Lesson 12: Estimation of the parameters of an ARMA model

Umberto Triacca

Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica
Università dell'Aquila,
umberto.triacca@univaq.it

Estimation of the parameters of an ARMA model

An $ARMA(p, q)$ model

$$x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} = u_t + \theta u_{t-1} + \ldots + \theta u_{t-p}$$

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

is characterized by $p + q + 1$ unknown parameters

- $\phi = (\phi_1, \ldots, \phi_p)'$
- $\theta = (\theta_1, \ldots, \theta_q)'$
- σ^2

that need to be estimated.

This lesson considers three techniques for estimation of the parameters ϕ, θ and σ^2. They are:

1. Two-Step Regression Estimation
2. Yule-Walker Estimation
3. Maximum Likelihood Estimation

Estimation for ARMA(p, q) process using two-step regression

This method works as follows:

1. We start by regressing x_t on its past $x_{t-1}, ..., x_{t-m}$. We derive the OLS estimates of the coefficients π_j, $j = 1, ..., m$ and of the estimation residuals as well

$$\hat{u}_t = x_t - \sum_{j=1}^{m} \hat{\pi}_j x_{t-j}$$

2. We turn to the ARMA representation of the process by writing it in the form

$$x_t = -\phi_1 x_{t-1} - ... - \phi_p x_{t-p} + \theta_1 u_{t-1} + ... + \theta_q u_{t-q} + u_t$$

This expression suggests to us to regress x_t on $x_{t-1}, ..., x_{t-p}, \hat{u}_{t-1}, ..., \hat{u}_{t-q}$ estimating the coefficients by OLS.

Estimation for ARMA(p, q) process using two-step regression

The regression coefficients so obtained provide consistent estimate of $-\phi_1, ... - \phi_p, \theta_1, ...\theta_q$.

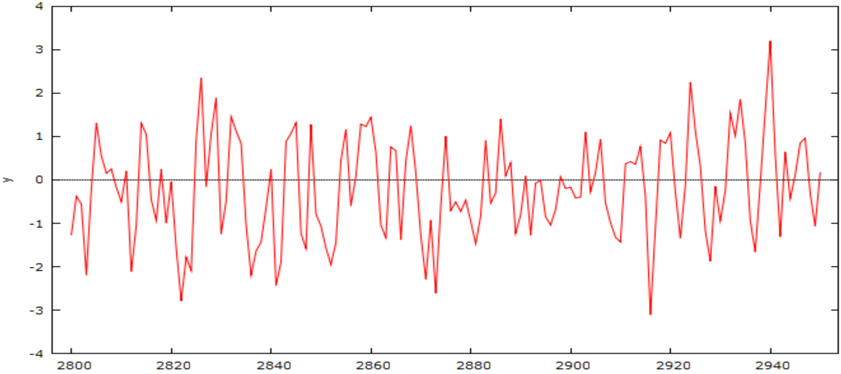
The sum of the squared corrisponding residuals divided by the number of observation corrected by the degrees of freddom is an estimator of σ^2

Estimation for ARMA(p, q) process using two-step regression

Example. We have simulated an MA(1) process defined by

$$x_t = u_t + .7u_{t-1}$$

with $u_t \sim i.i.d.N(0, 1)$



Estimation for ARMA(p, q) process using two-step regression

By using the two-step regression, with $m = 3$, we obtain the following estimates

$$\hat{\theta} = 0.765744$$

$$\hat{\sigma}^2 = 1,0233$$

The Yule-Walker Estimation

Consider an autoregressive stochastic process x_t of order p. It is well known that there is a link among the autoregressive coefficients and the autocovariances.In particular, we have

$$\Gamma\phi = \gamma$$

and

$$\sigma^2 = \gamma(0) - \phi'\gamma$$

where

$$\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix}$$

is the covariance matrix and

$$\gamma = (\gamma_1, ..., \gamma_p)'$$

The Sample Yule-Walker equation

If we replace the theoretical autocovariances by the corresponding sample autocovariances,we obtain

$$\hat{\Gamma}\phi = \hat{\gamma}$$

where

$$\hat{\Gamma} = \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{p-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{p-1} & \hat{\gamma}_{p-2} & \cdots & \hat{\gamma}_0 \end{bmatrix}$$

is the sample autocovariance matrix and

$$\hat{\gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_p)'$$

The Yule-Walker Estimation

We assume $\hat{\gamma}(0) > 0$.To obtain the Yule-Walker estimators as a function of the autocorrelation function, we divide the two sides of equation

$$\hat{\Gamma}\phi = \hat{\gamma}$$

by $\hat{\gamma}(0) > 0$.
We have

$$\hat{R}\phi = \hat{\rho}$$

where

$$\hat{R} = \begin{bmatrix} \hat{\rho}_0 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & \hat{\rho}_0 & \cdots & \hat{\rho}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & \hat{\rho}_0 \end{bmatrix}$$

is the sample autocorrelation matrix and

$$\hat{\rho} = (\hat{\rho}_1, ..., \hat{\rho}_p)'$$

The Yule-Walker Estimation

It is possible to show that

$$\hat{\gamma}(0) > 0 \Rightarrow \det \hat{R} \neq 0$$

The Yule-Walker Estimation

Thus we can solve the system

$$\hat{R}\phi = \hat{\rho}$$

obtaining the so-called Yule-Walker estimators, namely

$$\hat{\phi} = \hat{R}^{-1}\hat{\rho}$$

and

$$\hat{\sigma}^2 = \hat{\gamma}(0)\left[1 - \hat{\rho}'\hat{R}^{-1}\hat{\rho}\right]$$

The Yule-Walker Estimation

Theorem. If x_t is a zero-mean stationary autoregressive process of order p with $u_t \sim iid(0, \sigma^2)$, and $\hat{\phi}$ is the Yule-Walker estimator of ϕ, then

$$T^{1/2}(\hat{\phi} - \phi)$$

has a limiting normal distribution with mean $\mathbf{0}$ and covariance matrix $\sigma^2\Gamma^{-1}$. Moreover

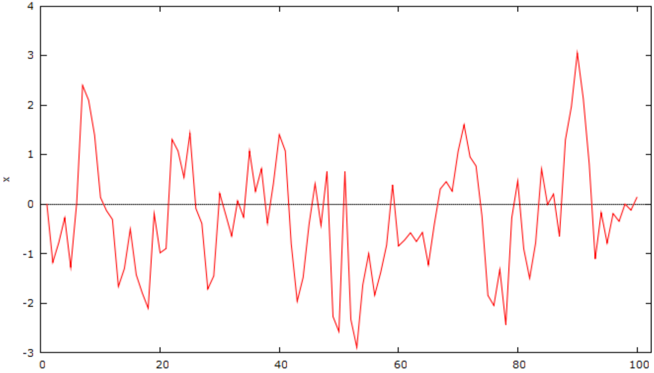
$$\hat{\sigma}^2 \xrightarrow{P} \sigma^2$$

Thus, under the assumption that the order p of the fitted model is the correct value, we can use the asymptotic distribution of ϕ to derive approximate large-sample confidence regions for ϕ and for each of its components.

Numerical example. We have simulated the following AR(1) process:

$$x_t = 0.7x_{t-1} + u_t$$

with $u_t \sim i.i.d.N(0, 1)$



The Yule-Walker Estimation

By using the Yule-Walker estimator we obtain the following estimates

$$\hat{\phi}_1 = \hat{\rho}_1 = 0.6877$$

$$\hat{\sigma}^2 = \hat{\gamma}_0(1 - \hat{\rho}_1) = 0.97989$$

The Yule-Walker estimators with $q > 0$

When $q > 0$ the Yule-Walker estimators are obtained solving the following system

$$\hat{\gamma}_k - \phi_1 \hat{\gamma}_{k-1} - ... - \phi_p \hat{\gamma}_{k-p} = \sigma^2 \sum_{j=k}^{q} \theta_j \psi_{j-k}, \quad 0 \leq k \leq p+q$$

with $\psi_j = 0$ for $j < 0$, $\theta_0 = 1$ and $\theta_j = 0$ for $j \notin \{0, 1..., q\}$.

The Yule-Walker equations with $q > 0$

We note that the equations of the system are nonlinear in the unknown coefficients. This can lead to possible nonexistence and nonuniqueness of solutions.

The Yule-Walker equations with $q > 0$

Example. Consider an MA(1) process, The sample Yule-Walker equation are:

$$\hat{\gamma}_0 = \hat{\sigma}^2(1 + \theta_1^2)$$

$$\hat{\rho}_1 = \frac{\theta_1}{1 + \theta_1^2}$$

We note that if $|\hat{\rho}_1| > .5$, there is no real solution.

The Yule-Walker equations with $q > 0$

If $|\hat{\rho}_1| \leq .5$, then the solution (with $|\hat{\theta}_1| \leq 1$) is

$$\hat{\theta}_1 = \frac{1 - \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1}$$

$$\hat{\sigma}^2 = \frac{\hat{\gamma}_0}{1 + \hat{\theta}_1^2}$$

The Yule-Walker equations with $q > 0$

Numerical Example. Consider again the MA(1) process defined by

$$x_t = u_t + .7u_{t-1}$$

with $u_t \sim i.i.d.N(0, 1)$

The Yule-Walker equations with $q > 0$

In this case $|\hat{\rho}_1| = 0.4751 \leq .5$ Thus the Yule-Walker estimates are

$$\hat{\theta}_1 = 0.16352$$

$$\hat{\sigma}^2 = 1.51791$$

Maximum Likelihood Estimation of the Parameters of ARMA Models

Let $\boldsymbol{\theta} = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \sigma^2)'$ denote the vector of population parameters.
Suppose we have observed a sample of size T

$$\mathbf{x} = (x_1, ..., x_T)$$

Maximum Likelihood Estimation of the Parameters of ARMA Models

Let the joint probability density function (p.d.f.) of these data be denoted

$$f(x_T, x_{T-1}, ..., x_1; \boldsymbol{\theta})$$

The likelihood function is this joint density treated as a function of the parameters $\boldsymbol{\theta}$ given the data $\mathbf{x}$:

$$L(\boldsymbol{\theta}|\mathbf{x}) = f(x_T, x_{T-1}, ..., x_1; \boldsymbol{\theta})$$

Maximum Likelihood Estimation of the Parameters of ARMA Models

The maximum likelihood estimator (MLE) is

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}|\mathbf{x})$$

where $\boldsymbol{\Theta}$ is the parameter space.

Maximum Likelihood Estimation of the Parameters of ARMA Models

For simplifying calculations, it is customary to work with the natural logarithm of L, given by

$$\log L(\boldsymbol{\theta}|\mathbf{x}) = l(\boldsymbol{\theta}|\mathbf{x}).$$

This function is commonly referred to as the log-likelihood.

Maximum Likelihood Estimation of the Parameters of ARMA Models

Since the logarithm is a monotone transformation the values that maximize $L(\boldsymbol{\theta}|\mathbf{x})$ are the same as those that maximize $l(\boldsymbol{\theta}|\mathbf{x})$, that is

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L(\boldsymbol{\theta}|\mathbf{x}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l(\boldsymbol{\theta}|\mathbf{x})$$

but the the log-likelihood is computationally more convenient.

Maximum Likelihood Estimation of the Parameters of ARMA Models

Now, we assume that the derivative of $l(\boldsymbol{\theta}|\mathbf{x})$ (w.r. $\boldsymbol{\theta}$) exists and is continuous for all $\boldsymbol{\theta}$.

The necessary condition for maximizing $l(\boldsymbol{\theta}|\mathbf{x})$ is

$$\frac{\delta l(\boldsymbol{\theta}|\mathbf{x})}{\delta \boldsymbol{\theta}} = \mathbf{0}$$

which is called likelihood equation.

Maximum Likelihood Estimation of the Parameters of ARMA Models

The maximum likelihood estimate, $\hat{\boldsymbol{\theta}}_{MLE}$, will be the solution of

$$\frac{\delta l(\boldsymbol{\theta}|\mathbf{x})}{\delta \boldsymbol{\theta}} = \mathbf{0}$$

Properties of Maximum Likelihood Estimators

Maximum Likelihood Estimators are most attractive because of their asymptotic properties.
Under regularity conditions, the Maximum Likelihood Estimator, $\hat{\boldsymbol{\theta}}_{MLE}$, will have the following asymptotic properties:

1. It is consistent
2. It is asymptotically normally distributed
3. It is asymptotically efficient

These three properties explain the prevalence of the maximum likelihood technique in time series analysis

The exact Gaussian likelihood of an ARMA process

To write down the likelihood function for an ARMA process, one must assume a particular distribution for the white noise process u_t. Here, we assume that u_t is a Gaussian white noise:

$$u_t \sim i.i.d.N(0, \sigma^2)$$

The exact Gaussian likelihood of an ARMA process

This implies that the exact Gaussian likelihood of $\mathbf{x}=(x_1, x_2, ..., x_T)'$ is given by

$$L(\boldsymbol{\theta}|\mathbf{x}) = (2\pi)^{-T/2} |\Gamma(\boldsymbol{\theta})|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{x}'\Gamma(\boldsymbol{\theta})^{-1}\mathbf{x}\right\}$$

where $\Gamma(\boldsymbol{\theta}) = E(\mathbf{x}\mathbf{x}')$ is the $T \times T$ covariance matrix of $\mathbf{x}$ depending on $\boldsymbol{\theta}$.

The exact Gaussian likelihood of an ARMA process

The exact Gaussian log-likelihood is then given by

$$l(\boldsymbol{\theta}|\mathbf{x}) = -\frac{1}{2}\left[T\log(2\pi) + \log|\Gamma(\boldsymbol{\theta})| + \mathbf{x}'\Gamma(\boldsymbol{\theta})^{-1}\mathbf{x}\right]$$

The exact Gaussian likelihood of an AR(1) process

A Gaussian AR(1) process takes the form

$$x_t = \phi_1 x_{t-1} + u_t$$

with

$$u_t \sim i.i.d.N(0, \sigma^2)$$

For this case, the vector of popolation parameters to be estimated consists of $\boldsymbol{\theta} = (\phi_1, \sigma^2)'$.

The exact Gaussian likelihood of an AR(1) process

The exact Gaussian likelihood of $\mathbf{x}=(x_1, x_2, ..., x_T)'$ is given by

$$L(\boldsymbol{\theta}|\mathbf{x}) = (2\pi)^{-T/2} |\Gamma(\boldsymbol{\theta})|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{x}'\Gamma(\boldsymbol{\theta})^{-1}\mathbf{x}\right\}$$

where

$$\Gamma(\boldsymbol{\theta}) = \frac{\sigma^2}{1-\phi_1^2}\begin{bmatrix} 1 & \phi_1 & \phi_1^2 & \cdots & \phi_1^{T-1} \\ \phi_1 & 1 & \phi_1 & \cdots & \phi_1^{T-2} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \phi_1^{T-1} & \phi_1^{T-2} & \phi_1^{T-3} & \cdots & 1 \end{bmatrix}$$

In fact we recall that the j-th autovariance for an $AR(1)$ process is given by

$$E(x_t x_{t-j}) = \frac{\sigma^2 \phi_1^j}{1-\phi_1^2}$$

The exact Gaussian likelihood of an MA(1) process

The exact Gaussian likelihood of $\mathbf{x}=(x_1, x_2, ..., x_T)'$ is given by

$$L(\boldsymbol{\theta}|\mathbf{x}) = (2\pi)^{-T/2} \left|\Gamma(\boldsymbol{\theta})\right|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{x}'\Gamma(\boldsymbol{\theta})^{-1}\mathbf{x}\right\}$$

where

$$\Gamma(\boldsymbol{\theta}) = \sigma^2 \begin{bmatrix} (1+\theta_1) & \theta_1 & 0 & \cdots & 0 \\ \theta_1 & (1+\theta_1) & \theta_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & (1+\theta_1) \end{bmatrix}$$

Non-zero mean μ

Consider an ARMA process $\{x_t; t \in \mathbb{Z}\}$ with mean $\mu \neq 0$, defined by the equation

$$x_t - \phi_1 x_{t-1} - ... - \phi_p x_{t-p} = c + u_t + \theta u_{t-1} + ... + \theta u_{t-p}$$

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

where $\phi^{-1}(1)c = \mu$. The unknown parameters in this model

are

- $\phi = (\phi_1, ..., \phi_p)'$
- $\theta = (\theta_1, ..., \theta_q)'$
- σ^2
- c

Non-zero mean μ

The equation

$$x_t - \phi_1 x_{t-1} - ... - \phi_p x_{t-p} = c + u_t + \theta u_{t-1} + ... + \theta u_{t-p}$$

can be rewritten as

$$(x_t - \mu) - \phi_1(x_{t-1} - \mu) - ... - \phi_p(x_{t-p} - \mu) = u_t + \theta u_{t-1} + ... + \theta u_{t-p}$$

Non-zero mean μ

We estimate μ by

$$\bar{x}_T = \sum_{t=1}^{T} x_t$$

and proceed to analyze the demeaned series

$$\{(x_t - \bar{x}_T); t = 1, ..., T\}$$