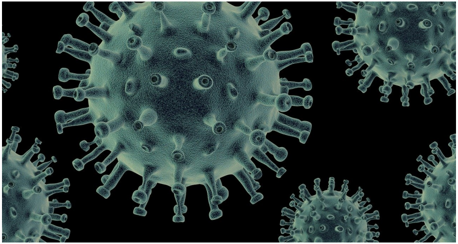
Identification of ARMA models and final estimation of parameters.

8 grudnia 2020



Suppressed sine waves

Definition of suppressed sine wave

The function $f : [0, \infty) \mapsto \mathbb{R}$ in the form

$$f(x) = Aa^{-x} \sin(\omega x + \varphi) \quad \omega, \varphi, A \in \mathbb{R}, a \in (-1, 1)$$

is called **suppressed sine waves** or **suppressed sinusoid**.

Suppressed sine wave

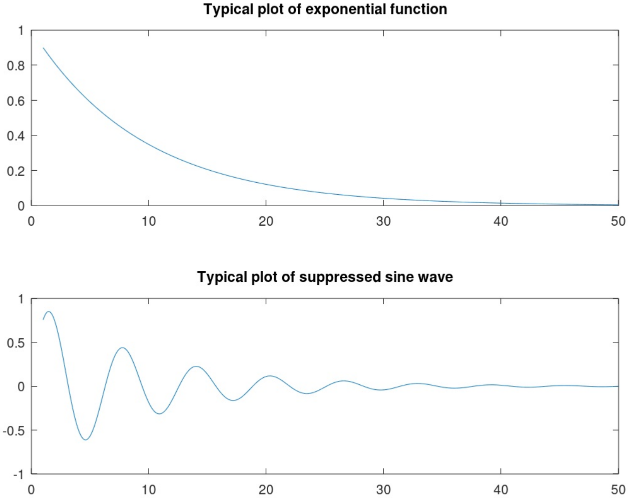
Basic properties of suppressed sine waves:

- the part Aa^{-x} is called **suppressed amplitude** of the sine wave and it tends to 0;
- the value $\omega \in \mathbb{R}_+$ is called **frequency**;
  - if $\omega = 0$, then the suppressed sine waves is in fact **the exponential function**;
- the value φ is called **phase**;
  - if $\varphi = \frac{\pi}{2} + k\pi$, then

    $$f(x) = Aa^{-x}\sin\left(\omega x + \frac{\pi}{2} + k\pi\right) = A(-1)^{k+1}a^{-x}\cos(\omega x),$$

    hence sin is exchanged for cos;

Illustration



Consider the time series $(X_t)_{t\in\mathbb{Z}}$.

- We observe only n values $X_1, X_2, \ldots, X_n$ of this process, and we believe the observed values are only a part the process above lasting forever;
- We compute ACF $(\hat{\rho}_h)_{h\in\mathbb{Z}}$ and PACF $(\hat{\phi}^{YW}_{h,h})_{h\in\mathbb{Z}}$, and imagine that we have noticed that so many values $|\hat{\phi}^{YW}_{h,h}|$ are significant;
- The purpose of ACF and PACF is to identify the lags of processes;
- The pair of plots ACF and PACF is called *correlogram*;
  - **We will see that the shapes of ACF and PACF often look like suppressed sine waves;**
- Using the maximal likelihood method we estimate the parameters of the model.

The premises for MA(q)

MA(q) - repetition

The time series X_t is MA(q) if can be written as

$$X_t = \epsilon_t - \sum_{k=1}^{q} \theta_k \epsilon_{t-k},$$

where ϵ_t is a white noise and θ_k $(k = 1, 2, \ldots, q)$ are unknown parameters.

Basic premises for identifying X_t as the MA(q) model

The typical correlogram of MA(q) looks like

- the plot of PACF looks like an exponential function or suppressed sine wave;
- the plot of ACF breaks down at the lag p (go back to the lecture 24.11.2020):
  - $\hat{\rho}_q$ is significant;
  - All but 5% of the observations $\{\hat{\rho}_h : h > q\}$ are insignificant;
  - Any of significant observations from this set above is relatively close to the significance barrier.

The premises for AR(p)

AR(q) - repetition

The time series X_t is AR(p) if can be written as

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + \epsilon_t,$$

where ϵ_t is a white noise and ϕ_k $(k = 1, 2, \ldots, p)$ are unknown parameters.

Basic premises for identifying X_t as the AR(p) model

The typical correlogram of AR(p) looks like

- the plot of ACF looks like an exponential function or suppressed sine wave;
- the plot of PACF breaks down at the lag p (go back to the lecture 24.11.2020):
  - $\hat{\phi}_{p,p}^{YW}$ is significant;
  - All but 5% of the observations $\{\hat{\phi}_{h,h}^{YW} : h > p\}$ are insignificant;
  - Any of significant observations from this set above is relatively close to the significance barrier.

The premises for ARMA(p,q)

ARMA(p,q) - repetition

The time series X_t is MA(q) if can be written as

$$X_t = \sum_{k=1}^{p} \theta_k X_{t-k} + \epsilon_t - \sum_{k=1}^{p} \theta_k \epsilon_{t-k},$$

where ϵ_t is a white noise and θ_k $(k = 1, 2, \ldots, q)$ and ϕ_k $(k = 1, 2..., p)$ are unknown parameters.

Basic premises for identifying X_t as the ARMA(p,q) model

The typical correlogram of ARMA(p,q) looks like

- the plot of PACF looks like an exponential function or suppressed sine wave or breaks down at very large p;
- the plot of ACF looks like an exponential function or suppressed sine wave or breaks down at very large q;

The premises for white noise

Basic premises for identifying X_t as the white noise ϵ_t

The typical correlogram of ϵ_t looks like

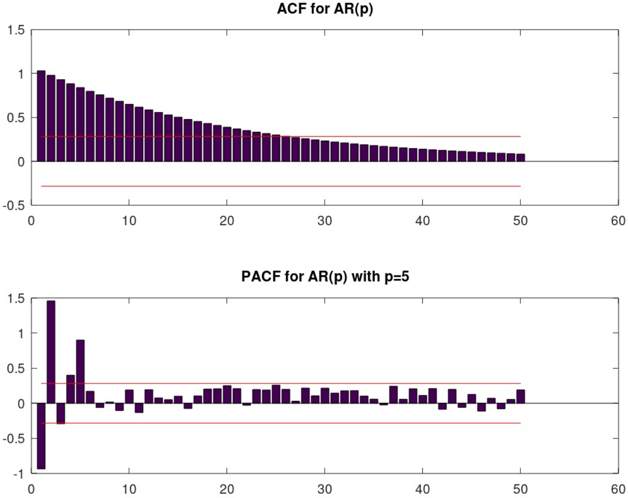
- the plot of ACF breaks down immediately (that is breaks down at 0);
    - All but 5% of the observations $\{\hat{\rho}_h : h = 1, 2, \ldots, n\}$ are insignificant;
    - Any of significant observations from this set above is relatively close to the significance barrier.
- the plot of PACF breaks down immediately;
    - All but 5% of the observations $\{\hat{\phi}^{YW}_{h,h} : h = 1, 2, \ldots, n\}$ are insignificant;
    - Any of significant observations from this set above is relatively close to the significance barrier.

Identification of models by corellogram

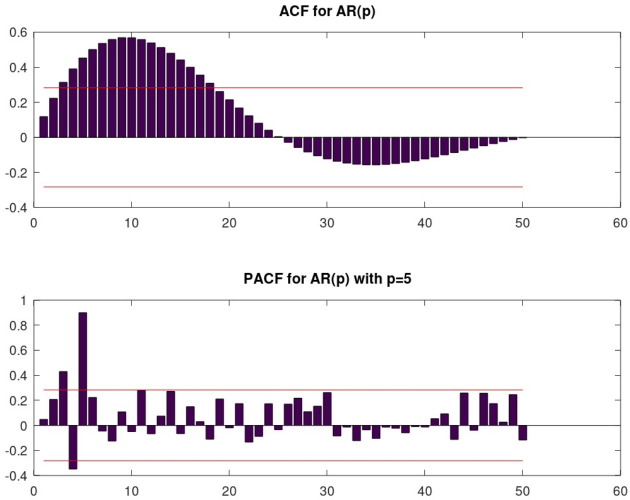
Process	Shape of ACF	Shape of PACF
AR(p)	exponential or suppressed sine wave	breaks down at p
MA(q)	breaks down at q	exponential or suppressed sine wave
ARMA(p,q)	exponential or suppressed sine wave	exponential or suppressed sine wave
white noise	breaks down at 0	breaks down at 0

(VERTE) for illustration.

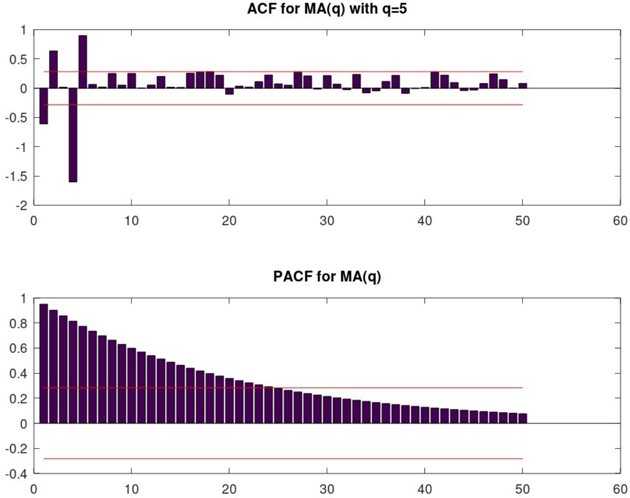
Identification of AR(p), ACF - exponential, PACF - breaks down



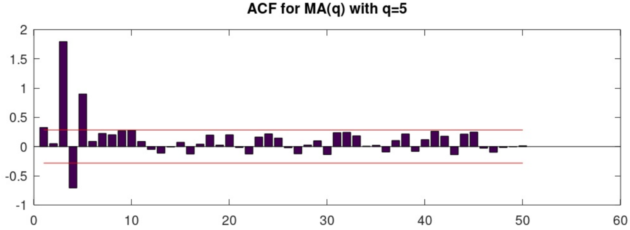
Identification of AR(p), ACF - suppressed sine wave, PACF - breaks down

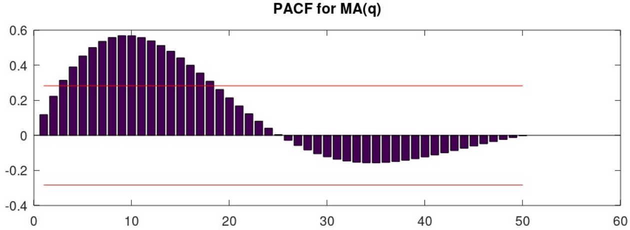


Identification of MA(q), PACF - exponential, ACF - breaks down

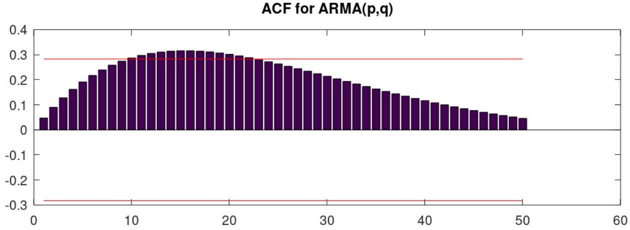


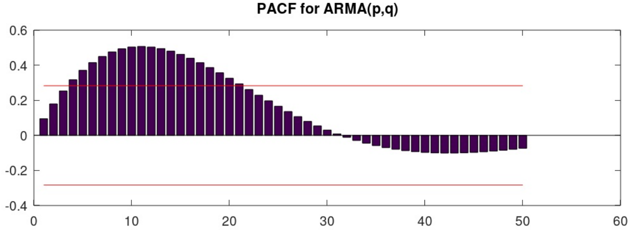
Identification of MA(q), PACF - suppressed sine wave, ACF - breaks down





Identification of ARMA(p,q), both ACF and PACF are exponential or suppressed sines





Final estimation of parameters in AR,MA or ARMA models

When we have already identified one of the following model AR(p), MA(q) we need to know:

- The estimators of parameters $\phi_1, \phi_2, \dots, \phi_p$ of the model AR(p) in the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \epsilon_t$$

  obtained by *Yule-Walker* formula are preliminary, not final;
- The estimators of parameters $\theta_1, \theta_2, \dots, \theta_q$ of the model AR(p) in the form

$$X_t = \epsilon_t - \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

  obtained by *innovation* formula are preliminary as well;
- Final estimators of the parameters of the models AR, MA and ARMA will be find by *maximum likelihood* method.

Suppose that $(X_1, X_2, \dots, X_n)$ is a Gaussian process with mean 0, that is:

- $EX_t = 0$ for all $t = 1, 2, \dots, n$;
- $Var(X_t)$ exists for all $t = 1, 2, \dots, n$ and Γ is its covariance matrix:

$$\Gamma = \begin{bmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & Cov(X_1, X_3) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Cov(X_2, X_2) & Cov(X_2, X_3) & \dots & Cov(X_2, X_n) \\ Cov(X_3, X_1) & Cov(X_3, X_2) & Cov(X_3, X_3) & \dots & Cov(X_3, X_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & Cov(X_n, X_3) & \dots & Cov(X_n, X_n) \end{bmatrix}$$

- This process is Gaussian, that is its density has the form

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det(\Gamma))^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T * \Gamma^{-1} * \mathbf{x}\right) \quad \text{for } \mathbf{x} \in \mathbb{R}.$$

Suppose that the Gaussian time series $(X_1, X_2, \dots, X_n)$ has one of three form AR(p), MA(q) or ARMA(p,q):

- Then in the function $f(\mathbf{x})$ depends on the parameters by Γ:
  - for AR(p), Γ depends on $(\phi_1, \phi_2, \dots, \phi_p)$;
  - for MA(q), Γ depends on $(\theta_1, \theta_2, \dots, \theta_q)$;
  - for ARMA(p,q), Γ depends on $(\phi_1, \phi_2, \dots, \phi_p)$ and $(\theta_1, \theta_2, \dots, \theta_q)$;
  - generally we can write $\Gamma = \Gamma(\vartheta)$, where ϑ is the common notation of vector of the parameters provided one of the model;
- for any $\mathbf{x} \in \mathbb{R}$, the density of this time series is expressed as

$$f_\vartheta(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det(\Gamma(\vartheta)))^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T * \Gamma(\vartheta)^{-1} * \mathbf{x}\right).$$

Hence we compute the estimator $\hat{\vartheta}$ of ϑ as follows: substituting $\mathbf{x}$ by the time series, i.e. $\mathbf{x} = [X_1, X_2, \dots, X_n]^T$

- find $\hat{\vartheta}$ by *maximum likelihood method*, i.e. $\hat{\vartheta}$ maximizes

$$\vartheta \mapsto f_\vartheta(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det(\Gamma(\vartheta)))^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T * \Gamma(\vartheta)^{-1} * \mathbf{x}\right);$$

- equivalently, the maximization problem of $f_\vartheta(\mathbf{x})$ is equivalent to the maximization of $\ln(f_\vartheta)$, hence $\hat{\vartheta}$ solves the *minimization* problem of

$$\vartheta \mapsto \mathbf{x}^T * \Gamma(\vartheta)^{-1} * \mathbf{x} + \ln(\det(\Gamma(\vartheta))).$$

VERTE for derivation of this formula,

- First, let us derive the formula of $\ln\left(f_\vartheta(\mathbf{x})\right)$.

$$\begin{aligned}
\ln(f_\vartheta(\mathbf{x})) &= \ln\left(\frac{1}{(2\pi)^{n/2}(\det(\Gamma(\vartheta)))^{1/2}}\exp\left(-\frac{1}{2}\mathbf{x}^T * \Gamma(\vartheta)^{-1} * \mathbf{x}\right)\right) \\
&= -\frac{1}{2}\left(\mathbf{x}^T * \Gamma(\vartheta)^{-1} * \mathbf{x} + \ln(\det(\Gamma(\vartheta))\right) + \underbrace{\ln\left(\frac{1}{(2\pi)^{n/2}}\right)}_{\text{does not depend on } \vartheta}
\end{aligned}$$

- Hence the problem of maximization of $\ln\left(f_\vartheta(\mathbf{x})\right)$ is equivalent to the minimization of

$$\vartheta \mapsto \mathbf{x}^T * \Gamma(\vartheta)^{-1} * \mathbf{x} + \ln(\det(\Gamma(\vartheta));$$

- The only one problem is to express the formula of Γ by ϑ together with its determinant.

For the model ARMA(p,q) (here $\vartheta := [\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q]^T$):

- The density has the form

$$f_\vartheta(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{1/2}\sigma r_0 r_1 \ldots r_{n-1}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\frac{(x_t - \hat{x}_t)^2}{r_{j-1}}\right)$$

where σ^2 is the constant variance of ϵ_t, $\hat{x}_1 = 0$ and for $t \geq 0$

$$\hat{x}_{t+1} = \begin{cases} \sum_{j=1}^{t}(x_{t+1-j} - \hat{x}_{t+1-j})\tilde{\theta}_{t,j} & \text{if} \quad 1 \leq t < \max(p,q) \\ \sum_{j=1}^{t}(x_{t+1-j} - \hat{x}_{t+1-j})\tilde{\theta}_{t,j} + \sum_{j=1}^{p}\phi_j X_{t+1-j} & \text{if} \quad t \geq \max(p,q), \end{cases}$$

where r_j, and $\tilde{\theta}_{t,j}$ are coefficients provided by innovations algorithm (Brockwell and Davis, *Introduction to Time Series and Forecasting* page 150 for details);

- The estimators of ϑ can be expressed numerically (VERTE).

Gaussian time series - estimation of parameters (sketch)

- The estimator of *distorting* parameter σ^2 has the following form
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\frac{(x_t - \hat{x}_t)^2}{r_{j-1}};$$
- and the parameters $\hat{\vartheta} = [\hat{\phi}, \hat{\theta}]^T$ are expressed numerically as a solution of the minimization problem of the following function
$$\vartheta = [\phi, \theta]^T \mapsto \ln(\hat{\sigma}^2) + \frac{1}{n}\sum_{t=1}^{n}\ln(r_{t-1}).$$