Stat 804

Lecture 13 Notes

Non Gaussian series.

The fitting methods we have studied are based on the likelihood for a normal fit. However, the estimates work reasonably well even if the errors are not normal.

Example: AR(1) fit. We fit $X_t - \mu = \rho(X_{t-1} - \mu) + \epsilon_t$ using $\hat{\mu} = \bar{X}$ which is consistent for non-Gaussian errors. (In fact

$$(1-\rho)\sum_0^{T-1} X_t + \rho X_{T-1} - X_0 = (T-1)(1-\rho)\mu + \sum_0^{T-1} \epsilon_t - \epsilon_0;$$

divide by T and apply the law of large numbers to $\bar{\epsilon}$ to see that $\bar{X}$ is consistent.)

Here is an outline of the logic of what follows. We will assume that the errors are iid mean 0, variance σ^2 and finite fourth moment $\mu_4 = \mathrm{E}(\epsilon_t^4)$. We will **not** assume that the errors have a normal distribution.

1. The estimates of ρ and σ are consistent.

2. The score function satisfies

   $$T^{-1/2}U(\theta_0) \Rightarrow MVN(0, B)$$

   where

   $$B = \left[\begin{array}{cc} \frac{1}{1-\rho^2} & 0 \\ 0 & \frac{\mu_4 - \sigma^4}{\sigma^6} \end{array}\right]$$

3. The matrix of second derivatives satisfies

   $$\lim_{T\to\infty} -\frac{1}{T}\frac{\partial U}{\partial \theta} = \lim_{T\to\infty} -\frac{1}{T}\mathrm{E}\left(\frac{\partial U}{\partial \theta}\right) = A$$

   where

   $$A = \left[\begin{array}{cc} \frac{1}{1-\rho^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{array}\right]$$

4. If $\mathcal{I}$ is the (conditional) Fisher information then

   $$\lim \lim_{T\to\infty} \frac{1}{T}\mathcal{I} = A$$

5. We can expand $U(\hat{\theta})$ about θ_0 and get

   $$\hat{\theta} - \theta = \left[\frac{1}{T}I(\theta_0)\right]^{-1}\left[T^{-1/2}U(\theta_0)\right] + \text{negligible remainder}$$

6. So

$$\hat\theta - \theta \approx MVN(0, A^{-1}BA^{-1}) = MVN(0,\Sigma)$$

where

$$\Sigma = A^{-1}BA^{-1} = \begin{bmatrix} 1-\rho^2 & 0 \\ 0 & \frac{\mu_4-\sigma^4}{4\sigma^2} \end{bmatrix}$$

7. So $T^{1/2}(\hat\rho - \rho) \Rightarrow N(0, 1-\rho^2)$ even for non-normal errors.

8. On the other hand the estimate of σ has a limiting distribution which will be different for non-normal errors (because it depends on μ_4 which is $3\sigma^4$ for normal errors and something else in general for non-normal errors).

Here are details.

Consistency: One of our many nearly equivalent estimates of ρ is

$$\hat\rho = \frac{\sum(X_t-\bar X)(X_{t-1}-\bar X)}{\sum(X_t-\bar X)^2}$$

Divide both top and bottom by T. You need essentially to prove

$$T^{-1}\sum(X_t-\mu)(X_{t-1}-\mu) \to C(1)$$

and

$$T^{-1}\sum(X_t-\mu)^2 \to C(0)$$

Each of these is correct and hinges on the fact that these linear processes are ergodic — long time averages converge to expected values. For these particular averages it is possible to compute means and variances and prove that the mean squared error converges to 0.

Score function: asymptotic normality

The score function is

$$U(\rho,\sigma) = \begin{bmatrix} \frac{\sum X_{t-1}(X_t-\rho X_{t-1})}{\sigma^2} \\ \frac{\sum(X_t-\rho X_{t-1})^2}{\sigma^3} - \frac{T-1}{\sigma} \end{bmatrix}$$

If ρ and σ are the true values of the parameters then

$$U(\rho,\sigma) = \begin{bmatrix} \frac{\sum X_{t-1}\epsilon_t}{\sigma^2} \\ \frac{\sum \epsilon_t^2}{\sigma^3} - \frac{T-1}{\sigma} \end{bmatrix}$$

I claim that $T^{-1/2}U(\rho,\sigma) \Rightarrow MVN(0,B)$. This is proved by the martingale central limit theorem. Technically you fix an $a \in R^2$ and study $T^{-1/2}a^tU(\rho,\sigma)$, proving that the limit is $N(0, a^tBa)$. I do here only the special cases $a=(1,0)^t$ and $a=(0,1)^t$. The second of these is simply

$$T^{-1/2}\sum(\epsilon_i^2-\sigma^2)/\sigma^3$$

which converges by the usual CLT to $N(0,(\mu_4-\sigma^4)/\sigma^6)$. For $a=(1,0)^t$ the claim is that

$$T^{-1/2}\sum X_{t-1}\epsilon_t \Rightarrow N(0, C(0)\sigma^2)$$

because $C(0) = \sigma^2/(1-\rho^2)$.

To prove this assertion we define for each T a martingale $M_{T,k}$ for $k = 1, \ldots, T$ where

$$M_{T,k} = \sum_1^T D_{T,i}$$

with

$$D_{T,i} = T^{-1/2} X_{i-1}\epsilon_i$$

The martingale property is that

$$\mathrm{E}(M_{T,k+1} | \epsilon_{t-1}, \epsilon_{t-2}, \ldots) = M_{T,k}$$

The martingale central limit theorem (Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* New York: Academic Press.) states that

$$M_{T,T} \Rightarrow N(0,b)$$

provided that

$$\sum_k D_{T,k}^2 \to b$$

and provided that an analogue of Lindeberg's condition holds. Here I check only the former condition:

$$\sum_k D_{T,k}^2 = \frac{1}{T} \sum X_{t-1}^2 \epsilon_t^2 \to \mathrm{E}(X_0^2 \epsilon_1^2) = C(0)\sigma^2$$

(by the ergodic theorem or you could compute means and variances).

Second derivative matrix and Fisher information: the matrix of negative second derivatives is

$$-\frac{\partial U}{\partial \theta} = \begin{bmatrix} \frac{\sum X_{t-1}^2}{\sigma^2} & 2\frac{\sum X_{t-1}(X_t - \rho X_{t-1})}{\sigma^3} \\ 2\frac{\sum X_{t-1}(X_t-\rho X_{t-1})}{\sigma^3} & 3\frac{\sum (X_t - \rho X_{t-1})^2}{\sigma^4} - \frac{T-1}{\sigma^2} \end{bmatrix}$$

If you evaluate at the true parameter value and divide by T the matrix and the expected value of the matrix converge to

$$A = \begin{bmatrix} \frac{C(0)}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$$

(Again this uses the ergodic theorem or a variance calculation.)

Taylor expansion: In the next step we are supposed to prove that a random vector has a MVN limit. The usual tactic to prove this uses the so called Cramér-Wold device — you prove that each linear combination of the entries in the vector has a univariate normal limit. Then $U(\hat\rho, \hat\sigma) = 0$ and Taylor's theorem is that

$$0 = U(\hat\rho, \hat\sigma) = U(\rho, \sigma) + \left[\frac{\partial U(\theta)}{\partial \theta}\right](\hat\theta - \theta) + R$$

(Here we are using $\theta^t = (\rho, \sigma)$ and R is a remainder term — a random variable with the property that

$$P(||R||/||U(\theta)|| > \eta) \to 0$$

for each $\eta > 0$.) Multiply through by

$$\left[\frac{\partial U(\theta)}{\partial \theta}\right]^{-1}$$

and get

$$T^{1/2}(\hat\theta - \theta) = \left[-T^{-1}\frac{\partial U(\theta)}{\partial \theta}\right]^{-1}(T^{-1/2}U(\rho,\sigma) + T^{-1/2}R)$$

It is possible with care to prove that

$$\left[-T^{-1}\frac{\partial U(\theta)}{\partial \theta}\right]^{-1}(T^{-1/2}R) \to 0$$

Asymptotic normality: This is a consequence of Slutsky's theorem applied to the Taylor expansion and the results above for U and I. According to Slutsky's theorem the asymptotic distribution of $T^{1/2}(\hat\theta - \theta)$ is the same as that of

$$A^{-1}(T^{-1/2}U(\rho,\sigma))$$

which converges in distribution to $MVN(0, A^{-1}B(A^{-1})^t)$. Now since $C(0) = \sigma^2/(1-\rho^2)$

$$A^{-1}B(A^{-1})^t = \left[\begin{array}{cc} 1-\rho^2 & 0 \\ 0 & \frac{\mu_4-\sigma^4}{4\sigma^4} \end{array}\right]$$

Behaviour of $\hat\rho$: pick off the first component and find

$$T^{1/2}(\hat\rho - \rho) \Rightarrow N(0, 1-\rho^2)$$

Notice that this answer is the same for normal and non-normal errors.
Behaviour of $\hat\sigma$: on the other hand

$$T^{1/2}(\hat\sigma - \sigma) \Rightarrow N(0, (\mu_4 - \sigma^4)/(4\sigma^2))$$

which has μ_4 in it and will match the normal theory limit if and only if $\mu_4 = 3\sigma^4$.
More general models: For an ARMA(p,q) model the parameter vector is

$$\theta = (a_1, \ldots, a_p, b_1, \ldots, b_q, \sigma)^t .$$

In general the matrices B and A are of the form

$$B = \left[\begin{array}{cc} B_1 & 0 \\ 0 & \frac{\mu_4-\sigma^4}{\sigma^6} \end{array}\right]$$

and

$$A = \left[\begin{array}{cc} A_1 & 0 \\ 0 & \frac{2}{\sigma^2} \end{array}\right]$$

where $A_1 = B_1$ and A_1 is a function of the parameters $a_1, \ldots, a_p, b_1, \ldots, b_q$ only and is the same for both normal and non-normal data.

Model assessment.

Having fitted an ARIMA model you get (essentially automatically) fitted residuals $\hat{\epsilon}$. Most of the fitting methods lead to fewer residuals than there were in the original series. Since the parameter estimates are consistent (if the model fitted is correct, of course) the fitted residuals should be essentially the true ϵ_t which is white noise. We will assess this by plotting the estimated ACF of $\hat{\epsilon}$ and then seeing if the estimates are all close enough to 0 to pass for white noise.

To judge close enough we need asymptotic distribution theory for autocovariance estimates.