Lecture 2 – Asymptotic Theory: Review of Some Basic Concepts
(Reference – 2.1, Hayashi)

Before we develop a large sample theory for time series regressions, it will be useful to quickly review some fundamental concepts that you should be familiar with from Econ 671.

1. Convergence of sequences of random variables

2. Strong and weak laws of large numbers

3. Central Limit Theorem
Convergence of Sequences of Random Variables (and Random Vectors)

Let \{z_n\} denote a sequence of random variables, \(z_1, z_2, \ldots\), with cumulative distribution functions (c.d.f.) \(F_1, F_2, \ldots\), respectively.

Definition: Convergence in Probability

\{z_n\} converges in probability to the constant \(\alpha\) if for any \(\varepsilon > 0\),

\[
\lim_{n \to \infty} \Pr(\{|z_n - \alpha| > \varepsilon\}) = 0
\]

In this case, we write \(z_n \to \alpha\) or \(\text{plim } z_n = \alpha\).
Definition: Almost Sure Convergence

\{z_n\} converges almost surely (a.s) to the constant \(\alpha\) if

\[
\Pr\left(\lim_{n\to\infty} z_n = \alpha\right) = 1
\]

In this case, we write \(z_n \xrightarrow{a.s.} \alpha\).
Note: $z_n \to \alpha_{a.s.}$ implies that $z_n \to \alpha$, though the reverse is not true.

In econometric applications, where the $z_n$’s will be estimators, the distinction is not of practical importance. That is, convergence in probability will be sufficient for our needs. However, it is sometimes easier to formulate a set of assumptions to prove a.s. convergence, so that in econometric theory the stronger convergence criteria is often more useful.

Note: Both of these convergence criteria extend to the case where $z_n$ and $\alpha$ are k-dimensional by applying the definition element by element.
Definition: Convergence in Distribution

\{z_n\} \textbf{converges in distribution} to the random variable \(z\) if the sequence of c.d.f.’s \(\{F_n\}\) converges to \(F\), the c.d.f. of \(z\), at all continuity points of \(F\).

In this case, we write \(z_n \xrightarrow{d} z\) and we call \(F\) the asymptotic (or limiting) distribution of \(z_n\).

Note: If \(z_n\) is a sequence of \(k\)-dimensional random vectors with joint c.d.f.s \(F_n\), then \(\{z_n\}\) converges in distribution to the \(k\)-dimensional random vector \(z\) if \(\{F_n\}\) converges to \(F\), the joint c.d.f. of \(z\), at all continuity points of \(F\).
The role of these convergence criteria in econometrics –

Let \( \hat{\theta}_n \) denote an estimator of a parameter \( \theta \), constructed from a sample of size \( n \). For a given \( n \), \( \hat{\theta}_n \) is a random variable (since its value depends upon the particular sample drawn). If we consider varying the sample size, we can generate a sequence of estimators or random variables \( \{ \hat{\theta}_n \} \).

If we allow the sample size to increase without bound and if \( \hat{\theta}_n \xrightarrow{p} \theta \), then we say that \( \hat{\theta}_n \) is a weakly consistent estimator of \( \theta \). If \( \hat{\theta}_n \xrightarrow{a.s.} \theta \), then we say that \( \hat{\theta}_n \) is a strongly consistent estimator of \( \theta \). (Clearly, strong consistency implies weak consistency since a.s. convergence implies convergence in probability.)
Consistency is a desirable property for an estimator because it means that with a sufficiently large sample your estimator will be “very likely” to be “very close” to the actual parameter value. (Compare to unbiasedness, which says that if you average $\hat{\theta}_n$ across a sufficient number of samples of size $n$, you will get very close to the actual value, $\theta$.)
Suppose that $\hat{\theta}_n$ is a (weakly or strongly) consistent estimator of $\theta$. It follows that $\hat{\theta}_n \xrightarrow{d} \theta$. That is, the limiting distribution of $\hat{\theta}_n$ is the distribution of the trivial random variable $\theta$.

This is not a very use approximate distribution to use in practice to draw inferences about $\theta$.

However, if we scale $\hat{\theta}_n$ by a factor $f(n)$, it may be that $f(n)\hat{\theta}_n$ converges in distribution to a nontrivial random variable and that nontrivial limiting distribution can be used in practice to draw inferences about $\theta$.

In particular, suppose that $\hat{\theta}_n$ is a consistent estimator of $\theta$ and that

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma)$$
In this case:

- $\hat{\theta}_n$ is $\sqrt{n}$-consistent (because $\hat{\theta}_n$ must be scaled by a this factor to attain a nontrivial limiting distribution; put another way, $\hat{\theta}_n$ is converging to $\theta$ at the rate $\sqrt{n}$).

- $N(0,\Sigma)$ is called the asymptotic distribution of $\hat{\theta}_n$ and we say that $\hat{\theta}_n$ is asymptotically normal with asymptotic variance matrix $\Sigma$.

If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,\Sigma)$ then hypothesis testing and confidence interval construction are very straightforward, provided we can find a consistent estimator of $\Sigma$. (More on this later.)
We establish the consistency and asymptotic normality of an estimator by applying a Law of Large Numbers (for consistency) and a Central Limit Theorem (for asymptotic normality).

Laws of large numbers (LLNs) are concerned with the convergence in probability (weak LLNs) or almost surely (strong LLNs) of the sequence of sample means associated with \( \{z_n\} \) to some constant. Central Limit Theorems (CLTs) are concerned with the convergence in distribution of a properly scaled sequence of sample means associated with \( \{z_n\} \), generally to a normal distribution.
Kolmogorov’s Strong LLN:

Let \{z_n\} be an i.i.d. sequence of r.v.’s with \(\text{E}(z_n) = \mu\) for all \(n\). Then \(\bar{z}_n \to \mu\) a.s.

Note – This LLN states that if the mean of the i.i.d. sequence exists then the sample mean is a strongly consistent estimator of the population mean.
Application (Theorem 3.5, White (2001)): Sufficient Conditions for Consistency of OLS

Assume –

\begin{align*}
  \text{i)} & \quad y_t = X_t' \beta + \epsilon_t, \; t = 1,2,... \\
  \text{ii)} & \quad E(X_t \epsilon_t) = 0, \; t = 1,2,... \\
  \text{iii)} & \quad E(X_t X_t') = M, \text{ where } M \text{ is a finite p.d. matrix, } t = 1,2,... \\
  \text{iv)} & \quad \{X_t', \epsilon_t\} \text{ is an i.i.d. sequence}
\end{align*}

Then

\[ \hat{\beta}_n \text{ exists a.s. for sufficiently large } n \text{ and } \hat{\beta}_n \rightarrow \beta \]
Note that:

(i) is just the usual linearity assumption (A.1)

(ii) replaces the strict exogeniety condition (A.2) with the weaker and more “time series friendly” assumption that the disturbances and the regressors are “contemporaneously uncorrelated”: \( E(\varepsilon_t \mid X_1, \ldots, X_n) \) vs. \( E(\varepsilon_t \mid X_t) \)

(iii) is essentially an asymptotic version of the no multicollinearity assumption (A.3)

(iv) replaces the spherical disturbances assumption (A.4) with a stronger and even less “time series friendly” assumption: the disturbances and the regressors are i.i.d.!
This assumption rules out serial correlation and conditional heteroskedasticity in the disturbances. It also rules out serial correlation and conditional heteroskedasticity in the regressors!
Lindeberg-Levy CLT:

Let \( \{z_n\} \) be an i.i.d. sequence of r.v.'s with \( E(z_n) = \mu \) and \( \text{var}(z_n) = \Sigma \). Then

\[
\sqrt{n}(z_n - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (z_i - \mu) \xrightarrow{d} N(0, \Sigma)
\]

That is, for an i.i.d. sequence with finite mean and variance, the sample mean is a \( \sqrt{n} \) - consistent estimator of the population mean and is asymptotically normal, with asymptotic variance equal to the variance of \( z_n \).

Note that the Lindeberg-Levy CLT, like the Kolmogorov LLN, requires an i.i.d. sequence with finite mean but also requires a finite variance and, in this sense, relies on a stronger set of conditions than the Kolmogorov LLN.
Application (Theorem 5.3, White (2001)): Sufficient Conditions for Asymptotic Normality of OLS

Assume –
\(i\) \(y_t = X_t'\beta + \epsilon_t, \ t = 1,2,...\)
\(ii\) \(E(X_t\epsilon_t) = 0, \ t = 1,2,...\)
\(iii\) \(E(X_tX_t') = M, \) where \(M\) is a finite p.d. matrix, \(t = 1,2,...\)
\(iv\) \(\{X_t', \epsilon_t\} \) is an i.i.d. sequence
\(v\) \(V_n \equiv \text{var}(X'\epsilon / \sqrt{n}) = V, \) where \(V\) is a finite p.d. matrix for all \(n > N.\)

Then
\[ D^{-1/2} \sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0,I) \] where \(D \equiv M^{-1}VM^{-1} \)

Notice that we have simply added assumption (v), a stronger second moment condition, to go from the consistency of the OLS estimator to the asymptotic normality of the estimator.
More important for our purposes, however, is that this theorem relies on the assumption that the X’s and ε’s are i.i.d.

Our next task is to develop a set of conditions under which the OLS estimator is consistent and asymptotically normal in a time series setting.

To do this, we will first have to introduce some new concepts to describe the behavior of time series in ways that will allow us to formulate an LLN and CLT that are suitable for time series regressions.