Lecture 10: Choosing the best ARMA model

Josep Vives

May 11, 2023

1 Introduction

Given a stationary time series, assume we have good estimations $\hat{\mu}$, $\hat{\sigma}^2$ and $\hat{\rho}$ of μ, σ^2 and ρ, respectively. The question that arises now is, what is the best ARMA(p,q) model for our data?

2 Choosing the best ARMA model

We know

1. In the MA(q) case, we have $\rho(l) = 0$ for all $l \geq q+1$ and $\alpha(l)$ decreases to 0 exponentially.
2. In the AR(p) case, we have $\rho(l)$ decreases exponentially to 0 and $\alpha(l) = 0$ for all $l \geq p+1$.
3. In an ARMA(p,q) model, $\rho(l)$ decreases exponentially to 0 for $l \geq q+1$ and $\alpha(l)$ decreases exponentially to 0 for $l \geq p+1$.

This three facts are the basis to identify heuristically the best ARMA model.

In practice, computationally, the system to fit the best ARMA(p,q) model consist in fitting a net of models for

$$(p,q) \in \{0,\ldots,N\} \times \{0,\ldots,N\}$$

and afterwards, to choose the one that minimizes a certain criterion that search for an equilibrium trying to improve fitting with the minimal number of necessary parameters. Recall that too much parameters make models too complicated and increases the estimation errors.

The most well-known criteria are

- Akaike Information Criterion (AIC):

$$\text{AIC} := -2\log \widehat{L} + 2(p+q+1)$$

- Corrected Akaike Information Criterion (AICC):

$$\text{AICC} := -2\log \widehat{L} + \frac{2n(p+q+1)}{n-p-q-2}$$

- Bayesian Information Criterion (BIC):

$$\begin{aligned}\text{BIC} &:= (n-p-q)\log\frac{n\widehat{\sigma}^2}{n-p-q} + n(1+\log\sqrt{2\pi}) \\ &+ (p+q)\,\log\left(\frac{1}{p+q}\sum_{j=1}^{n}Y_j^2 - \frac{n}{p+q}\widehat{\sigma}^2\right)\end{aligned}$$

- Schwartz Criterion

$$\text{Schwartz} := \log\widehat{\sigma}^2 + \frac{1}{n}(p+q)\log n.$$

Recall that $\widehat{\sigma}^2$ is the maximum likelihood estimator of σ^2 and $\widehat{L}$ is the value of the likelihood function applied to maximum likelihood estimations of the parameters of the model, that is,

$$\widehat{L} := L(y_1,\cdots,y_n,\,\widehat{\mu},\,\widehat{\sigma}^2,\,\widehat{\phi}_1,\cdots,\widehat{\phi}_p,\,\widehat{\theta}_1,\cdots,\widehat{\theta}_q).$$

So, the only pending problem is to estimate parameters ϕ_i and θ_j for $i=1,\ldots,p$ and $j=1,\ldots,q$. This is what we do in the next section.

3 Estimating the parameters of an ARMA(p,q) model

Consider an ARMA model with associated centered white noise Z with variance σ^2. We can always invert the model and write

$$Z_j = \sum_{i=0}^{\infty}\beta_i Y_{j-i}$$

for a certain coefficients β_i. Note that in the AR case the sum is finite and in all other cases the sum is infinite.

The general goal is to choose parameters ϕ and θ such that

$$S(\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q) := \sum_{j=1}^{n}Z_j^2$$

is minimal.

Assume first of all we have an AR(1) model:

$$Y_j = \phi Y_{j-1} + Z_j$$

with $Y_0 = Z_0 = 0$.

To estimate ϕ means to minimize

$$S(\phi) := \sum_{j=1}^{n}(Y_j - \phi Y_{j-1})^2.$$

Note we have

$$S(\phi) = \phi^2\sum_{j=1}^{n}Y_{j-1}^2 - 2\phi\sum_{j=1}^{n}Y_jY_{j-1} + \sum_{j=1}^{n}Y_j^2$$

This is a parabola with a minimum because the coefficient of ϕ^2 is positive. The minimum is attained at value $\hat{\phi}$ such that $S'(\hat{\phi}) = 0$. So,

$$\widehat{\phi} = \frac{\sum_{j=1}^{n} Y_j Y_{j-1}}{\sum_{j=1}^{n} Y_{j-1}^2}.$$

For a general AR(p) model $Y = \{Y_j,\ j \in \mathbb{Z}\}$ we can write

$$Y_j = (Y_{j-1}, \cdots, Y_{j-p}) \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} + Z_j = \bar{Y}_j^t \cdot \phi + Z_j.$$

where

$$\phi^t := (\phi_1, \ldots, \phi_p)$$

and

$$\bar{Y}_j^t := (Y_{j-1}, \ldots, Y_{j-p}).$$

In this case the least squares estimator of ϕ is

$$\widehat{\phi} := \left(\sum_{j=1}^{n} \bar{Y}_j\ \bar{Y}_j^t \right)^{-1} \sum_{j=1}^{n} \bar{Y}_j Y_j$$

where

$$\left(\sum_{j=1}^{n} \bar{Y}_j\ \bar{Y}_j^t \right)^{-1}$$

denotes the inverse matrix.

When $Z \sim \mathrm{GWN}(0, \sigma^2)$, $\widehat{\phi}$ coincide with the maximum likelihood estimator.

The good properties of $\hat{\phi}$ as estimator of ϕ are given by the following theorem:

Theorem 3.1

$$\sqrt{n}(\widehat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2 \Gamma_p^{-1})$$

where

$$\Gamma_p = \mathbb{E} \left\{ \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} (Y_1, \cdots, Y_p) \right\} = \begin{bmatrix} \gamma(0) & \cdots & \gamma(p-1) \\ \vdots & & \vdots \\ \gamma(p-1) & \cdots & \gamma(0) \end{bmatrix}.$$

In the AR(1) case, the previous theorem says

$$\sqrt{n}(\widehat{\phi} - \phi) \xrightarrow{d} N\left(0, \frac{\sigma^2}{\gamma(0)}\right)$$

and using that

$$\gamma(0) = \frac{\sigma^2}{1 - \phi^2},$$

we have

$$\widehat{\phi} \sim \text{AN}\left(\phi, \frac{1-\phi^2}{n}\right). \tag{3.1}$$

Note that formula (3.1) allows to built a confidence interval for ϕ. We have

$$\mathbb{P}(-z(\gamma) \leq \frac{\hat{\phi}-\phi}{\sqrt{1-\phi^2}}\sqrt{n} \leq z(\gamma)) = \gamma$$

where $z(\gamma) = \Phi^{-1}(\frac{1+\gamma}{2})$ and Φ is the cumulative probability function of the standard Normal law.

Isolating ϕ in the previous formula we obtain that

$$I_\gamma(\phi) = [\hat{\phi} - z(\gamma)\frac{\sqrt{1-\phi^2}}{\sqrt{n}}, \hat{\phi} + z(\gamma)\frac{\sqrt{1-\phi^2}}{\sqrt{n}}]$$

is an interval with a $100\gamma\%$ of confidence. Therefore, using that $1-\phi^2 \in (0,1)$, the confidence of the interval

$$I_\gamma(\phi) = [\hat{\phi} - z(\gamma)\frac{1}{\sqrt{n}}, \hat{\phi} + z(\gamma)\frac{1}{\sqrt{n}}]$$

is greater than γ.

An alternative method to estimate ϕ is based on Yule-Walker equations. Recall that we have

$$\rho(k) = \phi_1\rho(k-1) + \cdots + \phi_p\rho(k-p),\ k \geq 1,$$

and in matrix language this is equivalent to

$$\begin{bmatrix} \rho_1 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \cdots & \rho(p-1) \\ \vdots & & \vdots \\ \rho(p-1) & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix}$$

or,

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

Therefore,

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

is an estimator of ϕ.

Statistical packages use a third alternative method called Durbin-Levinson algorithm.

What happens in the MA(q) case? Here things are more complicated. To fix ideas consider first the MA(1) case:

$$Y_j = Z_j - \theta Z_{j-1}.$$

We know that

$$\rho_1 = -\frac{\theta}{1+\theta^2}$$

and this is equivalent to

$$\theta = \frac{-1 \pm \sqrt{1-4\rho_1^2}}{2\rho_1}.$$

So, a natural estimator would be

$$\widehat{\theta} := \frac{-1 \pm \sqrt{1-4\widehat{\rho}_1^2}}{2\widehat{\rho}_1}$$

but this is a complicated nonlinear function.

An alternative method is the following recursive procedure. Note that fixed $\theta \in (-1,1)$ we can compute recursively

$$\begin{aligned}
Z_0 &= 0 \\
Z_1 &= Y_1 \\
Z_2 &= Y_2 + \theta Z_1 \\
Z_3 &= Y_3 + \theta Z_2 = \cdots \\
\vdots & \quad \vdots \\
Z_n &= Y_n + \theta Z_{n-1} = \cdots
\end{aligned}$$

and

$$S(\theta) := \sum_{j=1}^{n} Z_j^2.$$

So, we can search numerically the value θ that minimizes $S(\theta)$. The idea is to do the previous computations for a grid of values of θ in the interval $(-1,1)$ and select the value that minimizes $S(\theta)$.

For an ARMA(1,1) model we do the same with a grid of values of (ϕ,θ) in the region $(-1,1)\times(-1,1)$ minimizing

$$S(\phi,\theta) = \sum_{j=1}^{n} Z_j^2$$

where we compute different values of Z_j recursively by the formula

$$Z_j = Y_j - \phi Y_{j-1} + \theta Z_{j-1}.$$

For the general case ARMA(p,q) we could proceed similarly.