ARMA models

- Extending auto-regressive models with smoother noise.

  In AR model for each t, we associate an independent noise w_t

Rice production in Maharashtra

AR(1)

MA terms.

$$x_t = \eta + \phi_1 x_{t-1} + w_t + \theta_1 w_{t-1}$$

$$x_t = \eta + \phi_1 x_{t-1} + \dots \phi_p x_{t-p} + \boxed{\theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots \theta_p w_{t-p}}$$

$+ w_t$

Need smoother handling of noise.

  A moving average (MA) model provides that.

- ARMA models: AR models + MA models

Moving average models (MA models)

- A value x_t in a time-series sometimes cannot be explained just in terms of its past values.
- External (unknown) variables might be influencing the values
    - Example: Total wheat export of India in 2023 can be determined by wheat export in 2022, but also other external factors like weather patterns, war, exchange rates, etc.
- External variables are also time-varying → errors at each position cannot be independent.
- Moving average models capture dependency on such external unknowns.

Properties of a series following MA(1) model

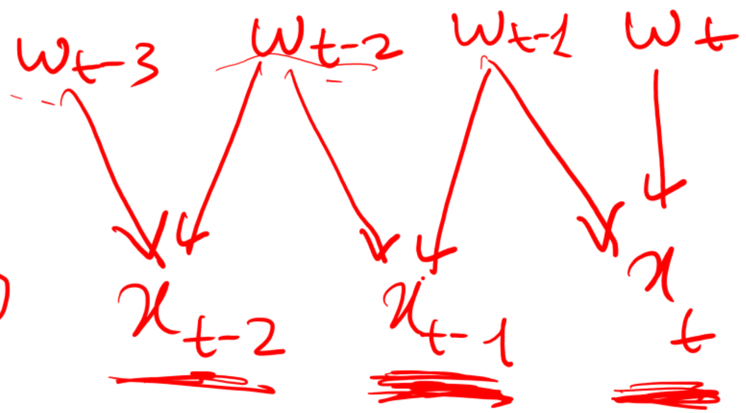
$p = 0,\ q = 1$

$x_t = \eta + \theta_1 w_{t-1} + w_t$

$x_{t-1} = \eta + \theta_1 w_{t-2} + w_{t-1}$

$x_{t-2} = \eta + \theta_1 w_{t-3} + w_{t-2}$

$E(x_t) =$



- Mean is $E(x_t) = \mu$
- Variance is $Var(x_t) = \sigma_w^2(1 + \theta_1^2)$
- Autocorrelation function (ACF) is:

$E(x_t x_{t-1})$

$$\rho_1 = \frac{\theta_1}{1 + \theta_1^2}, \text{ and } \rho_h = 0 \text{ for } h \geq 2$$

Proofs here: https://online.stat.psu.edu/stat510/lesson/2/2.1#paragraph--264

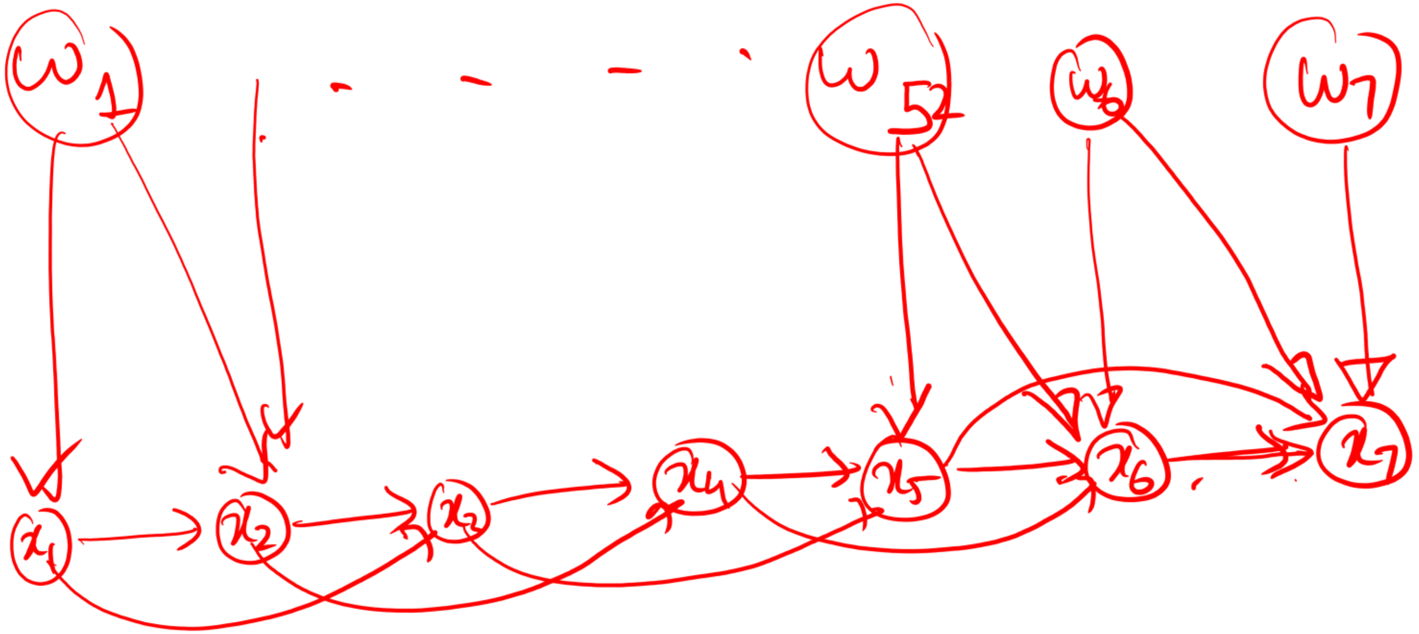
Pictorial representation of dependency.

$$E(x_t x_{t-1}) = E((n + \theta_1 w_{t-1} + w_t) x_{t-1})$$

$$= n E(x_{t-1}) + \theta_1 E(w_{t-1} \cdot x_{t-1}) + E(w_t x_{t-1})$$

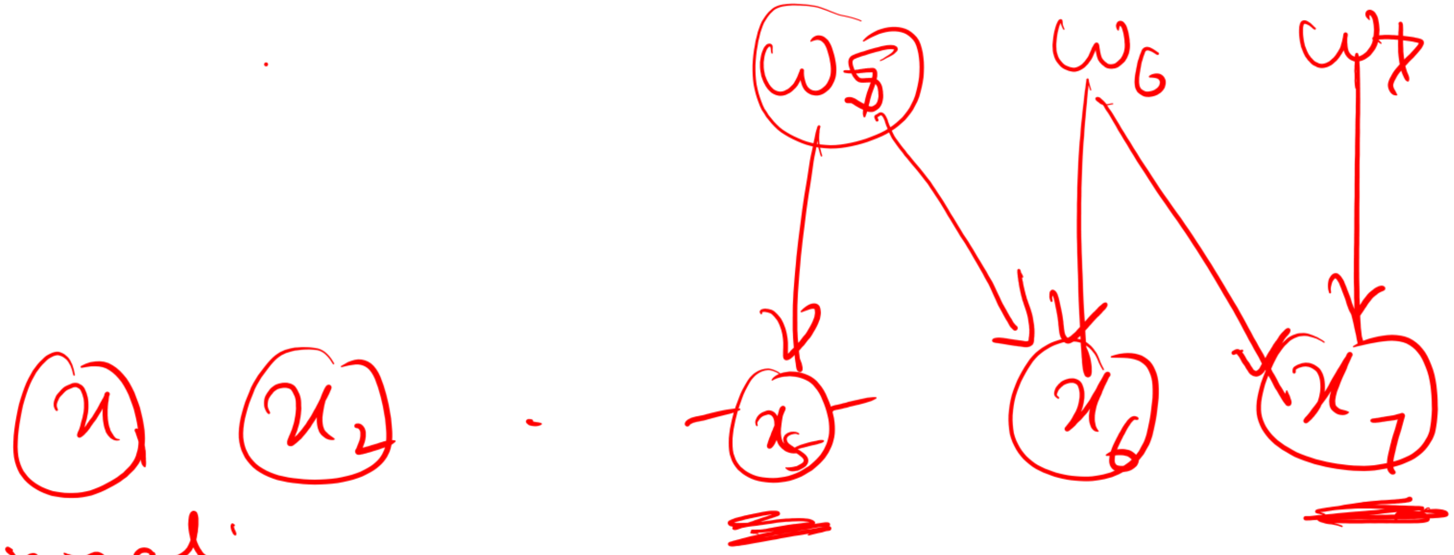
$$=$$

MA( )



AR(2)

$p = 0, \quad q = 1$



Original:

$x_t = \theta_1 w_{t-1} + w_t + \eta$: To determine PACF$(x_t, x_{t-2}) = \phi_2$

$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \tilde{\eta}$

ARMA (p,q) model

Each xt depends on p previous x-values, and q-previous error values

$$x_t = \eta + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

Estimating all the parameters of this model is not as straightforward as least-square regression since the w_t values are not observed (Not covered)

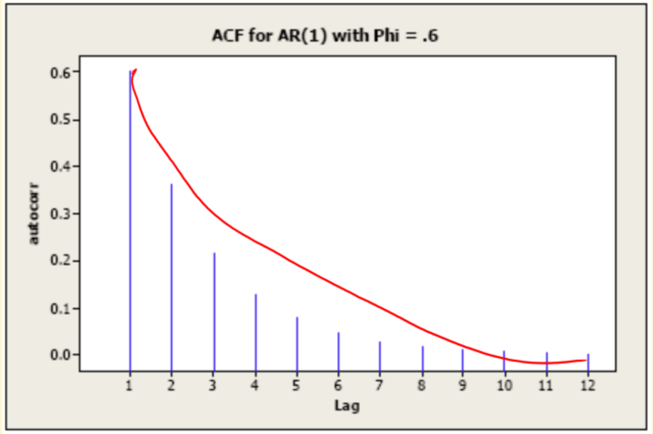
Comparing AR(1) and MA(1) on ACF and PACF

- ACF=plain correlation
- PACF(x_t, x_{t-2})=conditional correlation or what extra contribution you get from x_{t-2} after you x_{t-1}

Shape of ACF and PACF of a series following AR(1) model

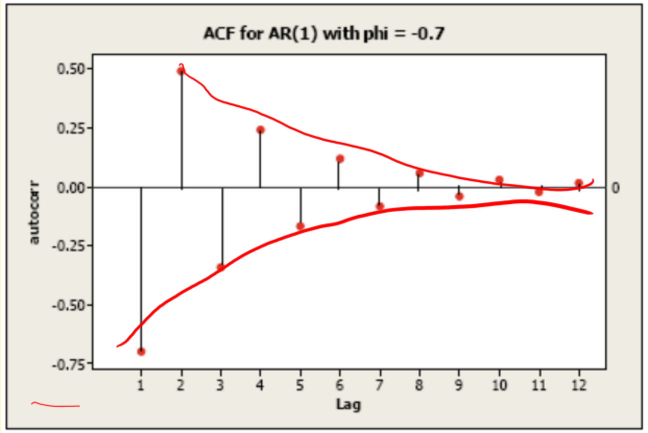
Following is the ACF of an AR(1) with $\phi_1 = 0.6$, for the first 12 lags.

Note!
The tapering pattern:

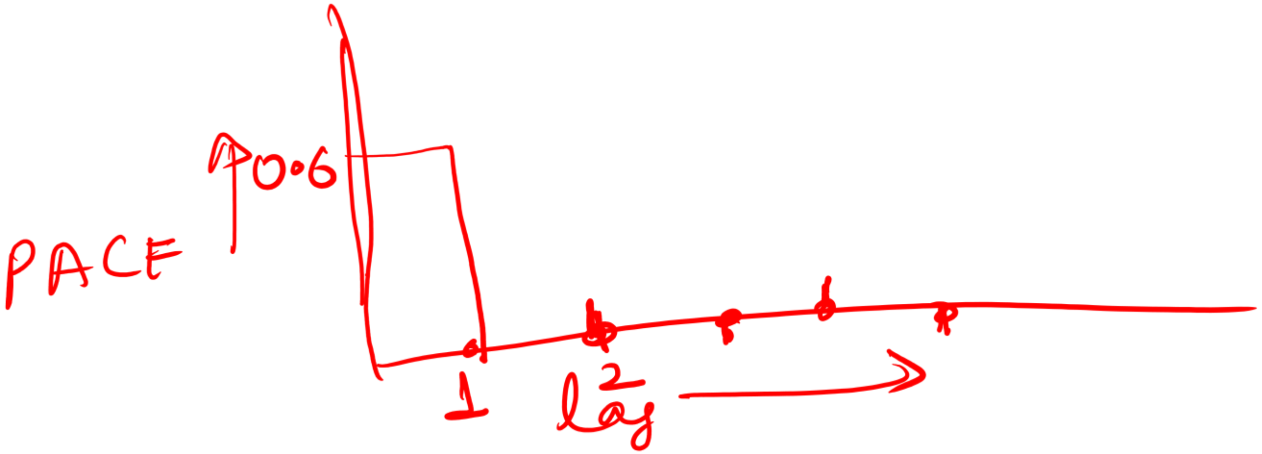


The ACF of an AR(1) with $\phi_1 = -0.7$ follows.

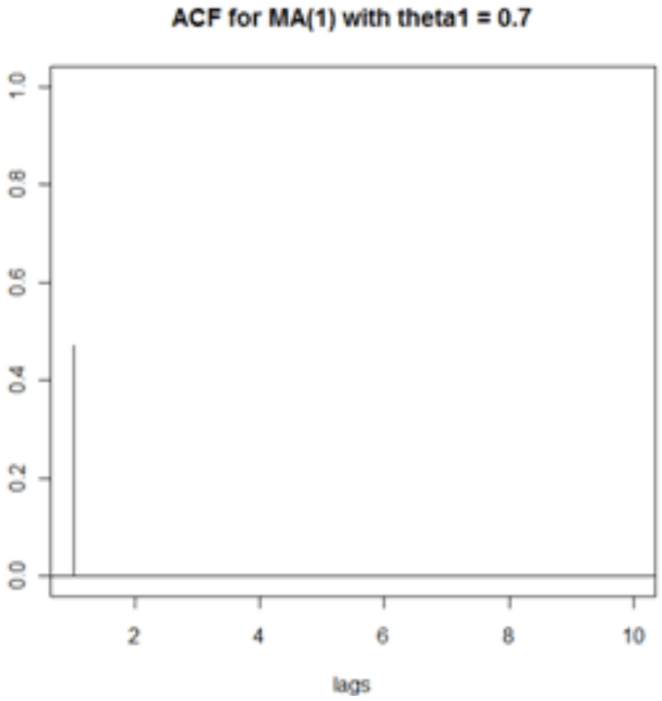
Note!
The alternating and tapering pattern.

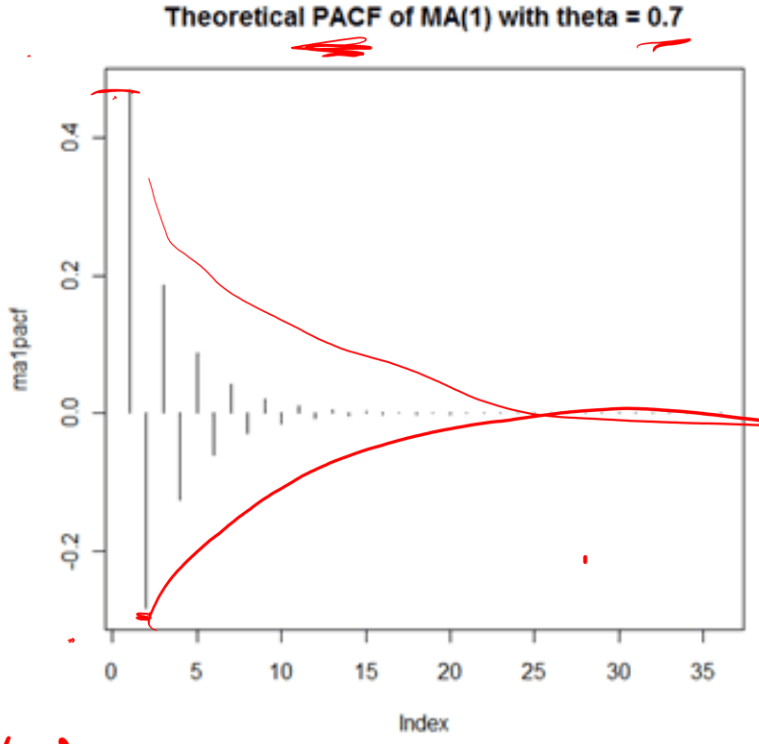


$$PACF(x_t, x_{t-1}) = ACF(x_t, x_{t-1}) \; ; \; PACF(x_t, x_{t-2}) = 0$$



Shape of ACF and PACF of a series following MA(1) model





$$\text{PACF}(x_t, x_{t+2}) = -\theta_1^2$$

$$x_t = \theta_1 w_{t-1} + w_t \qquad ; \qquad w_t = x_t - \theta_1 w_{t-1}$$

$$= \theta_1 (x_{t-1} - \theta_1 w_{t-2}) + w_t$$

$$= \theta_1 (x_{t-1} - \theta_1 (x_{t-2} - \theta_1 w_{t-3})) + w_t$$

$$= \boxed{\theta_1 x_{t-1} - \theta_1^2 x_{t-2} + \theta_1^3 w_{t-3} + w_t}$$

Choosing p,q

- Data may follow an ARIMA(p,d,0) model if the ACF and PACF plots of the differenced data show the following patterns:
  - the ACF is exponentially decaying or sinusoidal;
  - there is a significant spike at lag p in the PACF, but none beyond lag p.

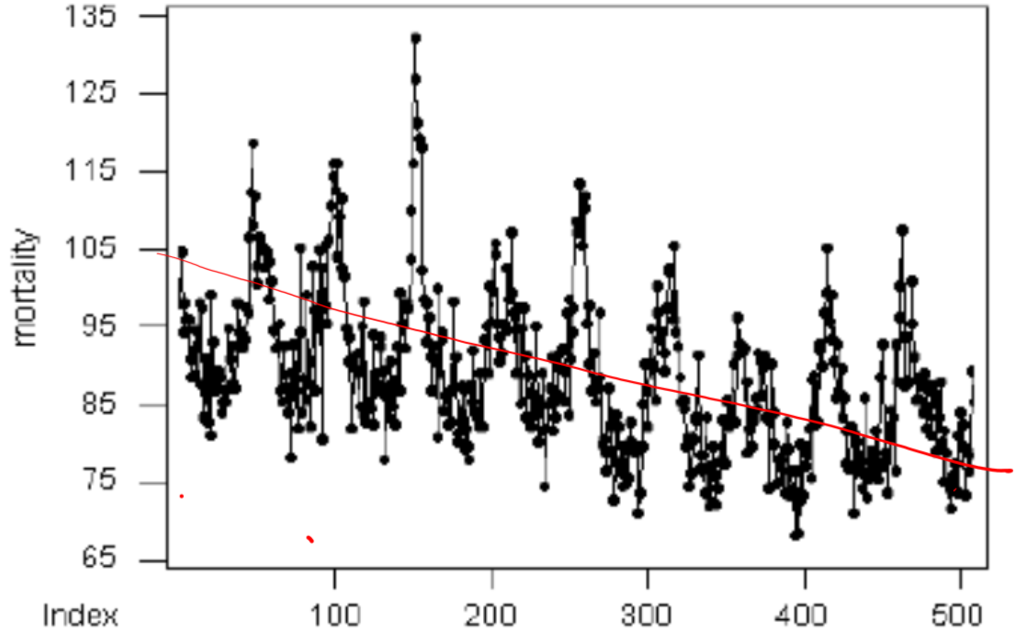
- The data may follow an ARIMA(0,d,q) model if the ACF and PACF plots of the differenced data show the following patterns:
  - The PACF is exponentially decaying or sinusoidal;
  - There is a significant spike at lag q in the ACF, but none beyond lag q.

. [Tips given in this link]()

Handling trend in time-series.

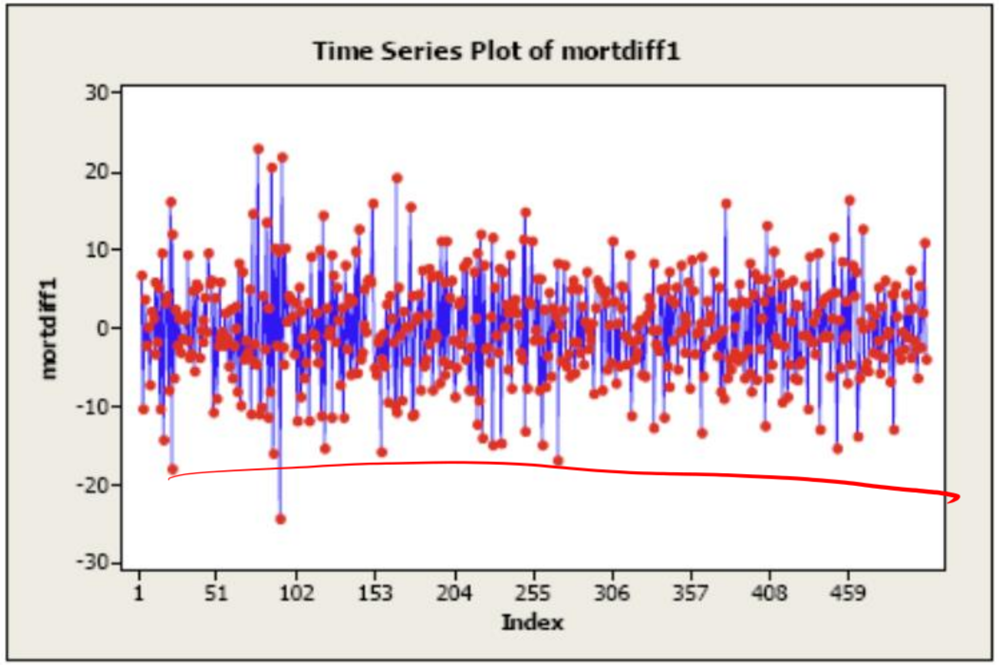
$x_t = \alpha t + \phi_1 x_{t-1} + \theta_1 \omega_{t-1} + \omega_t$

- If a time-series has a linear trend, then replace each value x_t with difference of consecutive x-values
  - $y_t = x_t - x_{t-1}$



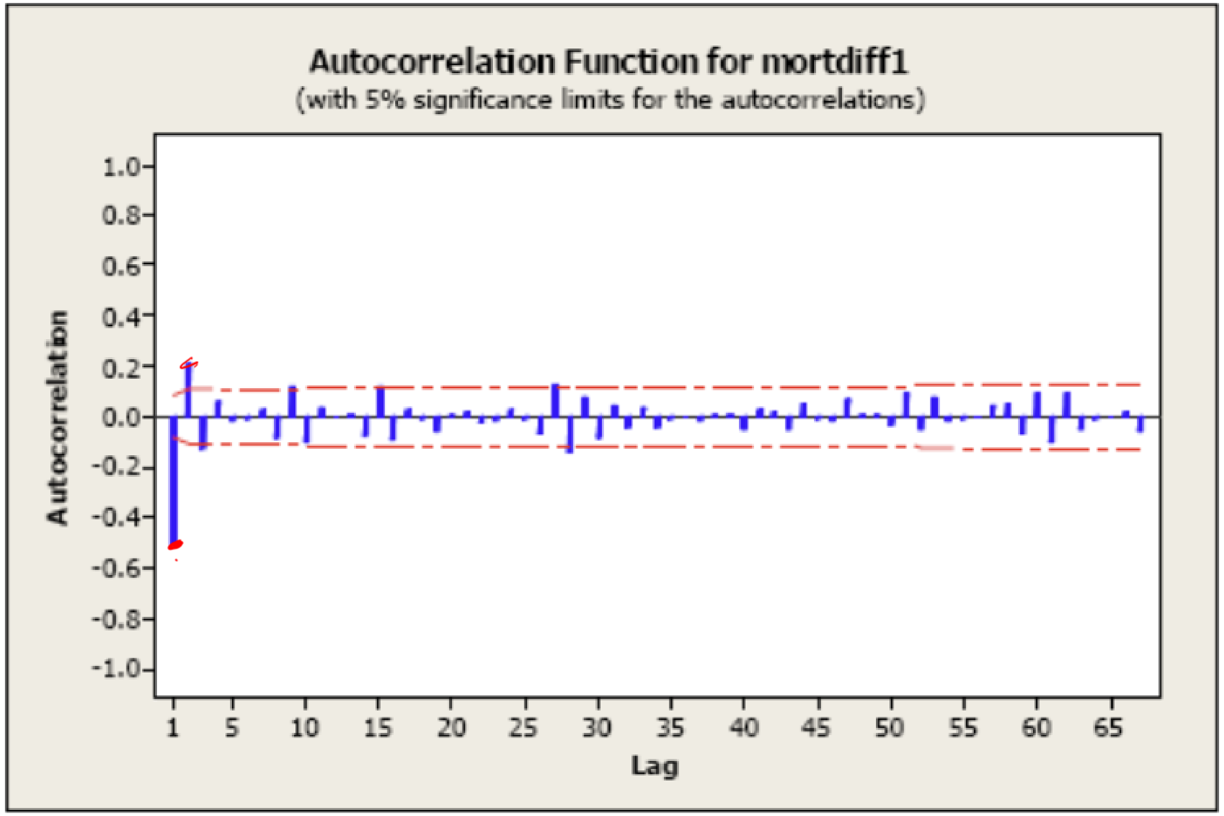
Daily cardiovascular mortality rate in Los Angeles County, 1970-1979.

Clear downward trend.



Plot of first differences

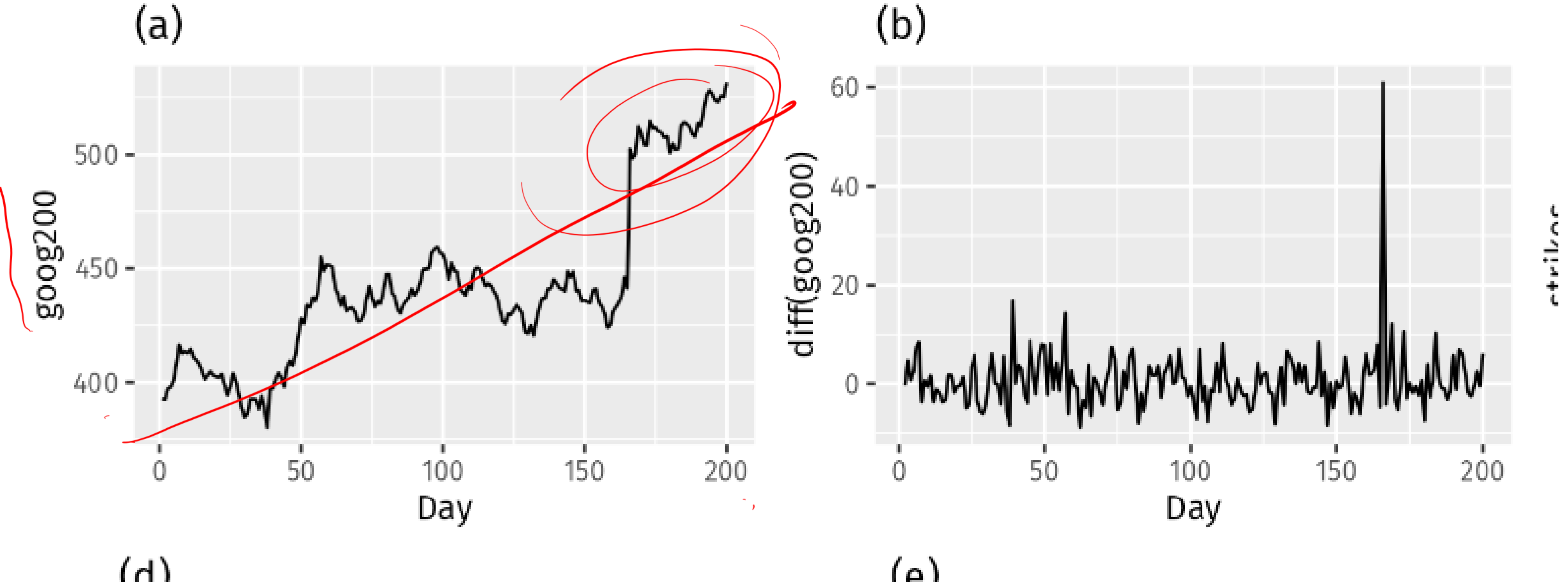
ACF of first differences.



Lag.	ACF
1.	-0.506029
2.	0.205100
3.	-0.126110
4.	0.062476
5.	-0.015190

$$\hat{y}_t = -0.04627 - 0.50636 y_{t-1}$$

Another example



ARIMA(p,d,q) models

- p is the order of the autoregressive part,

- d is the degree of first difference involved,

- q is the order of the moving average part.

Example: ARIMA(2,1,1) model

Incorporating seasonality.

- Seasonality in a time series is a regular pattern that repeats over S time periods.
  - Example: monthly seasonality repeats over S=12 (months of the year)
  - Example: quarter seasonality repeats over S=4 period

- Extending ARIMA to handle seasonality. One or more of the above might work
  - Introduce a AR term x_{t-S} in the model for every period S.
  - Introduce MA term w_{t-S} in the model for every period S.
  - Create seasonal differences $y_t = x_t - x_{t-S}$

Demo

- https://colab.research.google.com/drive/1Z4zNI_bVXoFQBsCHUtxBDCBno6yhXceB?usp=sharing#scrollTo=deWKK_D1mNlr