Chapter 2 - Inference for Stationarity and Ergodicity processes

Sample Moments

Sample mean

$$\overline{Y} = \frac{1}{T} \sum_{t=1}^{T} Y_t$$

Sample autocovariance

$$\widehat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} \left(Y_t - \overline{Y} \right) \left(Y_{t-j} - \overline{Y} \right)$$

Sample autocorrelation

$$\widehat{\rho}_j = \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0}$$

LLN and CLT for stationary and ergodic processes

Consider again model MA(1) $Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$ where $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is an independent, identically distributed process, with $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2$. Then,

$$Y_t + Y_{t-1} = 2\mu + \varepsilon_t + (1+\theta)\varepsilon_{t-1} + \theta\varepsilon_{t-2}$$

$$Y_t + Y_{t-1} + Y_{t-2} = \\ 3\mu + \varepsilon_t + (1+\theta)\varepsilon_{t-1} + (1+\theta)\varepsilon_{t-2} + \theta\varepsilon_{t-3}$$

$$\sum_{t=1}^{T} Y_t = T\mu + \varepsilon_T + (1+\theta)\sum_{t=1}^{T-1} \varepsilon_t + \theta\varepsilon_0$$

For the LLN,

$$\frac{\sum_{t=1}^{T} Y_t}{T} = \mu + (1+\theta)\frac{\sum_{t=1}^{T-1} \varepsilon_t}{T} + \frac{\varepsilon_T}{T} + \frac{\theta\varepsilon_0}{T}$$

and notice that $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ meets the conditions for the LLN, therefore

$$\frac{\sum_{t=1}^{T-1} \varepsilon_t}{T} \rightarrow_p 0$$

Moreover, since ε_t is stochastically bounded, it also follows that

$$\frac{\varepsilon_0}{T} \rightarrow_p 0, \ \frac{\varepsilon_T}{T} \rightarrow_p 0$$

and therefore

$$\frac{\sum_{t=1}^{T} Y_t}{T} \rightarrow_p \mu$$

Proceeding in the same way, for the CLT,

$$\sqrt{T}\frac{\sum_{t=1}^{T}(Y_t - \mu)}{T} = (1+\theta)\sqrt{T}\frac{\sum_{t=1}^{T-1}\varepsilon_t}{T} + \frac{\varepsilon_T}{\sqrt{T}} + \frac{\theta\varepsilon_0}{\sqrt{T}}$$

and notice that $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ meets the conditions for the CLT, therefore

$$\sqrt{T}\frac{\sum_{t=1}^{T-1}\varepsilon_t}{T} \to_d N(0,\sigma^2)$$

Moreover, since ε_t is stochastically bounded, it also follows that

$$\frac{\varepsilon_0}{\sqrt{T}} \to_p 0, \ \frac{\varepsilon_T}{\sqrt{T}} \to_p 0$$

and therefore

$$\sqrt{T}\frac{\sum_{t=1}^{T}(Y_t - \mu)}{T} \to_d N(0,\sigma^2(1+\theta)^2)$$

Heuristically, the LLN and CLT still hold because we decomposed our sum $\sum_{t=1}^{T}(Y_t - \mu)$ into the sum of an independent process $\sum_{t=1}^{T-1}(1+\theta)\varepsilon_t$, for which we know the LLN and CLT hold, and residuals that are stochastically negligible in a large sample. This technique is called **Beveridge-Nelson decomposition**.

Obviously, the same argument holds for the (MA(2)) model $Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$.

Heuristically, the same argument based on the Beveridge and Nelson decomposition also holds for any

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

provided that $\psi_j \to 0$ fast enough, as j gets large. Sufficient condition for this is

$$\sum_{j=0}^{\infty} j^{1/2} |\psi_j| < \infty$$

yielding the LLN $\frac{\sum_{t=1}^{T} Y_t}{T} \to_p \mu$ and the CLT

$$\sqrt{T} \frac{\sum_{t=1}^{T} (Y_t - \mu)}{T} \to_d N(0, \sigma^2 (\sum_{j=0}^{\infty} \psi_j)^2)$$

- Conditon $\sum_{j=0}^{\infty} j^{1/2} |\psi_j| < \infty$ imply that both $|\psi_j|$ and $|\gamma_j|$ drop to zero very quickly as j gets large.

Using again the Beveridge Nelson decomposition and similar conditions we can also establish the LLN and CLT for the sample autocovariance and the sample autocorrelation,

$$\begin{aligned} \widehat{\gamma}_j &\to_p \gamma_j \\ \widehat{\rho}_j &\to_p \rho_j \end{aligned}$$

and

$$\sqrt{T}\left(\widehat{\rho}_j - \rho_j\right) \to_d N\left(0, W_j\right)$$

for a known variance W_j.

- Notice that these results are for a fixed j. Clearly, these results do not hold, for example, for $j = T - 1$, as this would imply estimating estimating γ_j and ρ_j using only one observation, and the estimate is not consistent; by the same argument, it is clear that estimates for j proportional to T are also inconsistent.

Autocorrelation robust inference on the sample mean

Notice that

$$\sigma^2(\sum_{j=0}^{\infty}\psi_j)^2 = \sum_{j=-\infty}^{\infty}\gamma_j = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j$$

To test hypothesis on the sample mean, we need to know the term $\sigma^2(\sum_{j=0}^{\infty}\psi_j)^2$, which is known as "**long run variance**".

When we do not know the long run variance, we must estimate it, in order to test hypotheses on μ.
Both parametric and nonparametric methods can be used.

non-parametric estimation of the long run variance.

- As we noticed before, we only have $T-1$ estimates of γ_j. Fortunately, from stationarity and ergodicity and from the regularity conditions given with the Beveridge-Nelson decomposition, we know that ψ_j and γ_j decline to 0 very fast as j gets large, so as a first approximation we set $\widehat{\gamma}_j = 0$ as estimates for $j \geq T$. Hopefully, the mistake that we make with this assumption is small when T is large, and negligible as $T \to \infty$.
- This, however, is not enough, as estimates $\widehat{\gamma}_j$ are inconsistent for large j. To limit its potentially disruptive effect, we weight these estimates with a small weight.

These two arguments lead to estimates of the type

$$\widehat{\gamma_0} + 2\sum_{j=1}^{T-1} k_j \widehat{\gamma}_j$$

where k_j is a weight called *kernel* and it is such that $k_j \to 0$ as $j \to T$.

Two such estimates are

$$\widehat{\gamma_0} + 2\sum_{j=1}^{M} \widehat{\gamma}_j,\ \ M/T \to 0,\ \ \textit{rectangular kernel estimate}$$

$$\widehat{\gamma_0} + 2\sum_{j=1}^{M} \frac{M-j}{M}\widehat{\gamma}_j,\ \ M/T \to 0,\ \ \textit{triangular kernel estimate}$$

The triangular kernel estimate is also known as Bartlett (kernel) estimate, or as Newey-West estimate

- One problem with these non-parametric estimates is that, depending on the kernel, it is possible that the estimate is negative: obviously, this is not desirable for an estimate of a variance. As a matter of fact, the rectangular kernel estimate may be negative, and it is therefore rarely used, whereas the triangular kernel estimate is always non-negative.

- term M is known as *bandwdith*. As a rule of thumb, $M = \sqrt{T}$ is used, and theory for "optimal" choice of M has been developed.

parametric estimation of the long run variance.

Using a parametric model we can characterize the long run variance as a function of a small number of parameters only. For example,

- when Y_t is defined by the MA(1) model, $\sigma^2(\sum_{j=0}^{\infty}\psi_j)^2 = \sigma^2(1+\theta)^2$

- when Y_t is defined by the AR(1) model, $\sigma^2(\sum_{j=0}^{\infty}\psi_j)^2 = \sigma^2(\frac{1}{1-\phi})^2$

parametric or non parametric estimation?

- Adantages of parametric estimation. When we do not know the long run variance, it is sufficient to consistently estimate a small number of parameters to estimate the long run variance. Thus, the estimate of the long run variance may be more precise (making inference more reliable).
- Adantages of non-parametric estimation. Parametric estimation may be inconsistent (if we specify the wrong parametric model). Non-parametric estimation is not subject to this risk.
- Adantages of non-parametric estimation. Parametric estimation may be more time consuming, both for the time to select the correct model and for the effective estimation time in the parametric context.

On balance, non-parametric estimation seems to be more frequently used.