QBUS6840 Lecture 7

ARIMA Models

Professor Junbin Gao

The University of Sydney Business School

Outline

ARIMA Models

- Box-Jenkins Method: Part I

Readings

- Online Textbook Sections 8.1-8.6 (`https://otexts.com/fpp2/arima.html`); and/or
- BOK Ch 9 and Ch 10

Box-Jenkins Method

- Formal statistical time series models.
- Can capture changing components.
- Heavily rely on finding a stationary data transform.
- Time Series Analysis forecasting and control (ed. Box and Jenkins), 1976.

Time Series verse Stochastic Processes

- We have discussed so many time series. Each is a sequence of numbers (sales, production, etc)
- We introduced a number of ways to treat them: Smoothing, Modelling and Forecasting
- We rely on the patterns to decide what models to use and project the patterns into future as our forecasts.
- From now on, we will move further in theory, by considering a (concrete) time series as a "product" from a "factory"
- The factory is called a P[illegible] which is

$$Y_1, Y_2, Y_3, \cdots, Y_t, \cdots, \cdots$$

  where each Y_t $(t = 1, 2, ...)$ is a *Random Variable*.
- When we observe a (concrete) value y_t for each Y_t, we have obtained a time series.

Stationarity

Definition

A time series process is **strictly** stationary when the joint distribution (of the data) does not depend on time. That is, the joint distribution of

$$Y_t, Y_{t+1}, \ldots, Y_{t+k}$$

does not depend on t for any k.

Stationarity

Definition

A time series process is **strictly** stationary when the joint distribution (of the data) does not depend on time. That is, the joint distribution of

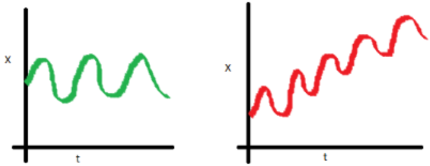
$$Y_t, Y_{t+1}, \ldots, Y_{t+k}$$

does not depend on t for any k.

Think about the case of $k = 0$: For any t, Y_t has the same distribution.

Visually Checking Stationarity

The mean of series should not be a function of time.

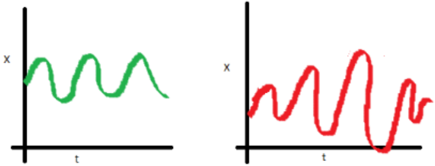


Picture is stolen from

http://www.blackarbs.com/blog/time-series-analysis-in-python-linear-models-to-garch/11/1/2016

Visually Checking Stationarity

The variance of the series should not be a function of time.

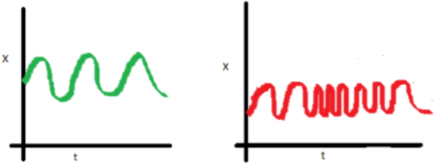


Picture is stolen from

http://www.blackarbs.com/blog/time-series-analysis-in-python-linear-models-to-garch/11/1/2016

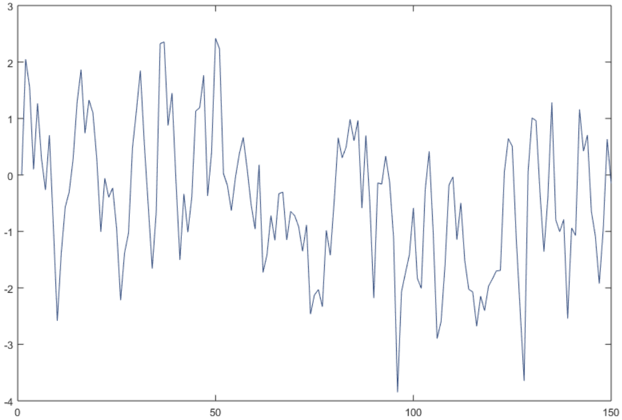
Visually Checking Stationarity

The covariance of the i-th term and the $(i+k)$-th term should not be a function of time.



Picture is stolen from

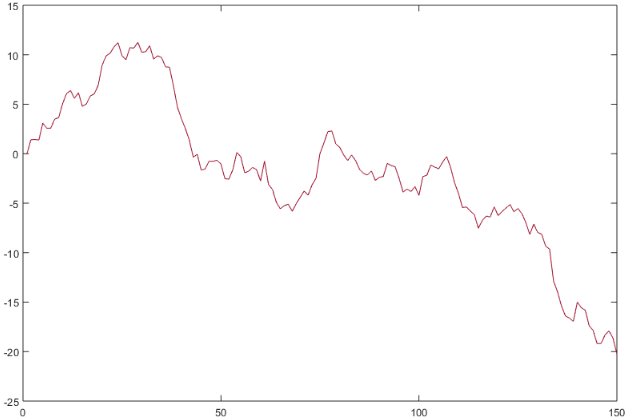
http://www.blackarbs.com/blog/time-series-analysis-in-python-linear-models-to-garch/11/1/2016



3
2
1
0
-1
-2
-3
-4
0
50
100
150

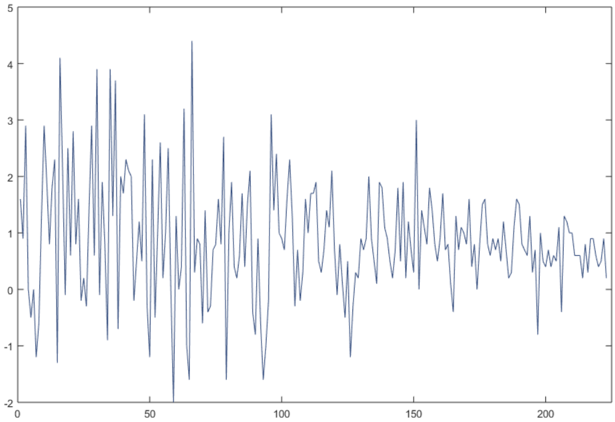
Non-stationarity

Illustration



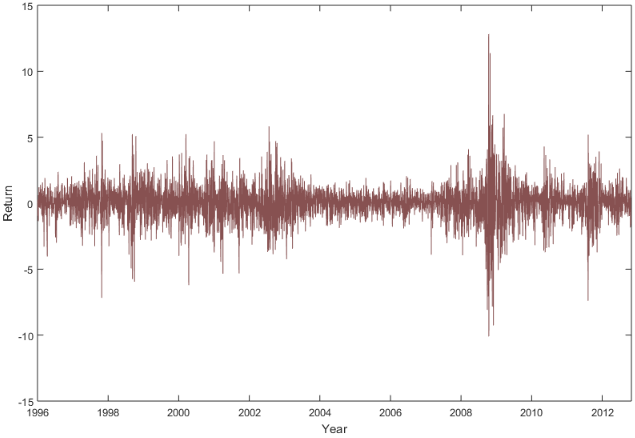
Australian seasonally adjusted quarterly GDP growth (1959-2015)

Stationary or non-stationary?



S&500 returns

Stationary or non-stationary?



Weak stationarity

Definition

A process $\{Y_t\}$ is **weakly** stationary if its mean, variance and covariance functions 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.

Weak stationarity

Definition

A process $\{Y_t\}$ is **weakly** stationary if its mean, variance and covariance functions do not change over time. That is,

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

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

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

for all t and k.

The covariance or correlation depends on the time gap, i.e., $k = t - (t - k)$

- If the mean, variance and covariances are finite (which is a technical point really), then strict stationarity implies weak stationarity.

- Weak stationarity implies strict stationarity if and only if the data is normally distributed.

- Box and Jenkins advocate using the ACF and PACF plots to assess stationarity and identify a suitable model.

- We may need to apply a suitable variance stabilising transform first.

Autocorrelation function (ACF)

Definitions

ACF:

$$\rho_k = \frac{E\left[(Y_t - \mu)(Y_{t+(\text{or } -)k} - \mu)\right]}{\sqrt{\text{Var}(Y_t)\text{Var}(Y_{t+(\text{or } -)k})}} = \text{Corr}(Y_t, Y_{t+(\text{or } -)k}).$$

Sample ACF:

$$r_k = \frac{\sum_{t=1}^{N-k}(y_{t+k} - \overline{y})(y_t - \overline{y})}{\sum_{t=1}^{N}(y_t - \overline{y})^2}.$$

What we have done is to compare, e.g., when $k = 2$,

the curve $\{y_1, y_2, y_3, ..., y_{N-2}\}$ with the curve $\{y_3, y_4, y_5, ..., y_N\}$

For $k = 5$: (see `Lecture07_Example00.py`)

the curve $\{y_1, y_2, y_3, ..., y_{N-5}\}$ with the curve $\{y_6, y_7, y_8, ..., y_N\}$

- Given a time series $\{y_1, y_2, ..., y_N\}$ and a lag k, consider the following linear regression

$$\boxed{y_{t+k} - \overline{y} = \gamma(y_t - \overline{y})} \quad \text{think of it as } Y = \gamma X$$

- Consider data set

X	$y_1 - \overline{y}$	$y_2 - \overline{y}$	$y_3 - \overline{y}$	$\cdots$	$y_{N-k} - \overline{y}$
Y	$y_{1+k} - \overline{y}$	$y_{2+k} - \overline{y}$	$y_{3+k} - \overline{y}$	$\cdots$	$y_N - \overline{y}$

- Then according to the least square regression solution

$$\gamma = \frac{\sum_{t=1}^{N-k}(y_t - \overline{y})(y_{t+k} - \overline{y})}{\sum_{t=1}^{N-k}(y_t - \overline{y})^2}$$

which is close to r_k.

Define the following standard errors:
If $k = 1$,

$$s_{r_k} = \frac{1}{\sqrt{N}}.$$

If $k > 1$,

$$s_{r_k} = \frac{\sqrt{1 + 2\sum_{j=1}^{k-1} r_j^2}}{\sqrt{N}}.$$

For a Gaussian uncorrelated series (white noise),

$$s_{r_k} \sim N(0, 1/N)$$

The t-statistic is defined as

$$t_{r_k} = \frac{r_k}{s_{r_k}}$$

(Sample) ACF Plots

- What is the value of r_0?
- In theory we can calculate r_k for all $k = 0, 1, 2, 3, ...$, i.e.,

$$r_0, r_1, r_2, r_3, r_4, \cdots$$

- If the length of time series $\{y_t\}$ is N, we can only calculate (at most)

$$r_0, r_1, r_2, r_3, r_4, \cdots, r_{N-1}$$

- An ACF Plot is a bar plot, such that the height of bar at lag k is r_k.
- We can assess the stationarity of $\{y_t\}$ by assessing the (sample) ACF plot

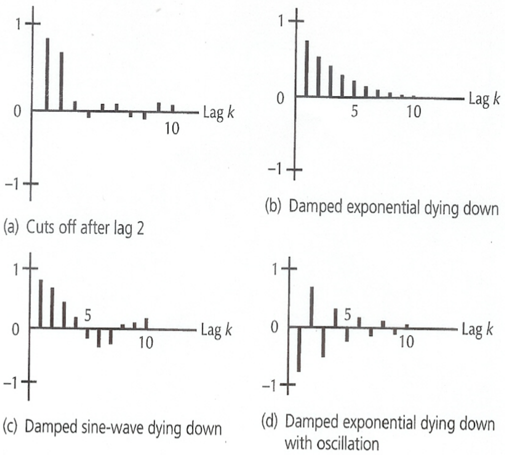
In general, it can be shown that for nonseasonal time series

- If the Sample ACF of a nonseasonal time series “dies down” or “cuts off” reasonably quickly, then the time series should be considered stationary.

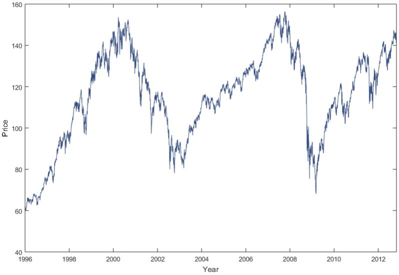
- If the Sample ACF of a nonseasonal time series “dies down” extremely slowly or not at all, then the time series should be considered nonstationary.

Autocorrelations:
Behaviour of ACFs

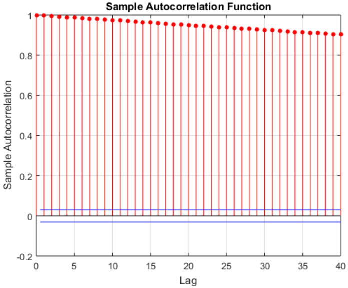
FIGURE 9.5 Examples of behavior for the SAC



S&P 500 index
ACF



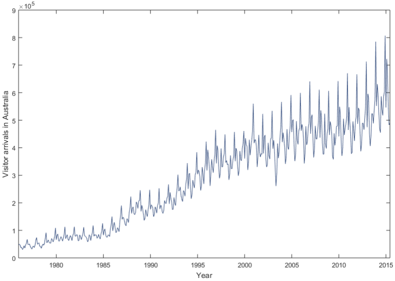
(a) Series



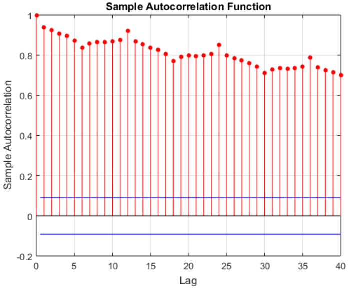
(b) ACF

Visitor arrivals in Australia

ACF



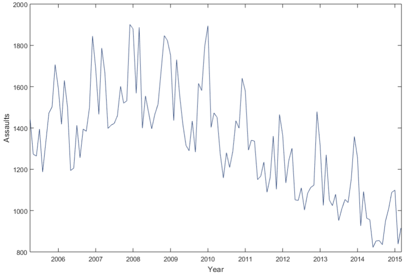
(c) Series



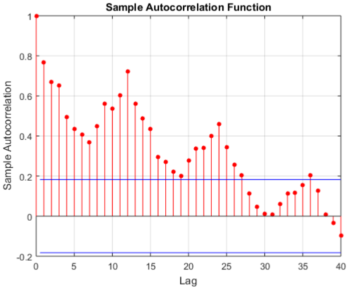
(d) ACF

Alcohol related assaults in NSW

ACF



(e) Series



(f) ACF

Tranforming

- If the ACF of a time series dies down extremely slowly, data transformation is necessary
- Trying first order differencing is always a good way. See example `Lecture07_Example01.py`
- If the ACF for the transformed data dies down extremely slowly, the transformed time series should be considered nonstationary. More transformations needed
- For nonseasonal data, first or second differencing will generally produce stationary time series values.

Partial ACF

- Partial autocorrelations measure the linear dependence of one variable after removing the effect of other variable(s) that affect to both variables.

$$
\begin{aligned}
Y_t &= \rho_{10} + \rho_{11} Y_{t-1} + \varepsilon_t \\
Y_t &= \rho_{20} + \rho_{21} Y_{t-1} + \rho_{22} Y_{t-2} + \varepsilon_t \\
Y_t &= \rho_{k0} + \rho_{k1} Y_{t-1} + \rho_{k2} Y_{t-2} + \ldots + \rho_{kk} Y_{t-k} + \varepsilon_t
\end{aligned}
$$

- ρ_{kk} is the correlation between y_t and y_{t-k} net of effects at times $t-1, t-2, \ldots, t-k+1$.

- ρ_{pp} is ϕ_p in an $AR(p)$ model (see this soon)

Partial ACF: Calculation Examples

- For example, the partial autocorrelation of 2nd order measures the effect (linear dependence) of Y_{t-2} on Y_t after removing the effect of Y_{t-1} on both Y_t and Y_{t-2}
- Each partial autocorrelation could be obtained as a series of regressions of the form:

$$
\begin{aligned}
Y_t &\approx \rho_{10} + \rho_{11} Y_{t-1} \\
Y_t &\approx \rho_{20} + \rho_{21} Y_{t-1} + \rho_{22} Y_{t-2} \\
&\vdots \\
Y_t &\approx \rho_{k0} + \rho_{k1} Y_{t-1} + \rho_{k2} Y_{t-2} + \ldots + \rho_{kk} Y_{t-k}
\end{aligned}
$$

- The estimate r_{kk} of ρ_{kk} will give the value of the partial autocorrelation of order k.
- The meaning of ACF coefficient ρ_k is

$$\boxed{Y_t = \rho_0 + \rho_k Y_{t-k} + \varepsilon_t}$$

without considering other $Y_{t-k+1}, \ldots, Y_{t-1}$.

Partial ACF: The Formula

- The Sample Partial ACF at lag k is

$$r_{kk} = \begin{cases} r_1 & \text{if } k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} & \text{if } k = 2, 3, \ldots \end{cases}$$

where

$$r_{k,j} = r_{k-1,j} - r_{kk} r_{k-1,k-j} \quad \text{for } j = 1, 2, \ldots, k-1$$

- The standard error of r_{kk} is

$$s_{r_{kk}} = \frac{1}{\sqrt{N}}$$

First Simple Process: White noise processes

- A sequence of independently and identically distributed random variables with mean 0 and finite variance σ^2.
- Model

$$y_t = \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim N(0, \sigma^2)$$

- What we hope and plan for which component in a times series model?
- What would the ACF plot look like for a white noise process? See `Lecture07_Example02.py`

Autoregressive (AR) processes

$AR(p)$ process:

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

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

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

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

$$E(Y_t) = c + \phi_1 E(Y_{t-1}),$$

Under the assumption of stationarity $E(Y_t) = E(Y_{t-1})$, so

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

$AR(1)$ process
Properties

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

$$\text{Var}(Y_t) = \phi_1^2 \text{Var}(Y_{t-1}) + \sigma^2,$$

Under the assumption of stationarity $\text{Var}(Y_t) = \text{Var}(Y_{t-1})$, so

$$\text{Var}(Y_t) = \frac{\sigma^2}{1 - \phi_1^2}.$$

In general, we have

$$\text{Cov}(Y_t, Y_{t-k}) = \phi_1^k \frac{\sigma^2}{1 - \phi_1^2}$$

Example: $AR(1)$ process

Properties

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t-1}) &= \mathrm{Cov}(c + \phi_1 Y_{t-1} + \varepsilon_t, Y_{t-1}) \\
&= \mathrm{Cov}(c, Y_{t-1}) + \mathrm{Cov}(\phi_1 Y_{t-1}, Y_{t-1}) + \mathrm{Cov}(\varepsilon_t, Y_{t-1}) \\
&= 0 + \phi_1 \mathrm{Var}(Y_{t-1}) + 0 = \phi_1 \mathrm{Var}(Y_{t-1}). \quad \text{Why?}
\end{aligned}
$$

$$
\rho_1 = \frac{\mathrm{Cov}(Y_t, Y_{t-1})}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_{t-1})}} \overset{Why?}{=\!=} \frac{\mathrm{Cov}(Y_t, Y_{t-1})}{\mathrm{Var}(Y_{t-1})} = \phi_1.
$$

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t-2}) &= \mathrm{Cov}(c + \phi_1 Y_{t-1} + \varepsilon_t, Y_{t-2}) \\
&= \mathrm{Cov}(\phi_1(c + \phi_1 Y_{t-2} + \varepsilon_{t-1}), Y_{t-2}) \\
&= \phi_1^2 \mathrm{Var}(Y_{t-2}).
\end{aligned}
$$

Thus, noting that $\mathrm{Var}(Y_{t-2}) = \mathrm{Var}(Y_{t-1}) = \mathrm{Var}(Y_t)$,

$$
\rho_2 = \frac{\mathrm{Cov}(Y_t, Y_{t-2})}{\mathrm{Var}(Y_t)} = \phi_1^2,
$$

$$
\vdots \quad \text{(Similarly)}
$$

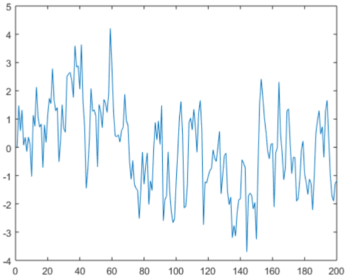
$$
\rho_k = \frac{\mathrm{Cov}(Y_t, Y_{t-k})}{\mathrm{Var}(Y_t)} = \phi_1^k.
$$

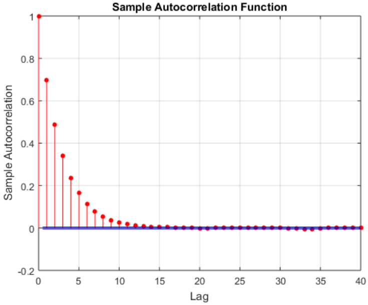
Example: $AR(1)$ process

- What happens to the ACF when $-1 < \phi_1 < 1$ and k increases?
- What happens when $\phi_1 = 1$?
- `Lecture07_Example03.py`

Example: $AR(1)$ process

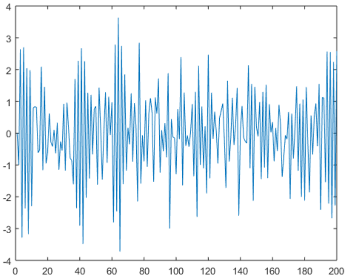
$\phi = 0.7$

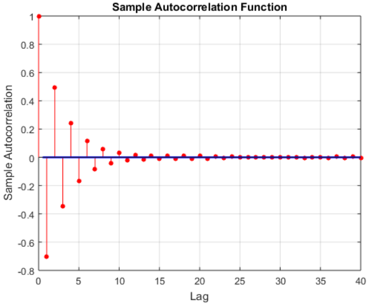




Example: $AR(1)$ process

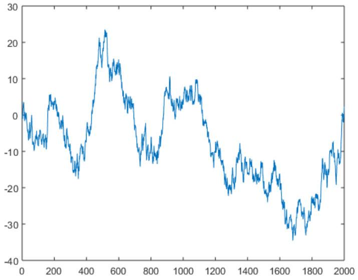
$\phi = -0.7$

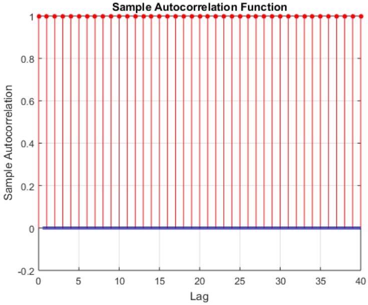




Example: $AR(1)$ process

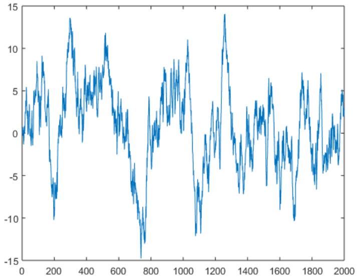
$\phi = 1$

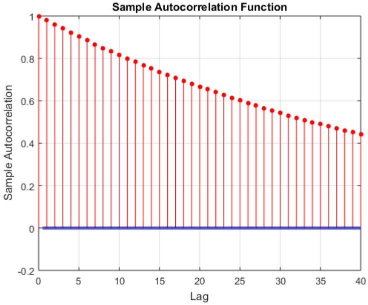




Example: $AR(1)$ process

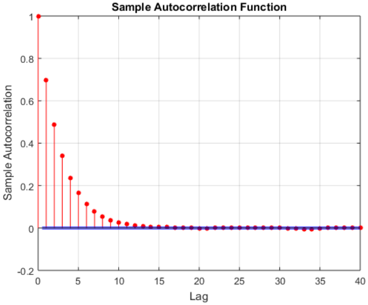
$\phi = 0.98$

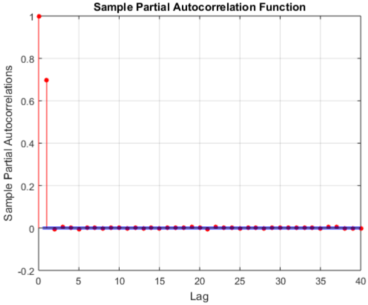




Example: $AR(1)$ process

$\phi = 0.7$ ACF (left) and Partial ACF (right)





Example: $AR(1)$ process

Stationarity

- When $|\phi_1| < 1$, the AR(1) process is weakly stationary

Example: $AR(1)$ process
Conditional Expectation and Variance

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

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

$$\begin{aligned}
E(Y_{t+1}|y_{1:t}) &= E(Y_{t+1}|y_1, \ldots, y_t) \\
&= E(Y_{t+1}|y_t) = E(c + \phi_1 y_t + \varepsilon_{t+1}|y_t) \\
&= c + \phi_1 y_t + E(\varepsilon_{t+1}) = c + \phi_1 y_t
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(Y_{t+1}|y_{1:t}) &= \mathrm{Var}(Y_{t+1}|y_1, \ldots, y_t) = \mathrm{Var}(Y_{t+1}|y_t) \\
&= \mathrm{Var}(c + \phi_1 y_t + \varepsilon_{t+1}|y_t) \\
&= 0 + \mathrm{Var}(\varepsilon_{t+1}) = \sigma^2
\end{aligned}$$

Example: $AR(1)$ process

Forecasting

$$
\begin{aligned}
E(Y_{t+h}|y_{1:t}) &= E(c + \phi_1 Y_{t+h-1} + \varepsilon_{t+h}|y_{1:t}) \\
&= c + \phi_1 E(Y_{t+h-1}|y_{1:t}) + 0 \\
&= c + \phi_1(c + \phi_1 E(Y_{t+h-2}|y_{1:t})) \\
&= ....
\end{aligned}
$$

Until we know

$$
\begin{aligned}
&\widehat{Y}_{t+1} := E(Y_{t+1}|y_{1:t}) = c + \phi_1 y_t, \\
&\mathrm{Var}(Y_{t+1}|y_{1:t}) = \mathrm{Var}(c + \phi_1 y_t + \varepsilon_{t+1}|y_{1:t}) = 0 + \mathrm{Var}(\varepsilon_{t+1}) = \sigma^2.
\end{aligned}
$$

Denote by $\widehat{Y}_{t+h} = E(Y_{t+h}|y_{1:t})$, then the above equation (second) says

$$
\widehat{Y}_{t+h} = c + \phi_1 \widehat{Y}_{t+h-1}
$$

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

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

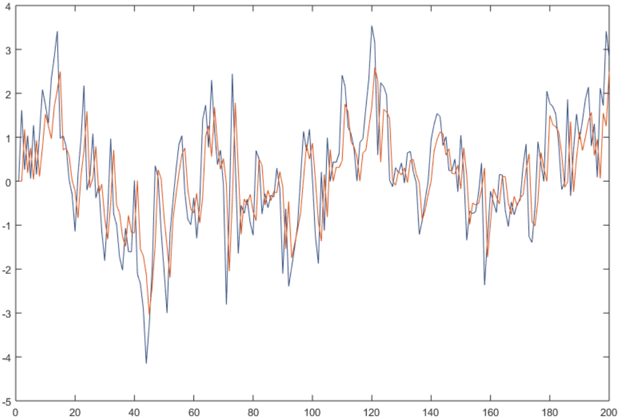
$$
\begin{aligned}
\widehat{Y}_{t+h} &= c + \phi_1 \widehat{Y}_{t+h-1} \\
&= c(1 + \phi_1 + \phi_1^2 + \ldots + \phi_1^{h-1}) + \phi_1^h y_t
\end{aligned}
$$

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

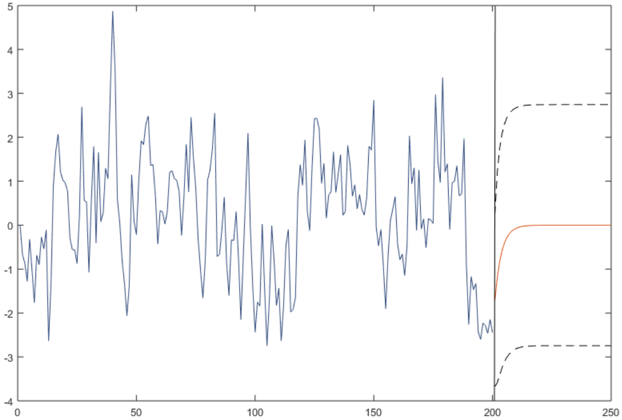
What happens as h gets larger?

Example: $AR(1)$ process

In-sample fit illustration



Example: $AR(1)$ process
Forecasting illustration



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

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

Suppose it is stationary, then

$$\begin{aligned} E(Y_t) &= \frac{c}{1 - \phi_1 - \phi_2 - \ldots - \phi_p} \\ &= \frac{c}{1 - \sum_{i=1}^{p} \phi_i} \end{aligned}$$

$AR(p)$ processes
Properties

$$Y_t = c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t,$$
$$\mathrm{Var}(Y_t) = \mathrm{Var}(c + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \varepsilon_t)$$

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

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

Can we continue like this?

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

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

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

Can we continue like this?

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

NO! because all

$$\text{Cov}(Y_{t-1}, Y_{t-2}) \neq 0$$

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

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

Can we continue like this?

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

NO! because all

$$\mathrm{Cov}(Y_{t-1}, Y_{t-2}) \neq 0$$

Under the stationary condition, it can be proved that

$$\mathrm{Var}(Y_t) = \frac{\sigma^2}{(1-\rho_{11}^2)(1-\rho_{22}^2)\ldots(1-\rho_{pp}^2)}$$

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t-1}) &= \mathrm{Cov}(c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, Y_{t-1}) \\
&= \phi_1 \mathrm{Var}(Y_{t-1}) + \phi_2 \mathrm{Cov}(Y_{t-2}, Y_{t-1})
\end{aligned}
$$

Under the stationary condition we have

$$
\mathrm{Cov}(Y_t, Y_{t-1}) = \mathrm{Cov}(Y_{t-2}, Y_{t-1}) = \frac{\phi_1}{1-\phi_2} \mathrm{Var}(Y_{t-1}).
$$

$$
\rho_1 = \frac{\mathrm{Cov}(Y_t, Y_{t-1})}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_{t-1})}} = \frac{\phi_1}{1-\phi_2}.
$$

where we have used $\mathrm{Var}(Y_t) = \mathrm{Var}(Y_{t-1})$.

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t-2}) &= \mathrm{Cov}(c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, Y_{t-2}) \\
&= \phi_2 \mathrm{Var}(Y_{t-2}) + \phi_1 \mathrm{Cov}(Y_{t-1}, Y_{t-2}) \\
&= \left(\phi_2 + \frac{\phi_1^2}{1 - \phi_2} \right) \mathrm{Var}(Y_{t-2}).
\end{aligned}
$$

$$
\rho_2 = \frac{\mathrm{Cov}(Y_t, Y_{t-2})}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_{t-2})}} = \phi_2 + \frac{\phi_1^2}{1 - \phi_2}.
$$

where we have used $\mathrm{Var}(Y_t) = \mathrm{Var}(Y_{t-2})$.

Example: $AR(2)$ processes

Properties

$$
\begin{aligned}
\text{Cov}(Y_t, Y_{t-3}) &= \text{Cov}(c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, Y_{t-3}) \\
&= \phi_1 \text{Cov}(Y_{t-1}, Y_{t-3}) + \phi_2 \text{Cov}(Y_{t-2}, Y_{t-3}) \\
&= \phi_1 \rho_2 \text{Var}(Y_{t-3}) + \phi_2 \rho_1 \text{Var}(Y_{t-3}).
\end{aligned}
$$

where we have used $\rho_2 = \frac{\text{Cov}(Y_{t-1}, Y_{t-3})}{\text{Var}(Y_{t-3})}$ and $\rho_1 = \frac{\text{Cov}(Y_{t-2}, Y_{t-3})}{\text{Var}(Y_{t-3})}$.

$$\rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1$$

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2},$$

$k > 2$.

- ρ_k dies down exponentially. (ACF)
- ρ_{kk} cuts off to zero after lag p. (PACF)
- This can be theoretically approved.

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

where

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

For example, consider AR(3),

$$Y_{t+1} = c + \phi_1 Y_t + \phi_2 Y_{t-1} + \phi_3 Y_{t-2} + \varepsilon_{t+1}$$

then

$$\begin{aligned} \widehat{y}_{t+1} &= c + \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2} \\ \widehat{y}_{t+2} &= c + \phi_1 \widehat{y}_{t+1} + \phi_2 y_t + \phi_3 y_{t-1} \\ \widehat{y}_{t+3} &= c + \phi_1 \widehat{y}_{t+2} + \phi_2 \widehat{y}_{t+1} + \phi_3 y_t \end{aligned}$$

Hence

$$\begin{aligned}
\widehat{y}_{t+1} =& c + \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2} \\
\widehat{y}_{t+2} =& c + \phi_1 \widehat{y}_{t+1} + \phi_2 y_t + \phi_3 y_{t-1} \\
=& c + \phi_1 (c + \phi_1 y_t + \phi_2 y_{t-1} + \phi_3 y_{t-2}) + \phi_2 y_t + \phi_3 y_{t-1} \\
=& c(1 + \phi_1) + (\phi_1^2 + \phi_2) y_t + (\phi_1 \phi_2 + \phi_3) y_{t-1} + \phi_1 \phi_3 y_{t-2} \\
\widehat{y}_{t+3} =& c + \phi_1 \color{red}{\widehat{y}_{t+2}} \color{black} + \phi_2 \color{red}{\widehat{y}_{t+1}} \color{black} + \phi_3 y_t \\
=& \cdots\cdots
\end{aligned}$$

Finally what about the variance?