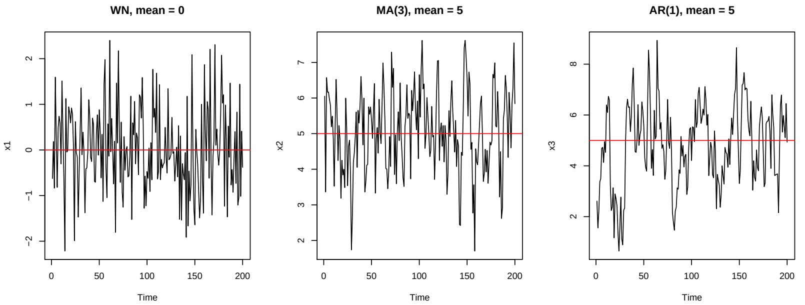
01 Stationary time series

Andrius Buteikis, andrius.buteikis@mif.vu.lt
http://web.vu.lt/mif/a.buteikis/

Introduction

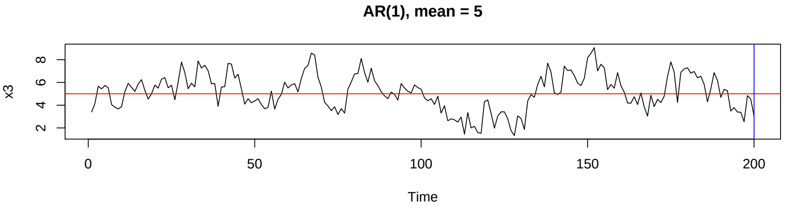
All time series may be divided into two big classes - stationary and non-stationary.

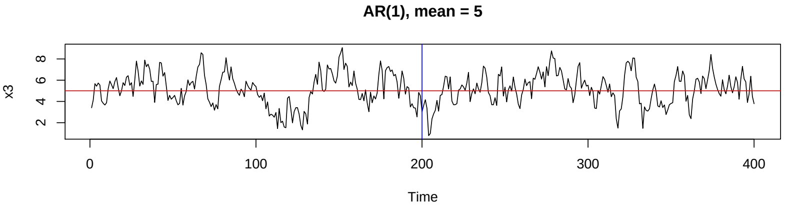
- **Stationary process** - a *random* process with a constant mean, variance and covariance. Examples of stationary time series:



The three example processes fluctuate around their constant mean values. Looking at the graphs, the fluctuations of the first two graphs seem to be constant, however the third one is not so apparent.

If we were to examine a longer time period of the last time series:





We can see that the fluctuations are indeed around a constant mean and the variance does not appear to change throughout the period.

Some **non-stationary** time series examples:

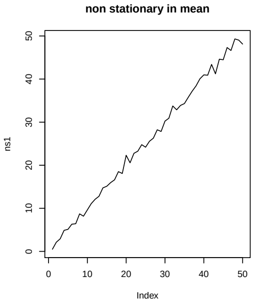
- $Y_t = t + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, 1)$;
- $Y_t = \epsilon_t \cdot t$, where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$;
- $Y_t = \sum_{j=1}^{t} Z_j$, where each independent variable Z_j is either 1 or -1, with a 50% probability for either value.

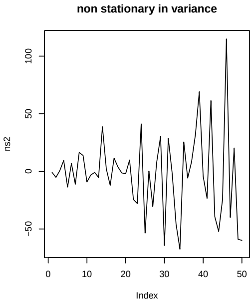
The reasons for their non-stationarity are as follows:

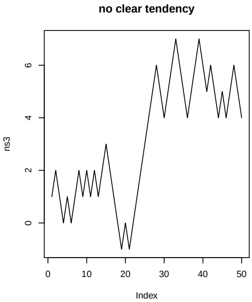
- The first time series is not stationary because its mean is not constant: $\mathbb{E}Y_t = t$ - depends on t;
- The second time series is not stationary because its variance is not constant: $\mathbb{V}\mathrm{ar}(Y_t) = t^2 \cdot \sigma^2$ - depends on t.
  However, $\mathbb{E}Y_t = 0 \cdot t = 0$ is constant;
- The third time series is not stationary because even though $\mathbb{E}Y_t = \sum_{j=1}^{t} (0.5 + (-0.5)) = 0$, the variance $\mathbb{V}\mathrm{ar}(Y_t) = \mathbb{E}(Y_t^2) - (\mathbb{E}(Y_t))^2 = \mathbb{E}(Y_t^2) = t$ where:

$$\mathbb{E}(Y_t^2) = \sum_{j=1}^{t} \mathbb{E}(Z_j^2) + 2\sum_{j \neq k} \mathbb{E}(Z_j Z_k) = t \cdot (0.5 \cdot 1 + 0.5 \cdot (-1)^2) = t$$

The sample data graphs are provided below:





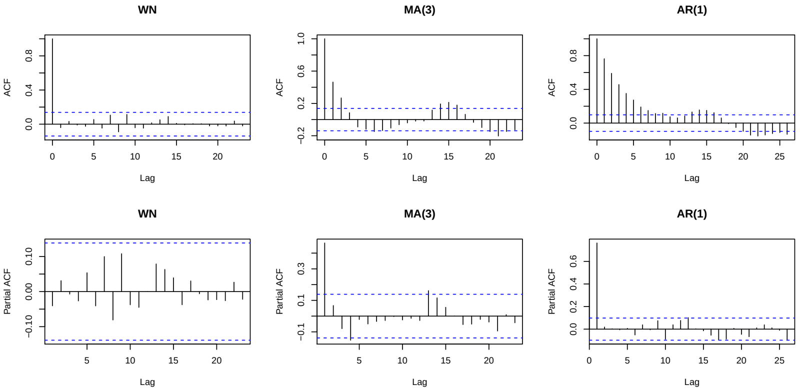


- **White noise** (*WN*) - a **stationary** process of uncorrelated (sometimes we may demand a stronger property of *independence*) random variables with zero mean and constant variance. White noise is a model of an absolutely chaotic process of uncorrelated observations - it is a process that immediately forgets its past.

How can we know which of the previous three stationary graphs are not *WN*? Two functions help us determine this:

- ACF - Autocorrelation function
- PACF - Partial autocorrelation function

If all the bars (except the 0th in the ACF) are within the blue band - the stationary process is *WN*.

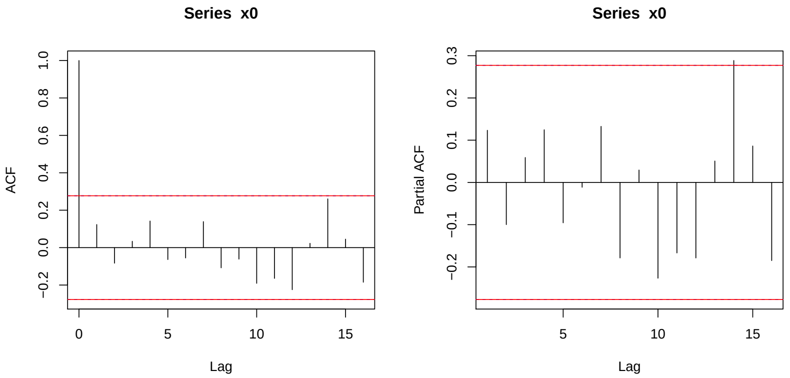


The 95% confidence intervals are calculated from:

```
qnorm(p = c(0.025, 0.975))/sqrt(n)
```

(more details on the confidence interval calculation are provided later in these slides)

```r
par(mfrow = c(1, 2))
set.seed(10)
n = 50
x0 <- rnorm(n)
acf(x0)
abline(h = qnorm(c(0.025, 0.975))/sqrt(n), col = "red")
pacf(x0)
abline(h = qnorm(c(0.025, 0.975))/sqrt(n), col = "red")
```



To decide whether a time series is stationary, examine its graph.

To decide whether a stationary time series is WN, examine its ACF and PACF.

Covariance-Stationary Time Series

- In **cross-sectional** data different observations were assumed to be uncorrelated;
- In **time series** we require that there be some dynamics, some persistence, some way in which the present is linked to the past and the future - to the present. Having historical data then would allow us to forecast the future.

If we want to forecast a series - at a minimum we would like its mean and covariance structure to be stable over time. In that case, we would say that the series is **covariance stationary**. There are two requirements for this to be true:

1. The mean of the series is stable over time: $\mathbb{E}Y_t = \mu$;
2. The covariance structure is stable over time.

In general, the (auto)covariance between Y_t and $Y_{t-\tau}$ is:

$$\gamma(t, \tau) = cov(Y_t, Y_{t-\tau}) = \mathbb{E}(Y_t - \mu)(Y_{t-\tau} - \mu)$$

If the covariance structure is stable, then the covariance depends on τ but not on t: $\gamma(t, \tau) = \gamma(\tau)$. Note: $\gamma(0) = \mathbb{C}\text{ov}(Y_t, Y_t) = \mathbb{V}\text{ar}(Y_t) < \infty$.

Remark

When observing/measuring time series we obtain numbers $y_1, ..., y_T$ which are the realization of random variables $Y_1, ..., Y_T$.
Using probabilistic concepts, we can give a more precise definition of a **(weak) stationary** series:

- If $\mathbb{E}Y_t = \mu$ - the process is called **mean-stationary**;
- If $\mathbb{V}\mathrm{ar}(Y_t) = \sigma^2 < \infty$ - the process is called **variance-stationary**;
- If $\gamma(t, \tau) = \gamma(\tau)$ - the process is called **covariance-stationary**.

In other words, a time series Y_t is stationary if its mean, variance and covariance do not depend on t.

If at least one of the three requirements is not met, then the process is **not-stationary**.

Since we often work with the (auto)correlation between Y_t and $Y_{t-\tau}$ rather than the (auto)covariance (because they are easier to interpret), we can calculate the autocorrelation function (**ACF**):

$$\rho(\tau) = \frac{cov(Y_t, Y_{t-\tau})}{\sqrt{\mathbb{V}\mathrm{ar}(Y_t)\mathbb{V}\mathrm{ar}(Y_{t-\tau})}} = \frac{\gamma(\tau)}{\gamma(0)}$$

Note: $\rho(0) = 1, |\rho(\tau)| \leq 1$.

The partial autocorrelation function (**PACF**) measures the association between Y_t and Y_{t-k}:

$$p(k) = \beta_k, \quad \textit{where} \quad Y_t = \alpha + \beta_1 Y_{t-1} + ... + \beta_k Y_{t-k} + \epsilon_t$$

The variance of the autocorrelation coefficient at lag k, r_k, is normally distributed at the limit, and the variance can be approximated: $\mathbb{V}\mathrm{ar}(r_k) \sim \frac{1}{T}$ (where T is the number of observations).

As such, we want to create lower and upper 95% confidence bounds for the normal distribution $\mathcal{N}\left(0, \frac{1}{T}\right)$, whose standard deviation is $\frac{1}{\sqrt{T}}$.

The 95% confidence interval (of a *WN* time series) is:

$$\Delta = 0 \pm \frac{1.96}{\sqrt{T}}$$

In general, the critical value of a standard normal distribution and its confidence interval can be found in these steps:

- Compute $\alpha = \frac{1-Q}{2}$, where Q is the confidence level;
- To express the critical value as a $z - score$, find the $z_{1-\alpha}$ value.

For example, if $Q = 0.95$, then $\alpha = 0.05$. Then, the standard normal distributions $1 - \alpha$ quantile is $z_{0.025} \approx 1.96$.

White Noise

White noise processes are the fundamental building blocks of all stationary time series.

We denote it as $\epsilon_t \sim WN(0, \sigma^2)$ - a zero mean, constant variance and serially uncorrelated ($\rho(t, \tau) = 0$, for $\tau > 0$ and any t) random variable process.

Sometimes we demand a stronger property of *independence*.

From the definition it follows that:

- $\mathbb{E}(\epsilon_t) = 0$;
- $\mathbb{V}\text{ar}(\epsilon_t) = \sigma^2 < \infty$;
- $\gamma(t, \tau) = \mathbb{E}(\epsilon_t - \mathbb{E}\epsilon_t)(\epsilon_{t-\tau} - \mathbb{E}\epsilon_{t-\tau}) = \mathbb{E}(\epsilon_t \epsilon_{t-\tau})$, where:

$$\mathbb{E}(\epsilon_t \epsilon_{t-\tau}) = \begin{cases} 0, & \text{if } \tau \neq 0 \\ \sigma^2, & \text{if } \tau = 0 \end{cases}$$

Checking if a process is stationary.

Let us check if $Y_t = \epsilon_t + \beta_1 \epsilon_{t-1}$, where $\epsilon_t \sim WN(0, \sigma^2)$ is stationary:

1. $\mathbb{E} Y_t = \mathbb{E}(\epsilon_t + \beta_1 \epsilon_{t-1}) = 0 + \beta_1 \cdot 0 = 0$;
2. $\mathbb{V}\mathrm{ar}(Y_r) = \mathbb{V}\mathrm{ar}(\epsilon_t + \beta_1 \epsilon_{t-1}) = \sigma^2 + \beta_1^2 \sigma^2 = \sigma^2(1 + \beta_1)$;
3. The autocovariance for $\tau > 0$:

$$\begin{aligned}
\gamma(t, \tau) &= \mathbb{E}(Y_t Y_{t-\tau}) = \mathbb{E}(\epsilon_t + \beta_1 \epsilon_{t-1})(\epsilon_{t-\tau} + \beta_1 \epsilon_{t-\tau-1}) \\
&= \mathbb{E}\epsilon_t \epsilon_{t-\tau} + \beta_1 \mathbb{E}\epsilon_t \epsilon_{t-\tau-1} + \beta_1 \mathbb{E}\epsilon_{t-1}\epsilon_{t-\tau} + \beta_1^2 \mathbb{E}\epsilon_{t-1}\epsilon_{t-\tau-1} \\
&= \beta_1 \mathbb{E}\epsilon_{t-1}\epsilon_{t-\tau} = \begin{cases} \beta_1 \sigma^2, & \text{if } \tau = 1 \\ 0, & \text{if } \tau > 1 \end{cases}
\end{aligned}$$

None of these characteristics depend on t, which means that the process is *stationary*. This process has a very short memory (i.e. if Y_t and $Y_{t+\tau}$ are separated by more than one time period - they are uncorrelated).
On the other hand, this process is **not** a *WN*.

The Lag Operator

The lag operator L is used to lag a time series: $LY_t = Y_{t-1}$. Similarly: $L^2 Y_t = L(LY_t) = L(Y_{t-1}) = Y_{t-2}$ etc. In general, we can write:

$$L^p Y_t = Y_{t-p}$$

Typically, we operate on a time series with a polynomial in the lag operator. A lag operator polynomial of degree m is:

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + ... + \beta_m L^m$$

For example, if $B(L) = 1 + 0.9L - 0.6L^2$, then:

$$B(L)Y_t = Y_t + 0.9Y_{t-1} - 0.6Y_{t-2}$$

A well known operator - the first-difference operator Δ - is a first-order polynomial in the lag operator: $\Delta Y_t = Y_t - Y_{t-1} = (1-L)Y_t$, i.e. $B(L) = 1 - L$.

We can also write an infinite-order lag operator polynomial as:

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + ... = \sum_{j=0}^{\infty} \beta_j L^j$$

The General Linear Process

Wold's representation theorem points to the appropriate model for **stationary processes**.

Wold's Representation Theorem

Let $\{Y_t\}$ be any *zero-mean* covariance-stationary process. Then we can write it as:

$$Y_t = B(L)\epsilon_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j}, \quad \epsilon_t \sim WN(0, \sigma^2)$$

where $\beta_0 = 1$ and $\sum_{j=0}^{\infty} \beta_j^2 < \infty$. On the other hand, any process of the above form is stationary.

- If $\beta_1 = \beta_2 = ... = 0$ (and $\beta_0 \neq 0$)- this corresponds to a *WN* process. This shows once again that *WN* is a stationary process.
- If $\beta_k = \phi^k$, then since $1 + \phi + \phi^2 + ... = 1/(1-\phi) < \infty$ we have that if $|\phi| < 1$, then the process $Y_t = \epsilon + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + ...$ is a stationary process.

- In Wold's theorem, we assumed a zero mean, though this is not as restrictive as it may seem. Whenever you see Y_t, you can analyse the process $Y_t - \mu$, so that the process is expressed in deviations from its mean. **The deviation from the mean has a zero mean by construction**. So, there is no generality loss when analyzing zero-mean processes.
- Wold's representation theorem points to the importance of models with infinite distributed (weighted) lags. Although infinite distributed lag models are not of immediate practical use since they contain **infinite** parameters, however, this may not always be the case.
  - As an example, from the previous slide, we may have $\beta_k = \phi^k$ in the infinite polynomial $B(L)$ - which is only a single (unknown) parameter.

Estimation and Inference for the Mean, ACF and PACF

Suppose we have a data *sample* of a *stationary* time series but we do not know the true model that generated the data (we only know that it was a polynomial $B(L)$), nor the mean, ACF or PACF associated with the model.

We want to use the **data** to *estimate* the mean, ACF and PACF, which we might use to help us decide on a suitable model to fit the data.

Sample Mean

The mean of a stationary series is $\mathbb{E}Y_t = \mu$. A fundamental principle of estimation, called the *analog principle*, suggests that we develop estimators by replacing expectations with sample averages. Thus, our estimator of the population mean, given a sample of size T is the sample mean:

$$\bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

Typically, we are not interested in estimating the mean but it is needed for estimating the autocorrelation function.

Sample Autocorrelations

The autocorrelation at displacement, or *lag*, τ for the covariance stationary series $\{Y_t\}$ is:

$$\rho(\tau) = \frac{\mathbb{E}\left(Y_t - \mu\right)\left(Y_{t-\tau} - \mu\right)}{\mathbb{E}\left(Y_t - \mu\right)^2}$$

Application of the analog principle yields a natural estimator of $\rho(\tau)$:

$$\hat{\rho}(\tau) = \frac{\frac{1}{T}\sum_{t=1}^{T}\left[\left(Y_t - \bar{Y}\right)\left(Y_{t-\tau} - \bar{Y}\right)\right]}{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \bar{Y}\right)^2}$$

This estimator is called the *sample* autocorrelation function (**sample ACF**).

Checking whether the autocorrelations are statistically significantly different from zero

It is often of interest to assess whether a series is reasonably approximated as white noise, i.e. whether all of its autocorrelations are zero in population.

If a series is white noise, then the sample autocorrelations $\hat{\rho}(\tau)$, $\tau = 1, ..., K$ in large samples are independent and have the $\mathcal{N}(0, 1/\sqrt{T})$ distribution.

Thus, if the series is *WN*, ~95% of the sample autocorrelations should fall in the interval of $\pm 1.96/\sqrt{T}$.

Exactly the same holds for both sample ACF and sample PACF. We typically plot the sample ACF and sample PACF along with their error bands.

> The aforementioned error bands provide 95% confidence bounds for only the sample autocorrelation taken **one** at a time.

Ljung-Box Test

We are often interested in whether a series is white noise, i.e. whether **all** its autocorrelations are **jointly zero**. Because of the sample size, we can only take a finite number of autocorrelations. We want to test:

$$H_0 : \rho(1) = 0, \rho(2) = 0, ..., \rho(k) = 0$$

Under the null hypothesis the **Ljung-Box statistic**:

$$Q = T(T+2) \sum_{\tau=1}^{k} \frac{\hat{\rho}^2(\tau)}{T - \tau}$$

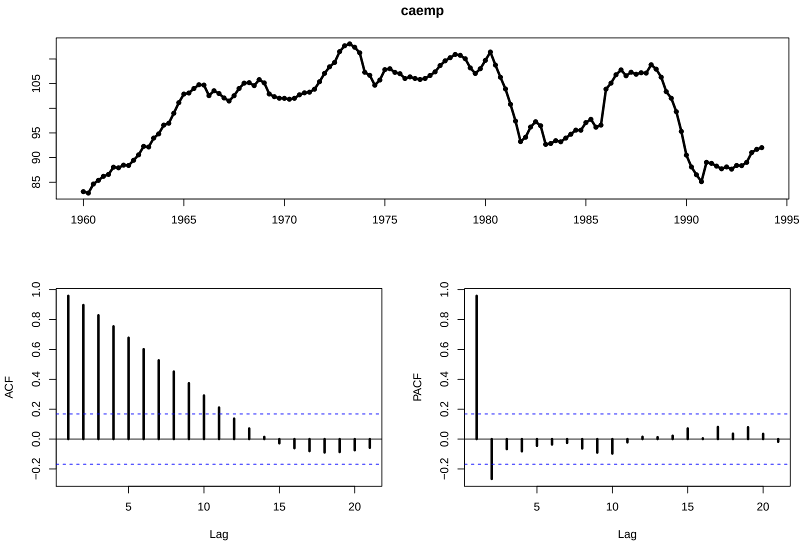
is approximately distributed as a χ^2_k random variable.

To test the null hypothesis, we have to calculate the $p - value = P(\chi^2_K > q)$: if $p - value < 0.05$ - we *reject* the null hypothesis, H_0, and assume that Y_t is not white noise.

Example: Canadian unemployment data

We will illustrate the provided ideas by examining quarterly Canadian employment index data. The data is seasonally adjusted and displays no trend (more on what this means in a later lecture), however it does appear to be highly serially correlated...

```
suppressPackageStartupMessages({require("forecast")})
txt1 <- "http://uosis.mif.vu.lt/~rlapinskas/(data%20R&GRETL/"
txt2 <- "caemp.txt"
caemp <- read.csv(url(paste0(txt1, txt2)),
                  header = TRUE, as.is = TRUE)
caemp <- ts(caemp, start = c(1960, 1), freq = 4)
tsdisplay(caemp)
```



- The sample ACF are large and display a slow one-sided decay;
- The sample PACF are large at first, but are statistically negligible beyond displacement $\tau = 2$.

We shall once again test the *WN* hypothesis, this time using the Ljung-Box test statistic.

```
Box.test(caemp, lag = 1, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  caemp
## X-squared = 127.73, df = 1, p-value < 2.2e-16
```

with `p < 0.05`, we reject the null hypothesis $H_0 : \rho(1) = 0$.

```
Box.test(caemp, lag = 2, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  caemp
## X-squared = 240.45, df = 2, p-value < 2.2e-16
```

with `p < 0.05`, we reject the null hypothesis $H_0 : \rho(1) = 0, \rho(2) = 0$, and so on. **We can see that the time series is not *WN*.**

We will now present a few more examples of stationary processes.

Moving-Average (MA) Models

- Finite-order moving-average processes are approximations to the Wold representation (which is an infinite-order moving average process).
- The variation in time series, one way of another, is driven by shocks of various sorts. This suggests the possibility of modelling time series directly as distributed lags of current and past shocks - i.e. as moving-average processes.

The MA(1) Process

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

$$Y_t = \epsilon_t + \theta\epsilon_{t-1} = (1-\theta L)\epsilon_t, \quad -\infty < \theta < \infty, \quad \epsilon \sim WN(0, \sigma^2)$$

Defining characteristics of an MA process: the current value of the observed series can be expressed as a function of current and lagged unobservable shocks ϵ_t.

Whatever the value of θ (as long as $|\theta| < \infty$), **MA(1) is always a stationary process** and:

- $\mathbb{E}(Y_t) = \mathbb{E}(\epsilon_t) + \theta\mathbb{E}(\epsilon_{t-1}) = 0$;
- $\mathbb{V}\text{ar}(Y_t) = \mathbb{V}\text{ar}(\epsilon_t) + \theta^2\mathbb{V}\text{ar}(\epsilon_{t-1}) = (1+\theta^2)\sigma^2$;
- $\boldsymbol{\rho(\tau)} = \begin{cases} 1, \text{ if } \tau = 0 \\ \theta/(1+\theta^2), \text{ if } \tau = 1 \\ 0, \text{ otherwise} \end{cases}$

Key feature of MA(1): (sample) ACF has a sharp cutoff beyond $\tau = 1$.

We can write MA(1) another way:

Since:

$$Y_t = (1 - \theta L)\epsilon_t \Rightarrow \epsilon_t = \frac{1}{1 - \theta L} Y_t$$

Recalling the formula of a geometric series, if $|\theta| < 1$:

$$\begin{aligned}
\epsilon_t &= (1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + ...)Y_t \\
&= Y_t - \theta Y_{t-1} + \theta^2 Y_{t-2} - \theta^3 Y_{t-3} + ...
\end{aligned}$$

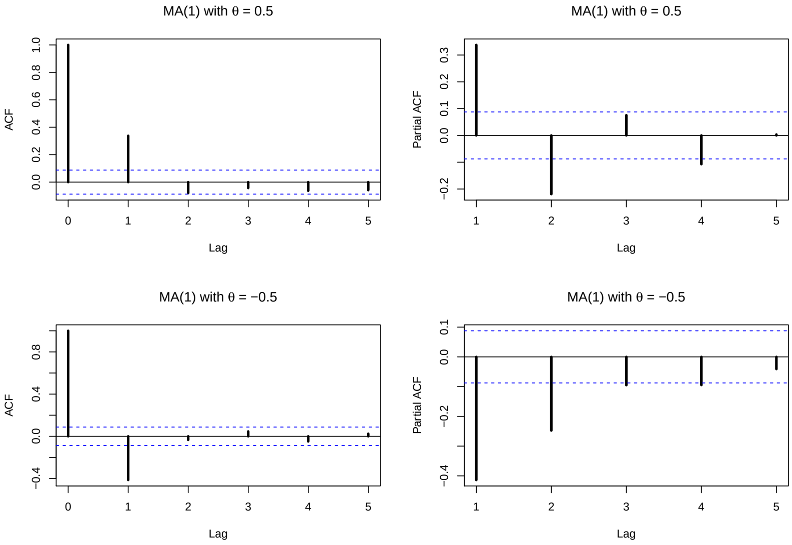
and we can express Y_t as an infinite AR, $AR(\infty)$, process:

$$\begin{aligned}
Y_t &= \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 Y_{t-3} - ... + \epsilon_t \\
&= \sum_{j=1}^{\infty} (-1)^{j+1} \theta^j Y_{t-j} + \epsilon_t
\end{aligned}$$

Remembering the definition of a PACF we have that for an MA(1) process it will decay *gradually* to zero. Furthermore:

- If $\theta < 0$, then the pattern of decay will be one-sided
- If $0 < \theta < 1$, then the pattern of decay will be oscillating.

An example on how the sample ACF and PACF would look like for some $MA(1)$ processes:



The MA(q) Process

We will now consider a general finite-order moving average process of order q, MA(q):

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} = \Theta(L)\epsilon_t, \quad -\infty < \theta < \infty, \quad \epsilon \sim WN(0, \sigma^2)$$

where

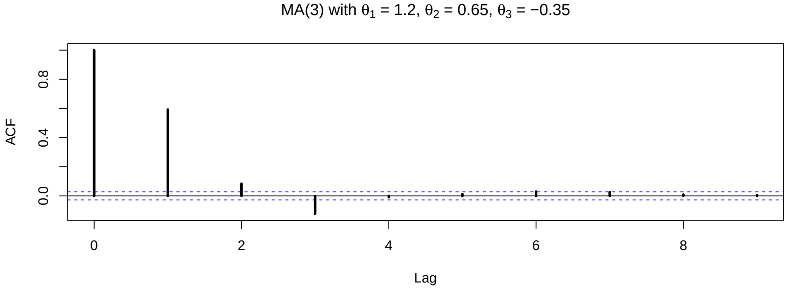
$$\Theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$$

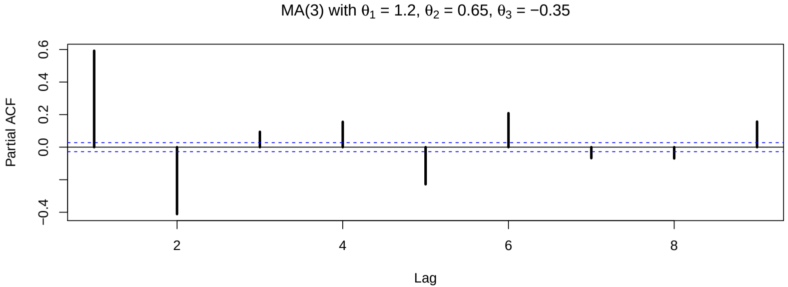
is the qth-order lag polynomial. The $MA(q)$ process is a generalization of the $MA(1)$ process. Compared to $MA(1)$, the $MA(q)$ can capture richer dynamic patterns which can be used for improved forecasting.

The properties of an MA(q) processes are parallel to those of an MA(1) process in all respects:

- The finite-order MA(q) process is covariance stationary *for any value of its parameters* ($|\theta_j| < \infty, \quad j = 1, ..., q$);
- In MA(q) case, all autocorrelations *in ACF beyond displacement q are 0* (a distinctive property of the MA process);
- The *PACF* of the MA(q) *decays gradually* in accordance with the infinite autoregressive representation, similar to MA(1): $Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + ... + \epsilon_t$ (with certain conditions for a_j).

An example on how the sample ACF and PACF would look like for a specific MA(3) process:





Autoregressive (AR) Models

- The autoregressive process is also a natural approximation of the Wold representation.
- We have seen that, under certain conditions, a moving-average process has an autoregressive representation.
- Consequently, an autoregressive process is, in a sense, the same as a moving average process.

The AR(1) Process

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

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma^2)$$

or:

$$(1 - \phi L) Y_t = \epsilon_t \Rightarrow Y_t = \frac{1}{1 - \phi L} \epsilon_t$$

Note the special interpretation of the errors, or disturbances, or shocks ϵ_t in time series theory: in contrast to the regression theory where they were understood as the summary of all unobserved X's, now they are treated as *economic effects* which have developed in period t.

As we will see when analyzing ACF, the AR(1) model is capable of capturing much more **persistent** dynamics (depending on its parameter value) than the MA(1) model, which has a very short memory regardless of its parameter value.

Recall that a finite-order moving-average process is always covariance stationary, but that certain conditions must be satisfied for AR(1) to be stationary. The AR(1) process can be rewritten as:

$$Y_t = \frac{1}{1-\phi L}\epsilon_t = (1 + \phi L + \phi^2 L^2 + ...)\epsilon_t = \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + ...$$

This Wold's moving-average representation for Y is convergent if $|\phi| < 1$, thus:

AR(1) is stationary is $|\phi| < 1$

Equivalently, the condition for covariance stationarity is that the root, z_1, of the autoregressive lag operator polynomial (i.e. $1 - \phi z_1 = 0 \Leftrightarrow z_1 = 1/\phi$) be **greater than 1** in absolute value (a similar condition on the roots is important for the AR(p) case).

We can also get the above equation by recursively applying the equation of $AR(1)$ to get the infinite MA process:

$$Y_t = \phi Y_{t-1} + \epsilon_t = \phi(\phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t$$
$$= \epsilon_t + \phi\epsilon_{t-1} + \phi^2 Y_{t-2} = ... = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

From the moving average representation of the covariance stationary AR(1) process:

- $\mathbb{E}(Y_t) = \mathbb{E}(\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + ...) = 0$;
- $\mathbb{V}\mathrm{ar}(Y_t) = \mathbb{V}\mathrm{ar}(\epsilon_t) + \phi^2\mathbb{V}\mathrm{ar}(\epsilon_{t-1}) + ... = \sigma^2/(1-\phi^2)$;

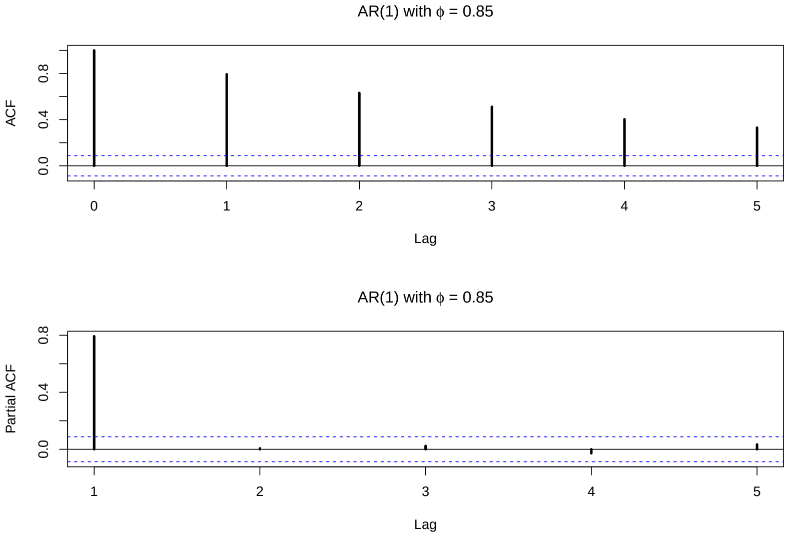
Or, alternatively: when $|\phi| < 1$ - the process is stationary, i.e. $\mathbb{E}Y_t = m$, therefore $\mathbb{E}Y_t = \phi\mathbb{E}Y_{t-1} + \mathbb{E}\epsilon_t \Rightarrow m = \phi m + 0 \Rightarrow m = 0$.

This allows us to easily estimate the mean of the *generalized AR*(1) process: if $Y_t = \alpha + \phi Y_{t-1} + \epsilon_t$, then $m = \alpha/(1-\phi)$.

The **correlogram** (ACF & PACF) of AR(1) is in a sense symmetric to that of MA(1):

- $\boldsymbol{\rho(\tau)} = \phi^\tau$, $\tau = 0, 1, 2...$ - ACF decays exponentially;
- $\boldsymbol{p(\tau)} = \begin{cases} \phi, \tau = 1 \\ 0, \tau > 1 \end{cases}$ - PACF cuts off abruptly.

An example on how the sample ACF and PACF would look like for some AR(1) process:



The AR(p) Process

The general pth order autoregressive process, AR(p) is:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma^2)$$

In lag operator form, we write:

$$\Phi(L) Y_t = (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) Y_t = \epsilon_t$$

Similar to the AR(1) case, **the AR(p) process is covariance stationary** if and only if all the roots z_i of the autoregressive lag operator polynomial $\Phi(z)$ are **outside** the complex unit circle:

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p = 0 \Rightarrow |z_i| > 1$$

So:

AR(p) is stationary if all the roots $|z_i| > 1$

For a quick check of stationarity, use the following **rule of thumb**:

If $\sum_{i=1}^{p} \phi_i \geq 1$, the process **isn't** stationary

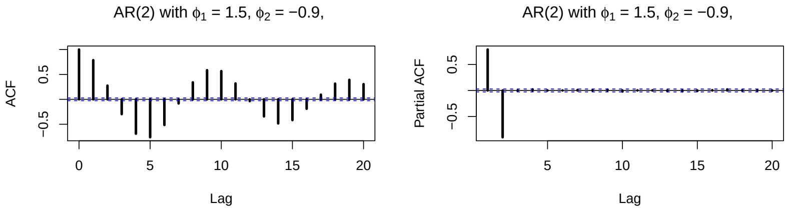
(Note: a rule of thumb and not a guarantee)

In the covariance stationary case, we can write the process in the infinite moving average $MA(\infty)$ form:

$$Y_t = \frac{1}{\Phi(L)}\epsilon_t$$

- The ACF for the general AR(p) process decays gradually when the lag increases;
- The PACF for the general AR(p) process has a sharp cutoff at displacement p.

An example on how the sample ACF and PACF would look like for an AR(2) process $Y_t = 1.5Y_{t-1} - 0.9Y_{t-2} + \epsilon_T$:



The corresponding lag operator polynomial is $1 - 1.5L + 0.9L^2$ with two complex conjugate roots: $z_{1,2} = 0.83 \pm 0.65i$,
$|z_{1,2}| = \sqrt{0.83^2 + 0.65^2} = 1.05423 > 1$ - thus the process is stationary.

The ACF for an AR(2) is:

$$\rho(\tau) = \begin{cases} 0, & \tau = 0 \\ \phi_1/(1-\phi_2), & \tau = 1 \\ \phi_1\rho(\tau-1) + \phi_2\rho(\tau-2), & \tau = 2, 3, ... \end{cases}$$

Because the roots are complex, the ACF **oscillates** and because the roots are close to the unit circle, the oscillation damps *slowly*.

Stationarity and Invertibility

The AR(p) is a generalization of the AR(1) strategy for approximating the Wold representation. The moving-average representation associated with the **stationary** $AR(p)$ process:

$$Y_t = \frac{1}{\Phi(L)}\epsilon_t \text{ where } \frac{1}{\Phi(L)} = \sum_{j=0}^{\infty} \psi_j L^j, \quad \psi_0 = 1$$

depends on p parameters only. This gives us the infinite process from Wold's Representation Theorem:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

which is known as the infinite moving-average process, $MA(\infty)$. Because AR is stationary, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ and Y_t take finite values.

Thus, a **stationary** AR process can be rewritten as an $MA(\infty)$ process.

Stationarity and Invertibility

In some cases the AR form of a stationary process is preferred to that of MA. Just as we can write an AR process as an $MA(\infty)$, we can written an MA process as an $AR(\infty)$. The necessary definition says that *the MA process is called* ***invertible*** if it can be expressed as an AR process. So, the MA(q) process:

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} = \Theta(L)\epsilon_t, \quad -\infty < \theta_i < \infty, \quad \epsilon_t \sim WN(0, \sigma^2)$$

is invertible if all the roots of $\Theta(x) = 1 + \theta_1 x + ... + \theta_q x^q$ lie **outside** the unit circle:

$$1 + \theta_1 x + ... + \theta_q x^q = 0 \Rightarrow |x_i| > 1$$

Stationarity and Invertibility

Then we can write the process as:

$$\epsilon_t = \frac{1}{\Theta(L)} Y_t, \text{ where } \frac{1}{\Theta(L)} = \sum_{j=0}^{\infty} \pi_j L^j, \quad \pi_0 = 1$$

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j} = Y_t + \sum_{j=1}^{\infty} \pi_j Y_{t-j}$$

which gives us the infinite-order autoregressive process, $AR(\infty)$:

$$Y_t = \sum_{j=1}^{\infty} \widetilde{\pi}_j Y_{t-j} + \epsilon_t$$

Because the MA process is invertible, the infinite series converges to a finite value.

For example, MA(1) of the form $Y_t = \epsilon_t - \epsilon_{t-1}$ is **not invertible** since $1 - x = 0 \Rightarrow x = 1$.

Causality

A process $\{Y_t\}$ is **causal**, or a causal function of $\{\epsilon_t\}$, if Y_t can be expressed in terms of the current and past values of ϵ_t.

So, by definition:

- a **stationary** $AR(p)$ process is causal;
- **any** $MA(q)$ process is causal.

Note, for an $AR(1)$ process $Y_t = \phi Y_{t-1} + \epsilon_t$, $\epsilon_t \sim WN(0, \sigma^2)$:

- If $|\phi| < 1$, then Y_t is **causal**, because a **stationary** $AR(1)$ process can be expressed in terms of its shocks: $Y_t = 1/(1 - \phi L)\epsilon_t$;
- If $|\phi| > 1$, then Y_t is **non-causal**.

To recap:

- The $AR(p)$ process is **always invertible** (it contains no MA terms);
- The $MA(q)$ process is **invertible** if all the roots of $\theta(x) = 1 + \theta_1 x + ... + \theta_q x^q$ lie **outside** the unit circle;
- An **invertible** $MA(q)$ process can be rewritten as an $AR(\infty)$ process;
- The $MA(q)$ process is **always stationary** (it contains no AR terms);
- The $AR(p)$ process is **stationary** if all the roots of $\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$ lie **outside** the unit circle;
- A **stationary** $AR(p)$ process can be rewritten as an $MA(\infty)$ process.
- A **stationary** $AR(p)$ process is **causal**;
- **Any** $MA(q)$ process is **causal**.

Autoregressive Moving-Average (ARMA) Models

AR and MA models are often combined in attempts to obtain better approximations to the Wold representation. This results in the **ARMA(p,q)** process. The motivation for using ARMA models is as follows:

- If the random shock that drives and AR process is itself a MA process, then we obtain an ARMA process;
- ARMA processes arise from aggregation - sums of AR processes, sums of AR and MA processes;
- AR processes observed subject to measurement error also turn out to be ARMA processes.

ARMA(1,1) process

The simplest ARMA process, that is not a pure AR or a pure MA, is the ARMA(1,1) process:

$$Y_t = \phi Y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}, \quad \epsilon_t \sim WN(0, \sigma^2)$$

or in lag operator form:

$$(1 - \phi L) Y_t = (1 + \theta L) \epsilon_t$$

where:

1. $|\phi| < 1$ - required for stationarity;
2. $|\theta| < 1$ - required for invertibility.

If the covariance stationarity conditions are satisfied, then we have the MA representation:

$$Y_t = \frac{(1 - \phi L)}{(1 + \theta L)} \epsilon_t = \epsilon_t + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2} + ...$$

which is an infinite distributed lag of current and past innovations.

Similarly, we can rewrite it in the infinite AR form:

$$Y_t + a_1 Y_{t-1} + a_2 Y_{t-2} + ... = \frac{(1 + \theta L)}{(1 - \phi L)} Y_t = \epsilon_t$$

ARMA(p,q) process

A natural generalization of the ARMA(1,1) is the ARMA(p,q) process that allows for multiple moving-average and autoregressive lags. We can write it as:

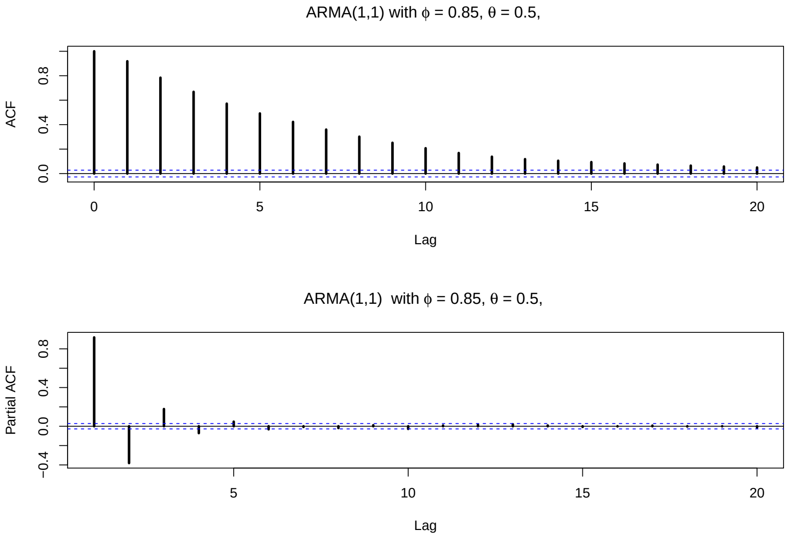
$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \epsilon_t + \theta_q \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}, \quad \epsilon_t \sim WN(0, \sigma^2)$$

or:

$$\Phi(L) Y_t = \Theta(L) \epsilon_t$$

- If all the roots of $\Phi(L)$ are outside the unit circle, then the process is stationary and has a **convergent** infinite moving average representation: $Y_t = (\Phi(L)/\Theta(L))\, \epsilon_t$;
- If all roots of $\Theta(L)$ are outside the unit circle, then the process is invertible and can be expressed as the **convergent** infinite autoregression: $(\Phi(L)/\Theta(L))\; Y_t = \epsilon_t$.

An example of an ARMA(1,1) process: $Y_t = 0.85Y_{t-1} + \epsilon_t + 0.5\epsilon_{t-1}$:



Choosing between AR, MA and ARMA (Part I: Order Selection)

ARMA models are often both highly accurate and highly parsimonious.

- In a particular situation, for example, it might take an $AR(5)$ model to get the same approximation accuracy as could be obtained with an $ARMA(1,1)$, but the $AR(5)$ has five parameters to be estimated, whereas the $ARMA(1,1)$ has only two.

The rule to determine the number of AR and MA terms:
- AR(p) - ACF declines, $PACF = 0$ if $\tau > p$;
- MA(q) - $ACF = 0$ if $\tau > q$, PACF declines;
- ARMA(p,q) - both ACF and PACF decline.

- Generally, when choosing a model order for $ARMA$, we choose the order that results in the smallest BIC. Note that there are alternative information criterions - AIC, which does not penalize higher order models, as well as $AICc$, which does take into account the number of parameters.

Stationarity and Invertibility in Model Specification

By inverting and truncating the appropriate backshift operator function (i.e. either $\phi(L)$, or $\theta(L)$):

- a stationary $\mathrm{AR(p)}$ process can be approximated with an arbitrary precision by truncating its infinite MA representation at some **high order** $\mathrm{MA(q^*)}$;
- an invertible $\mathrm{MA(q)}$ process can be approximated with an arbitrary precision by truncating its infinite AR representation by some **high order** $\mathrm{AR(p^*)}$;
- a stationary and invertible $\mathrm{ARMA(p, q)}$ process can be closely approximated by either a **high order** AR, or a **high order** MA process.

Consequently:

- a higher order $\mathrm{AR(p)}$ process can be well approximated by a lower order $ARMA(\tilde{p}, \tilde{q})$, where $\tilde{p} + \tilde{q} < p$;
- a higher order $\mathrm{MA(q)}$ process can be well approximated by a lower order $ARMA(\tilde{p}, \tilde{q})$, where $\tilde{p} + \tilde{q} < q$;

In an empirical applications, where the data sample may be small, the correlation structure (ACF and PACF) may be such that in order to achieve a good model by a pure AR (or pure MA) requires a high order p (or high order q).

On the other hand, an approximation with a lower order ARMA (e.g. with $p, q \in \{0, 1, 2\}$) may be reasonable for a given series.

Based on the $ARMA$ definition, we can write $ARMA(p, 0) = AR(p)$ and $ARMA(0, q) = MA(q)$. This idea leads to universal functions in R and Python, which allow estimation of AR/MA/ARMA models with various lag orders.

Choosing between AR, MA and ARMA (Part II: Diagnostic Checking)

While the $AIC/AICc/BIC$ is a good indicator of the adequacy of the model, another important part is the model **residuals**.

Remember that one of the primary assumptions about the model is that the shocks (i.e. residuals) are white noise - $\epsilon_t \sim WN(0, \sigma^2)$.

If this does not hold true for the selected model, then the stationarity conditions may not hold and we need to specify a different model. Consequently:

The **Ljung-Box test** is commonly used to check whether the residuals of an ARMA model have no autocorrelation. In such cases, the degrees of freedom need to be adjusted to reflect the parameter estimation. For example, if $\widehat{\epsilon}_t$ are the residuals of an $ARMA(p, q)$ model, we want to test: $H_0 : \rho_{\hat{\epsilon}}(1) = 0, ..., \rho_{\hat{\epsilon}}(k) = 0$. Then, under the null hypothesis, the statistic:

$$Q = T(T+2) \sum_{\tau=1}^{k} \frac{\hat{\rho}^2_{\hat{\epsilon}}(\tau)}{T - \tau} \sim \chi^2_{1-\alpha, k-p-q}$$

For a given α significance level.

Estimation

Autoregressive process parameter estimation

Let say we want to estimate the parameters of our AR(1) process:

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

- The OLS estimator of ϕ for the AR(1) case:

$$\hat{\phi} = \frac{\sum_{t=1}^{T} Y_t Y_{t-1}}{\sum_{t=1}^{T} Y_{t-1}^2}$$

- Yule-Walker estimator of ϕ for AR(1) can be calculated by multiplying $Y_t = \phi_1 Y_{t-1} + \epsilon_t$ by Y_{t-1} and taking its expectation. We will get the equation:

$$\gamma(1) = \phi\gamma(0)$$

  Recall that $\gamma(\tau)$ is the **covariance** between Y_t and $Y_{t-\tau}$.

For the AR(p) case, we would need p different equations,i.e.:

$$\gamma(k) = \theta_1 \gamma(t-1) + ... + \theta_p \gamma(k-p), \quad k = 1, ..., p$$

Moving-average process parameter estimation

Let say we want to estimate the parameter of our invertible MA(1) process (i.e. $|\theta| < 1$):

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} \Rightarrow \epsilon_t = Y_t - \theta Y_{t-1} + ...$$

Let $S(\theta) = \sum_{t=1}^{T} \epsilon_t$ and $\epsilon_0 = 0$. We can find the parameter θ by minimizing $S(\theta)$.

ARMA process parameter estimation

For the ARMA(1,1): $Y_t = \phi Y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$ we would need to minimize $S(\theta, \phi) = \sum_{t=1}^{T} \epsilon_t^2$ with $\epsilon_0 = Y_0 = 0$.
For the ARMA(p,q), we would need to minimize $S(\theta, \phi)$ by setting $\epsilon_k = Y_k = 0$ for $k \leq 0$.
We can also estimate the parameters using the maximum likelihood method.

Forecasting: The General Idea

So far we thought of the information set as containing the available past history of the series, $\Omega_T = \{Y_T, Y_{T-1}, ...\}$, where we imagined the history as having begun in an infinite past. Based on that information set, we want to find the optimal forecast of Y at some future time $T+h$.

> If Y_t is a *stationary* process, then the forecast tends to the process *mean*, as h increases. Therefore, the forecast is only interesting for several *small* values of h.

The basic idea of the forecast method is always the same: write out the process for the future time period, $T+h$ and project it on what is known at time T when the forecast is made. We denote the forecast as $Y_{T+h|T}, \quad h \geq 1$.

Point forecasts can be calculated using the following three steps.

1. If needed, expand the equation so that Y_t is on the left hand side and all other terms are on the right;
2. Rewrite the equation by replacing T by $T+h$;
3. On the right hand side of the equation, replace future observations by their forecasts, future errors ($\epsilon_{T+j}, 0 < j \leq h$) by **zero**, and past errors by the corresponding **residuals**, $\widehat{\epsilon}_t$, $t \leq T$.

Forecasting MA(q) process

Consider, for example, an MA(1) process:

$$Y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}, \quad \epsilon_t \sim WN(0, \sigma^2)$$

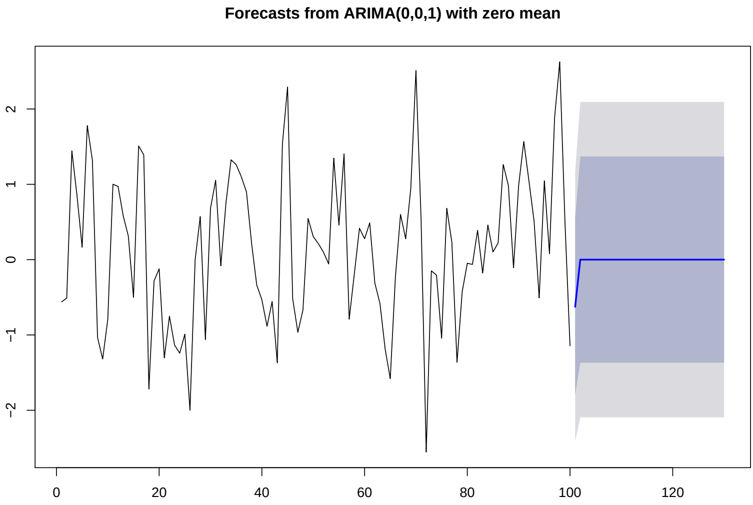
We then calculate the **forecasts** for periods $T+1$, ..., $T+h$ as:

$$\begin{aligned}
Y_{T+1} = \mu + \epsilon_{T+1} + \theta\epsilon_T &\Rightarrow Y_{T+1|T} = \mu + 0 + \theta\epsilon_T \\
Y_{T+2} = \mu + \epsilon_{T+2} + \theta\epsilon_{T+1} &\Rightarrow Y_{T+2|T} = \mu + 0 + 0 \\
&\vdots \\
Y_{T+h} = \mu + \epsilon_{T+h} + \theta\epsilon_{T+h-1} &\Rightarrow Y_{T+h|T} = \mu
\end{aligned}$$

The forecast can be generalized for $MA(q)$ as follows:

- The forecast quickly approaches the (sample) mean of the process and for $h \geq q+1$ - coincides with it.
- When h increases, the accuracy of the forecast diminishes up to the moment $h = q+1$, whereupon it becomes a constant.

An example of an MA(1) process: $Y_t = \epsilon_t + 0.5\epsilon_{t-1}$:



Forecasting AR(p) process

Consider, for example, an AR(1) process:

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim WN(0, \sigma^2)$$

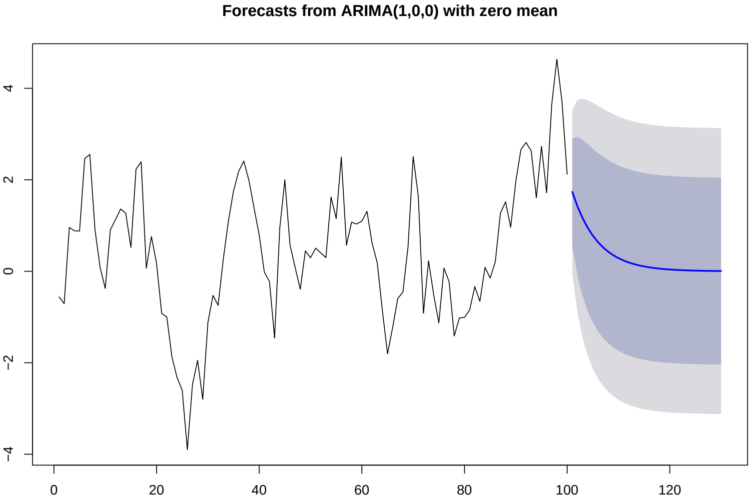
We then calculate the **forecasts** for periods $T+1, \dots, T+h$ as:

$$Y_{T+1} = \phi Y_T + \epsilon_{T+1} \Rightarrow Y_{T+1|T} = \phi Y_T + 0$$
$$Y_{T+2} = \phi Y_{T+1} + \epsilon_{T+2} \Rightarrow Y_{T+2|T} = \phi Y_{T+1} + 0 = \phi^2 Y_T$$
$$\vdots$$
$$Y_{T+h} = \phi Y_{T+h-1} + \epsilon_{T+h} \Rightarrow Y_{T+h|T} = \phi Y_{T+h-1} + 0 = \phi^h Y_T$$

The forecast can be generalized for $AR(p)$ as follows:

- When h increases, the forecast tends to the (sample) mean exponentially fast, but never reaches it.

An example of an AR(1) process: $Y_t = 0.85Y_{t-1} + \epsilon_t$:



Forecasting ARMA(p,q) process

Consider, for example, an ARMA(1,1) process:

$$Y_t = \phi Y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}, \quad \epsilon_t \sim WN(0, \sigma^2)$$

We then calculate the **forecasts** for periods $T+1$, ..., $T+h$ as:

$$Y_{T+1} = \phi Y_T + \epsilon_{T+1} + \theta \epsilon_T \Rightarrow Y_{T+1|T} = \phi Y_T + 0 + \theta \epsilon_T$$

$$Y_{T+2} = \phi Y_{T+1} + \epsilon_{T+2} + \theta \epsilon_{T+1} \Rightarrow Y_{T+2|T} = \phi Y_{T+1} + 0 + 0 = \phi^2 Y_T + \phi\theta\epsilon_T$$

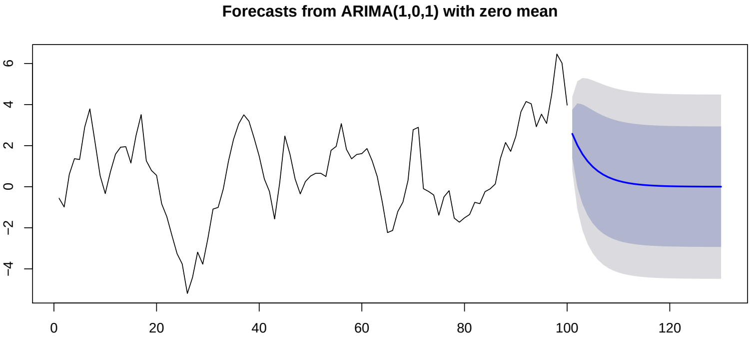
$$\vdots$$

$$Y_{T+h} = \phi Y_{T+h-1} + \epsilon_{T+h} + \theta \epsilon_{T+h-1} \Rightarrow Y_{T+h|T} = \phi^h Y_T + \phi^{h-1}\theta\epsilon_T$$

The forecast can be generalized for $ARMA(p,q)$ as follows:

- Similar to the $AR(p)$ process, the $ARMA(p,q)$ process tends to the average, but never reaches it.

An example of an ARMA(1,1) process: $Y_t = 0.85Y_{t-1} + \epsilon_t + 0.5\epsilon_{t-1}$:



- The forecast $Y_{T+h|T}$ of an MA(q) process in $h = q$ steps reaches its average and then does not change anymore;
- The forecast $Y_{T+h|T}$ of an AR(p) or ARMA(p,q) process tends to the average, but never reaches it. The speed of convergence depends on the coefficients;

ARIMA models and interpretation

ARMA models are *atheoretic* models (i.e. not concerned with (economic) theory). We are selecting an appropriate model purely based on the information criteria and residual diagnostics. The goal is usually to get an adequate *ARMA* model for forecasting.

Nevertheless, there are a couple of ways to examine the coefficients of the model.

Short-run and Long-run coefficients

If we look at the $MA(1)$ and $AR(1)$ models separately (and apply the **Wold's decomposition theorem** to the $AR(1)$ model):

- $MA(1)$: $Y_t = \epsilon_t + \theta_1 \epsilon_{t-1}$;
- $AR(1)$: $Y_t = \phi_1 Y_{t-1} + \epsilon_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + ...$

We see that in the AR model, the lagged terms of Y can be expressed as a an **infinite sum of the past** values of ϵ, whereas in the MA model, the number of past error terms depends on the model lag order.

To generalize for an $ARMA(p, q)$ model, the error terms $\epsilon_t, ..., \epsilon_{t-q}$ explain the **short-term** influence of the past, whereas $Y_{t-1}, ..., Y_{t-p}$ explain the **long-term** influence.

Note: Invertibility is a restriction in the software used to estimate the coefficients of models with *MA* terms. It is not something that we check for in the data analysis separately (unlike stationarity).

Interpreting ARMA coefficients as regression parameters

Consider an $AR(2)$ model for the inflation rate, π_t:

$$\pi_t = \alpha + \phi_1 \pi_{t-1} + \phi_2 \pi_{t-2} + \epsilon_t$$

If we were to attempt to interpret the coefficients, as we would have in a cross-sectional regression case, we might say that *"inflation today depends on the level of inflation yesterday and on the level of inflation the day before yesterday"*. The **problem** with this type of interpretation is two-fold:

- it would be harder to describe for higher order lags, even more so, if we include lags of ϵ;
- when we interpret the regression coefficients, we make use of the **ceteris paribus** condition. This is much harder to do, when a unit increase in inflation the day before yesterday, would also affect the inflation rate yesterday.

Impulse-Response Functions

Instead of attempting to interpret the estimated coefficients, which are often too difficult to interpret in *ARMA* models, it is better to try to understand the **dynamics** of the system itself. This can be done in two ways:

- By looking at the forecast dynamics (remember the differences between *AR* and *MA* model forecasts);
- By looking at the impulse-response function or time path associated with the model.

Before examining impulse-response functions, we define:

- **Momentum** - the tendency to continue moving in the same direction. The momentum effect can offset the force of regression (convergence) toward the mean and can allow a variable to move away from its historical mean, for some time, **but not indefinitely**;
- **Persistence** - a persistence variable will hang around where it is and converge slowly only to the historical mean.

The impulse-response functions allow us to ask the question: *Suppose that a variable is at its historical mean and it receives a* ***temporary*** *one unit shock in a single period.* ***How will the variable respond in future periods?***

Example

Consider an $MA(1)$:

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1}$$

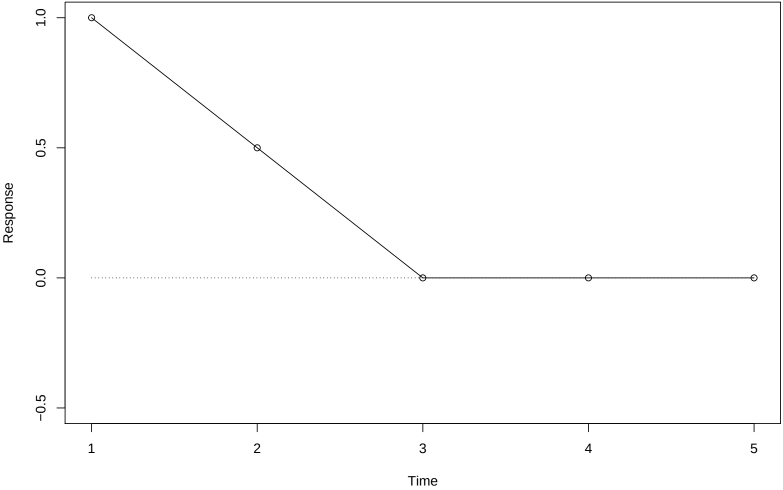
Assume that a unit shock arrives at $t = 0$ (we can do this equivalently at any other moment, for example $t = T$), so that $\epsilon_0 = \sigma$. We then have that the effect of this unit shock on Y, as $t \to \infty$, is:

$$\begin{aligned} t &= 0 : \sigma \\ t &= 1 : \theta_1 \times \sigma \\ t &= 2 : 0 \\ &\vdots \\ t &= h : 0 \end{aligned}$$

The shock completely disappears after two periods. **This means that if we have a momentary shock in an $MA(1)$, it only lasts for two periods**.

> In practical applications, this could be used in analysing a temporary shock to advertising, expenditure, interest rate and so on.

If we have $\sigma = 1$ and $\theta_1 = 0.5$, then the *IRF* is:



Example

$AR(1)$:

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

Assume that a unit shock arrives at $t = 0$, so that $\epsilon_0 = \sigma$. We then have that the effect of this unit shock on Y, as $t \to \infty$, is:

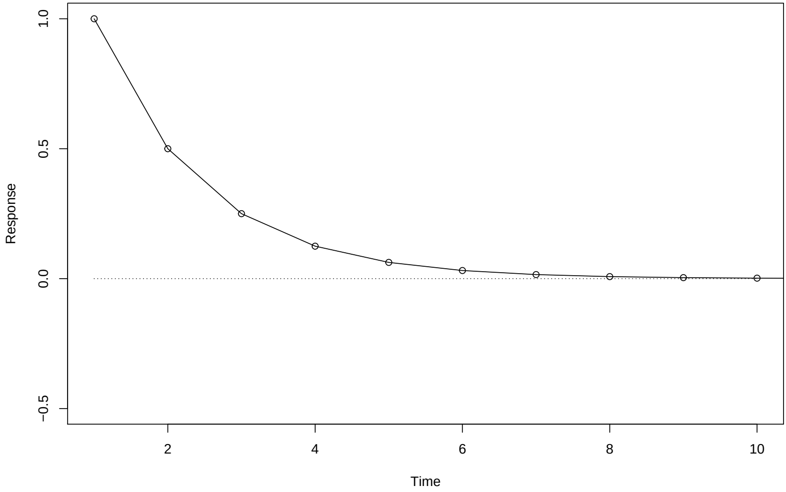
$$\begin{aligned} t &= 0 : \sigma \\ t &= 1 : \phi_1 \times \sigma \\ t &= 2 : \phi_1^2 \times \sigma \\ &\vdots \\ t &= h : \phi_1^h \times \sigma \end{aligned}$$

Note: if we rewrite the stationary $AR(1)$ process as an $MA(\infty)$:

$$Y_t = \sum_{j=1}^{\infty} \phi_1^j \epsilon_{t-j}$$

We see that the unit shock to ϵ_0 will have the same effect as in the $AR(1)$ specification.

If we have $\sigma = 1$ and $\phi_1 = 0.5$, then the *IRF* is:



P.S. We are examining **momentary** shocks, but we could similarly examine the effects of a **permanent shock** (we leave this for later lectures).

Example

ARMA(2, 1):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1},$$

Assuming that the model has no constant, it may be easier to set $Y_0 = \epsilon_0$ and examine how the value of Y changes, when $\epsilon_j = 0, \forall j > 0$. Then, a unit shock $\epsilon_0 = \sigma$ will have the following effect on Y, as $t \to \infty$, is:

$$\begin{aligned}
Y_0 &= \sigma \\
Y_1 &= \phi_1 \times Y_0 + \theta_1 \times \sigma \\
Y_2 &= \phi_1 \times Y_1 + \phi_2 \times Y_0 \\
&\vdots \\
Y_h &= \phi_1 \times Y_{h-1} + \phi_2 \times Y_{h-2}
\end{aligned}$$

Note that this is similar to how the forecasts are calculated. However, instead of using Y_T, the beginning period is set $Y_0 = \sigma$. If we were to have a constant term, then we would need to set $Y_0 = \alpha + \sigma$.

Sometimes, for convenience (and depending on the model complexity), it may be easier to set $\sigma = 1$ and then re-scale the resulting *IRF* values for various different initial shocks $\tilde{\sigma}$.

For example, a unit shock to

$$Y_t = 0.3Y_{t-1} - 0.1Y_{t-2} + \epsilon_t + 0.05\epsilon_{t-1}, \quad \epsilon_t \sim WN(0,1)$$

results in the following *IRF*: