Economics 583: Econometric Theory I
A Primer on Asymptotics: Time Series Concepts

Eric Zivot

January 16, 2013

Time Series Concepts

A *stochastic process* $\{Y_t\}_{t=1}^{\infty}$ is a sequence of random variables indexed by time t :

$$\{\ldots, Y_1, Y_2, \ldots, Y_t, Y_{t+1}, \ldots\}$$

A realization of a stochastic process is the sequence of observed data $\{y_t\}_{t=1}^{\infty}$:

$$\{\ldots, Y_1 = y_1, Y_2 = y_2, \ldots, Y_t = y_t, Y_{t+1} = y_{t+1}, \ldots\}$$

We are interested in the conditions under which we can treat the stochastic process like a random sample, as the sample size goes to infinity. Under such conditions, at any point in time t_0, the *ensemble average*

$$\frac{1}{N}\sum_{k=1}^{N} Y_{t_0}^{(k)}$$

will converge to the sample *time average*

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

as N and T go to infinity. If this result occurs then the stochastic process is called *ergodic*.

Stationary Stochastic Processes

Definition 1 *Strict stationarity*

A stochastic process $\{Y_t\}_{t=1}^{\infty}$ is *strictly stationary* if, for any given finite integer r and for any set of subscripts $t_1, t_2, \ldots, t_r$ the joint distribution of

$$(Y_t, Y_{t_1}, Y_{t_2}, \ldots, Y_{t_r})$$

depends only on $t_1 - t, t_2 - t, \ldots, t_r - t$ but not on t.

Remarks

1. For example, the distribution of (Y_1, Y_5) is the same as the distribution of (Y_{12}, Y_{16}).

2. For a strictly stationary process, Y_t has the same mean, variance (moments) for all t.

3. Any function/transformation $g(\cdot)$ of a strictly stationary process, $\{g(Y_t)\}$ is also strictly stationary.

Example 1 *iid sequence*

If $\{Y_t\}$ is an iid sequence, then it is strictly stationary.

Let $\{Y_t\}$ be an iid sequence and let $X \sim N(0,1)$ independent of $\{Y_t\}$. Let $Z_t = Y_t + X$. Then the sequence $\{Z_t\}$ is strictly stationary.

Since $\{Z_t\}$ is strictly stationary, $\{Z_t^2\}$ is also strictly stationary.

Definition 2 *Covariance (Weak) stationarity*

A stochastic process $\{Y_t\}_{t=1}^{\infty}$ is *covariance stationary* (weakly stationary) if

1. $E[Y_t] = \mu$ does not depend on t
2. cov$(Y_t.Y_{t-j}) = \gamma_j$ exists, is finite, and depends only on j but not on t for $j = 0, 1, 2, \ldots$

Remark:

A strictly stationary process is covariance stationary if the mean and variance exist and the covariances are finite.

For a weakly stationary process $\{Y_t\}_{t=1}^{\infty}$ define the following moments:

$$\begin{aligned}
\gamma_j &= \text{cov}(Y_t, Y_{t-j}) = \text{cov}(Y_{t+j}, Y_t) = \text{ j}^{th} \text{ order autocovariance} \\
\gamma_0 &= \text{var}(Y_t) = \text{ variance} \\
\rho_j &= \gamma_j / \gamma_0 = \text{ j}^{th} \text{ order autocorrelation}
\end{aligned}$$

Remark

A weakly stationary process is uniquely determined by its mean, variance and autocovariances.

Definition 3 *Ergodicity*

Loosely speaking, a stochastic process $\{Y_t\}_{t=1}^{\infty}$ is *ergodic* if any two collections of random variables partitioned far apart in the sequence are almost independently distributed. The formal definition of ergodicity is highly technical (see Hayashi 2000, p. 101 and note typo from errata).

Proposition 1 *LLN for Stationary and Ergodic Time Series (Hamilton, 1994 page 47).*

Let $\{Y_t\}$ be a covariance stationary process with mean $E[Y_t] = \mu$ and autocovariances $\gamma_j = \text{cov}(Y_t, Y_{t-j})$. If

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty$$

then $\{Y_t\}$ is *ergodic for the mean*. That is, $\bar{Y} \xrightarrow{p} E[Y_t] = \mu$.

Example 2 *MA(1)*

Let

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1},\ |\theta| < 1,\ \varepsilon_t \sim \text{iid } (0, \sigma^2)$$

Then

$$\begin{aligned} E[Y_t] &= \mu \\ \gamma_0 &= E[(Y_t - \mu)^2] = \sigma^2(1 + \theta^2) \\ \gamma_1 &= E[(Y_t - \mu)(Y_{t-1} - \mu)] = \sigma^2\theta \\ \gamma_k &= 0,\ k > 1 \end{aligned}$$

Clearly,

$$\sum_{j=0}^{\infty} |\gamma_j| = \sigma^2(1 + \theta^2 + |\theta|) < \infty$$

so that $\{Y_t\}$ is both weakly stationary and ergodic.

Theorem 2 *Ergodic Theorem*

Let $\{Y_t\}$ be stationary and ergodic with $E[Y_i] = \mu$. Then

$$\bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t \xrightarrow{p} E[Y_t] = \mu$$

Remarks

1. The ergodic theorem says that for a stationary and ergodic sequence $\{Y_t\}$ the time average converges to the ensemble average as the sample size gets large. That is, the ergodic theorem is a LLN for stochastic processes.

2. The ergodic theorem is a substantial generalization of Kolmongorov's LLN because it allows for serial dependence in the time series.

3. Any transformation $g(\cdot)$ of a stationary and ergodic process $\{Y_t\}$ is also stationary and ergodic. That is, $\{g(Y_t)\}$ is stationary and ergodic. Therefore, if $E[g(Y_t)]$ exists then the ergodic theorem gives

$$\bar{g} = \frac{1}{T}\sum_{t=1}^{T} g(Y_t) \xrightarrow{p} E[g(Y_t)]$$

This is a very useful result. For example, we may use it to prove that the sample autocovariances

$$\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T} (Y_t - \bar{Y})(Y_{t-j} - \bar{Y})$$

converge in probability to the population autocovariances $\gamma_j = E[(Y_t - \mu)(Y_{t-j} - \mu)] = cov(Y_t, Y_{t-j})$.

Example 3 *Stationary but not ergodic process (White, 1984)*

Let $\{Y_t\}$ be an iid sequence with $E[Y_t] = \mu$ and let $X \sim N(0, 1)$ independent of $\{Y_t\}$. Let $Z_t = Y_t + X$. Note that $E[Z_t] = \mu$.

Claim: Z_t is stationary but not ergodic.

(Show proof on white board)

Martingales and Martingale Difference Sequences

Let $\{Y_t\}$ be a sequence of random variables and let $\{I_t\}$ be a sequence of information sets ($\sigma-$fields) with $I_t \subset I$ for all t and I the universal information set. For example,

$$\begin{aligned} I_t &= \{Y_1, Y_2, \ldots, Y_t\} = \text{ past history of } Y_t \\ I_t &= \{(Y_s, Z_s)_{s=1}^t\},\ \{Z_t\} = \text{ auxiliary variables} \end{aligned}$$

Definition 4 *Conditional Expectation*

Let Y_t be a random variable with conditional pdf $f(y_t|I_s)$, where I_s is an information set with $s < t$. Then

$$E[Y_t|I_s] = \int_{-\infty}^{\infty} y_t f(y_t|I_s) dy_t$$

Proposition 3 *Law of Iterated Expectation*

Let I_1 and I_2 be information sets such that $I_1 \subseteq I_2$, and let Y be a random variable such that $E[Y|I_1]$ and $E[Y|I_2]$ are defined. Then

$$E[Y|I_1] = E[E[Y|I_2]|I_1] \text{ (smaller set wins)}$$

If $I_1 = \varnothing$ (empty set) then

$$\begin{aligned} E[Y|I_1] &= E[Y] \text{ (unconditional expectation)} \\ E[Y] &= E[E[Y|I_2]|\varnothing] = E[E[Y|I_2]] \end{aligned}$$

Definition 5 *Martingale*

The pair (Y_t, I_t) is a *martingale* (MG) if

1. $I_t \subset I_{t+1}$ (increasing sequence of information sets - a *filtration*)
2. $Y_t \subset I_t$ (Y_t is *adapted* to I_t; i.e., Y_t is an event in I_t)
3. $E[|Y_t|] < \infty$
4. $E[Y_t | I_{t-1}] = Y_{t-1}$ (MG property)

Example 4 *Random walk*

Let $Y_t = Y_{t-1} + u_t$ where $\{u_t\}$ is an iid sequence with mean zero and variance σ^2. Let $I_t = \{Y_1, Y_2, \ldots, Y_t\}$. Then

$$E[Y_t|I_{t-1}] = Y_{t-1}$$

Example 5 *Heteroskedastic random walk*

Let $Y_t = Y_{t-1} + u_t/t = Y_{t-1} + v_t$ where $\{u_t\}$ is an iid sequence with mean zero and variance σ^2 and $v_t = u_t/t$. Note that var$(v_t) = \sigma^2/t$. Let $I_t = \{Y_1, Y_2, \ldots, Y_t\}$. Then

$$E[Y_t|I_{t-1}] = Y_{t-1}$$

Remark

If (Y_t, I_t) is a MG, then

$$E[Y_{t+m}|I_t] = Y_t \text{ for all } t \geq 1$$

To see this, let $m = 2$ and note that by iterated expectations

$$E[Y_{t+2}|I_t] = E[E[Y_{t+2}|I_{t+1}]|I_t]$$

By the MG property

$$E[Y_{t+2}|I_{t+1}] = Y_{t+1}$$

which leave us with

$$E[Y_{t+2}|I_t] = E[Y_{t+1}|I_t] = Y_t$$

Definition 6 *Martingale Difference Sequence (MDS)*

The pair (u_t, I_t) is a *martingale difference sequence* (MDS) if (u_t, I_t) is an adapted sequence and

$$E[u_t|I_{t-1}] = 0$$

Remarks

1. If (Y_t, I_t) is a MG and we define

   $$u_t = Y_t - E[Y_t|I_{t-1}]$$

   we have, by virtual construction,

   $$E[u_t|I_{t-1}] = 0$$

   so that (u_t, I_t) is a MDS.

2. The sequence $\{u_t\}$ is sometime referred to as a sequence of *nonlinear innovations*. The term arises because if Z_t is any function of the past history of Y_t, and thus $Z_t \subset I_t$, we have by iterated expectations

$$\begin{aligned} E[u_t Z_{t-1}] &= E[E[u_t Z_{t-1} | I_{t-1}]] \\ &= E[Z_{t-1} E[u_t | I_{t-1}]] \\ &= 0 \end{aligned}$$

so that u_t is orthogonal to any function of the past history of Y_t.

3. If (u_t, I_t) is a MDS then

$$E[u_{t+m} | I_t] = 0$$

Example 6 *ARCH process*

Consider the first order *autoregressive conditional heteroskedasticity* (ARCH(1)) process

$$\begin{aligned} u_t &= Z_t \sigma_t \\ & \quad Z_t \sim \text{iid } N(0,1) \\ \sigma_t^2 &= \omega + \alpha u_{t-1}^2, \ 0 < \alpha < 1, \ \omega > 0 \end{aligned}$$

The process was proposed by Nobel Lauret Robert Engle to describe and predict time varying volatility in macroeconomic and financial time series. If $I_t = \{u_t, u_{t-1}, \ldots, u_1\}$ then (u_t, I_t) is a stationary and ergodic conditionally heteroskedastic MDS.

(Derive properties on white board)

Theorem 4 *Multivariate CLT for stationary and ergodic MDS (Billingsley, 1961)*

Let $(\mathbf{u}_t, I_t)$ be a vector MDS that is stationary and ergodic with $k \times k$ covariance matrix $E[\mathbf{u}_t \mathbf{u}_t'] = \mathbf{\Sigma}$. Let

$$\bar{\mathbf{u}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{u}_t$$

Then

$$\sqrt{T} \bar{\mathbf{u}} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{u}_t \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma})$$

or, equivalently,

$$\bar{\mathbf{u}} \overset{A}{\sim} N(\mathbf{0}, T^{-1} \mathbf{\Sigma})$$

Large Sample Distribution of Least Squares Estimator

As an application of the previous results, consider estimation and inference for the linear regression model

$$y_t = \underset{(1\times k)}{\mathbf{x}_t'} \underset{(k\times 1)}{\boldsymbol{\beta}} + \varepsilon_t, \; t = 1, \ldots, T$$

under the following assumptions:

Assumption 1 (Linear regression model with stochastic regressors)

1. $\{\mathbf{x}_t, \varepsilon_t\}$ is jointly stationary and ergodic

2. $E[\mathbf{x}_t\mathbf{x}_t'] = \mathbf{\Sigma}_{xx}$ is positive definite (full rank k)

3. $E[x_{jt}\varepsilon_t] = 0$ for all j, t $(j = 1, \ldots, k)$ (i.e., predetermined regressors)

4. The process $\{\mathbf{g}_t\} = \{\mathbf{x}_t\varepsilon_t\}$ is a MDS with $E[\mathbf{g}_t\mathbf{g}_t'] = E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2] = \mathbf{S}$ nonsingular.

Remarks

1. Part 1 implies that ε_t is stationary so that $E[\varepsilon_t^2] = \sigma^2$ is the unconditional variance.

2. Part 4 allows for general conditional heteroskedasticity; e.g. var$(\varepsilon_t|\mathbf{x}_t) = f(\mathbf{x}_t)$. In this case,

$$\begin{aligned} E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2] &= E[E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2|\mathbf{x}_t]] \\ &= E[\mathbf{x}_t\mathbf{x}_t'E[\varepsilon_t^2|\mathbf{x}_t]] = E[\mathbf{x}_t\mathbf{x}_t'f(\mathbf{x}_t)] = \mathbf{S} \end{aligned}$$

If the errors are conditionally homoskedastic then var$(\varepsilon_t|x_t) = \sigma^2$ and

$$\begin{aligned} E[\mathbf{x}_t\mathbf{x}_t'\varepsilon_t^2] &= E[\mathbf{x}_t\mathbf{x}_t'E[\varepsilon_t^2|\mathbf{x}_t]] \\ &= \sigma^2 E[\mathbf{x}_t\mathbf{x}_t'] = \sigma^2\mathbf{\Sigma}_{xx} = \mathbf{S} \end{aligned}$$

The least squares estimator of $\boldsymbol{\beta}$ is

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{y}_t \\
&= \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t (\mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t) \\
&= \boldsymbol{\beta} + \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t
\end{aligned}$$

For the asymptotic analysis, we subtract $\boldsymbol{\beta}$ from both sides giving

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t = \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \sum_{t=1}^{T} \mathbf{g}_t$$

Proposition 5 *Consistency and asymptotic normality of the least squares estimator*

Under Assumption 1, as $T \rightarrow \infty$

1. $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$

2. $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}\boldsymbol{\Sigma}_{xx}^{-1})$ which implies $\hat{\boldsymbol{\beta}} \overset{A}{\sim} N(\boldsymbol{\beta}, T^{-1}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{S}\boldsymbol{\Sigma}_{xx}^{-1})$

Remarks

1. Typically, $\mathbf{\Sigma}_{xx}$ and S are unknown and must be estimated. Hence, the practically useful result is

$$\hat{\boldsymbol{\beta}} \overset{A}{\sim} N(\boldsymbol{\beta}, T^{-1}\hat{\mathbf{\Sigma}}_{xx}^{-1}\hat{\mathbf{S}}\hat{\mathbf{\Sigma}}_{xx}^{-1})$$

where $\hat{\mathbf{\Sigma}}_{xx} \xrightarrow{p} \mathbf{\Sigma}_{xx}$and $\hat{\mathbf{S}} \xrightarrow{p} \mathbf{S}$. Question: How to consistently estimate $\mathbf{\Sigma}_{xx}$ and $\mathbf{S}$?

2. The estimator $\widehat{\mathsf{avar}}(\hat{\boldsymbol{\beta}}) = T^{-1}\hat{\mathbf{\Sigma}}_{xx}^{-1}\hat{\mathbf{S}}\hat{\mathbf{\Sigma}}_{xx}^{-1}$ is often referred to as the "White" or heteroskedasticity consistent (HC) estimator. The square root of the diagonal elements of the HC estimator are known as the White or HC standard errors for $\hat{\beta}_i$:

$$\widehat{\mathsf{SE}}_{\mathsf{HC}}(\hat{\beta}_i) = \sqrt{\left[T \cdot (\mathbf{X}'\mathbf{X})^{-1}\hat{\mathbf{S}}(\mathbf{X}'\mathbf{X})^{-1}\right]_{ii}},\ i = 1, \ldots, k$$

3. If the errors are conditionally homoskedastic, then $var(\varepsilon_t^2|\mathbf{x}_t) = \sigma^2$, $\mathbf{S} = \sigma^2\mathbf{\Sigma}_{xx}$ and

$$\text{avar}(\hat{\boldsymbol{\beta}}) = T^{-1}\sigma^2\mathbf{\Sigma}_{xx}^{-1}$$

Using

$$\begin{aligned}
\mathbf{S}_{xx} &= T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t' = T^{-1}\mathbf{X}'\mathbf{X} \xrightarrow{p} \mathbf{\Sigma}_{xx} \\
\hat{\sigma}^2 &= \frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}})^2 \xrightarrow{p} \sigma^2
\end{aligned}$$

then gives the usual textbook result

$$\widehat{\text{avar}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

Sketch of Proof (Do one white board)

Proposition 6 *Consistent estimation of* $\mathbf{S} = \mathbf{E}[\varepsilon_t^2 \mathbf{x}_t \mathbf{x}_t']$

Assume $E[\left(x_{ik} x_{ij}\right)^2]$ exists and is finite for all k, j $(i = 1, 2, \ldots, k)$ Then as $T \to \infty$

$$\hat{\mathbf{S}} = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^2 \mathbf{x}_t \mathbf{x}_t' \xrightarrow{p} \mathbf{S}$$

where $\hat{\varepsilon}_t = y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}}$.

Sketch of Proof. (Do on White Board)

Remark:

Davidson and MacKinnon (1993) highly recommend using the following degrees of freedom corrected estimate for $\mathbf{S}$

$$\hat{\mathbf{S}} = (T-k)^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^2 \mathbf{x}_t \mathbf{x}_t' \xrightarrow{p} \mathbf{S}$$

They show that the HC standard errors based on this estimator have better finite sample properties than the HC standard errors based on $\hat{\mathbf{S}}$ that doesn't use a degrees of freedom correction.