ECONOMICS 7395, Spring, 2005
Bent E. Sørensen
March 1, 2005

1 Asymptotic theory and testing.

Maximum likelihood theory was developed for models with independent identical observations, but it turns out that most of the standard results and asymptotic formulae emerge in exactly the same way for time series models as they do for iid models. This statement is only true for stable models, so until further notice it is implicitly assumed that all the models are stable. The results below are asymptotic so initial values will likewise be ignored.

A general result in maximum likelihood theory is that if ψ is a vector of parameters and $L(\psi)$ is the likelihood function for a single observation then the maximum likelihood estimator $\tilde{\psi}$ has an asymptotically normal distribution with the true parameter as the mean and the inverse information matrix as the variance.
The information matrix is defined as

$$I(\psi) \ = \ -E(\frac{\partial^2 logL}{\partial\psi\partial\psi'}) \ ,$$

where L is the likelihood function corresponding to a single observation, and in the iid case one can show that

$$\sqrt{T}(\psi - \tilde{\psi}) \ \Rightarrow \ N(0, I(\psi)^{-1}) \ ,$$

subject to regularity conditions. The most important regularity conditions are sufficient differentiability of the likelihood function (typically 2 times differentiable with continuous second derivative) and that the likelihood function has a unique solution for ψ, i.e. that ψ is identified.

In the time series case we define the *asymptotic information matrix*

$$IA(\psi) \ = \ -plimT^{-1}(\frac{\partial^2 logL_T}{\partial\psi\partial\psi'}) \ ,$$

where $L_T(\psi)$ is the likelihood function corresponding to the full set of observations. Under regularity conditions (see below) one can then show that

$$\sqrt{T}(\psi - \tilde{\psi}) \ \Rightarrow \ N(0, IA(\psi)^{-1}) \ ,$$

where $\Rightarrow$ indicates convergence in distribution for the sample size (T) going to infinity. Following Harvey (1980,1989) I will express this as

$$\tilde{\psi} \sim AN(\psi, T^{-1} IA(\psi)^{-1}) \ .$$

Example For an AR(1) model with N(0,σ^2)errors, it is straightforward to find the asymptotic information matrix as

$$IA(a, \sigma^2) = \begin{pmatrix} \frac{1}{1-a^2} & \\ & \frac{1}{2\sigma^4} \end{pmatrix} .$$

I will show it in detail for this example. The contribution from a single observation x_t is:

$$logL(x_t|x_{t-1}) = -\frac{1}{2} log\sigma^2 - \frac{(x_t - ax_{t-1})^2}{2\sigma^2} \ .$$

We find that

$$\begin{aligned}
\frac{\partial logL}{\partial a} &= \frac{(x_t - ax_{t-1})x_{t-1}}{\sigma^2} \\
\frac{\partial logL}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} + \frac{(x_t - ax_{t-1})^2}{2\sigma^4} \\
\frac{\partial^2 logL}{\partial a \partial a} &= -\frac{x_{t-1}^2}{\sigma^2} \\
\frac{\partial^2 logL}{\partial \sigma^2 \partial a} &= -\frac{(x_t - ax_{t-1})x_{t-1}}{\sigma^4} \\
\frac{\partial^2 logL}{\partial \sigma^2 \partial \sigma^2} &= \frac{1}{2\sigma^4} - \frac{(x_t - ax_{t-1})^2}{\sigma^6} \ .
\end{aligned}$$

Now, by the law of large numbers, the asymptotic information matrix can be found by taking the mean of the contribution to the likelihood function from a single observation.

In the general ARMA case it will still be the case that the estimator of the error variance will be distributed independently of the parameters in the lag polynomials. If ψ is the vector of parameters $a_1, ..., a_k, b_1, ..., b_l$ of an ARMA(k,l) model with normally distributed error terms then (as in the example above) we find

$$\frac{\partial^2 logL}{\partial \psi \partial \psi'} = -\frac{z_t z_t'}{\sigma^2} \ ,$$

where $z_t = \frac{-\partial u_t}{\partial \psi}$.

Example For the MA(1) we have

$$u_t = x_t - bu_{t-1} \ ,$$

from which we find the recursion equation

$$\frac{\partial u_t}{\partial b} \;=\; -b\frac{\partial u_{t-1}}{\partial b} \;-\; u_{t-1} \; .$$

Now use the following very elegant argument. The recursion implies that z_t follows an AR(1) process

$$z_t \;= b z_{t-1} \;+\; u_{t-1} \; ,$$

where u_{t-1} is uncorrelated with z_{t-1} (why?). Now from our results fom AR(1) models we find that

$$E(z_t^2) \;=\; var(z_t) \;=\; \frac{\sigma^2}{1-b^2} \; ,$$

from which we conclude that $\hat{b}$ is asymptotically normally distributed with mean b and variance $1-b^2$.

For the ARMA(1,1) the same type of reasoning can be applied, see Harvey (1980) p. 131 where it is shown that the asymptotic variance Avar(a,b) of $\sqrt{T}(a,b)$ is

$$Avar(a,b) \;=\; \frac{1+ab}{(a+b)^2} \begin{pmatrix} (1-a^2)(1+ab) & -(1-a^2)(1-b^2) \\ -(1-a^2)(1-b^2) & (1-b^2)(1+ab) \end{pmatrix} \; .$$

The formulae become more complicated for higher order processes, and one way around this is to evaluate the asymptotic covariance matrix numerically, rather than analytically. Harvey (1989) p. 140-143, shows how one can extend the Kalman filter to include the derivatives of the likelihood function in the updating and predicting loops so as to arrive at an estimate of the asymptotic variance.