Estimation of ARMA Models

Eric Zivot

April 6, 2005

1 Maximum Likelihood Estimation of ARMA Models

For iid data with marginal pdf $f(y_t;\boldsymbol{\theta})$, the joint density function for a sample $\mathbf{y} = (y_1, \ldots, y_T)$ is simply the product of the marginal densities for each observation

$$f(\mathbf{y};\boldsymbol{\theta}) = f(y_1, \ldots, y_T;\boldsymbol{\theta}) = \prod_{t=1}^{T} f(y_t;\boldsymbol{\theta})$$

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

$$L(\boldsymbol{\theta}|\mathbf{y}) = L(\boldsymbol{\theta}|y_1, \ldots, y_T) = \prod_{t=1}^{T} f(y_t;\boldsymbol{\theta})$$

The log-likelihood then as the simple form

$$\ln L(\boldsymbol{\theta}|\mathbf{y}) = \sum_{t=1}^{T} \ln f(y_t;\boldsymbol{\theta})$$

For a sample from a covariance stationary time series $\{y_t\}$, the construction of the log-likelihood give above doesn't work because the random variables in the sample $(y_1, \ldots, y_T)$ are not iid. One solution is to try to determine the joint density $f(y_1, \ldots, y_T;\boldsymbol{\theta})$ directly, which requires, among other things, the $T \times T$ variance-covariance matrix $var(\mathbf{y})$. Hamilton describes this approach in detail for Gaussian ARMA processes. An alternative approach relies on factorization of the joint density into a series of conditional densities and the density of a set of initial values. To illustrate this approach, consider the joint density of two adjacent observations $f(y_2, y_1;\boldsymbol{\theta})$ from a covariance stationary time series. The joint density can always be factored as the product of the conditional density of y_2 given y_1 and the marginal density of y_1 :

$$f(y_2, y_1;\boldsymbol{\theta}) = f(y_2|y_1;\boldsymbol{\theta})f(y_1;\boldsymbol{\theta})$$

For three observations, the factorization becomes

$$f(y_3, y_2, y_1; \boldsymbol{\theta}) = f(y_3|y_2, y_1; \boldsymbol{\theta}) f(y_2|y_1; \boldsymbol{\theta}) f(y_1; \boldsymbol{\theta})$$

In general, the conditional marginal factorization has the form

$$f(y_T, \ldots, y_1; \boldsymbol{\theta}) = \left(\prod_{t=p+1}^{T} f(y_t|I_{t-1}, \boldsymbol{\theta}) \right) \cdot f(y_p, \ldots, y_1; \boldsymbol{\theta})$$

where $I_t = \{y_t, \ldots, y_1\}$ denotes the information available at time t, and $y_p, \ldots, y_1$ denotes the initial values. The log-likelihood function may then be expressed as

$$\ln L(\boldsymbol{\theta}|\mathbf{y}) = \sum_{t=p+1}^{T} \ln f(y_t|I_{t-1}, \boldsymbol{\theta}) + \ln f(y_p, \ldots, y_1; \boldsymbol{\theta})$$

The full log-likelihood function is called the exact log-likelihood. The first term is called the conditional log-likelihood, and the second term is called the marginal log-likelihood for the initial values.

In the maximum likelihood estimation of time series models, two types of maximum likelihood estimates (mles) may be computed. The first type is based on maximizing the conditional log-likelihood function. These estimates are called conditional mles and are defined by

$$\hat{\boldsymbol{\theta}}_{cmle} = \arg\max_{\theta} \sum_{t=p+1}^{T} \ln f(y_t|I_{t-1}, \boldsymbol{\theta})$$

The second type is based on maximizing the exact log-likelihood function. These estimates are called exact mles, and are defined by

$$\hat{\boldsymbol{\theta}}_{mle} = \arg\max_{\theta} \sum_{t=p+1}^{T} \ln f(y_t|I_{t-1}, \boldsymbol{\theta}) + \ln f(y_p, \ldots, y_1; \boldsymbol{\theta})$$

For stationary models, $\hat{\boldsymbol{\theta}}_{cmle}$ and $\hat{\boldsymbol{\theta}}_{mle}$ are consistent and have the same limiting normal distribution. In finite samples, however, $\hat{\boldsymbol{\theta}}_{cmle}$ and $\hat{\boldsymbol{\theta}}_{mle}$ are generally not equal and my differ by a substantial amount if the data are close to being non-stationary or non-invertible.

Example 1 *Maximum likelihood estimation of an AR(1) model*

Consider the stationary AR(1) model

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim iid\ N(0, \sigma^2), \ t = 1, \ldots, T \\ \boldsymbol{\theta} &= (c, \phi, \sigma^2)', \ |\phi| < 1 \end{aligned}$$

Conditional on I_{t-1}

$$y_t|I_{t-1} \sim N(c + \phi y_{t-1}, \sigma^2),\ t = 2, \ldots, T$$

which only depends on y_{t-1}. The conditional density $f(y_t|I_{t-1}, \theta)$ is then

$$f(y_t|y_{t-1}, \boldsymbol{\theta}) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(y_t - c - \phi y_{t-1})^2\right),\ t = 2, \ldots, T$$

To determine the marginal density for the initial value y_1, recall that for a stationary AR(1) process

$$\begin{aligned} E[y_1] &= \mu = \frac{c}{1-\phi} \\ var(y_1) &= \frac{\sigma^2}{1-\phi^2} \end{aligned}$$

It follows that

$$\begin{aligned} y_1 &\sim N\left(\frac{c}{1-\phi}, \frac{\sigma^2}{1-\phi^2}\right) \\ f(y_1; \boldsymbol{\theta}) &= \left(2\pi\frac{\sigma^2}{1-\phi^2}\right)^{-1/2} \exp\left(-\frac{1-\phi^2}{2\sigma^2}\left(y_1 - \frac{c}{1-\phi}\right)^2\right) \end{aligned}$$

The conditional log-likelihood function is

$$\begin{aligned} \sum_{t=2}^{T} \ln f(y_t|y_{t-1}, \boldsymbol{\theta}) &= \frac{-(T-1)}{2}\ln(2\pi) - \frac{(T-1)}{2}\ln(\sigma^2) \\ &\quad -\frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - c - \phi y_{t-1})^2 \end{aligned}$$

Notice that the conditional log-likelihood function has the form of the log-likelihood function for a linear regression model with normal errors. It follows that the conditional mles for c and ϕ are identical to the least squares estimates from the regression

$$y_t = c + \phi y_{t-1} + \varepsilon_t, t = 2, \ldots, T$$

and the conditional mle for σ^2 is

$$\hat{\sigma}^2_{cmle} = (T-1)^{-1}\sum_{t=2}^{T}(y_t - \hat{c}_{cmle} - \hat{\phi}_{cmle} y_{t-1})^2$$

The marginal log-likelihood for the initial value y_1 is

$$\ln f(y_1; \boldsymbol{\theta}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\frac{\sigma^2}{1-\phi^2}\right) - \frac{1-\phi^2}{2\sigma^2}\left(y_1 - \frac{c}{1-\phi}\right)^2$$

The exact log-likelihood function is then

$$
\begin{aligned}
\ln L(\boldsymbol{\theta}|\mathbf{y}) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\frac{\sigma^2}{1-\phi^2}\right) - \frac{1-\phi^2}{2\sigma^2}\left(y_1 - \frac{c}{1-\phi}\right)^2 \\
&\quad -\frac{(T-1)}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}(y_t - c - \phi y_{t-1})^2
\end{aligned}
$$

The exact log-likelihood function is a non-linear function of the parameters θ, and so there is no closed form solution for the exact mles. The exact mles must be determined by numerically maximizing the exact log-likelihood function. Usually, a Newton-Raphson type algorithm is used for the maximization which leads to the interative scheme

$$
\hat{\boldsymbol{\theta}}_{mle,n} = \hat{\boldsymbol{\theta}}_{mle,n-1} - \hat{\mathbf{H}}(\hat{\boldsymbol{\theta}}_{mle,n-1})^{-1}\hat{\mathbf{s}}(\hat{\boldsymbol{\theta}}_{mle,n-1})
$$

where $\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})$ is an estimate of the Hessian matrix (2nd derivative of the log-likelihood function), and $\hat{\mathbf{s}}(\hat{\boldsymbol{\theta}})$ is an estimate of the score vector (1st derivative of the log-likelihood function). The estimates of the Hessian and Score may be computed numerically (using numerical derivative routines) or they may be computed analytically (if analytic derivatives are known).

1.1 Prediction Error Decomposition

For general time series models, the log-likelihood function may be computed using an algorithm known as the *prediction error decomposition*. To illustrate this algorithm, consider again the simple AR(1) model. Recall,

$$
y_t|I_{t-1} \sim N(c + \phi y_{t-1}, \sigma^2),\ t = 2, \ldots, T
$$

from which it follows that

$$
\begin{aligned}
E[y_t|I_{t-1}] &= c + \phi y_{t-1} \\
var(y_t|I_{t-1}) &= \sigma^2
\end{aligned}
$$

The 1-step ahead prediction errors may then be defined as

$$
v_t = y_t - E[y_t|I_{t-1}] = y_t - c + \phi y_{t-1},\ t = 2, \ldots T
$$

The variance of the prediction error at time t is

$$
f_t = var(v_t) = var(\varepsilon_t) = \sigma^2,\ t = 2, \ldots T
$$

For the initial value, the first prediction error and its variance are

$$
\begin{aligned}
v_1 &= y_1 - E[y_1] = y_1 - \frac{c}{1-\phi} \\
f_1 &= var(v_1) = \frac{\sigma^2}{1-\phi^2}
\end{aligned}
$$

Using the prediction errors and the prediction error variances, the exact log-likelihood function may be re-expressed as

$$\ln L(\boldsymbol{\theta}|\mathbf{y}) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln f_t - \frac{1}{2}\sum_{t=1}^{T}\frac{v_t^2}{f_t}$$

which is the prediction error decomposition.

Remarks

1. A further simplification may be achieved by writing

$$\begin{aligned} var(v_t) &= \sigma^2 f_t^* \\ &= \sigma^2 \cdot \frac{1}{1-\phi^2} \text{ for } t = 1 \\ &= \sigma^2 \cdot 1 \text{ for } t > 1 \end{aligned}$$

   That is $f_t^* = 1/(1-\phi^2)$ for $t = 1$ and $f_t^* = 1$ for $t > 1$. Then the log-likelihood becomes

$$\ln L(\boldsymbol{\theta}|\mathbf{y}) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\sum_{t=1}^{T}\ln f_t^* - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\frac{v_t^2}{f_t^*}$$

2. For general time series models, the prediction error decomposition may be conveniently computed as a by product of the *Kalman filter algorithm* if the time series model can be cast in *state space form*.