MONASH University

MONASH BUSINESS SCHOOL

ETF3231/5231 Business forecasting

Week 8: ARIMA models

https://bf.numbat.space/

AACSB ACCREDITED EQUIS ACCREDITED AMBA ACCREDITED

Outline

ARIMA models

AR: autoregressive (lagged observations as inputs)
I: integrated (differencing to make series stationary)
MA: moving average (lagged errors as inputs)

ARIMA models

AR: autoregressive (lagged observations as inputs)
I: integrated (differencing to make series stationary)
MA: moving average (lagged errors as inputs)

An ARIMA model is rarely interpretable in terms of visible data structures like trend and seasonality. But it can capture a huge range of time series patterns.

ARIMA models

AR: autoregressive (lagged observations as inputs)
I: integrated (differencing to make series stationary)
MA: moving average (lagged errors as inputs)

An ARIMA model is rarely interpretable in terms of visible data structures like trend and seasonality. But it can capture a huge range of time series patterns.

Make data stationary (variance & mean), fit model, reverse, forecast.

Outline

AR(1) model including a constant

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t, \quad |\phi_1| < 0 \quad (P.T.O)$$

- When $\phi_1 = 0$ and $c = 0$, y_t is equivalent to WN;
- When $\phi_1 = 1$ and $c = 0$, y_t is equivalent to a RW;
- When $\phi_1 = 1$ and $c \neq 0$, y_t is equivalent to a RW with drift;
- When $\phi_1 > 0$, y_t tends to hang below or above the mean of y_t.
- When $\phi_1 < 0$, y_t tends to oscillate below and above the mean of y_t.
- If $E(y_t) = \mu$, $\mu = \frac{c}{1-\phi_1}$, c is related to the mean of y_t.

For stationarity we require $-1 < \phi_1 < 1$

Let's set $\phi_1 = 2$ $\quad y_t = 2y_{t-1} + \varepsilon_t$ and see what happens:

$$y_1 = 2y_0 + \varepsilon_1$$

$$y_2 = 2y_1 + \varepsilon_2 = 4y_0 + 2\varepsilon_1 + \varepsilon_2$$

$$y_3 = 2y_2 + \varepsilon_3 = 8y_0 + 4\varepsilon_1 + 2\varepsilon_2 + \varepsilon_3 \quad \ldots \text{ and so on}$$

$$\vdots$$

- Getting larger & larger
- Hence, we only allow for the new obs. to be a fraction of the previous ones.

Autoregressive models

A multiple regression with lagged values of y_t as predictors.

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \\ &= c + (\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) y_t + \varepsilon_t \end{aligned}$$

Autoregressive models

A multiple regression with lagged values of y_t as predictors.

$$\begin{aligned}
y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \\
&= c + (\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) y_t + \varepsilon_t
\end{aligned}$$

$$\begin{aligned}
(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) y_t &= c + \varepsilon_t \\
\phi(B) y_t &= c + \varepsilon_t
\end{aligned}$$

- ε_t is white noise.
- $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$

Stationarity conditions

We normally restrict autoregressive models to stationary data, and then some constraints on the values of the parameters are required.

General condition for stationarity

Complex roots of $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$ lie outside the unit circle on the complex plane.

Stationarity conditions

We normally restrict autoregressive models to stationary data, and then some constraints on the values of the parameters are required.

General condition for stationarity

Complex roots of $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ lie outside the unit circle on the complex plane.

- For $p = 1$: $-1 < \phi_1 < 1$.
- For $p = 2$:

  $$-1 < \phi_2 < 1 \qquad \phi_2 + \phi_1 < 1 \qquad \phi_2 - \phi_1 < 1.$$
- More complicated conditions hold for $p \geq 3$.
- fable takes care of this.

Moving Average (MA) models

A multiple regression with past *errors* as predictors.

$$\begin{aligned} y_t &= c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} \\ &= c + (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)\varepsilon_t \\ &= c + \theta(B)\varepsilon_t \end{aligned}$$

Moving Average (MA) models

A multiple regression with past *errors* as predictors.

$$
\begin{aligned}
y_t &= c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} \\
&= c + (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)\varepsilon_t \\
&= c + \theta(B)\varepsilon_t
\end{aligned}
$$

- ε_t is white noise.
- $\theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)$

Invertibility

General condition for invertibility

Complex roots of $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ lie outside the unit circle on the complex plane.

Invertibility

General condition for invertibility

Complex roots of $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ lie outside the unit circle on the complex plane.

- For $q = 1$: $-1 < \theta_1 < 1$.
- For $q = 2$:

  $-1 < \theta_2 < 1 \qquad \theta_2 + \theta_1 > -1 \qquad \theta_1 - \theta_2 < 1.$
- More complicated conditions hold for $q \geq 3$.
- fable takes care of this.

PROPERTIES OF AR & MA models

- AR(1) $\quad y_t = c + \phi\, y_{t-1} + \varepsilon_t\,;\quad \varepsilon_t \sim NID(0, \sigma^2)\,;\quad |\phi| < 1\,;\quad y_0$ (starting value).

$$y_t = c + \phi\, y_{t-1} + \varepsilon_t$$

$$= c + \phi\left(c + \phi\, y_{t-2} + \varepsilon_{t-1}\right) + \varepsilon_t$$

$$= c + \phi c + \phi^2 y_{t-2} + \phi\,\varepsilon_{t-1} + \varepsilon_t$$

$$= c + \phi c + \phi^2\left(c + \phi\, y_{t-3} + \varepsilon_{t-2}\right) + \phi\,\varepsilon_{t-1} + \varepsilon_t$$

$$= c + \phi c + \phi^2 c + \phi^3 y_{t-3} + \phi^2 \varepsilon_{t-2} + \phi\,\varepsilon_{t-1} + \varepsilon_t$$

$$\vdots$$

$$\Rightarrow y_t = c\left(1 + \phi + \phi^2 + \dots\right) + \phi^t y_0 + \varepsilon_t + \phi\,\varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots$$

for $t \to \infty$ $\quad = \dfrac{c}{1-\phi} + \varepsilon_t + \phi\,\varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots\quad$ MA(∞)

PROPERTIES OF AR & MA models

- AR(1) $\quad y_t = c + \phi\, y_{t-1} + \varepsilon_t$; $\varepsilon_t \sim NID(0, \sigma^2)$; $|\phi| < 1$; y_0 (starting value).

$$
\begin{aligned}
y_t &= c + \phi y_{t-1} + \varepsilon_t \\
&= c + \phi\,(c + \phi\, y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= c + \phi c + \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\
&= c + \phi c + \phi^2 (c + \phi\, y_{t-3} + \varepsilon_{t-2}) + \phi \varepsilon_{t-1} + \varepsilon_t \\
&= c + \phi c + \phi^2 c + \phi^3 y_{t-3} + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\
&\ \ \vdots
\end{aligned}
$$

$$\Rightarrow y_t = c\,(1 + \phi + \phi^2 + \dots) + \phi^t y_0 + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots$$

for $t \to \infty$ $\quad = \dfrac{c}{1-\phi} + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots \qquad MA(\infty)$

- $E(y_t) = \dfrac{c}{1-\phi} = \mu$

- $var(y_t) = \gamma_0 = E(y_t - \mu)^2 = \dfrac{\sigma^2}{1-\phi^2}$ (stat)

$$
\begin{aligned}
\gamma_0 &= E\left(\frac{c}{1-\phi} + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots - \frac{c}{1-\phi}\right)^2 \\
&= E(\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots)^2 \\
&= E(\varepsilon_t)^2 + \phi^2 E(\varepsilon_{t-1})^2 + \phi^4 E(\varepsilon_{t-2})^2 + \dots \quad \text{cross-products} \\
&= \sigma^2 + \phi^2 \sigma^2 + (\phi^2)^2 \sigma^2 + (\phi^2)^3 \sigma^2 + \dots \\
&= \sigma^2 \left(1 + \phi^2 + (\phi^2)^2 + (\phi^2)^3 + \dots\right) = \frac{\sigma^2}{1-\phi^2}
\end{aligned}
$$

$$\mathrm{Cov}(y_t, y_{t-1}) = \gamma_1 = \phi \frac{\sigma^2}{1-\phi^2} = \phi \gamma_0$$

$$\gamma_1 = E(y_t - \mu)(y_{t-1} - \mu)$$

$$= E(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \ldots)(\varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \phi^3\varepsilon_{t-4} + \ldots)$$

$$= \phi E(\varepsilon_{t-1})^2 + \phi^3 E(\varepsilon_{t-2})^2 + \phi^5 E(\varepsilon_{t-3})^2 + \ldots$$

$$= \phi\sigma^2 + \phi\phi^2\sigma^2 + \phi\phi^4\sigma^2 + \ldots$$

$$= \phi\sigma^2 \left(1 + \phi^2 + (\phi^2)^2 + (\phi^2)^3 + \ldots\right)$$

$$= \phi \frac{\sigma^2}{1-\phi^2}$$

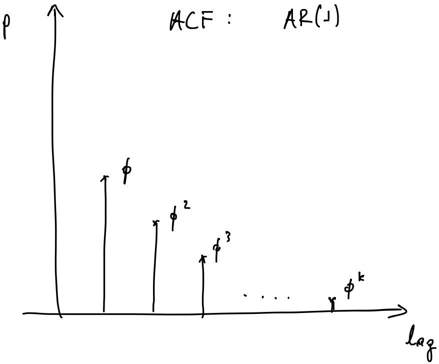
In general $\gamma_k = \phi^k \gamma_0$ (cov stat)

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\phi\gamma_0}{\gamma_0} = \phi$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi^2\gamma_0}{\gamma_0} = \phi^2$$

$$\vdots$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi^k\gamma_0}{\gamma_0} = \phi^k$$



Properties of an MA(1); $y_t = c + \varepsilon_t + \theta \varepsilon_{t-1}$, $\varepsilon_t \sim NID(0, \sigma^2)$ $|\theta| < 1$

$E(y_t) = \mu = c$

$var(y_t) = \gamma_0 = E(y_t - \mu)^2 = E(\varepsilon_t)^2 + \theta^2 E(\varepsilon_{t-1})^2 + \cancel{cov(\varepsilon_t, \varepsilon_{t-1})} \xrightarrow{} 0 = \sigma^2 + \theta^2 \sigma^2 = (1 + \theta^2)\sigma^2$

$cov(y_t, y_{t-1}) = \gamma_1 = E(y_t - \mu)(y_{t-1} - \mu) = E(\varepsilon_t + \underline{\theta \varepsilon_{t-1}})(\underline{\varepsilon_{t-1}} + \theta \varepsilon_{t-2}) = \theta E(\varepsilon_{t-1})^2 = \theta \sigma^2$

$\gamma_2 = E(y_t - \mu)(y_{t-2} - \mu) = E(\varepsilon_t + \theta \varepsilon_{t-1})(\varepsilon_{t-2} - \mu) = 0$

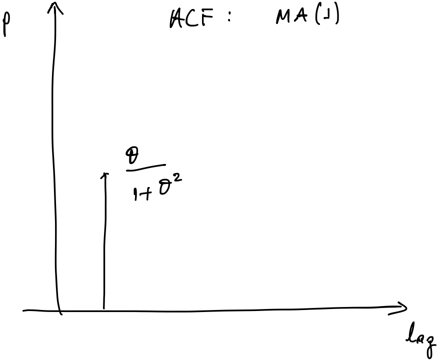
In fact for an MA(q) all $\gamma_j = 0$ $j > q$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta \cancel{\sigma^2}}{(1+\theta^2)\cancel{\sigma^2}}$$

$$\rho_2 = 0$$

$$\vdots$$

$$\rho_k = 0$$



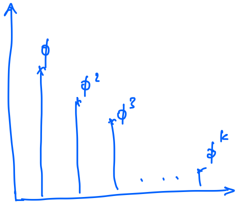
AR(1) $\quad y_t = c + \phi y_{t-1} + \varepsilon_t \qquad |\phi| < 1$ (stationary)

$$E(y_t) = \mu = \frac{c}{1-\phi}$$

$$var(y_t) = \gamma_0 = \frac{\sigma^2}{1-\phi^2}$$

$$\gamma_k = \phi^k \gamma_0 \qquad \rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k$$

ACF



MA(1) $\quad y_t = c + \theta \varepsilon_{t-1} + \varepsilon_t \qquad |\theta| < 1$ (invertible)

$$E(y_t) = \mu = c$$

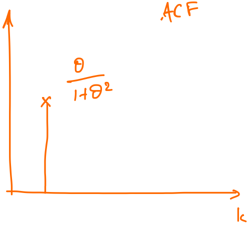
$$var(y_t) = \gamma_0 = (1+\theta^2)\sigma^2$$

$$\gamma_1 = \theta\sigma^2 \qquad \gamma_2 = \dots = \gamma_k = 0$$

$$\boxed{\text{in general } \gamma_j = 0 \quad j > q}$$

$$\rho_1 = \frac{\theta}{1+\theta^2}\ ; \quad \rho_2 = \dots = \rho_k = 0$$

ACF



AR(1) $\quad y_t = c + \phi y_{t-1} + \varepsilon_t \qquad |\phi| < 1$

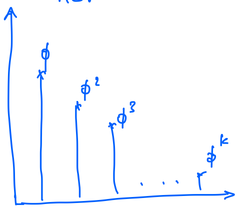
(stationary)

$$E(y_t) = \mu = \frac{c}{1-\phi}$$

$$var(y_t) = \gamma_0 = \frac{\sigma^2}{1-\phi^2}$$

$$\gamma_k = \phi^k \gamma_0 \qquad \rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k$$

ACF



MA(1) $\quad y_t = c + \theta \varepsilon_{t-1} + \varepsilon_t \qquad |\theta| < 1$

(invertible)

$$E(y_t) = \mu = c$$

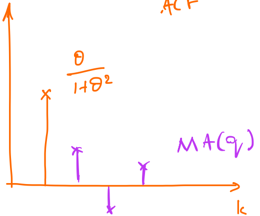
$$var(y_t) = \gamma_0 = (1+\theta^2)\sigma^2$$

$$\gamma_1 = \theta\sigma^2 \qquad \gamma_2 = \ldots = \gamma_k = 0$$

$$\boxed{\text{in general } \gamma_j = 0 \quad j > q}$$

$$\rho_1 = \frac{\theta}{1+\theta^2} \;;\quad \rho_2 = \ldots = \rho_k = 0$$

ACF



ARIMA models

ARIMA(p, d, q) model: $\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t$

AR:	p = order of the autoregressive part
I:	d = degree of first differencing involved
MA:	q = order of the moving average part.

ARIMA models

ARIMA(p, d, q) model: $\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t$

AR: p = order of the autoregressive part
I: d = degree of first differencing involved
MA: q = order of the moving average part.

- Conditions on AR coefficients ensure stationarity.
- Conditions on MA coefficients ensure invertibility.
- White noise model: ARIMA(0,0,0) $y_t = \varepsilon_t$
- $I(1)$ Random walk: ARIMA(0,1,0) with no constant $y_t' = \varepsilon_t \Rightarrow y_t = y_{t-1} + \varepsilon_t$
- Random walk with drift: ARIMA(0,1,0) with const. $y_t' = c + \varepsilon_t \Rightarrow y_t = c + y_{t-1} + \varepsilon_t$
- AR(p): ARIMA(p,0,0)
- MA(q): ARIMA(0,0,q)

R model

Intercept form

*

$$(1 - \phi_1 B - \cdots - \phi_p B^p) y'_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

Mean form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(y'_t - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

* Other books use this. Don't get confused.

- $y'_t = (1 - B)^d y_t$
- μ is the mean of y'_t.
- $c = \mu(1 - \phi_1 - \cdots - \phi_p)$.
- `ARIMA()` in the fable package uses intercept form.

Recall : the mean of the (differenced) data is not the contront.

Understanding ARIMA models

* VERY IMPORTANT TO UNDERSTAND

(p, q) → short-run

(c, d) → long-run

- If $c = 0$ and $d = 0$, the long-term forecasts will go to zero.
- If $c = 0$ and $d = 1$, the long-term forecasts will go to a non-zero constant.
- If $c = 0$ and $d = 2$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 0$, the long-term forecasts will go to the mean of the data.
- If $c \neq 0$ and $d = 1$, the long-term forecasts will follow a straight line.
- If $c \neq 0$ and $d = 2$, the long-term forecasts will follow a quadratic trend.

SUMMARY

constant

$c = 0$	$d = 0$	$\hat{y}_{T+\infty} \to 0$	ARMA stat around	$E(y_t) = 0$
$c \neq 0$	$d = 0$	$\hat{y}_{T+\infty} \to E(y_t)$	" " "	$E(y_t) \neq 0$
$c = 0$	$d = 1$	$\hat{y}_{T+\infty} \to$ const	RW + ARMA	• diff of RW is stat and the ARMA part will converge to const.

linear trend

$c \neq 0$	$d = 1$		RW + drift + ARMA
		$\hat{y}_{T+\infty} \to t$	
$c = 0$	$d = 2$		Two unit roots

quadratic trend $c \neq 0$ $d = 2$

DO IT AT YOUR OWN RISK.

Understanding ARIMA models

Forecast variance and d

- The higher the value of d, the more rapidly the prediction intervals increase in size.
- For $d = 0$, the long-term forecast standard deviation will go to the standard deviation of the historical data.

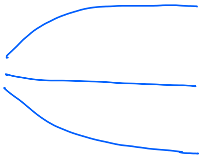
Cyclic behaviour

- For cyclic forecasts, $p \geq 2$ and some restrictions on coefficients are required.
- If $p = 2$, we need $\phi_1^2 + 4\phi_2 < 0$. Then average cycle of length

$$(2\pi)/\left[\text{arc}\cos(-\phi_1(1-\phi_2)/(4\phi_2))\right].$$

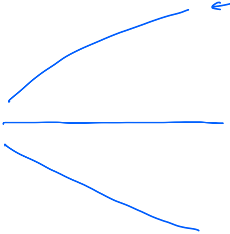
$d = 0$

constant (in sample st. dev).



$d = 1$

keep increasing



Outline

Partial autocorrelations

Partial autocorrelations measure relationship between y_t and y_{t-k}, when the effects of other time lags — $1, 2, 3, \ldots, k-1$ — are removed. why. PTO →

Partial autocorrelations

Partial autocorrelations measure relationship between y_t and y_{t-k}, when the effects of other time lags — $1, 2, 3, \dots, k-1$ — are removed.

α_k = kth partial autocorrelation coefficient

= equal to the estimate of ϕ_k in regression:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_k y_{t-k}. \neq \rho_1, \dots, \rho_k \text{ in ACF}$$

Partial autocorrelations

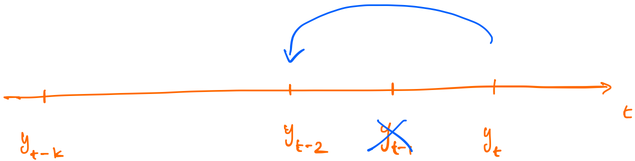
Partial autocorrelations measure relationship between y_t and y_{t-k}, when the effects of other time lags — $1, 2, 3, \dots, k-1$ — are removed.

$$\alpha_k = k\text{th partial autocorrelation coefficient}$$
$$= \text{equal to the estimate of } \phi_k \text{ in regression:}$$
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_k y_{t-k}.$$

- Varying number of terms on RHS gives α_k for different values of k.
- $\boxed{\alpha_1 = \rho_1}$ ALWAYS
- same critical values of $\pm 1.96/\sqrt{T}$ as for ACF.
- * Last significant α_k indicates the order of an AR model. *

AR(1) $\quad y_t = \phi y_{t-1} + \varepsilon_t \quad \Rightarrow \quad y_{t-1} = \phi y_{t-2} + \varepsilon_{t-1}$

$$\Rightarrow y_t = \phi(\phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t$$



- correlation between y_t & y_{t-k} conditional on all other regressors
- this is what regression does

ACF

$y_t = c + \rho_1 y_{t-1} + \varepsilon_t$

$y_t = c + \rho_2 y_{t-2} + \varepsilon_t$

PACF

$y_t \quad \boxed{\begin{matrix} y_{t-1} \\ y_{t-1} \end{matrix}} \quad y_{t-2}$

$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_2$

$\alpha_1 = \phi_1 = \rho_1$ (Nothing between y_t & y_{t-1}) ALWAYS

$\alpha_2 = \phi_2 < \rho_2$ (takes out the chaining / link of y_{t-1})

$\alpha_k = \phi_k = 0$ for $k > p$, hence max $\phi_k \neq 0$ gives you the AR order

Note: the purpose of the PACF is largely to select the AR order. If ACF shows WN, PACF will also show WN ($\alpha_1 = \phi_1 = 0$).

ACF and PACF interpretation

AR(1)

$$\rho_k = \phi_1^k \qquad \text{for } k = 1, 2, \dots;$$
$$\alpha_1 = \phi_1 \qquad \alpha_k = 0 \qquad \text{for } k = 2, 3, \dots.$$

So we have an AR(1) model when

- autocorrelations exponentially decay
- there is a single significant partial autocorrelation.

ACF and PACF interpretation

AR(p)

- ACF dies out in an exponential or damped sine-wave manner
- PACF has all zero spikes beyond the pth spike

So we have an AR(p) model when

- the ACF is exponentially decaying or sinusoidal
- there is a significant spike at lag p in PACF, but none beyond p

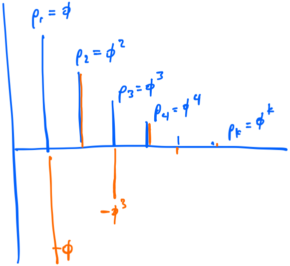
ACF

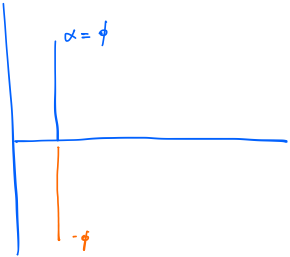
$\boxed{\text{PACF}}$ *

$0 < \phi < 1$

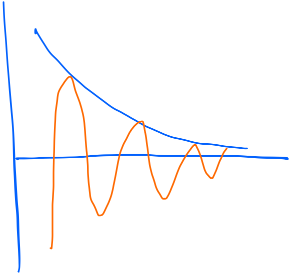
$-1 < \phi < 0$

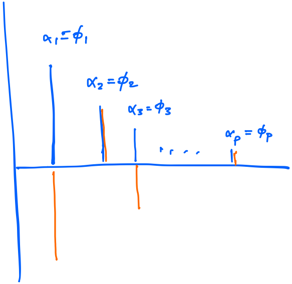
AR(1)





AR(p)





ACF and PACF interpretation

MA(1)

$$\rho_1 = \theta_1 \qquad \rho_k = 0 \qquad \text{for } k = 2, 3, \ldots;$$
$$\alpha_k = -(-\theta_1)^k$$

So we have an MA(1) model when

- the PACF is exponentially decaying and
- there is a single significant spike in ACF

ACF and PACF interpretation

MA(q)

- PACF dies out in an exponential or damped sine-wave manner
- ACF has all zero spikes beyond the qth spike

So we have an MA(q) model when

- the PACF is exponentially decaying or sinusoidal
- there is a significant spike at lag q in ACF, but none beyond q

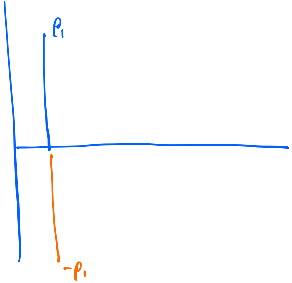
$\boxed{\text{ACF}}$ *

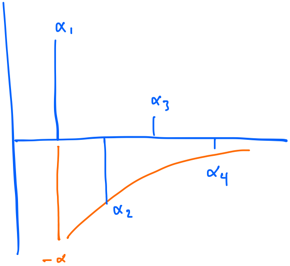
PACF

$0 < \theta < 1$

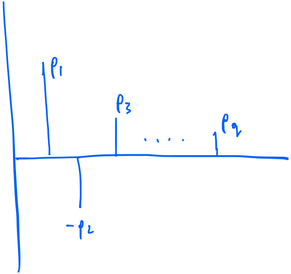
$-1 < \theta < 0$

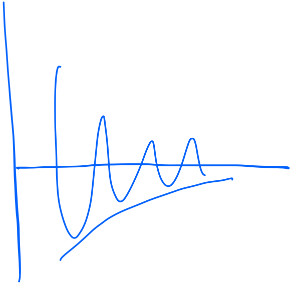
MA(1)



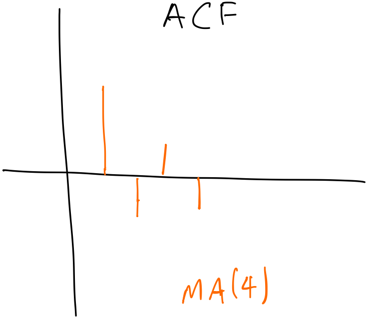


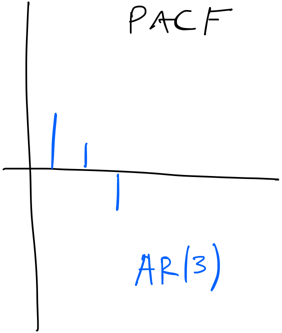
MA(q)





EXTRA JUST TO RECAP *





* from theory / reading ACF / PACF we can tell either AR or MA orders

EGYPTIAN EXPORTS

ARIMA(2,0,1) with constant.

$$(1 - \phi_1 B - \phi_2 B^2)\, y_t = c + (1 + \theta_1 B)\, \varepsilon_t$$

EGYPTIAN EXPORTS

ARIMA(2,0,1) with constant.

$$(1 - \phi_1 B - \phi_2 B^2)\, y_t = c + (1 + \theta_1 B)\, \varepsilon_t$$

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$\Rightarrow \quad y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$\Rightarrow \quad y_t = 2.562 + 1.676\, y_{t-1} - 0.803\, y_{t-2} - 0.69\, \varepsilon_{t-1} + \varepsilon_t$$

where $\varepsilon_t \sim NID(0, 8.046)$

or $\hat{\sigma} = \sqrt{8.046} = 2.837$