A guide to the two-step regression method for estimating ARMA(p,q) parameters

Michael Hauser

This short article gives a quick and easy to follow method describing a technique to solve for the autoregressive moving average (ARMA) model parameters. The technique described here will be a two-step regression process, where the first step estimates the noise, and the second step estimates the model. There are two other common techniques besides this two-step regression, namely using Yule-Walker and Maximum Likelihood, but again this article will only focus on estimating the parameters via a two-step regression.

An ARMA(p,q) process is a stochastic process given as follows:

\[ x_t = \phi_1 \cdot x_{t-1} + \phi_2 \cdot x_{t-2} + \cdots + \phi_p \cdot x_{t-p} + u_t + \theta_1 \cdot u_{t-1} + \theta_2 \cdot u_{t-2} + \cdots + \theta_q \cdot u_{t-q} \]  

(1)

where the \( u_n \sim N(0, \sigma^2) \) is white Gaussian noise.

It is thus a linear model on the time series history, in which the autoregressive AR(p) portion, namely \( x_t^{AR} := \sum_{n=1}^{p} \phi_n \cdot x_{t-n} \), models the system and the moving average MA(q) portion, namely \( x_t^{MA} := u_t + \sum_{m=1}^{q} \theta_m \cdot u_{t-m} \), models the noise, so that \( x_t = x_t^{AR} + x_t^{MA} \).

Given an N-long time series to train from, \( x_1, x_2, \ldots, x_N \), we can regress over the time series, through the \( \hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p \), to estimate the noise as the error residuals \( u_{p+1}, u_{p+2}, u_{p+3}, \ldots, u_{N} \).

\[
\begin{bmatrix}
  x_p & x_{p-1} & \cdots & x_1 \\
  x_{p+1} & x_p & \cdots & x_2 \\
  x_{p+2} & x_{p+1} & \cdots & x_3 \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{N-1} & x_{N-2} & \cdots & x_{N-p}
\end{bmatrix}
\begin{bmatrix}
  \hat{\phi}_1 \\
  \hat{\phi}_2 \\
  \vdots \\
  \hat{\phi}_p
\end{bmatrix} =
\begin{bmatrix}
  x_{p+1} \\
  x_{p+2} \\
  \vdots \\
  x_{N}
\end{bmatrix}
\]  

(2)

Note that we will not use the \( \hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p \) after this point; the only purpose of this step was to estimate the noise.

The residual errors of the least squares solution to this equation then define the noise \( u_{p+1}, u_{p+2}, \ldots, u_{N} \). Specifically, first we solve Equation 2 in the least-squares sense for the parameters \( \hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p \), and then once we have
these parameters we define \( u_n := (\hat{\phi}_1 \cdot x_{n-1} + \cdots + \hat{\phi}_p \cdot x_{n-p}) - x_n \), for \( n = p + 1, \ldots, N \).

Because these errors \( u_n \sim N(0, \sigma^2) \), the unbiased estimate for the variance is found:

\[
\sigma^2 := \frac{1}{N - p - 1} \sum_{n=p+1}^{N} u_n^2
\]  

Having estimated the \( u_n \)'s, we are now in a position to do the second step and estimate the parameters \( \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q \) via a second linear regression:

\[
\begin{bmatrix}
  x_{p+1} - u_{p+1} \\
  x_{p+2} - u_{p+2} \\
  \vdots \\
  x_N - u_N
\end{bmatrix} =
\begin{bmatrix}
  \phi_1 \\
  \phi_2 \\
  \vdots \\
  \theta_q
\end{bmatrix}
\]  

We assumed \( p > q \), which is usually true when one chooses these parameters (for example typical values are, \( p = 8 \) and \( q = 2 \)). This matrix equation then defines the linear regression learned by least squares for the parameters to learn, \( \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q \), and so we have an estimate of the ARMA\((p, q)\) model.