

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

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.

$ARMA(p, q)$ 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. The order of ACF will be the lag q of the ARIMA and the order of PACF will be the lag p 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

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

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

- 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.

- The corrected AIC described in FPP has better performance in small samples.

ARIMA order selection

Corrected Akaike's Information Criterion

- 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 has penalises extra parameters more heavily has better performance in small samples.

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

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

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

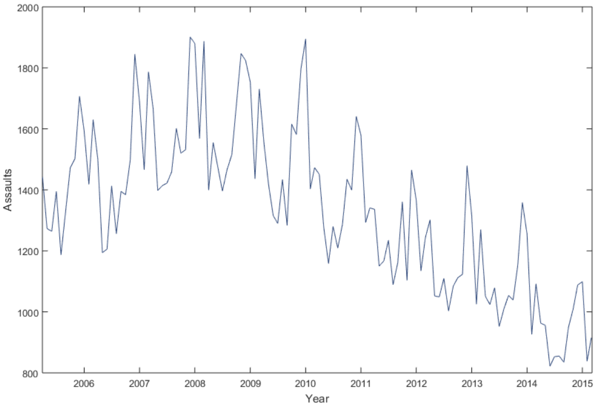
$$\text{BIC} = AIC + \log(n)(p + q + k - 1).$$

- 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 paramteres. This is because BIC penalizes the SSE more heavily than the AIC.

- Many statisticians like to use BIC because it has the feature that if there is a true underlying model, then with enough data the BIC will select that model.

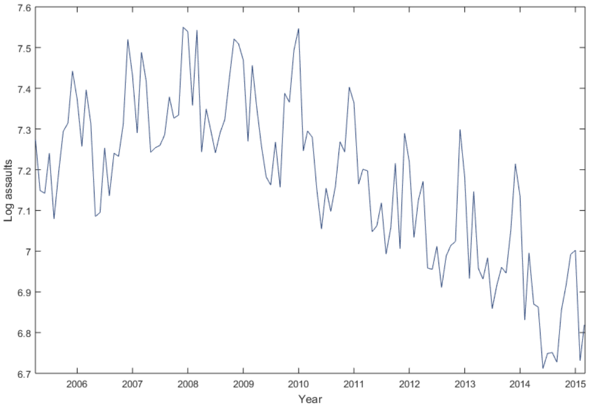
ARIMA model selection: example

Alcohol related assaults in NSW



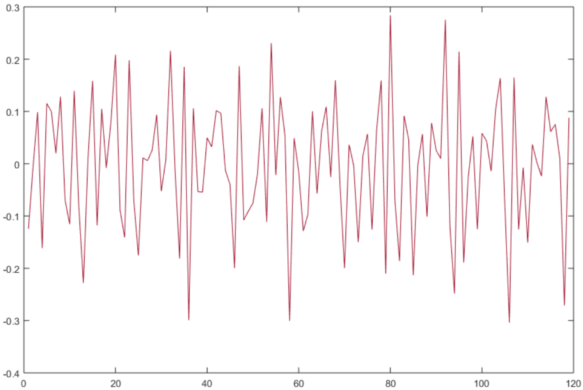
ARIMA model selection: example

Log series



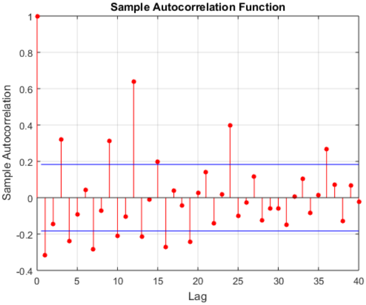
ARIMA model selection: example

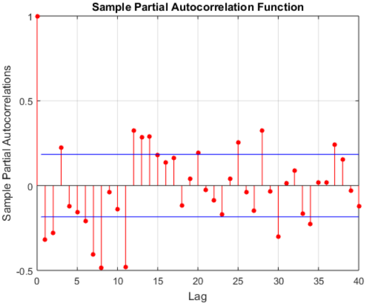
First differenced log series



ARIMA model selection: example

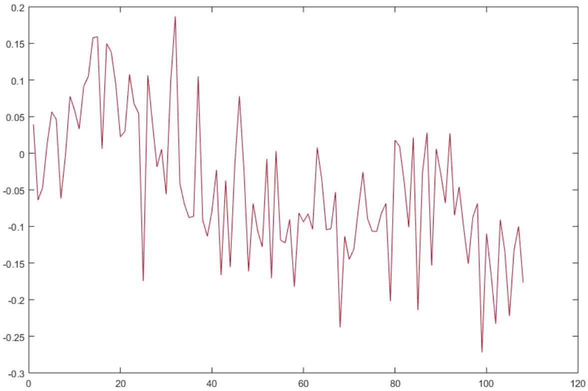
ACF and PACF for the first differenced series





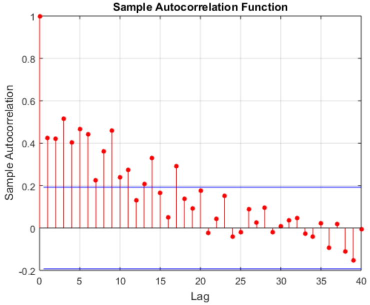
ARIMA model selection: example

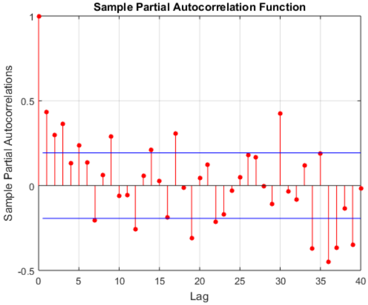
Seasonally differenced log series



ARIMA model selection: example

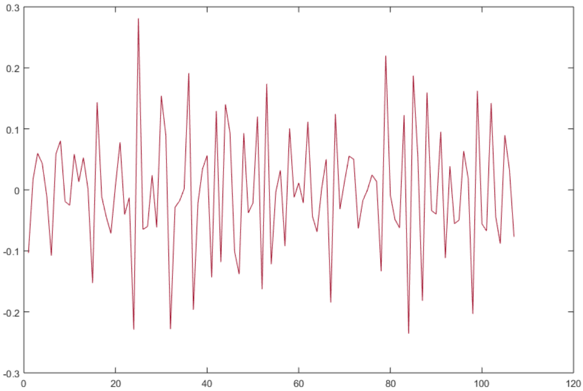
ACF and PACF for the seasonally differenced series





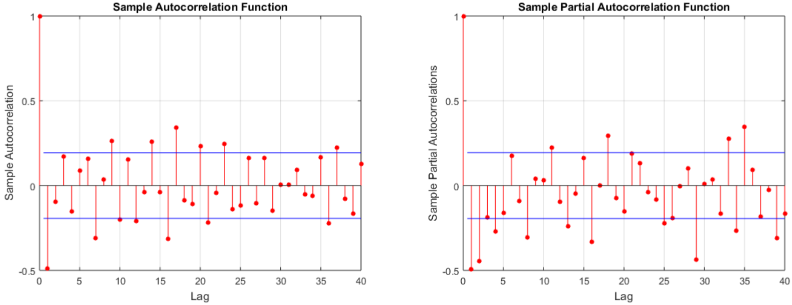
ARIMA model selection: example

First and seasonally differenced log series



ARIMA model selection: example

ACF and PACF for the first and seasonally differenced series



ARIMA model selection: example

Estimation of tentative models

```
ARIMA(0,1,1)(0,1,0)[12]

Coefficients:
          ma1
      -0.8042
s.e.   0.0535

sigma^2 estimated as 0.005896:  log likelihood=122.29
AIC=-240.58   AICc=-240.47   BIC=-235.24


ARIMA(0,1,1)(1,1,0)[12]

Coefficients:
          ma1     sar1
      -0.7900  -0.2467
s.e.   0.0525   0.0964

sigma^2 estimated as 0.005525:  log likelihood=125.41
AIC=-244.82   AICc=-244.59   BIC=-236.8
```

ARIMA model selection: example

Estimation of tentative models

```
ARIMA(0,1,1)(2,0,0)[12]

Coefficients:
          ma1    sar1    sar2
      -0.7771  0.6745  0.1648
s.e.   0.0581  0.0923  0.0985

sigma^2 estimated as 0.005177:  log likelihood=137.23
AIC=-266.45   AICc=-266.1   BIC=-255.34


ARIMA(0,1,1)(1,0,1)[12]

Coefficients:
          ma1    sar1     sma1
      -0.7311  0.9847  -0.7174
s.e.   0.0586  0.0173   0.1503

sigma^2 estimated as 0.004567:  log likelihood=140.83
AIC=-273.65   AICc=-273.3   BIC=-262.54
```

ARIMA model selection: example

Forecasting with the best model in terms of AICc