# Modeling cycles: AR( ), MA( ), and ARMA( )

Dr. Natalia Khorunzhina

CBS

# Cycles

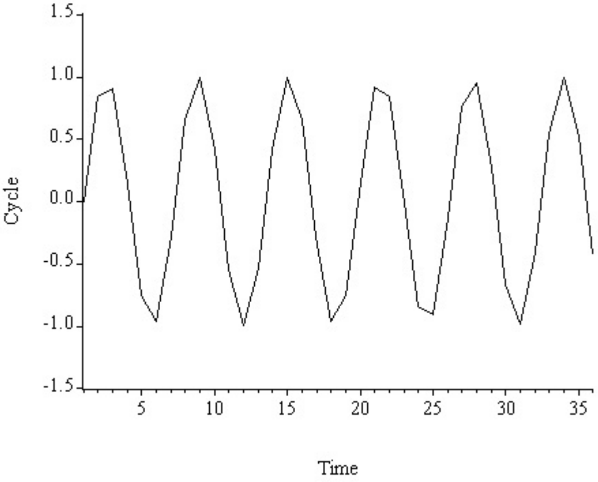
- Recall the component decomposition

$$\mu_t = T_t + S_t + C_t$$

- The cycle component $C_t$ should be free of trend and seasonal
- We will focus on pure cycle models

$$\mu_t = C_t$$

# A Rigid Cyclical Pattern



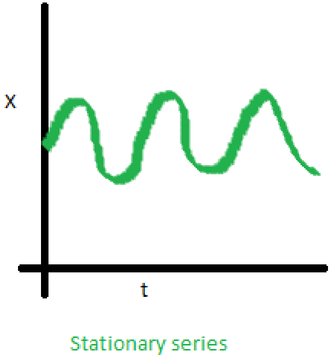
# Mean Stationary

- **Definition:** A time series $Y_t$ has a constant mean, or is **mean stationary**, if
$$E(Y_t) = \mu$$
is constant (stable) over time.

- Counter-example:
  - A trended time series is not mean stationary

- We assume the cyclical component $C_t$ is mean stationary

# Mean Stationary




# Variance Stationary

- **Definition:** A time series $Y_t$ has a constant variance, or is **variance stationary**, if

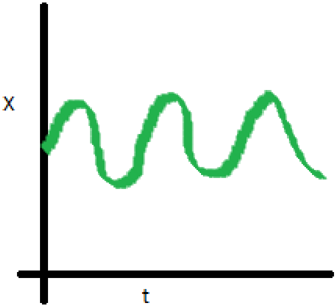
$$var(Y_t) = \sigma^2$$

  is constant (stable) over time.

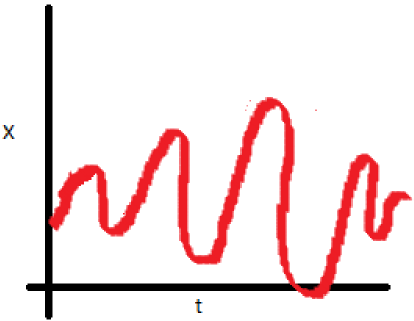
- Counter-example:
  - A time-series with trended (increasing) variance is not variance stationary

- We assume the cyclical component $C_t$ is variance stationary

# Variance Stationary




Stationary series




Non-Stationary series

# Covariance

- The covariance of two random variables X and Z is

$$cov(X, Z) = E((X - EX)(Z - EZ))$$

- The covariance measures the linear dependence between X and Z

# Autocovariance

- The first **autocovariance** of a time series $Y_t$ is the covariance of $Y_t$ with its value in the preceding time period $Y_{t-1}$
- We call $Y_{t-1}$ the first **first lag** of $Y_t$
- We write the first autocovariance as

$$\begin{aligned}\gamma(1) &= cov(Y_t, Y_{t-1}) \\ &= E((Y_t - \mu)(Y_{t-1} - \mu))\end{aligned}$$

# Autocovariances

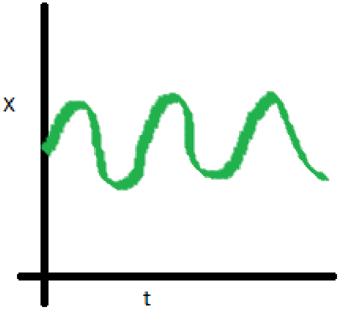
- The k'th **autocovariance** of a time series $Y_t$ is the covariance of $Y_t$ with its lag $Y_{t-k}$
- It is written as

$$\begin{aligned}\gamma(k) &= cov(Y_t, Y_{t-k}) \\ &= E((Y_t - \mu)(Y_{t-k} - \mu))\end{aligned}$$

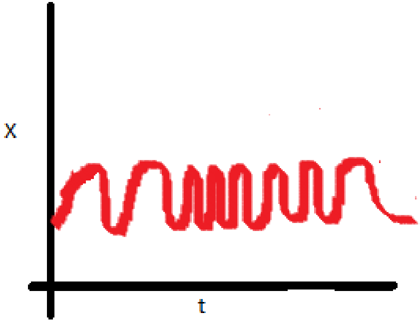
# Covariance Stationary

- **Definition**: A time series $Y_t$ is **covariance stationary** if its mean $E(Y_t)$, variance, and autocovariance function $\gamma(k)$ are constant (stable) over time
- Counter-example:
  - A time-series with changing correlations is not covariance stationary
- We assume the cyclical component $C_t$ is covariance stationary

# Covariance Stationary




Stationary series




Non-Stationary series

# Correlation

- The correlation normalizes the covariance

$$corr(X,Z) = \frac{cov(X,Z)}{\sqrt{var(X)var(Z)}}$$

- Correlations lie between -1 and 1
  - $corr(X,Z) = 0$ means no linear association
  - $corr(X,Z) = 1$ means $X = Z$
  - $corr(X,Z) = -1$ means $X = -Z$

# Autocorrelation

- The first **autocorrelation** of a time series $Y_t$ is the correlation of $Y_t$ with $Y_{t-1}$
- We write the first autocorrelation as

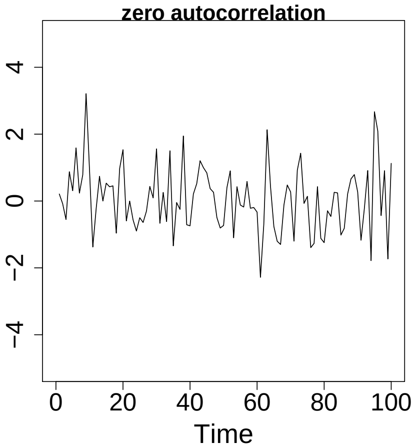
$$\begin{aligned}\rho(1) &= corr(Y_t, Y_{t-1}) \\ &= \frac{cov(Y_t, Y_{t-1})}{\sqrt{var(Y_t)var(Y_{t-1})}} \\ &= \frac{cov(Y_t, Y_{t-1})}{var(Y_t)}\end{aligned}$$

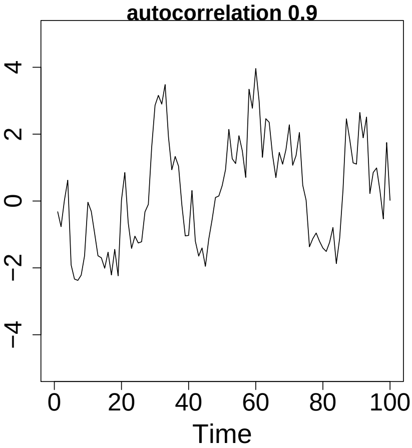
- The third equality holds by variance stationarity

# Autocorrelation

- The autocorrelation $\rho(1)$ lies between -1 and 1
- $\rho(1)$ is close to 1 for highly correlated series
- $\rho(1)$ is close to -1 if the correlation is negative - if there are movements back and forth
- $\rho(1) = 0$ if the series is uncorrelated

# Autocorrelation: two examples



# Autocorrelations

- The k'th **autocorrelation** of a time series $Y_t$ is the correlation of $Y_t$ with $Y_{t-k}$

- It is written as

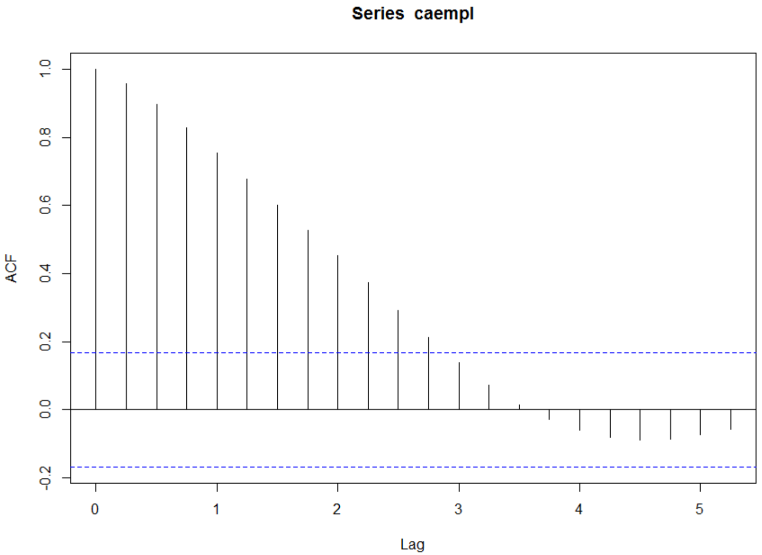
$$\rho(k) = \frac{cov(Y_t, Y_{t-k})}{\sqrt{var(Y_t)var(Y_{t-k})}} = \frac{cov(Y_t, Y_{t-k})}{var(Y_t)}$$

- Autocorrelations lie between -1 and 1

# Autocorrelation Function

- The autocovariance $\gamma(k)$ and autocorrelation $\rho(k)$ are functions of the lag $k$.
- We call $\rho(k)$ the **autocorrelation function.**
- Plotted as a function of $k$ it shows us how the dependence pattern alters with the lag.

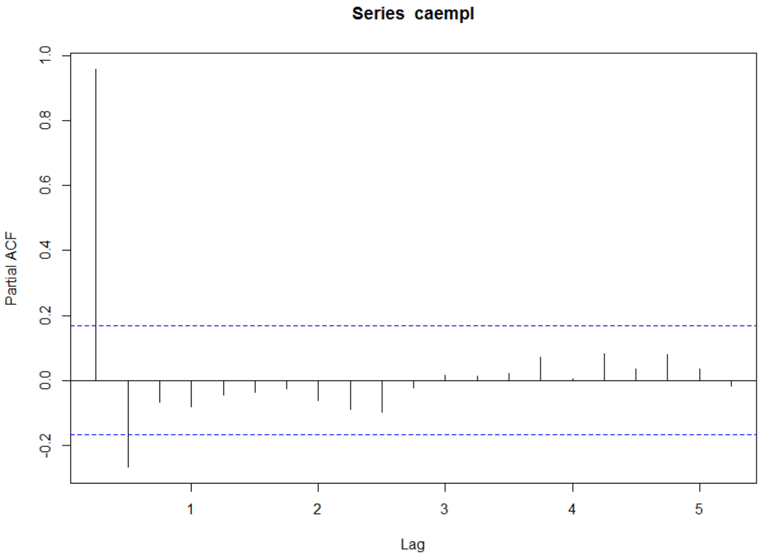
# Autocorrelation Function



# Partial Autocorrelation Function

- **Partial autocorrelation function**, $p(k)$ is sometimes useful.
- $p(k)$ is just the coefficient on $y_{t-k}$ in a population linear regression of $y_t$ on $y_{t-1}$, ..., $y_{t-k}$.
- We call such a regression an **autoregression**, because the variable is regressed on lagged values of itself.
- The partial autocorrelations measure the association between $y_t$ and $y_{t-k}$ after *controlling* for the effects of $y_{t-1}$, ..., $y_{t-k+1}$.

# Partial Autocorrelation Function



# MA(1) Process

- The **first-order moving average** process, or **MA(1)** process, is

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

  where $\varepsilon_t$ is $WN(0, \sigma^2)$

- The MA coefficient $\theta$ controls the degree of serial correlation. It may be positive or negative.

- The innovations $\varepsilon_t$ impact $y_t$ over two periods
  - An contemporaneous (same period) impact
  - A one-period delayed impact

# Mean of MA(1)

- The unconditional mean of $y_t$ is

$$
\begin{aligned}
E(y_t) &= E(\varepsilon_t + \theta\varepsilon_{t-1}) \\
&= E(\varepsilon_t) + \theta E(\varepsilon_{t-1}) \\
&= 0
\end{aligned}
$$

# Variance of MA(1)

- The unconditional variance of $y_t$ is

$$\begin{aligned} var(y_t) &= var(\varepsilon_t + \theta\varepsilon_{t-1}) \\ &= var(\varepsilon_t) + var(\theta\varepsilon_{t-1}) + 2cov(\varepsilon_t, \theta\varepsilon_{t-1}) \\ &= \sigma^2 + \theta^2\sigma^2 + 0 \\ &= (1+\theta^2)\sigma^2 \end{aligned}$$

- This is a function of both the innovation variance $\sigma^2$ and the MA coefficient $\theta$.

# Autocovariance of MA(1)

- The first autocovariance is

$$
\begin{aligned}
\gamma(1) &= E(y_t y_{t-1}) \\
&= E((\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})) \\
&= E(\varepsilon_t\varepsilon_{t-1}) + \theta E(\varepsilon_{t-1}^2) + \theta E(\varepsilon_t\varepsilon_{t-2}) + \theta^2 E(\varepsilon_{t-1}\varepsilon_{t-2}) \\
&= 0 + \theta E(\varepsilon_{t-1}^2) + 0 + 0 \\
&= \theta\sigma^2
\end{aligned}
$$

# Autocovariance of MA(1)

- The autocovariance for $k > 1$ are

$$
\begin{aligned}
\gamma(k) &= E(y_t y_{t-k}) \\
&= E((\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-k} + \theta\varepsilon_{t-k-1})) \\
&= E(\varepsilon_t\varepsilon_{t-k}) + \theta E(\varepsilon_{t-1}\varepsilon_{t-k}) + \theta E(\varepsilon_t\varepsilon_{t-k-1}) + \theta^2 E(\varepsilon_{t-1}\varepsilon_{t-k-1}) \\
&= 0 + 0 + 0 + 0 \\
&= 0
\end{aligned}
$$

- Thus the autocovariance function is zero for $k > 1$

# Autocorrelations of MA(1)

- Since

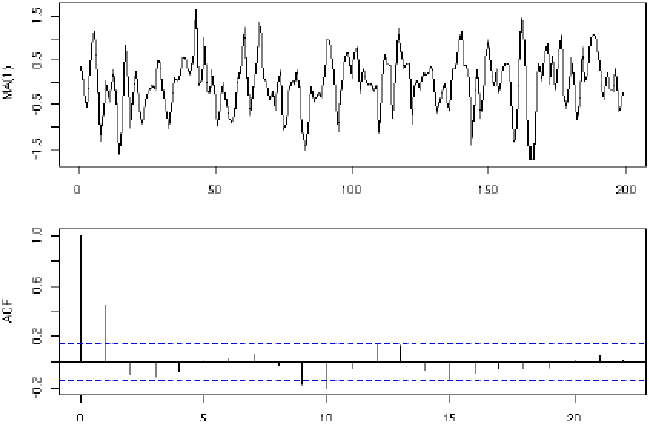
$$\gamma(0) = var(y_t) = (1+\theta^2)\sigma^2$$
$$\gamma(1) = \theta\sigma^2$$
$$\gamma(k) = 0, k \geq 2$$

then

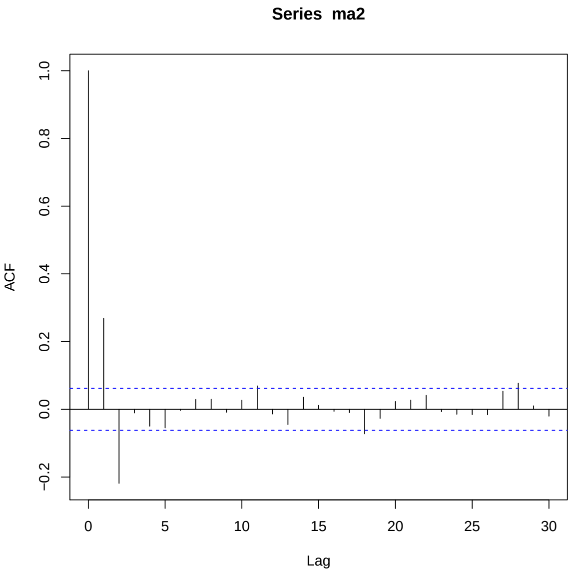
$$\rho(1) = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2}$$
$$\rho(k) = 0, k \geq 2$$

- The autocorrelation function of an MA(1) is zero after the first lag.

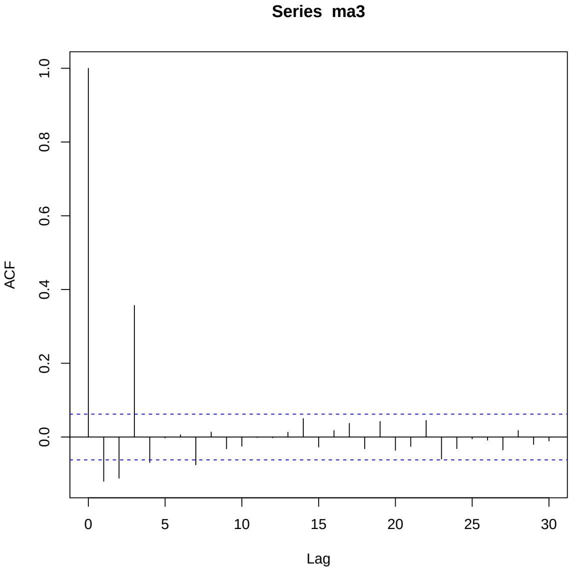
# Autocorrelation Function of MA(1)



# ACF for MA(2)



# ACF for MA(3)



# AR(1) Process

- The first-order autoregressive process, AR(1) is

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

where $\varepsilon_t$ is $WN(0, \sigma^2)$

# Variance of AR(1)

- Take variance of both sides of

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

- Thus

$$\begin{aligned} var(y_t) &= var(\beta y_{t-1} + \varepsilon_t) \\ &= var(\beta y_{t-1}) + var(\varepsilon_t) \\ &= \beta^2 var(y_{t-1}) + \sigma^2 \end{aligned}$$

- If y is variance stationary, we solve and find

$$var(y_t) = var(y_{t-1}) = \frac{\sigma^2}{1 - \beta^2}$$

# $|\beta| < 1$

- We calculated that

$$var(y_t) = \beta^2 var(y_{t-1}) + \sigma^2$$

- When $\mid \beta \mid = 1$, then

$$var(y_t) = var(y_{t-1}) + \sigma^2 > var(y_{t-1})$$

so the variance is increasing with t

- $\mid \beta \mid = 1$ is inconsistent with variance stationarity.
- $\mid \beta \mid < 1$ is necessary for stationarity.

# Random Walk

- An AR(1) with $\beta = 1$ is known as a random walk or unit root process

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

- By back-substitution

$$y_t = y_0 + \sum_{i=0}^{t} \varepsilon_{t-i}$$

- The past never disappears. Shocks have permanent effects.

# Autocovariance of AR(1)

- Take the equation

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

- And then multiply both sides by $y_{t-k}$

$$y_{t-k} y_t = \beta y_{t-k} y_{t-1} + y_{t-k} \varepsilon_t$$

- Then take expectations. Since $\varepsilon_t$ is white noise, it is uncorrelated with

$$E(y_{t-k} y_t) = \beta E(y_{t-k} y_{t-1}) + E(y_{t-k} \varepsilon_t)$$

or

$$\gamma(k) = \beta\gamma(k-1)$$

# Autocorrelations of AR(1)

- Dividing by the variance, this implies

$$\rho(k) = \beta\rho(k-1)$$

- We know

$$\rho(0) = 1$$

- Then

$$\begin{aligned}\rho(1) &= \beta\rho(0) = \beta \\ \rho(2) &= \beta\rho(1) = \beta^2 \\ &\vdots \\ \rho(k) &= \beta^k\end{aligned}$$

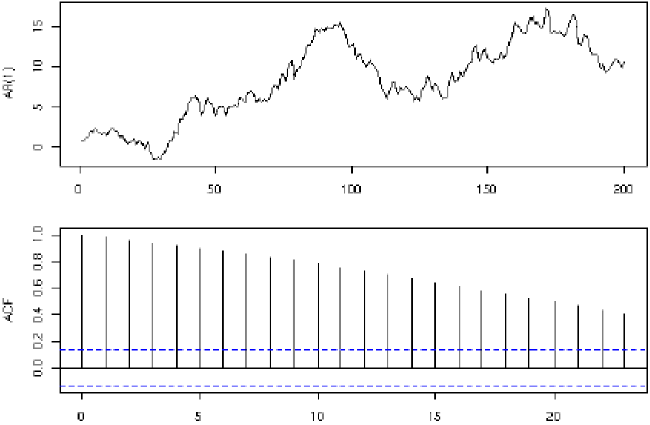
# Autocorrelations of AR(1)

- We have derived

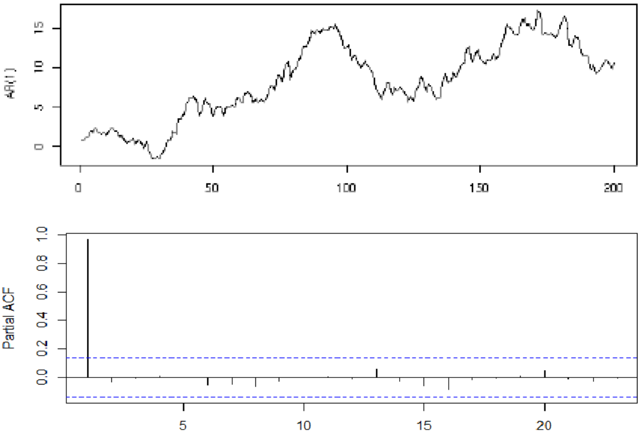
$$\rho(k) = \beta^k$$

- The autocorrelation of the stationary AR(1) is a simple geometric decay ($|\beta| < 1$)
- If $\beta$ is small, the autocorrelations decay rapidly to zero with k
- If $\beta$ is large (close to 1), then the autocorrelations decay moderately
- The AR(1) parameter describes the persistence in the time series

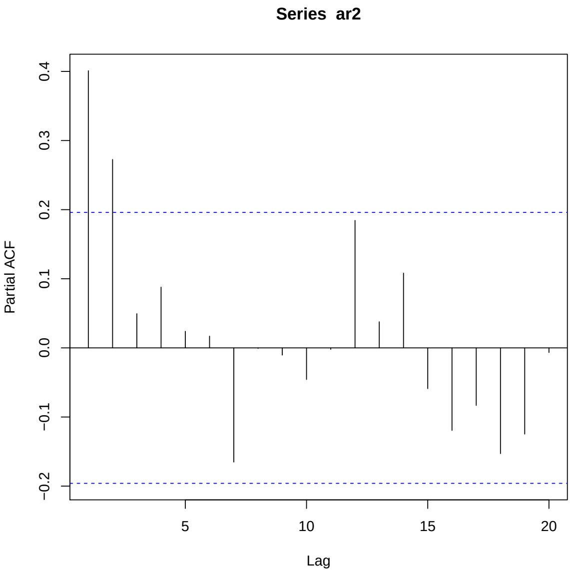
# Autocorrelation Function of AR(1)



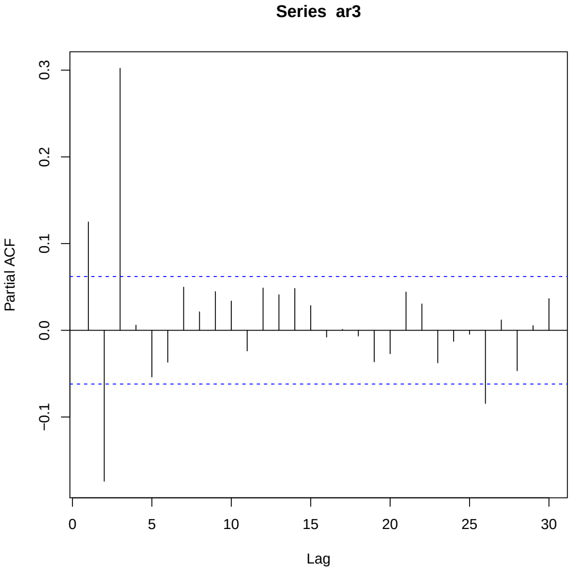
# Autocorrelation Function of AR(1)



# PACF of AR(2)



# PACF of AR(3)



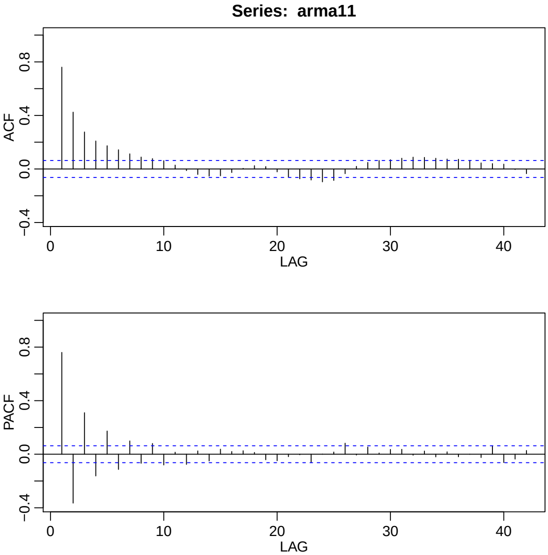
# ARMA(1,1) Process

- The random shocks that drives an autoregressive process is itself a moving average process.
- It can arise from aggregation (sum of AR or of AR and MA).
- AR processes observed with measurement error.
- ARMA(1,1)

$$y_t = \beta y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

# ACF and PACF ARMA(1,1)



# ARMA(p,q) Process

- ARMA models have a fixed unconditional mean but a time-varying conditional mean.
- Neither the autocorrelation nor partial autocorrelation functions of ARMA process cut off at any particular displacement.
- Autocorrelation and partial autocorrelation functions damp gradually, with the precise pattern depending on the process.

# Akaike Information criterion (AIC)

The Akaike Information criterion is effectively an estimate of the out-of-sample forecast error variance:

$$AIC = \exp\left\{\frac{2k}{T}\right\}\frac{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}{T}$$

where $k$ is the number of estimated parameters.

- Larger models have smaller sum of squared residuals, but larger $k$.
- AIC is designed to find models with low forecast risk.
- AIC assumes all models are approximations, and is trying to find the model which makes the best forecast.
- Akaike recommended selecting forecasting models by finding the one model with the smallest AIC.

# Best fitting ARMA(p,q) model

- Neither the ACF nor PACF of ARMA process cut off at any particular displacement. Therefore, ACF and PACF may not be helpful in determining the order $p$ and $q$ for the model.
- One way of choosing the best fitting model is to try them all and compare.
- We can estimate as many ARMA(p,q) models as we can and choose the one with the smallest AIC.
- If you let both $p$ and $q$ vary between 1 and 4, you will estimate and consider 16 different models.

# AIC for ARMA(p,q)

| | MA(q) | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| AR(p) | | | | | |
| 1 | | 502.0 | 497.5 | 498.4 | 499.1 |
| 2 | | 494.8 | 496.6 | 497.8 | 499.8 |
| 3 | | 497.0 | 498.7 | 499.8 | 501.4 |
| 4 | | 498.8 | 499.7 | 501.6 | 503.6 |

# AIC for ARMA(p,q)

| | MA(q) | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| AR(p) | | | | | |
| 1 | | 502.0 | 497.5 | 498.4 | 499.1 |
| 2 | | **494.8** | 496.6 | 497.8 | 499.8 |
| 3 | | 497.0 | 498.7 | 499.8 | 501.4 |
| 4 | | 498.8 | 499.7 | 501.6 | 503.6 |

# Best fitting ARMA(p,q) model using *auto.arima*

- We can let the algorithm in *auto.arima* to pick the best fitting model.
- The algorithm loops over pairwise values of $p \in \{0, 1, 2, 3, ...\}$ and $q \in \{0, 1, 2, 3, ...\}$ to estimate ARMA(p,q) and calculate the AIC.
- The algorithm will choose the order $p$ and $q$ in ARMA(p,q) such that the AIC is the smallest.