7 Asymptotic Distribution Theory

Suppose a sample of T observations $(y_1, y_2, \ldots, y_T)$ has been used to construct $\hat{\boldsymbol{\theta}}$, an estimate of the vector of population parameters. For example, the parameter vector $\boldsymbol{\theta} = (c, \phi_1, \phi_2, \ldots, \phi_p, \sigma^2)'$ for an $AR(p)$ process might have been estimated from an *OLS* regression of y_t on lagged y's. We would like to know how far this estimate $\hat{\boldsymbol{\theta}}$ is likely to be from the true value $\boldsymbol{\theta}$ and how to test a hypothesis about the true value based on the observed sample of y's.

Much of the distribution theory used to answer these questions is *asymptotic*: that is, it describes the properties of estimators as the sample size (T) goes to infinity. This chapter develops the basic asymptotic results that will be used in subsequent chapters. The first section summarizes the key tools of asymptotic analysis and presents limit theorems for the sample mean of a sequence of i.i.d. random variables. Section 7.2 develops limit theorems for serially dependent variables with time-varying marginal distributions.

7.1. Review of Asymptotic Distribution Theory

Limits of Deterministic Sequences

Let $\{c_T\}_{T=1}^{\infty}$ denote a sequence of deterministic numbers. The sequence is said to *converge* to c if for any $\varepsilon > 0$, there exists an N such that $|c_T - c| < \varepsilon$ whenever $T \geq N$; in other words, c_T will be as close as desired to c so long as T is sufficiently large. This is indicated as

$$\lim_{T\to\infty} c_T = c, \qquad [7.1.1]$$

or, equivalently,

$$c_T \to c.$$

For example, $c_T = 1/T$ denotes the sequence $\{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$, for which

$$\lim_{T\to\infty} c_T = 0.$$

A sequence of deterministic $(m \times n)$ matrices $\{\mathbf{C}_T\}_{T=1}^{\infty}$ converges to $\mathbf{C}$ if each element of $\mathbf{C}_T$ converges to the corresponding element of $\mathbf{C}$.

Convergence in Probability

Consider a sequence of scalar random variables, $\{X_T\}_{T=1}^{\infty}$. The sequence is said to *converge in probability* to c if for every $\varepsilon > 0$ and every $\delta > 0$ there exists a value N such that, for all $T \geq N$,

$$P\{|X_T - c| > \delta\} < \varepsilon. \qquad [7.1.2]$$

In words, if we go far enough along in the sequence, the probability that X_T differs from c by more than δ can be made arbitrarily small for any δ.

When [7.1.2] is satisfied, the number c is called the *probability limit*, or *plim*, of the sequence $\{X_T\}$. This is indicated as

$$\text{plim}\ X_T = c,$$

or, equivalently,

$$X_T \xrightarrow{p} c.$$

Recall that if $\{c_T\}_{T=1}^{\infty}$ is a deterministic sequence converging to c, then there exists an N such that $|c_T - c| < \delta$ for all $T \geq N$. Then $P\{|c_T - c| > \delta\} = 0$ for all $T \geq N$. Thus, if a deterministic sequence converges to c, then we could also say that $c_T \xrightarrow{p} c$.

A sequence of $(m \times n)$ matrices of random variables $\{\mathbf{X}_T\}$ converges in probability to the $(m \times n)$ matrix $\mathbf{C}$ if each element of $\mathbf{X}_T$ converges in probability to the corresponding element of $\mathbf{C}$.

More generally, if $\{\mathbf{X}_T\}$ and $\{\mathbf{Y}_T\}$ are sequences of $(m \times n)$ matrices, we will use the notation

$$\mathbf{X}_T \xrightarrow{p} \mathbf{Y}_T$$

to indicate that the difference between the two sequences converges in probability to zero:

$$\mathbf{X}_T - \mathbf{Y}_T \xrightarrow{p} \mathbf{0}.$$

An example of a sequence of random variables of interest is the following. Suppose we have a sample of T observations on a random variable $\{Y_1, Y_2, \ldots, Y_T\}$. Consider the sample mean,

$$\bar{Y}_T \equiv (1/T) \sum_{t=1}^{T} Y_t, \qquad [7.1.3]$$

as an estimator of the population mean,

$$\hat{\mu}_T = \bar{Y}_T.$$

We append the subscript T to this estimator to emphasize that it describes the mean of a sample of size T. The primary focus will be on the behavior of this estimator as T grows large. Thus, we will be interested in the properties of the sequence $\{\hat{\mu}_T\}_{T=1}^{\infty}$.

When the plim of a sequence of estimators (such as $\{\hat{\mu}_T\}_{T=1}^{\infty}$) is equal to the true population parameter (in this case, μ), the estimator is said to be *consistent*. If an estimator is consistent, then there exists a sufficiently large sample such that we can be assured with very high probability that the estimate will be within any desired tolerance band around the true value.

The following result is quite helpful in finding plims; a proof of this and some of the other propositions of this chapter are provided in Appendix 7.A at the end of the chapter.

Proposition 7.1: *Let* $\{\mathbf{X}_T\}$ *denote a sequence of* $(n \times 1)$ *random vectors with plim* $\mathbf{c}$, *and let* $\mathbf{g}(\mathbf{c})$ *be a vector-valued function,* $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$, *where* $\mathbf{g}(\cdot)$ *is continuous at* $\mathbf{c}$ *and does not depend on T. Then* $\mathbf{g}(\mathbf{X}_T) \xrightarrow{p} \mathbf{g}(\mathbf{c})$.

The basic idea behind this proposition is that, since $\mathbf{g}(\cdot)$ is continuous, $\mathbf{g}(\mathbf{X}_T)$ will be close to $\mathbf{g}(\mathbf{c})$ provided that $\mathbf{X}_T$ is close to $\mathbf{c}$. By choosing a sufficiently large value of T, the probability that $\mathbf{X}_T$ is close to $\mathbf{c}$ (and thus that $\mathbf{g}(\mathbf{X}_T)$ is close to $\mathbf{g}(\mathbf{c})$) can be brought as near to unity as desired.

Note that $\mathbf{g}(\mathbf{X}_T)$ depends on the *value* of $\mathbf{X}_T$ but cannot depend on the index T itself. Thus, $g(X_T, T) = T \cdot X_T^2$ is not a function covered by Proposition 7.1.

Example 7.1

If $X_{1T} \xrightarrow{p} c_1$ and $X_{2T} \xrightarrow{p} c_2$, then $(X_{1T} + X_{2T}) \xrightarrow{p} (c_1 + c_2)$. This follows immediately, since $g(X_{1T}, X_{2T}) \equiv (X_{1T} + X_{2T})$ is a continuous function of (X_{1T}, X_{2T}).

Example 7.2

Let $\{\mathbf{X}_{1T}\}$ denote a sequence of $(n \times n)$ random matrices with $\mathbf{X}_{1T} \xrightarrow{p} \mathbf{C}_1$, a nonsingular matrix. Let $\mathbf{X}_{2T}$ denote a sequence of $(n \times 1)$ random vectors with $\mathbf{X}_{2T} \xrightarrow{p} \mathbf{c}_2$. Then $[\mathbf{X}_{1T}]^{-1}\mathbf{X}_{2T} \xrightarrow{p} [\mathbf{C}_1]^{-1}\mathbf{c}_2$. To see this, note that the elements of the matrix $[\mathbf{X}_{1T}]^{-1}$ are continuous functions of the elements of $\mathbf{X}_{1T}$ at $\mathbf{X}_{1T} = \mathbf{C}_1$, since $[\mathbf{C}_1]^{-1}$ exists. Thus, $[\mathbf{X}_{1T}]^{-1} \xrightarrow{p} [\mathbf{C}_1]^{-1}$. Similarly, the elements of $[\mathbf{X}_{1T}]^{-1}\mathbf{X}_{2T}$ are sums of products of elements of $[\mathbf{X}_{1T}]^{-1}$ with those of $\mathbf{X}_{2T}$. Since each sum is again a continuous function of $\mathbf{X}_{1T}$ and $\mathbf{X}_{2T}$,

$$\text{plim } [\mathbf{X}_{1T}]^{-1}\mathbf{X}_{2T} = [\text{plim } \mathbf{X}_{1T}]^{-1} \text{ plim } \mathbf{X}_{2T} = [\mathbf{C}_1]^{-1}\mathbf{c}_2.$$

Proposition 7.1 also holds if some of the elements of $\mathbf{X}_T$ are deterministic with conventional limits as in expression [7.1.1]. Specifically, let $\mathbf{X}_T' = (\mathbf{X}_{1T}', \mathbf{c}_{2T}')$, where $\mathbf{X}_{1T}$ is a stochastic $(n_1 \times 1)$ vector and $\mathbf{c}_{2T}$ is a deterministic $(n_2 \times 1)$ vector. If plim $\mathbf{X}_{1T} \doteq \mathbf{c}_1$ and $\lim_{T\to\infty} \mathbf{c}_{2T} = \mathbf{c}_2$, then $\mathbf{g}(\mathbf{X}_{1T}, \mathbf{c}_{2T}) \xrightarrow{p} \mathbf{g}(\mathbf{c}_1, \mathbf{c}_2)$. (See Exercise 7.1.)

Example 7.3

Consider an alternative estimator of the mean given by $\bar{Y}_T^* \equiv [1/(T-1)] \times \Sigma_{t=1}^T Y_t$. This can be written as $c_{1T}\bar{Y}_T$, where $c_{1T} \equiv [T/(T-1)]$ and $\bar{Y}_T \equiv (1/T)\Sigma_{t=1}^T Y_t$. Under general conditions detailed in Section 7.2, the sample mean is a consistent estimator of the population mean, implying that $\bar{Y}_T \xrightarrow{p} \mu$. It is also easy to verify that $c_{1T} \to 1$. Since $c_{1T}\bar{Y}_T$ is a continuous function of c_{1T} and $\bar{Y}_T$, it follows that $c_{1T}\bar{Y}_T \xrightarrow{p} 1 \cdot \mu = \mu$. Thus, $\bar{Y}_T^*$, like $\bar{Y}_T$, is a consistent estimator of μ.

Convergence in Mean Square and Chebyshev's Inequality

A stronger condition than convergence in probability is *mean square convergence*. The random sequence $\{X_T\}$ is said to converge in mean square to c, indicated as

$$X_T \xrightarrow{m.s.} c,$$

if for every $\varepsilon > 0$ there exists a value N such that, for all $T \geq N$,

$$E(X_T - c)^2 < \varepsilon. \qquad [7.1.4]$$

Another useful result is the following.

Proposition 7.2: *(Generalized Chebyshev's inequality). Let X be a random variable with $E(|X|^r)$ finite for some $r > 0$. Then, for any $\delta > 0$ and any value of c,*

$$P\{|X - c| > \delta\} \leq \frac{E|X - c|^r}{\delta^r}. \qquad [7.1.5]$$

An implication of Chebyshev's inequality is that if $X_T \xrightarrow{m.s.} c$, then $X_T \xrightarrow{p} c$. To see this, note that if $X_T \xrightarrow{m.s.} c$, then for any $\varepsilon > 0$ and $\delta > 0$ there exists an N such that $E(X_T - c)^2 < \delta^2\varepsilon$ for all $T \geq N$. This would ensure that

$$\frac{E(X_T - c)^2}{\delta^2} < \varepsilon$$

for all $T \geq N$. From Chebyshev's inequality, this also implies

$$P\{|X_T - c| > \delta\} < \varepsilon$$

for all $T \geq N$, or that $X_T \xrightarrow{p} c$.

Law of Large Numbers for Independent and Identically Distributed Variables

Let us now consider the behavior of the sample mean $\bar{Y}_T = (1/T)\Sigma_{t=1}^{T} Y_t$ where $\{Y_t\}$ is i.i.d. with mean μ and variance σ^2. For this case, $\bar{Y}_T$ has expectation μ and variance

$$E(\bar{Y}_T - \mu)^2 = (1/T^2)\operatorname{Var}\left(\sum_{t=1}^{T} Y_t\right) = (1/T^2)\sum_{t=1}^{T}\operatorname{Var}(Y_t) = \sigma^2/T.$$

Since $\sigma^2/T \to 0$ as $T \to \infty$, this means that $\bar{Y}_T \xrightarrow{m.s.} \mu$, implying also that $\bar{Y}_T \xrightarrow{p} \mu$.

Figure 7.1 graphs an example of the density of the sample mean $f_{\bar{Y}_T}(\bar{y}_T)$ for three different values of T. As T becomes large, the density becomes increasingly concentrated in a spike centered at μ.

The result that the sample mean is a consistent estimate of the population mean is known as the *law of large numbers*.[1] It was proved here for the special case of i.i.d. variables with finite variance. In fact, it turns out also to be true of any sequence of i.i.d. variables with finite mean μ.[2] Section 7.2 explores some of the circumstances under which it also holds for serially dependent variables with time-varying marginal distributions.

Convergence in Distribution

Let $\{X_T\}_{T=1}^{\infty}$ be a sequence of random variables, and let $F_{X_T}(x)$ denote the cumulative distribution function of X_T. Suppose that there exists a cumulative distribution function $F_X(x)$ such that

$$\lim_{T\to\infty} F_{X_T}(x) = F_X(x)$$

[1] This is often described as the weak law of large numbers. An analogous result known as the strong law of large numbers refers to almost sure convergence rather than convergence in probability of the sample mean.

[2] This is known as *Khinchine's theorem*. See, for example, Rao (1973, p. 112).

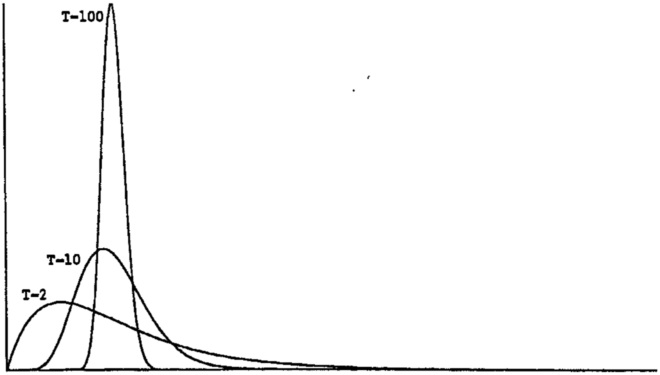

FIGURE 7.1 Density of the sample mean for a sample of size T.

at any value x at which $F_X(\cdot)$ is continuous. Then X_T is said to *converge in distribution* (or in law) to X, denoted

$$X_T \xrightarrow{L} X.$$

When $F_X(x)$ is of a common form, such as the cumulative distribution function for a $N(\mu, \sigma^2)$ variable, we will equivalently write

$$X_T \xrightarrow{L} N(\mu, \sigma^2).$$

The definitions are unchanged if the scalar X_T is replaced with an $(n \times 1)$ vector $\mathbf{X}_T$. A simple way to verify convergence in distribution of a vector is the following.[3] If the scalar $(\lambda_1 X_{1T} + \lambda_2 X_{2T} + \cdots + \lambda_n X_{nT})$ converges in distribution to $(\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_n X_n)$ for any real values of $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, then the vector $\mathbf{X}_T \equiv (X_{1T}, X_{2T}, \ldots, X_{nT})'$ converges in distribution to the vector $\mathbf{X} \equiv (X_1, X_2, \ldots, X_n)'$.

The following results are useful in determining limiting distributions.[4]

Proposition 7.3:

(*a*) *Let* $\{\mathbf{Y}_T\}$ *be a sequence of* $(n \times 1)$ *random vectors with* $\mathbf{Y}_T \xrightarrow{L} \mathbf{Y}$. *Suppose that* $\{\mathbf{X}_T\}$ *is a sequence of* $(n \times 1)$ *random vectors such that* $(\mathbf{X}_T - \mathbf{Y}_T) \xrightarrow{p} \mathbf{0}$. *Then* $\mathbf{X}_T \xrightarrow{L} \mathbf{Y}$; *that is,* $\mathbf{X}_T$ *and* $\mathbf{Y}_T$ *have the same limiting distribution.*

(*b*) *Let* $\{\mathbf{X}_T\}$ *be a sequence of random* $(n \times 1)$ *vectors with* $\mathbf{X}_T \xrightarrow{p} \mathbf{c}$, *and let* $\{\mathbf{Y}_T\}$ *be a sequence of random* $(n \times 1)$ *vectors with* $\mathbf{Y}_T \xrightarrow{L} \mathbf{Y}$. *Then the sequence constructed from the sum* $\{\mathbf{X}_T + \mathbf{Y}_T\}$ *converges in distribution to* $\mathbf{c} + \mathbf{Y}$ *and the sequence constructed from the product* $\{\mathbf{X}_T'\mathbf{Y}_T\}$ *converges in distribution to* $\mathbf{c}'\mathbf{Y}$.

(*c*) *Let* $\{\mathbf{X}_T\}$ *be a sequence of random* $(n \times 1)$ *vectors with* $\mathbf{X}_T \xrightarrow{L} \mathbf{X}$, *and let* $\mathbf{g}(\mathbf{X})$, $\mathbf{g}\colon \mathbb{R}^n \rightarrow \mathbb{R}^m$ *be a continuous function (not dependent on* T*). Then the sequence of random variables* $\{\mathbf{g}(\mathbf{X}_T)\}$ *converges in distribution to* $\mathbf{g}(\mathbf{X})$.

[3] This is known as the Cramér-Wold theorem. See Rao (1973, p. 123).

[4] See Rao (1973, pp. 122–24).

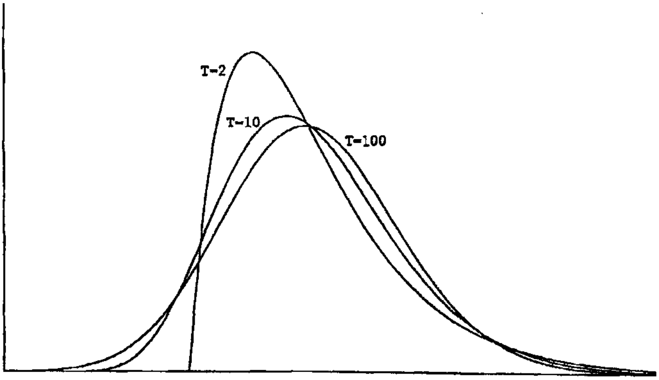

FIGURE 7.2 Density of $\sqrt{T}(\bar{Y}_T - \mu)$.

Example 7.4
Suppose that $X_T \xrightarrow{p} c$ and $Y_T \xrightarrow{L} Y$, where $Y \sim N(\mu, \sigma^2)$. Then, by Proposition 7.3(b), the sequence $X_T Y_T$ has the same limiting probability law as that of c times a $N(\mu, \sigma^2)$ variable. In other words, $X_T Y_T \xrightarrow{L} N(c\mu, c^2\sigma^2)$.

Example 7.5
Generalizing the previous result, let $\{\mathbf{X}_T\}$ be a sequence of random $(m \times n)$ matrices and $\{\mathbf{Y}_T\}$ a sequence of random $(n \times 1)$ vectors with $\mathbf{X}_T \xrightarrow{p} \mathbf{C}$ and $\mathbf{Y}_T \xrightarrow{L} \mathbf{Y}$, with $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega})$. Then the limiting distribution of $\mathbf{X}_T\mathbf{Y}_T$ is the same as that of $\mathbf{CY}$; that is, $\mathbf{X}_T\mathbf{Y}_T \xrightarrow{L} N(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Omega}\mathbf{C}')$.

Example 7.6
Suppose that $X_T \xrightarrow{L} N(0, 1)$. Then Proposition 7.3(c) implies that the square of X_T asymptotically behaves as the square of a $N(0, 1)$ variable: $X_T^2 \xrightarrow{p} \chi^2(1)$.

Central Limit Theorem

We have seen that the sample mean $\bar{Y}_T$ for an i.i.d. sequence has a degenerate probability density as $T \to \infty$, collapsing toward a point mass at μ as the sample size grows. For statistical inference we would like to describe the distribution of $\bar{Y}_T$ in more detail. For this purpose, note that the random variable $\sqrt{T}(\bar{Y}_T - \mu)$ has mean zero and variance given by $(\sqrt{T})^2 \operatorname{Var}(\bar{Y}_T) = \sigma^2$ for all T, and thus, in contrast to $\bar{Y}_T$, the random variable $\sqrt{T}(\bar{Y}_T - \mu)$ might be expected to converge to a nondegenerate random variable as T goes to infinity.

The *central limit theorem* is the result that, as T increases, the sequence $\sqrt{T}(\bar{Y}_T - \mu)$ converges in distribution to a Gaussian random variable. The most familiar, albeit restrictive, version of the central limit theorem establishes that if Y_t is i.i.d. with mean μ and variance σ^2, then[5]

$$\sqrt{T}(\bar{Y}_T - \mu) \xrightarrow{L} N(0, \sigma^2). \qquad [7.1.6]$$

Result [7.1.6] also holds under much more general conditions, some of which are explored in the next section.

Figure 7.2 graphs an example of the density of $\sqrt{T}(\bar{Y}_T - \mu)$ for three different

[5] See, for example, White (1984, pp. 108–9).

values of T. Each of these densities has mean zero and variance σ^2. As T becomes large, the density converges to that of a $N(0, \sigma^2)$ variable.

A final useful result is the following.

Proposition 7.4: *Let $\{\mathbf{X}_t\}$ be a sequence of random $(n \times 1)$ vectors such that $\sqrt{T}(\mathbf{X}_T - \mathbf{c}) \xrightarrow{L} \mathbf{X}$, and let $\mathbf{g}: \mathbb{R}^n \to \mathbb{R}^m$ have continuous first derivatives with $\mathbf{G}$ denoting the $(m \times n)$ matrix of derivatives evaluated at $\mathbf{c}$:*

$$\mathbf{G} \equiv \left.\frac{\partial \mathbf{g}}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}}.$$

Then $\sqrt{T}[\mathbf{g}(\mathbf{X}_T) - \mathbf{g}(\mathbf{c})] \xrightarrow{L} \mathbf{G}\mathbf{X}$.

Example 7.7

Let $\{Y_1, Y_2, \ldots, Y_T\}$ be an i.i.d. sample of size T drawn from a distribution with mean $\mu \neq 0$ and variance σ^2. Consider the distribution of the reciprocal of the sample mean, $S_T = 1/\bar{Y}_T$, where $\bar{Y}_T \equiv (1/T)\Sigma_{t=1}^{T} Y_t$. We know from the central limit theorem that $\sqrt{T}(\bar{Y}_T - \mu) \xrightarrow{L} Y$, where $Y \sim N(0, \sigma^2)$. Also, $g(y) = 1/y$ is continuous at $y = \mu$. Let $G \equiv (\partial g/\partial y)|_{y=\mu} = (-1/\mu^2)$. Then $\sqrt{T}[S_T - (1/\mu)] \xrightarrow{L} G \cdot Y$; in other words, $\sqrt{T}[S_T - (1/\mu)] \xrightarrow{L} N(0, \sigma^2/\mu^4)$.

7.2. Limit Theorems for Serially Dependent Observations

The previous section stated the law of large numbers and central limit theorem for independent and identically distributed random variables with finite second moments. This section develops analogous results for heterogeneously distributed variables with various forms of serial dependence. We first develop a law of large numbers for a general covariance-stationary process.

Law of Large Numbers for a Covariance-Stationary Process

Let $(Y_1, Y_2, \ldots, Y_T)$ represent a sample of size T from a covariance-stationary process with

$$E(Y_t) = \mu \quad \text{for all } t \qquad [7.2.1]$$

$$E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_j \quad \text{for all } t \qquad [7.2.2]$$

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty. \qquad [7.2.3]$$

Consider the properties of the sample mean,

$$\bar{Y}_T = (1/T) \sum_{t=1}^{T} Y_t. \qquad [7.2.4]$$

Taking expectations of [7.2.4] reveals that the sample mean provides an unbiased estimate of the population mean,

$$E(\bar{Y}_T) = \mu,$$

while the variance of the sample mean is

$$
\begin{aligned}
&E(\overline{Y}_T - \mu)^2 \\
&\quad = E\left[(1/T)\sum_{t=1}^{T}(Y_t - \mu)\right]^2 \\
&\quad = (1/T^2)E\{[(Y_1 - \mu) + (Y_2 - \mu) + \cdots + (Y_T - \mu)] \\
&\qquad \times [(Y_1 - \mu) + (Y_2 - \mu) + \cdots + (Y_T - \mu)]\} \\
&\quad = (1/T^2)E\{(Y_1 - \mu)[(Y_1 - \mu) + (Y_2 - \mu) + \cdots + (Y_T - \mu)] \\
&\qquad + (Y_2 - \mu)[(Y_1 - \mu) + (Y_2 - \mu) + \cdots + (Y_T - \mu)] \\
&\qquad + (Y_3 - \mu)[(Y_1 - \mu) + (Y_2 - \mu) + \cdots + (Y_T - \mu)] \\
&\qquad + \cdots + (Y_T - \mu)[(Y_1 - \mu) + (Y_2 - \mu) + \cdots + (Y_T - \mu)]\} \\
&\quad = (1/T^2)\ \{[\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3 + \cdots + \gamma_{T-1}] \\
&\qquad + [\gamma_1 + \gamma_0 + \gamma_1 + \gamma_2 + \cdots + \gamma_{T-2}] \\
&\qquad + [\gamma_2 + \gamma_1 + \gamma_0 + \gamma_1 + \cdots + \gamma_{T-3}] \\
&\qquad + \cdots + [\gamma_{T-1} + \gamma_{T-2} + \gamma_{T-3} + \cdots + \gamma_0]\}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E(\overline{Y}_T - \mu)^2 = (1/T^2)\{&T\gamma_0 + 2(T-1)\gamma_1 \\
&+ 2(T-2)\gamma_2 + 2(T-3)\gamma_3 + \cdots + 2\gamma_{T-1}\}
\end{aligned}
$$

or

$$
\begin{aligned}
E(\overline{Y}_T - \mu)^2 = (1/T)\{&\gamma_0 + [(T-1)/T](2\gamma_1) + [(T-2)/T](2\gamma_2) \\
&+ [(T-3)/T](2\gamma_3) + \cdots + [1/T](2\gamma_{T-1})\}.
\end{aligned}
\qquad [7.2.5]
$$

It is easy to see that this expression goes to zero as the sample size grows—that is, that $\overline{Y}_T \xrightarrow{m.s.} \mu$:

$$
\begin{aligned}
T \cdot E(\overline{Y}_T - \mu)^2 &= |\gamma_0 + [(T-1)/T](2\gamma_1) + [(T-2)/T](2\gamma_2) \\
&\qquad + [(T-3)/T](2\gamma_3) + \cdots + [1/T](2\gamma_{T-1})| \\
&\le \{|\gamma_0| + [(T-1)/T]\cdot 2|\gamma_1| + [(T-2)/T]\cdot 2|\gamma_2| \\
&\qquad + [(T-3)/T]\cdot 2|\gamma_3| + \cdots + [1/T]\cdot 2|\gamma_{T-1}|\} \\
&\le \{|\gamma_0| + 2|\gamma_1| + 2|\gamma_2| + 2|\gamma_3| + \cdots\}.
\end{aligned}
\qquad [7.2.6]
$$

Hence, $T \cdot E(\overline{Y}_T - \mu)^2 < \infty$, by [7.2.3], and so $E(\overline{Y}_T - \mu)^2 \to 0$, as claimed.

It is also of interest to calculate the limiting value of $T \cdot E(\overline{Y}_T - \mu)^2$. Result [7.2.5] expresses this variance for finite T as a weighted average of the first $T - 1$ autocovariances γ_j. For large j, these autocovariances approach zero and will not affect the sum. For small j, the autocovariances are given a weight that approaches unity as the sample size grows. Thus, we might guess that

$$
\lim_{T\to\infty} T \cdot E(\overline{Y}_T - \mu)^2 = \sum_{j=-\infty}^{\infty} \gamma_j = \gamma_0 + 2\gamma_1 + 2\gamma_2 + 2\gamma_3 + \cdots. \qquad [7.2.7]
$$

This conjecture is indeed correct. To verify this, note that the assumption [7.2.3] means that for any $\varepsilon > 0$ there exists a q such that

$$
2|\gamma_{q+1}| + 2|\gamma_{q+2}| + 2|\gamma_{q+3}| + \cdots < \varepsilon/2.
$$

Now

$$\begin{aligned}\left|\sum_{j=-\infty}^{\infty}\gamma_j - T\cdot E(\overline{Y}_T-\mu)^2\right| &\\ &= |\{\gamma_0 + 2\gamma_1 + 2\gamma_2 + 2\gamma_3 + \cdots\}\\ &\quad - \{\gamma_0 + [(T-1)/T]\cdot 2\gamma_1 + [(T-2)/T]\cdot 2\gamma_2\\ &\quad + [(T-3)/T]\cdot 2\gamma_3 + \cdots + [1/T]\cdot 2\gamma_{T-1}\}|\\ &\le (1/T)\cdot 2|\gamma_1| + (2/T)\cdot 2|\gamma_2| + (3/T)\cdot 2|\gamma_3| + \cdots\\ &\quad + (q/T)\cdot 2|\gamma_q| + 2|\gamma_{q+1}| + 2|\gamma_{q+2}| + 2|\gamma_{q+3}| + \cdots\\ &\le (1/T)\cdot 2|\gamma_1| + (2/T)\cdot 2|\gamma_2| + (3/T)\cdot 2|\gamma_3| + \cdots\\ &\quad + (q/T)\cdot 2|\gamma_q| + \varepsilon/2.\end{aligned}$$

Moreover, for this given q, we can find an N such that

$$(1/T)\cdot 2|\gamma_1| + (2/T)\cdot 2|\gamma_2| + (3/T)\cdot 2|\gamma_3| + \cdots + (q/T)\cdot 2|\gamma_q| < \varepsilon/2$$

for all $T \ge N$, ensuring that

$$\left|\sum_{j=-\infty}^{\infty}\gamma_j - T\cdot E(\overline{Y}_T-\mu)^2\right| < \varepsilon,$$

as was to be shown.

These results can be summarized as follows.

Proposition 7.5: *Let Y_t be a covariance-stationary process with moments given by [7.2.1] and [7.2.2] and with absolutely summable autocovariances as in [7.2.3]. Then the sample mean [7.2.4] satisfies*

(*a*) $\overline{Y}_T \xrightarrow{m.s.} \mu$

(*b*) $\lim_{T\to\infty}\{T\cdot E(\overline{Y}_T-\mu)^2\} = \sum_{j=-\infty}^{\infty}\gamma_j.$

Recall from Chapter 3 that condition [7.2.3] is satisfied for any covariance-stationary $ARMA(p, q)$ process,

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)Y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t,$$

with roots of $(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) = 0$ outside the unit circle.

Alternative expressions for the variance in result (b) of Proposition 7.5 are sometimes used. Recall that the autocovariance-generating function for Y_t is defined as

$$g_Y(z) = \sum_{j=-\infty}^{\infty}\gamma_j z^j,$$

while the spectrum is given by

$$s_Y(\omega) = \frac{1}{2\pi}g_Y(e^{-i\omega}).$$

Thus, result (b) could equivalently be described as the autocovariance-generating function evaluated at $z = 1$,

$$\sum_{j=-\infty}^{\infty} \gamma_j = g_Y(1),$$

or as 2π times the spectrum at frequency $\omega = 0$,

$$\sum_{j=-\infty}^{\infty} \gamma_j = 2\pi s_Y(0),$$

the last result coming from the fact that $e^0 = 1$. For example, consider the $MA(\infty)$ process

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \equiv \mu + \psi(L)\varepsilon_t$$

with $E(\varepsilon_t \varepsilon_\tau) = \sigma^2$ if $t = \tau$ and 0 otherwise and with $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$. Recall that its autocovariance-generating fucntion is given by

$$g_Y(z) = \psi(z)\sigma^2\psi(z^{-1}).$$

Evaluating this at $z = 1$,

$$\sum_{j=-\infty}^{\infty} \gamma_j = \psi(1)\sigma^2\psi(1) = \sigma^2[1 + \psi_1 + \psi_2 + \psi_3 + \cdots]^2. \qquad [7.2.8]$$

Martingale Difference Sequence

Some very useful limit theorems pertain to *martingale difference sequences*. Let $\{Y_t\}_{t=1}^{\infty}$ denote a sequence of random scalars with $E(Y_t) = 0$ for all t.[6] Let Ω_t denote information available at date t, where this information includes current and lagged values of Y.[7] For example, we might have

$$\Omega_t = \{Y_t, Y_{t-1}, \ldots, Y_1, X_t, X_{t-1}, \ldots, X_1\},$$

where X_t is a second random variable. If

$$E(Y_t|\Omega_{t-1}) = 0 \qquad \text{for } t = 2, 3, \ldots, \qquad [7.2.9]$$

then $\{Y_t\}$ is said to be a *martingale difference sequence* with respect to $\{\Omega_t\}$.

Where no information set is specifies, Ω_t is presumed to consist solely of current and lagged values of Y:

$$\Omega_t = \{Y_t, Y_{t-1}, \ldots, Y_1\}.$$

Thus, if a sequence of scalars $\{Y_t\}_{t=1}^{\infty}$ satisfied $E(Y_t) = 0$ for all t and

$$E(Y_t|Y_{t-1}, Y_{t-2}, \ldots, Y_1) = 0, \qquad [7.2.10]$$

for $t = 2, 3, \ldots$, then we will say simply that $\{Y_t\}$ is a martingale difference sequence. Note that [7.2.10] is implied by [7.2.9] by the law of iterated expectations.

A sequence of $(n \times 1)$ vectors $\{\mathbf{Y}_t\}_{t=1}^{\infty}$ satisfying $E(\mathbf{Y}_t) = \mathbf{0}$ and $E(\mathbf{Y}_t|\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots, \mathbf{Y}_1) = \mathbf{0}$ is said to form a *vector martingale difference sequence*.

[6] Wherever an expectation is indicated, it is taken as implicit that the integral exists, that is, that $E|Y_t|$ is finite.

[7] More formally, $\{\Omega_t\}_{t=1}^{\infty}$ denotes an increasing sequence of σ-fields ($\Omega_{t-1} \subset \Omega_t$) with Y_t measurable with respect to Ω_t. See, for example, White (1984, p. 56).

Note that condition [7.2.10] is stronger than the condition that Y_t is serially uncorrelated. A serially uncorrelated sequence cannot be forecast on the basis of a linear function of its past values. No function of past values, linear or nonlinear, can forecast a martingale difference sequence. While stronger than absence of serial correlation, the martingale difference condition is weaker than independence, since it does not rule out the possibility that higher moments such as $E(Y_t^2|Y_{t-1}, Y_{t-2}, \ldots, Y_1)$ might depend on past Y's.

Example 7.8
If $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$, then $Y_t = \varepsilon_t\varepsilon_{t-1}$ is a martingale difference sequence but not serially independent.

L^1-*Mixingales*

A more general class of processes known as L^1-*mixingales* was introduced by Andrews (1988). Consider a sequence of random variables $\{Y_t\}_{t=1}^{\infty}$ with $E(Y_t) = 0$ for $t = 1, 2, \ldots$. Let Ω_t denote information available at time t, as before, where Ω_t includes current and lagged values of Y. Suppose that we can find sequences of nonnegative deterministic constants $\{c_t\}_{t=1}^{\infty}$ and $\{\xi_m\}_{m=0}^{\infty}$ such that $\lim_{m\to\infty} \xi_m = 0$ and

$$E\left|E(Y_t|\Omega_{t-m})\right| \le c_t\xi_m \qquad [7.2.11]$$

for all $t \ge 1$ and all $m \ge 0$. Then $\{Y_t\}$ is said to follow an L^1-mixingale with respect to $\{\Omega_t\}$.

Thus, a zero-mean process for which the m-period-ahead forecast $E(Y_t|\Omega_{t-m})$ converges (in absolute expected value) to the unconditional mean of zero is described as an L^1-mixingale.

Example 7.9
Let $\{Y_t\}$ be a martingale difference sequence. Let $c_t = E|Y_t|$, and choose $\xi_0 = 1$ and $\xi_m = 0$ for $m = 1, 2, \ldots$. Then [7.2.11] is satisfied for $\Omega_t = \{Y_t, Y_{t-1}, \ldots, Y_1\}$, so that $\{Y_t\}$ could be described as an L^1-mixingale sequence.

Example 7.10
Let $Y_t = \Sigma_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$, where $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$ and $\{\varepsilon_t\}$ is a martingale difference sequence with $E|\varepsilon_t| < M$ for all t for some $M < \infty$. Then $\{Y_t\}$ is an L^1-mixingale with respect to $\Omega_t = \{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$. To see this, notice that

$$E\left|E(Y_t|\varepsilon_{t-m}, \varepsilon_{t-m-1}, \ldots)\right| = E\left|\sum_{j=m}^{\infty}\psi_j\varepsilon_{t-j}\right| \le E\left\{\sum_{j=m}^{\infty}|\psi_j\varepsilon_{t-j}|\right\}.$$

Since $\{\psi_j\}_{j=0}^{\infty}$ is absolutely summable and $E|\varepsilon_{t-j}| < M$, we can interchange the order of expectation and summation:

$$E\left\{\sum_{j=m}^{\infty}|\psi_j\varepsilon_{t-j}|\right\} = \sum_{j=m}^{\infty}|\psi_j|\cdot E|\varepsilon_{t-j}| \le \sum_{j=m}^{\infty}|\psi_j|\cdot M.$$

Then [7.2.11] is satisfied with $c_t = M$ and $\xi_m = \Sigma_{j=m}^{\infty}|\psi_j|$. Moreover, $\lim_{m\to\infty} \xi_m = 0$, because of absolute summability of $\{\psi_j\}_{j=0}^{\infty}$. Hence, $\{Y_t\}$ is an L^1-mixingale.

Law of Large Numbers For L¹*-Mixingales*

Andrews (1988) derived the following law of large numbers for L^1-mixingales.[8]

Proposition 7.6: *Let $\{Y_t\}$ be an L^1-mixingale. If (a) $\{Y_t\}$ is uniformly integrable and (b) there exists a choice for $\{c_t\}$ such that*

$$\lim_{T\to\infty} (1/T)\sum_{t=1}^{T} c_t < \infty,$$

then $(1/T)\Sigma_{t=1}^{T} Y_t \xrightarrow{p} 0$.

To apply this result, we need to verify that a sequence is uniformly integrable. A sequence $\{Y_t\}$ is said to be *uniformly integrable* if for every $\varepsilon > 0$ there exists a number $c > 0$ such that

$$E(|Y_t|\cdot\delta_{[|Y_t|\geq c]}) < \varepsilon \qquad [7.2.12]$$

for all t, where $\delta_{[|Y_t|\geq c]} = 1$ if $|Y_t| \geq c$ and 0 otherwise. The following proposition gives sufficient conditions for uniform integrability.

Proposition 7.7: *(a) Suppose there exist an $r > 1$ and an $M' < \infty$ such that $E(|Y_t|^r) < M'$ for all t. Then $\{Y_t\}$ is uniformly integrable. (b) Suppose there exist an $r > 1$ and an $M' < \infty$ such that $E(|X_t|^r) < M'$ for all t. If $Y_t = \Sigma_{j=-\infty}^{\infty} h_j X_{t-j}$ with $\Sigma_{j=-\infty}^{\infty} |h_j| < \infty$, then $\{Y_t\}$ is uniformly integrable.*

Condition (a) requires us to find a moment higher than the first that exists. Typically, we would use $r = 2$. However, even if a variable has infinite variance, it can still be uniformly integrable as long as $E|Y_t|^r$ exists for some r between 1 and 2.

Example 7.11

Let $\overline{Y}_T$ be the sample mean from a martingale difference sequence, $\overline{Y}_T = (1/T)\Sigma_{t=1}^{T} Y_t$ with $E|Y_t|^r < M'$ for some $r > 1$ and $M' < \infty$. Note that this also implies that there exists an $M < \infty$ such that $E|Y_t| < M$. From Proposition 7.7(a), $\{Y_t\}$ is uniformly integrable. Moreover, from Example 7.9, $\{Y_t\}$ can be viewed as an L^1-mixingale with $c_t = M$. Thus, $\lim_{T\to\infty} (1/T)\Sigma_{t=1}^{T} c_t = M < \infty$, and so, from Proposition 7.6, $\overline{Y}_T \xrightarrow{p} 0$.

Example 7.12

Let $Y_t = \Sigma_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, where $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$ and $\{\varepsilon_t\}$ is a martingale difference sequence with $E|\varepsilon_t|^r < M' < \infty$ for some $r > 1$ and some $M' < \infty$. Then, from Proposition 7.7(b), $\{Y_t\}$ is uniformly integrable. Moreover, from Example 7.10, $\{Y_t\}$ is an L^1-mixingale with $c_t = M$, where M represents the largest value of $E|\varepsilon_t|$ for any t. Then $\lim_{T\to\infty} (1/T)\Sigma_{t=1}^{T} c_t = M < \infty$, establishing again that $\overline{Y}_T \xrightarrow{p} 0$.

Proposition 7.6 can also be applied to a double-indexed array $\{Y_{t,T}\}$; that is, each sample size T can be associated with a different sequence $\{Y_{1,T}, Y_{2,T}, \ldots, Y_{T,T}\}$. The array is said to be an L^1-mixingale with respect to an information set $\Omega_{t,T}$ that includes $\{Y_{1,T}, Y_{2,T}, \ldots, Y_{T,T}\}$ if there exist nonnegative constants ξ_m and $c_{t,T}$ such that $\lim_{m\to\infty} \xi_m = 0$ and

$$E|E(Y_{t,T}|\Omega_{t-m,T})| \leq c_{t,T}\xi_m$$

[8] Andrews replaced part (b) of the proposition with the weaker condition $\overline{\lim}_{T\to\infty} (1/T)\,\Sigma_{t=1}^{T} c_t < \infty$. See Royden (1968, p. 36) on the relation between "lim" and "$\overline{\text{lim}}$."

for all $m \geq 0$, $T \geq 1$, and $t = 1, 2, \ldots, T$. If the array is uniformly integrable with $\lim_{T\to\infty} (1/T)\Sigma_{t=1}^{T} c_{t,T} < \infty$, then $(1/T)\Sigma_{t=1}^{T} Y_{t,T} \xrightarrow{p} 0$.

Example 7.13

Let $\{\varepsilon_t\}_{t=1}^{\infty}$ be a martingale difference sequence with $E|\varepsilon_t|^r < M'$ for some $r > 1$ and $M' < \infty$, and define $Y_{t,T} \equiv (t/T)\varepsilon_t$. Then the array $\{Y_{t,T}\}$ is a uniformly integrable L^1-mixingale with $c_{t,T} = M$, where M denotes the maximal value for $E|\varepsilon_t|$, $\xi_0 = 1$, and $\xi_m = 0$ for $m > 0$. Hence, $(1/T)\Sigma_{t=1}^{T}(t/T)\varepsilon_t \xrightarrow{p} 0$.

Consistent Estimation of Second Moments

Next consider the conditions under which

$$(1/T)\sum_{t=1}^{T} Y_t Y_{t-k} \xrightarrow{p} E(Y_t Y_{t-k})$$

(for notational simplicity, we assume here that the sample consists of $T + k$ observations on Y). Suppose that $Y_t = \Sigma_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$, where $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$ and $\{\varepsilon_t\}$ is an i.i.d. sequence with $E|\varepsilon_t|^r < \infty$ for some $r > 2$. Note that the population second moment can be written[9]

$$\begin{aligned} E(Y_t Y_{t-k}) &= E\left(\sum_{u=0}^{\infty}\psi_u\varepsilon_{t-u}\right)\left(\sum_{v=0}^{\infty}\psi_v\varepsilon_{t-k-v}\right) \\ &= E\left(\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\psi_u\psi_v\varepsilon_{t-u}\varepsilon_{t-k-v}\right) \qquad [7.2.13] \\ &= \sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\psi_u\psi_v E(\varepsilon_{t-u}\varepsilon_{t-k-v}). \end{aligned}$$

Define $X_{t,k}$ to be the following random variable:

$$\begin{aligned} X_{t,k} &\equiv Y_t Y_{t-k} - E(Y_t Y_{t-k}) \\ &= \left(\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\psi_u\psi_v\varepsilon_{t-u}\varepsilon_{t-k-v}\right) - \left(\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\psi_u\psi_v E(\varepsilon_{t-u}\varepsilon_{t-k-v})\right) \\ &= \sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\psi_u\psi_v[\varepsilon_{t-u}\varepsilon_{t-k-v} - E(\varepsilon_{t-u}\varepsilon_{t-k-v})]. \end{aligned}$$

Consider a forecast of $X_{t,k}$ on the basis of $\Omega_{t-m} \equiv \{\varepsilon_{t-m}, \varepsilon_{t-m-1}, \ldots\}$ for $m > k$:

$$E(X_{t,k}|\Omega_{t-m}) = \sum_{u=m}^{\infty}\sum_{v=m-k}^{\infty}\psi_u\psi_v[\varepsilon_{t-u}\varepsilon_{t-k-v} - E(\varepsilon_{t-u}\varepsilon_{t-k-v})].$$

[9] Notice that

$$\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}|\psi_u\psi_v| = \sum_{u=0}^{\infty}|\psi_u|\sum_{v=0}^{\infty}|\psi_v| < \infty$$

and $E|\varepsilon_{t-u}\varepsilon_{t-k-v}| < \infty$, permitting us to move the expectation operator inside the summation signs in the last line of [7.2.13].

The expected absolute value of this forecast is bounded by

$$\begin{aligned} E\left|E(X_{t,k}|\Omega_{t-m})\right| &= E\left|\sum_{u=m}^{\infty}\sum_{v=m-k}^{\infty}\psi_u\psi_v[\varepsilon_{t-u}\varepsilon_{t-k-v} - E(\varepsilon_{t-u}\varepsilon_{t-k-v})]\right| \\ &\leq E\left(\sum_{u=m}^{\infty}\sum_{v=m-k}^{\infty}|\psi_u\psi_v|\cdot|\varepsilon_{t-u}\varepsilon_{t-k-v} - E(\varepsilon_{t-u}\varepsilon_{t-k-v})|\right) \\ &\leq \sum_{u=m}^{\infty}\sum_{v=m-k}^{\infty}|\psi_u\psi_v|\cdot M \end{aligned}$$

for some $M < \infty$. Define

$$\xi_m \equiv \sum_{u=m}^{\infty}\sum_{v=m-k}^{\infty}|\psi_u\psi_v| = \sum_{u=m}^{\infty}|\psi_u|\sum_{v=m-k}^{\infty}|\psi_v|.$$

Since $\{\psi_j\}_{j=0}^{\infty}$ is absolutely summable, $\lim_{m\to\infty}\Sigma_{u=m}^{\infty}|\psi_u| = 0$ and $\lim_{m\to\infty}\xi_m = 0$. It follows that $X_{t,k}$ is an L^1-mixingale with respect to Ω_t with coefficient $c_t = M$. Moreover, $X_{t,k}$ is uniformly integrable, from a simple adaptation of the argument in Proposition 7.7(b) (see Exercise 7.5). Hence,

$$(1/T)\sum_{t=1}^{T}X_{t,k} = (1/T)\sum_{t=1}^{T}[Y_tY_{t-k} - E(Y_tY_{t-k})] \xrightarrow{p} 0,$$

from which

$$(1/T)\sum_{t=1}^{T}Y_tY_{t-k} \xrightarrow{p} E(Y_tY_{t-k}). \qquad [7.2.14]$$

It is straightforward to deduce from [7.2.14] that the jth sample autocovariance for a sample of size T gives a consistent estimate of the population autocovariance,

$$(1/T)\sum_{t=k+1}^{T}(Y_t - \overline{Y}_T)(Y_{t-k} - \overline{Y}_T) \xrightarrow{p} E(Y_t - \mu)(Y_{t-k} - \mu), \qquad [7.2.15]$$

where $\overline{Y}_T = (1/T)\Sigma_{t=1}^{T}Y_t$; see Exercise 7.6.

Central Limit Theorem for a Martingale Difference Sequence

Next we consider the asymptotic distribution of $\sqrt{T}$ times the sample mean. The following version of the central limit theorem can often be applied.

Proposition 7.8: *(White, 1984, Corollary 5.25, p. 130). Let $\{Y_t\}_{t=1}^{\infty}$ be a scalar martingale difference sequence with $\overline{Y}_T = (1/T)\Sigma_{t=1}^{T}Y_t$. Suppose that (a) $E(Y_t^2) = \sigma_t^2 > 0$ with $(1/T)\Sigma_{t=1}^{T}\sigma_t^2 \to \sigma^2 > 0$, (b) $E|Y_t|^r < \infty$ for some $r > 2$ and all t, and (c) $(1/T)\Sigma_{t=1}^{T}Y_t^2 \xrightarrow{p} \sigma^2$. Then $\sqrt{T}\,\overline{Y}_T \xrightarrow{L} N(0, \sigma^2)$.*

Again, Proposition 7.8 can be extended to arrays $\{Y_{t,T}\}$ as follows. Let $\{Y_{t,T}\}_{t=1}^{T}$ be a martingale difference sequence with $E(Y_{t,T}^2) = \sigma_{t,T}^2 > 0$. Let $\{Y_{t,T+1}\}_{t=1}^{T+1}$ be a potentially different martingale difference sequence with $E(Y_{t,T+1}^2) = \sigma_{t,T+1}^2 > 0$. If (a) $(1/T)\Sigma_{t=1}^{T}\sigma_{t,T}^2 \to \sigma^2$, (b) $E|Y_{t,T}|^r < \infty$ for some $r > 2$ and all t and T, and (c) $(1/T)\Sigma_{t=1}^{T}Y_{t,T}^2 \xrightarrow{p} \sigma^2$, then $\sqrt{T}\,\overline{Y}_T \xrightarrow{L} N(0, \sigma^2)$.

Proposition 7.8 also readily generalizes to vector martingale difference sequences.

Proposition 7.9: *Let $\{\mathbf{Y}_t\}_{t=1}^{\infty}$ be an n-dimensional vector martingale difference sequence with $\bar{\mathbf{Y}}_T = (1/T)\Sigma_{t=1}^{T}\mathbf{Y}_t$. Suppose that (a) $E(\mathbf{Y}_t\mathbf{Y}_t') = \mathbf{\Omega}_t$, a positive definite matrix with $(1/T)\Sigma_{t=1}^{T}\mathbf{\Omega}_t \to \mathbf{\Omega}$, a positive definite matrix; (b) $E(Y_{it}Y_{jt}Y_{lt}Y_{mt}) < \infty$ for all t and all i, j, l, and m (including $i = j = l = m$), where Y_{it} is the ith element of the vector $\mathbf{Y}_t$; and (c) $(1/T)\Sigma_{t=1}^{T}\mathbf{Y}_t\mathbf{Y}_t' \xrightarrow{p} \mathbf{\Omega}$. Then $\sqrt{T}\,\bar{\mathbf{Y}}_T \xrightarrow{L} N(\mathbf{0}, \mathbf{\Omega})$.*

Again, Proposition 7.9 holds for arrays $\{\mathbf{Y}_{t,T}\}_{t=1}^{T}$ satisfying the stated conditions.

To apply Proposition 7.9, we will often need to assume that a certain process has finite fourth moments. The following result can be useful for this purpose.

Proposition 7.10: *Let X_t be a strictly stationary stochastic process with $E(X_t^4) = \mu_4 < \infty$. Let $Y_t = \Sigma_{j=0}^{\infty}h_jX_{t-j}$, where $\Sigma_{j=0}^{\infty}|h_j| < \infty$. Then Y_t is a strictly stationary stochastic process with $E|Y_tY_sY_uY_v| < \infty$ for all t, s, u, and v.*

Example 7.14

Let $Y_t = \phi_1Y_{t-1} + \phi_2Y_{t-2} + \cdots + \phi_pY_{t-p} + \varepsilon_t$, where $\{\varepsilon_t\}$ is an i.i.d. sequence and where roots of $(1 - \phi_1z - \phi_2z^2 - \cdots - \phi_pz^p) = 0$ lie outside the unit circle. We saw in Chapter 3 that Y_t can be written as $\Sigma_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$ with $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$. Proposition 7.10 states that if ε_t has finite fourth moments, then so does Y_t.

Example 7.15

Let $Y_t = \Sigma_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$ with $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$ and ε_t i.i.d. with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$, and $E(\varepsilon_t^4) < \infty$. Consider the random variable X_t defined by $X_t \equiv \varepsilon_tY_{t-k}$ for $k > 0$. Then X_t is a martingale difference sequence with variance $E(X_t^2) = \sigma^2\cdot E(Y_t^2)$ and with fourth moment $E(\varepsilon_t^4)\cdot E(Y_t^4) < \infty$, by Example 7.14. Hence, if we can show that

$$(1/T)\sum_{t=1}^{T} X_t^2 \xrightarrow{p} E(X_t^2), \qquad [7.2.16]$$

then Proposition 7.8 can be applied to deduce that

$$(1/\sqrt{T})\sum_{t=1}^{T} X_t \xrightarrow{L} N\Big(0,\ E(X_t^2)\Big)$$

or

$$(1/\sqrt{T})\sum_{t=1}^{T} \varepsilon_tY_{t-k} \xrightarrow{L} N\Big(0,\ \sigma^2\cdot E(Y_t^2)\Big). \qquad [7.2.17]$$

To verify [7.2.16], notice that

$$\begin{aligned}(1/T)\sum_{t=1}^{T} X_t^2 &= (1/T)\sum_{t=1}^{T} \varepsilon_t^2Y_{t-k}^2 \\ &= (1/T)\sum_{t=1}^{T} (\varepsilon_t^2 - \sigma^2)Y_{t-k}^2 + (1/T)\sum_{t=1}^{T} \sigma^2Y_{t-k}^2.\end{aligned} \qquad [7.2.18]$$

But $(\varepsilon_t^2 - \sigma^2)Y_{t-k}^2$ is a martingale difference sequence with finite second moment, so, from Example 7.11,

$$(1/T)\sum_{t=1}^{T} (\varepsilon_t^2 - \sigma^2)Y_{t-k}^2 \xrightarrow{p} 0.$$

It further follows from result [7.2.14] that

$$(1/T) \sum_{t=1}^{T} \sigma^2 Y_{t-k}^2 \xrightarrow{p} \sigma^2 \cdot E(Y_t^2).$$

Thus, [7.2.18] implies

$$(1/T) \sum_{t=1}^{T} X_t^2 \xrightarrow{p} \sigma^2 \cdot E(Y_t^2),$$

as claimed in [7.2.16].

Central Limit Theorem for Stationary Stochastic Processes

We now present a central limit theorem for a serially correlated sequence. Recall from Proposition 7.5 that the sample mean has asymptotic variance given by $(1/T)\Sigma_{j=-\infty}^{\infty}\gamma_j$. Thus, we would expect the central limit theorem to take the form $\sqrt{T}(\bar{Y}_T - \mu) \xrightarrow{L} N(0, \Sigma_{j=-\infty}^{\infty}\gamma_j)$. The next proposition gives a result of this type.

Proposition 7.11: *(Anderson, 1971, p. 429). Let*

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

where $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with $E(\varepsilon_t^2) < \infty$ and $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$. Then

$$\sqrt{T}(\bar{Y}_T - \mu) \xrightarrow{L} N(0, \sum_{j=-\infty}^{\infty} \gamma_j). \qquad [7.2.19]$$

A version of [7.2.19] can also be developed for $\{\varepsilon_t\}$ a martingale difference sequence satisfying certain restrictions; see Phillips and Solo (1992).

APPENDIX 7.A. *Proofs of Chapter 7 Propositions*

■ **Proof of Proposition 7.1.** Let $g_j(\mathbf{c})$ denote the jth element of $\mathbf{g}(\mathbf{c})$, g_j: $\mathbb{R}^n \to \mathbb{R}^1$. We need to show that for any $\delta > 0$ and $\varepsilon > 0$ there exists an N such that for all $T \geq N$,

$$P\{|g_j(\mathbf{X}_T) - g_j(\mathbf{c})| > \delta\} < \varepsilon. \qquad [7.\text{A}.1]$$

Continuity of $g_j(\cdot)$ implies that there exists an η such that $|g_j(\mathbf{X}_T) - g_j(\mathbf{c})| > \delta$ only if

$$[(X_{1T} - c_1)^2 + (X_{2T} - c_2)^2 + \cdots + (X_{nT} - c_n)^2] > \eta^2. \qquad [7.\text{A}.2]$$

This would be the case only if $(X_{iT} - c_i)^2 > \eta^2/n$ for some i. But from the fact that plim $X_{iT} = c_i$, for any i and specified values of ε and η we can find a value of N such that

$$P\{|X_{iT} - c_i| > \eta/\sqrt{n}\} < \varepsilon/n$$

for all $T \geq N$.

Recall the elementary addition rule for the probability of any events A and B,

$$P\{A \text{ or } B\} \leq P\{A\} + P\{B\},$$

from which it follows that

$$P\{(|X_{1T} - c_1| > \eta/\sqrt{n}) \text{ or } (|X_{2T} - c_2| > \eta/\sqrt{n}) \text{ or } \cdots \text{ or } (|X_{nT} - c_n| > \eta/\sqrt{n})\}$$
$$< (\varepsilon/n) + (\varepsilon/n) + \cdots + (\varepsilon/n).$$

Hence,

$$P\{[(X_{1T} - c_1)^2 + (X_{2T} - c_2)^2 + \cdots + (X_{nT} - c_n)^2] > \eta^2\} < \varepsilon$$

for all $T \geq N$. Since [7.A.2] was a necessary condition for $|g_j(\mathbf{X}_T) - g_j(\mathbf{c})|$ to be greater than δ, it follows that the probability that $|g_j(\mathbf{X}_T) - g_j(\mathbf{c})|$ is greater than δ is less than ε, which was to be shown. ■

■ **Proof of Proposition 7.2.** Let S denote the set of all x such that $|x - c| > \delta$, and let $\bar{S}$ denote its complement (all x such that $|x - c| \leq \delta$). Then, for $f_X(x)$ the density of x,

$$\begin{aligned} E|X - c|^r &= \int |x - c|^r f_X(x)\, dx \\ &= \int_S |x - c|^r f_X(x)\, dx + \int_{\bar{S}} |x - c|^r f_X(x)\, dx \\ &\geq \int_S |x - c|^r f_X(x)\, dx \\ &\geq \int_S \delta^r f_X(x)\, dx \\ &= \delta^r P\{|X - c| > \delta\}, \end{aligned}$$

so that

$$E|X - c|^r \geq \delta^r P\{|X - c| > \delta\},$$

as claimed. ■

■ **Proof of Proposition 7.4.** Consider any real $(m \times 1)$ vector $\boldsymbol{\lambda}$, and form the function h: $\mathbb{R}^n \rightarrow \mathbb{R}^1$ defined by $h(\mathbf{x}) \equiv \boldsymbol{\lambda}'\mathbf{g}(\mathbf{x})$, noting that $h(\cdot)$ is differentiable. The *mean-value theorem* states that for a differentiable function $h(\cdot)$, there exists an $(n \times 1)$ vector $\mathbf{c}_T$ between $\mathbf{X}_T$ and $\mathbf{c}$ such that[10]

$$h(\mathbf{X}_T) - h(\mathbf{c}) = \left.\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}_T} \times (\mathbf{X}_T - \mathbf{c})$$

and therefore

$$\sqrt{T}\,[h(\mathbf{X}_T) - h(\mathbf{c})] = \left.\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}_T} \times \sqrt{T}(\mathbf{X}_T - \mathbf{c}). \qquad [7.A.3]$$

Since $\mathbf{c}_T$ is between $\mathbf{X}_T$ and $\mathbf{c}$ and since $\mathbf{X}_T \xrightarrow{p} \mathbf{c}$, we know that $\mathbf{c}_T \xrightarrow{p} \mathbf{c}$. Moreover, the derivative $\partial h(\mathbf{x})/\partial \mathbf{x}'$ is itself a continuous function of $\mathbf{x}$. Thus, from Proposition 7.1,

$$\left.\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}_T} \xrightarrow{p} \left.\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}}.$$

Given that $\sqrt{T}(\mathbf{X}_T - \mathbf{c}) \xrightarrow{L} \mathbf{X}$, Proposition 7.3(b) applied to expression [7.A.3] gives

$$\sqrt{T}\,[h(\mathbf{X}_T) - h(\mathbf{c})] \xrightarrow{L} \left.\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}} \mathbf{X},$$

or, in terms of the original function $\mathbf{g}(\cdot)$,

$$\boldsymbol{\lambda}'\{\sqrt{T}\,[\mathbf{g}(\mathbf{X}_T) - \mathbf{g}(\mathbf{c})]\} \xrightarrow{L} \boldsymbol{\lambda}' \left.\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}} \mathbf{X}.$$

Since this is true for any $\boldsymbol{\lambda}$, we conclude that

$$\sqrt{T}\,[\mathbf{g}(\mathbf{X}_T) - \mathbf{g}(\mathbf{c})] \xrightarrow{L} \left.\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{c}} \mathbf{X},$$

as claimed. ■

[10] That is, for any given $\mathbf{X}_T$ there exists a scalar μ_T with $0 \leq \mu_T \leq 1$ such that $\mathbf{c}_T = \mu_T\mathbf{X}_T + (1 - \mu_T)\mathbf{c}$. See, for example, Marsden (1974, pp. 174–75).

■ **Proof of Proposition 7.7.** Part (a) is established as in Andrews (1988, p. 463) using *Hölder's inequality* (see, for example, White, 1984, p. 30), which states that for $r > 1$, if $E[|Y|^r] < \infty$ and $E[|W|^{r/(r-1)}] < \infty$, then

$$E|YW| \leq \{E[|Y|^r]\}^{1/r} \times \{E[|W|^{r/(r-1)}]\}^{(r-1)/r}.$$

This implies that

$$E(|Y_t|\cdot\delta_{[|Y_t|\geq c]}) \leq \{E[|Y_t|^r]\}^{1/r} \times \{E[(\delta_{[|Y_t|\geq c]})^{r/(r-1)}]\}^{(r-1)/r}. \qquad [7.A.4]$$

Since $\delta_{[|Y_t|\geq c]}$ is either 0 or 1, it follows that

$$(\delta_{[|Y_t|\geq c]})^{r/(r-1)} = \delta_{[|Y_t|\geq c]}$$

and so

$$E[(\delta_{[|Y_t|\geq c]})^{r/(r-1)}] = E[\delta_{[|Y_t|\geq c]}] = \int_{|Y_t|\geq c} 1\cdot f_{Y_t}(y_t)\,dy_t = P\{|Y_t| \geq c\} \leq \frac{E|Y_t|}{c}, \qquad [7.A.5]$$

where the last result follows from Chebyshev's inequality. Substituting [7.A.5] into [7.A.4],

$$E(|Y_t|\cdot\delta_{[|Y_t|\geq c]}) \leq \{E[|Y_t|^r]\}^{1/r} \times \left\{\frac{E|Y_t|}{c}\right\}^{(r-1)/r}. \qquad [7.A.6]$$

Recall that $E[|Y_t|^r] < M'$ for all t, implying that there also exists an $M < \infty$ such that $E|Y_t| < M$ for all t. Hence,

$$E(|Y_t|\cdot\delta_{[|Y_t|\geq c]}) \leq (M')^{1/r} \times (M/c)^{(r-1)/r}.$$

This expression can be made as small as desired by choosing c sufficiently large. Thus condition [7.2.12] holds, ensuring that $\{Y_t\}$ is uniformly integrable.

To establish (b), notice that

$$E(|Y_t|\cdot\delta_{[|Y_t|\geq c]}) = E\left|\sum_{j=-\infty}^{\infty} h_j X_{t-j}\cdot\delta_{[|Y_t|\geq c]}\right| \leq E\left\{\sum_{j=-\infty}^{\infty} |h_j|\cdot|X_{t-j}|\cdot\delta_{[|Y_t|\geq c]}\right\}. \qquad [7.A.7]$$

Since $E[|X_{t-j}|^r] < M'$ and since $\delta_{[|Y_t|\geq c]} \leq 1$, it follows that $E\{|X_{t-j}|\cdot\delta_{[|Y_t|\geq c]}\}$ is bounded. Since $\{h_j\}_{j=-\infty}^{\infty}$ is absolutely summable, we can bring the expectation operator inside the summation in the last expression of [7.A.7] to deduce that

$$\begin{aligned} E\left\{\sum_{j=-\infty}^{\infty} |h_j|\cdot|X_{t-j}|\cdot\delta_{[|Y_t|\geq c]}\right\} &= \sum_{j=-\infty}^{\infty} |h_j|\cdot E\{|X_{t-j}|\cdot\delta_{[|Y_t|\geq c]}\} \\ &\leq \sum_{j=-\infty}^{\infty} |h_j|\cdot\{E[|X_{t-j}|^r]\}^{1/r} \times \left\{\frac{E|Y_t|}{c}\right\}^{(r-1)/r}, \end{aligned}$$

where the last inequality follows from the same arguments as in [7.A.6]. Hence, [7.A.7] becomes

$$E(|Y_t|\cdot\delta_{[|Y_t|\geq c]}) \leq \sum_{j=-\infty}^{\infty} |h_j| \times (M')^{1/r} \times \left\{\frac{E|Y_t|}{c}\right\}^{(r-1)/r}. \qquad [7.A.8]$$

But certainly, $E|Y_t|$ is bounded:

$$E|Y_t| = E\left|\sum_{j=-\infty}^{\infty} h_j X_{t-j}\right| \leq \sum_{j=-\infty}^{\infty} |h_j|\cdot E|X_{t-j}| = K < \infty.$$

Thus, from [7.A.8],

$$E(|Y_t|\cdot\delta_{[|Y_t|\geq c]}) \leq (M')^{1/r}(K/c)^{(r-1)/r} \sum_{j=-\infty}^{\infty} |h_j|. \qquad [7.A.9]$$

Since $\Sigma_{j=-\infty}^{\infty}|h_j|$ is finite, [7.A.9] can again be made as small as desired by choosing c sufficiently large. ■

■ **Proof of Proposition 7.9.** Consider $Y_t = \boldsymbol{\lambda}'\mathbf{Y}_t$ for $\boldsymbol{\lambda}$ any real $(n \times 1)$ vector. Then Y_t is a martingale difference sequence. We next verify that each of the conditions of Proposition

7.8 is satisfied. (a) $E(Y_t^2) = \boldsymbol{\lambda}'\boldsymbol{\Omega}_t\boldsymbol{\lambda} \equiv \sigma_t^2 > 0$, by positive definiteness of $\boldsymbol{\Omega}_t$. Likewise,

$$(1/T)\sum_{t=1}^{T}\sigma_t^2 = \boldsymbol{\lambda}'(1/T)\sum_{t=1}^{T}\boldsymbol{\Omega}_t\boldsymbol{\lambda} \to \boldsymbol{\lambda}'\boldsymbol{\Omega}\boldsymbol{\lambda} \equiv \sigma^2,$$

with $\sigma^2 > 0$, by positive definiteness of $\boldsymbol{\Omega}$. (b) $E(Y_t^4)$ is a finite sum of terms of the form $\lambda_i\lambda_j\lambda_l\lambda_m E(Y_{it}Y_{jt}Y_{lt}Y_{mt})$ and so is bounded for all t by condition (b) of Proposition 7.9; hence, Y_t satisfies condition (b) of Proposition 7.8 for $r = 4$. (c) Define $S_T \equiv (1/T) \times \Sigma_{t=1}^T Y_t^2$ and $\mathbf{S}_T \equiv (1/T)\Sigma_{t=1}^T \mathbf{Y}_t\mathbf{Y}_t'$, noticing that $S_T = \boldsymbol{\lambda}'\mathbf{S}_T\boldsymbol{\lambda}$. Since S_T is a continuous function of $\mathbf{S}_T$, we know that $\operatorname{plim} S_T = \boldsymbol{\lambda}'\boldsymbol{\Omega}\boldsymbol{\lambda} \equiv \sigma^2$, where $\boldsymbol{\Omega}$ is given as the plim of $\mathbf{S}_T$. Thus Y_t satisfies conditions (a) through (c) of Proposition 7.8, and so $\sqrt{T}\,\bar{Y}_T \xrightarrow{L} N(0, \sigma^2)$, or $\sqrt{T}\,\bar{Y}_T \xrightarrow{L} \boldsymbol{\lambda}'\mathbf{Y}$, where $\mathbf{Y} \sim (\mathbf{0}, \boldsymbol{\Omega})$. Since this is true for any $\boldsymbol{\lambda}$, this confirms the claim that $\sqrt{T}\,\bar{\mathbf{Y}}_T \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega})$. ■

■ **Proof of Proposition 7.10.** Let $Y \equiv X_tX_s$ and $W \equiv X_uX_v$. Then Hölder's inequality implies that for $r > 1$,

$$E|X_tX_sX_uX_v| \leq \{E|X_tX_s|^r\}^{1/r} \times \{E|X_uX_v|^{r/(r-1)}\}^{(r-1)/r}.$$

For $r = 2$, this means

$$E|X_tX_sX_uX_v| \leq \{E(X_tX_s)^2\}^{1/2} \times \{E(X_uX_v)^2\}^{1/2} \leq \max\{E(X_tX_s)^2, E(X_uX_v)^2\}.$$

A second application of Hölder's inequality with $Y \equiv X_t^2$ and $W \equiv X_s^2$ reveals that

$$E(X_tX_s)^2 = E(X_t^2X_s^2) \leq \{E(X_t^2)^r\}^{1/r} \times \{E(X_s^2)^{r/(r-1)}\}^{(r-1)/r}.$$

Again for $r = 2$, this implies from the strict stationarity of $\{X_t\}$ that

$$E(X_tX_s)^2 \leq E(X_t^4).$$

Hence, if $\{X_t\}$ is strictly stationary with finite fourth moment, then

$$E|X_tX_sX_uX_v| \leq E(X_t^4) = \mu_4$$

for all t, s, u, and v.

Observe further that

$$\begin{aligned}
E|Y_tY_sY_uY_v| &= E\left|\sum_{i=0}^{\infty} h_iX_{t-i}\sum_{j=0}^{\infty} h_jX_{s-j}\sum_{l=0}^{\infty} h_lX_{u-l}\sum_{m=0}^{\infty} h_mX_{v-m}\right| \\
&= E\left|\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty} h_ih_jh_lh_mX_{t-i}X_{s-j}X_{u-l}X_{v-m}\right| \\
&\leq E\left\{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty} |h_ih_jh_lh_m|\cdot|X_{t-i}X_{s-j}X_{u-l}X_{v-m}|\right\}.
\end{aligned}$$

But

$$\begin{aligned}
\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty} |h_ih_jh_lh_m| &= \sum_{i=0}^{\infty}|h_i|\sum_{j=0}^{\infty}|h_j|\sum_{l=0}^{\infty}|h_l|\sum_{m=0}^{\infty}|h_m| \\
&< \infty
\end{aligned}$$

and

$$E|X_{t-i}X_{s-j}X_{u-l}X_{v-m}| < \mu_4$$

for any value of any of the indices. Hence,

$$\begin{aligned}
E|Y_tY_sY_uY_v| &< \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty} |h_ih_jh_lh_m|\cdot\mu_4 \\
&< \infty. \quad ■
\end{aligned}$$

Chapter 7 Exercises

7.1. Let $\{X_T\}$ denote a sequence of random scalars with $\operatorname{plim} X_T = \xi$. Let $\{c_T\}$ denote a sequence of deterministic scalars with $\lim_{T\to\infty} c_T = c$. Let $g\colon \mathbb{R}^2 \to \mathbb{R}^1$ be continuous at (ξ, c). Show that $g(X_T, c_T) \xrightarrow{p} g(\xi, c)$.

7.2. Let $Y_t = 0.8Y_{t-1} + \varepsilon_t$ with $E(\varepsilon_t\varepsilon_\tau) = 1$ for $t = \tau$ and zero otherwise.

(a) Calculate $\lim_{T\to\infty} T\cdot\text{Var}(\bar{Y}_T)$.

(b) How large a sample would we need in order to have 95% confidence that $\bar{Y}_T$ differed from the true value zero by no more than 0.1?

7.3. Does a martingale difference sequence have to be covariance-stationary?

7.4. Let $Y_t = \Sigma_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$, where $\Sigma_{j=0}^{\infty}|\psi_j| < \infty$ and $\{\varepsilon_t\}$ is a martingale difference sequence with $E(\varepsilon_t^2) = \sigma^2$. Is Y_t covariance-stationary?

7.5. Define $X_{t,k} \equiv \Sigma_{u=0}^{\infty}\Sigma_{v=0}^{\infty}\psi_u\psi_v[\varepsilon_{t-u}\varepsilon_{t-k-v} - E(\varepsilon_{t-u}\varepsilon_{t-k-v})]$, where ε_t is an i.i.d. sequence with $E|\varepsilon_t|^r < M^r$ for some $r > 2$ and $M^r < \infty$ with $\Sigma_{j=0}^{\infty} |\psi_j| < \infty$. Show that $X_{t,k}$ is uniformly integrable.

7.6. Derive result [7.2.15].

7.7. Let Y_t follow an $ARMA(p, q)$ process,

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)(Y_t - \mu) = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t,$$

with roots of $(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) = 0$ and $(1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q) = 0$ outside the unit circle. Suppose ε_t has mean zero and is independent of ε_τ for $t \neq \tau$ with $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_t^4) < \infty$ for all t. Prove the following:

(a) $(1/T) \sum_{t=1}^{T} Y_t \xrightarrow{p} \mu$

(b) $[1/(T - k)] \sum_{t=k+1}^{T} Y_t Y_{t-k} \xrightarrow{p} E(Y_t Y_{t-k})$.

Chapter 7 References

Anderson, T. W. 1971. *The Statistical Analysis of Time Series*. New York: Wiley.

Andrews, Donald W. K. 1988. "Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables." *Econometric Theory* 4:458–67.

Hoel, Paul G., Sidney C. Port, and Charles J. Stone. 1971. *Introduction to Probability Theory*. Boston: Houghton Mifflin.

Marsden, Jerrold E. 1974. *Elementary Classical Analysis*. San Francisco: Freeman.

Phillips, Peter C. B., and Victor Solo. 1992. "Asymptotics for Linear Processes." *Annals of Statistics* 20:971–1001.

Rao, C. Radhakrishna. 1973. *Linear Statistical Inference and Its Applications*, 2d ed. New York: Wiley.

Royden, H. L. 1968. *Real Analysis*, 2d ed. New York: Macmillan.

Theil, Henri. 1971. *Principles of Econometrics*. New York: Wiley.

White, Halbert. 1984. *Asymptotic Theory for Econometricians*. Orlando, Fla.: Academic Press.

8 Linear Regression Models

We have seen that one convenient way to estimate the parameters of an autoregression is with ordinary least squares regression, an estimation technique that is also useful for a number of other models. This chapter reviews the properties of linear regression. Section 8.1 analyzes the simplest case, in which the explanatory variables are nonrandom and the disturbances are i.i.d. Gaussian. Section 8.2 develops analogous results for ordinary least squares estimation of more general models such as autoregressions and regressions in which the disturbances are non-Gaussian, heteroskedastic, or autocorrelated. Linear regression models can also be estimated by generalized least squares, which is described in Section 8.3.

8.1. *Review of Ordinary Least Squares with Deterministic Regressors and i.i.d. Gaussian Disturbances*

Suppose that a scalar y_t is related to a $(k \times 1)$ vector $\mathbf{x}_t$ and a disturbance term u_t according to the regression model

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t. \qquad [8.1.1]$$

This relation could be used to describe either the random variables or their realization. In discussing regression models, it proves cumbersome to distinguish notationally between random variables and their realization, and standard practice is to use small letters for either.

This section reviews estimation and hypothesis tests about $\boldsymbol{\beta}$ under the assumptions that $\mathbf{x}_t$ is deterministic and u_t is i.i.d. Gaussian. The next sections discuss regression under more general assumptions. First, however, we summarize the mechanics of linear regression and present some formulas that hold regardless of statistical assumptions.

The Algebra of Linear Regression

Given an observed sample $(y_1, y_2, \ldots, y_T)$, the *ordinary least squares* (*OLS*) estimate of $\boldsymbol{\beta}$ (denoted $\mathbf{b}$) is the value of $\boldsymbol{\beta}$ that minimizes the residual sum of squares (*RSS*):

$$RSS \equiv \sum_{t=1}^{T} (y_t - \mathbf{x}_t'\boldsymbol{\beta})^2. \qquad [8.1.2]$$

We saw in Appendix 4.A to Chapter 4 that the *OLS* estimate is given by

$$\mathbf{b} = \left[\sum_{t=1}^{T} (\mathbf{x}_t\mathbf{x}_t')\right]^{-1}\left[\sum_{t=1}^{T} (\mathbf{x}_t y_t)\right], \qquad [8.1.3]$$

assuming that the $(k \times k)$ matrix $\Sigma_{t=1}^{T}(\mathbf{x}_t\mathbf{x}_t')$ is nonsingular. The *OLS* sample residual for observation t is

$$\hat{u}_t \equiv y_t - \mathbf{x}_t'\mathbf{b}. \qquad [8.1.4]$$

Often the model in [8.1.1] is written in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad [8.1.5]$$

where

$$\underset{(T\times 1)}{\mathbf{y}} \equiv \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \qquad \underset{(T\times k)}{\mathbf{X}} \equiv \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_T' \end{bmatrix} \qquad \underset{(T\times 1)}{\mathbf{u}} \equiv \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}.$$

Then the *OLS* estimate in [8.1.3] can be written as

$$\begin{aligned} \mathbf{b} &= \left\{ [\mathbf{x}_1 \quad \mathbf{x}_2 \cdots \mathbf{x}_T] \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_T' \end{bmatrix} \right\}^{-1} \left\{ [\mathbf{x}_1 \quad \mathbf{x}_2 \cdots \mathbf{x}_T] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \right\} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \end{aligned} \qquad [8.1.6]$$

Similarly, the vector of *OLS* sample residuals [8.1.4] can be written as

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\mathbf{b} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = [\mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} = \mathbf{M}_\mathbf{X}\mathbf{y}, \qquad [8.1.7]$$

where $\mathbf{M}_\mathbf{X}$ is defined as the following $(T \times T)$ matrix:

$$\mathbf{M}_\mathbf{X} \equiv \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \qquad [8.1.8]$$

One can readily verify that $\mathbf{M}_\mathbf{X}$ is symmetric:

$$\mathbf{M}_\mathbf{X} = \mathbf{M}_\mathbf{X}';$$

idempotent:

$$\mathbf{M}_\mathbf{X}\mathbf{M}_\mathbf{X} = \mathbf{M}_\mathbf{X};$$

and orthogonal to the columns of $\mathbf{X}$:

$$\mathbf{M}_\mathbf{X}\mathbf{X} = \mathbf{0}. \qquad [8.1.9]$$

Thus, from [8.1.7], the *OLS* sample residuals are orthogonal to the explanatory variables in $\mathbf{X}$:

$$\hat{\mathbf{u}}'\mathbf{X} = \mathbf{y}'\mathbf{M}_\mathbf{X}'\mathbf{X} = \mathbf{0}'. \qquad [8.1.10]$$

The *OLS* sample residual $(\hat{u}_t)$ should be distinguished from the population residual (u_t). The sample residual is constructed from the sample estimate $\mathbf{b}$ $(\hat{u}_t = y_t - \mathbf{x}_t'\mathbf{b})$, whereas the population residual is a hypothetical construct based on the true population value $\boldsymbol{\beta}$ $(u_t = y_t - \mathbf{x}_t'\boldsymbol{\beta})$. The relation between the sample