Introduction to Statistical Time Series
WAYNE A. FULLER

CHAPTER 5

Some Large Sample Theory

So far, we have been interested in ways of representing time series and describing their properties. In most practical situations we have a portion of a realization, or of several realizations, and we wish a description (an estimate of the parameters) of the time series.

Most of the presently available results on the estimation of the covariance function, the parameters of autoregressive and moving average processes, and the spectral density rest on large sample theory. Therefore, we shall present some results in large sample statistics.

5.1. ORDER IN PROBABILITY

Concepts of relative magnitude or *order of magnitude* are useful in investigating limiting behavior of random variables. We first define the concepts of order as used in real analysis. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers and $\{g_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers.

Definition 5.1.1. We say a_n is of smaller order than g_n and write

$$a_n = o(g_n)$$

if

$$\lim_{n\to\infty} g_n^{-1} a_n = 0 .$$

Definition 5.1.2. We say a_n is at most of order g_n and write

$$a_n = O(g_n)$$

if there exists a real number M such that $g_n^{-1}|a_n| \le M$ for all n.

The properties of Lemma 5.1.1 are easily established using the definitions of order and the properties of limits.

Lemma 5.1.1. Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers. Let $\{f_n\}$ and $\{g_n\}$ be sequences of positive real numbers.

(i) If $a_n = o(f_n)$ and $b_n = o(g_n)$, then

$$\begin{aligned} a_n b_n &= o(f_n g_n)\,, \\ |a_n|^s &= o(f_n^s) \qquad \text{for} \quad s > 0\,, \\ a_n + b_n &= o(\max\{f_n, g_n\})\,. \end{aligned}$$

(ii) If $a_n = O(f_n)$ and $b_n = O(g_n)$, then

$$\begin{aligned} a_n b_n &= O(f_n g_n)\,, \\ |a_n|^s &= O(f_n^s) \qquad \text{for} \quad s \geq 0\,, \\ a_n + b_n &= O(\max\{f_n, g_n\})\,. \end{aligned}$$

(iii) If $a_n = o(f_n)$ and $b_n = O(g_n)$, then

$$a_n b_n = o(f_n g_n)\,.$$

Proof. Reserved for the reader. ▲

The concepts of order when applied to random variables are closely related to *convergence in probability*.

Definition 5.1.3. The sequence of random variables $\{X_n\}$ converges in probability to the random variable X, and we write

$$\text{plim}\, X_n = X$$

(the *probability limit* of X_n is X), if for every $\epsilon > 0$

$$\lim_{n\to\infty} P\{|X_n - X| > \epsilon\} = 0\,.$$

An equivalent definition is that for every $\epsilon > 0$ and $\delta > 0$ there exists an N such that for all $n > N$,

$$P\{|X_n - X| > \epsilon\} < \delta\,.$$

The notation

$$X_n \xrightarrow{P} X$$

is also frequently used to indicate that X_n converges in probability to X.

For sequences of random variables, definitions of *order in probability* were introduced by Mann and Wald (1943b). Let $\{X_n\}$ be a sequence of random variables and $\{g_n\}$ a sequence of positive real numbers.

Definition 5.1.4. We say X_n is of smaller order in probability than g_n and write

$$X_n = o_p(g_n)$$

if

$$p\lim g_n^{-1} X_n = 0 .$$

Definition 5.1.5. We say X_n is at most of order in probability g_n and write

$$X_n = O_p(g_n)$$

if, for every $\epsilon > 0$, there exists a positive real number M_ϵ such that

$$P\{|X_n| \geq M_\epsilon g_n\} \leq \epsilon$$

for all n.

If $X_n = O_p(g_n)$, we sometimes say that X_n is *bounded in probability* by g_n. We define a vector random variable to be $O_p(g_n)$ if every element of the vector is $O_p(g_n)$ as follows.

Definition 5.1.6. If $\mathbf{X}_n$ is a k-dimensional random variable, then $\mathbf{X}_n$ is at most of order in probability g_n and we write

$$\mathbf{X}_n = O_p(g_n)$$

if, for every $\epsilon > 0$, there exists a positive real number M_ϵ such that

$$P\{|X_{jn}| \geq M_\epsilon g_n\} \leq \epsilon , \qquad j = 1, 2, \ldots, k ,$$

for all n. We say $\mathbf{X}_n$ is of smaller order in probability than g_n and write

$$\mathbf{X}_n = o_p(g_n)$$

if, for every $\epsilon > 0$ and $\delta > 0$, there exists an N such that for all $n > N$,

$$P\{|X_{jn}| > \epsilon g_n\} < \delta , \qquad j = 1, 2, \ldots, k .$$

Note that k might be a function of n, and $\mathbf{X}_n$ could still satisfy the definition. However, it is clear that the M_ϵ of the definition is a function of ϵ only (and not of n).

A matrix random variable may be viewed as a vector random variable with the elements displayed in a particular manner, or as a collection of vector random variables. Therefore, we shall define the order of matrix random variables in an analogous manner.

Definition 5.1.7. A $k \times r$ matrix $\mathbf{B}_n$ of random variables is at most of order in probability g_n, and we write

$$\mathbf{B}_n = O_p(g_n),$$

if for every $\epsilon > 0$ there exists a positive real number M_ϵ such that

$$P\{|b_{ijn}| \geq M_\epsilon g_n\} \leq \epsilon, \qquad i = 1, 2, \ldots, k, \quad j = 1, 2, \ldots, r,$$

for all n, where the b_{ijn} are the elements of $\mathbf{B}_n$. We say that $\mathbf{B}_n$ is of smaller order in probability than g_n and write

$$\mathbf{B}_n = o_p(g_n)$$

if for every $\epsilon > 0$ and $\delta > 0$ there exists an N such that for all $n > N$,

$$P\{|b_{ijn}| > \epsilon g_n\} < \delta, \qquad i = 1, 2, \ldots, k, \quad j = 1, 2, \ldots, r.$$

For real numbers a_i, $i = 1, \ldots, n$, we know that $|\Sigma_{i=1}^n a_i| \leq \Sigma_{i=1}^n |a_i|$. The following lemma, resting on this property, furnishes a useful bound on the probability that the absolute value of a sum of random variables exceeds a given number.

Lemma 5.1.2. Let $\mathbf{X}_i$, $i = 1, 2, \ldots, n$, be k-dimensional random variables. Then, for every $\epsilon > 0$,

$$P\left\{\left|\sum_{i=1}^{n} \mathbf{X}_i\right| \geq \epsilon\right\} \leq \sum_{i=1}^{n} P\left\{|\mathbf{X}_i| \geq \frac{\epsilon}{n}\right\}.$$

Proof. Let $\epsilon > 0$ be arbitrary. We see that if $\Sigma_{i=1}^n |\mathbf{X}_i| \geq \epsilon$, then $|\mathbf{X}_i| \geq \epsilon/n$ for at least one $i \in \{1, 2, \ldots, n\}$. Therefore,

$$\begin{aligned} P\left\{\left|\sum_{i=1}^{n} \mathbf{X}_i\right| \geq \epsilon\right\} &\leq P\left\{\sum_{i=1}^{n} |\mathbf{X}_i| \geq \epsilon\right\} \\ &\leq \sum_{i=1}^{n} P\left\{|\mathbf{X}_i| \geq \frac{\epsilon}{n}\right\}. \end{aligned}$$ ▲

Definition 5.1.3 applies for vector random variables of fixed dimension as well as for scalar random variables, if it is understood that $|\mathbf{X}_n - \mathbf{X}|$ is the common Euclidean distance.

Lemma 5.1.3. Let $\mathbf{X}_n$ be a k-dimensional random variable such that

$$\text{plim}\, X_{jn} = X_j, \qquad j = 1, 2, \ldots, k,$$

where X_{jn} is the jth element of $\mathbf{X}_n$. Then, for k fixed,

$$\text{plim}\, \mathbf{X}_n = \mathbf{X}.$$

Proof. By hypothesis, for each j and for every $\epsilon > 0$ and $\delta > 0$, there exists an integer N_j such that for all $n > N_j$

$$P\{|X_{jn} - X_j| > k^{-1/2}\epsilon\} \le k^{-1}\delta .$$

Let N be the maximum of $\{N_1, N_2, \ldots, N_k\}$. Using Lemma 5.1.2, we have

$$P\{|\mathbf{X}_n - \mathbf{X}| > \epsilon\} \le \sum_{j=1}^{k} P\{|X_{jn} - X_j| > k^{-1/2}\epsilon\} \le \delta$$

for $n > N$. ▲

The proof of Lemma 5.1.3 should also help to make it clear that if k is not fixed, then the fact that $\mathbf{X}_n = o_p(1)$ does not necessarily imply that $\text{plim}|\mathbf{X}_n| = 0$. The vector random variable composed of n entries all equal to $n^{-1/2}$ furnishes a counterexample for $k = n$.

We shall demonstrate later (Theorems 5.1.5 and 5.1.6) that operations valid for order are also valid for order in probability. Since it is relatively easy to establish the properties analogous to those of Lemma 5.1.1, we do so at this time.

Lemma 5.1.4. Let $\{f_n\}$ and $\{g_n\}$ be sequences of positive real numbers, and let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables.

(i) If $X_n = o_p(f_n)$ and $Y_n = o_p(g_n)$, then

$$\begin{aligned} X_n Y_n &= o_p(f_n g_n) , \\ |X_n|^s &= o_p(f_n^s) \qquad \text{for} \quad s > 0 , \\ X_n + Y_n &= o_p(\max\{f_n, g_n\}) . \end{aligned}$$

(ii) If $X_n = O_p(f_n)$ and $Y_n = O_p(g_n)$, then

$$\begin{aligned} X_n Y_n &= O_p(f_n g_n) , \\ |X_n|^s &= O_p(f_n^s) \qquad \text{for} \quad s \ge 0 , \\ X_n + Y_n &= O_p(\max\{f_n, g_n\}) . \end{aligned}$$

(iii) If $X_n = o_p(f_n)$ and $Y_n = O_p(g_n)$, then

$$X_n Y_n = o_p(f_n g_n) .$$

Proof. We investigate only part i, leaving parts ii and iii as an exercise. By arguments similar to those of Lemma 5.1.3, $|X_n Y_n| > f_n g_n$ implies that $|X_n / f_n| > 1$ or (and) $|Y_n / g_n| > 1$. By hypothesis, given $\epsilon > 0$ and $\delta > 0$, there is an N such that

$$P\{|X_n| > \epsilon f_n\} < 0.5\delta\,,$$
$$P\{|Y_n| > \epsilon g_n\} < 0.5\delta$$

for $n > N$. Therefore,

$$\begin{aligned} P\{|X_n Y_n| > \epsilon^2 f_n g_n\} &\leq P\{|f_n^{-1} X_n| > \epsilon \text{ or } |g_n^{-1} Y_n| > \epsilon\} \\ &\leq P\{|f_n^{-1} X_n| > \epsilon\} + P\{|g_n^{-1} Y_n| > \epsilon\} \\ &< \delta \end{aligned}$$

for $n > N$.

The second equality in part i follows from

$$P\{|X_n| > \epsilon f_n\} = P\{|X_n|^s > \epsilon^s f_n^s\}\,,$$

which holds for all $\epsilon > 0$.

Let $q_n = \max\{f_n, g_n\}$. Given $\epsilon > 0$ and $\delta > 0$, there exists an n such that

$$P\{|X_n| > \tfrac{1}{2}\epsilon q_n\} < \tfrac{1}{2}\delta\,,$$
$$P\{|Y_n| > \tfrac{1}{2}\epsilon q_n\} < \tfrac{1}{2}\delta$$

for $n > N$. Hence, the third result of part i follows by Lemma 5.1.2. ▲

One of the most useful tools for establishing the order in probability of random variables is Chebyshev's inequality.

Theorem 5.1.1 (Chebyshev's inequality). Let $r > 0$, let X be a random variable such that $E\{|X|^r\} < \infty$, and let $F(x)$ be the distribution function of X. Then, for every $\epsilon > 0$ and finite A,

$$P\{|X - A| \geq \epsilon\} \leq \frac{E\{|X - A|^r\}}{\epsilon^r}\,.$$

Proof. Let us denote by S the set of x for which $|x - A| \geq \epsilon$ and by $\bar{S}$ the set of x for which $|x - A| < \epsilon$. Then,

$$\begin{aligned} \int |x - A|^r \, dF(x) &= \int_S |x - A|^r \, dF(x) + \int_{\bar{S}} |x - A|^r \, dF(x) \\ &\geq \epsilon^r \int_S dF(x) = \epsilon^r P\{|X - A| \geq \epsilon\}\,. \end{aligned}$$ ▲

It follows from Chebyshev's inequality that any random variable with finite variance is bounded in probability by the square root of its second moment about the origin.

Corollary 5.1.1.1. Let $\{X_n\}$ be a sequence of random variables and $\{a_n\}$ a sequence of positive real numbers such that

$$E\{X_n^2\} = O(a_n^2)\,.$$

Then

$$X_n = O_p(a_n)\,.$$

Proof. By assumption there exists an M_1 such that

$$E\{X_n^2\} < M_1^2 a_n^2$$

for all n. By Chebyshev's inequality, for any $M_2 > 0$,

$$P\{|X_n| \geqslant M_2 a_n\} \leqslant \frac{E\{X_n^2\}}{M_2^2 a_n^2}\,.$$

Hence, given $\epsilon > 0$, we choose $M_2 \geqslant M_1 \epsilon^{-1/2}$, and the result follows. ▲

If the sequence $\{X_n\}$ has zero mean or a mean whose order is less than or equal to the order of the standard error, then the order in probability of the sequence is the order of the standard error.

Corollary 5.1.1.2. Let the sequence of random variables $\{X_n\}$ satisfy

$$E\{(X_n - E\{X_n\})^2\} = O(a_n^2)$$

and

$$E\{X_n\} = O(a_n)\,,$$

where $\{a_n\}$ is a sequence of positive real numbers. Then

$$X_n = O_p(a_n)\,.$$

Proof. By the assumptions and by property ii of Lemma 5.1.1,

$$E\{X_n^2\} = E\{(X_n - E\{X_n\})^2\} + (E\{X_n\})^2 = O(a_n^2)\,,$$

and the result follows by Corollary 5.1.1.1. ▲

Let the probability limits of two sequences of random variables be defined. We now demonstrate that the sequences have a common probability limit if the probability limit of the sequence of differences is zero.

Theorem 5.1.2. Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables such that

$$p\lim|X_n - Y_n| = 0 .$$

If there exists a random variable X such that $p\lim X_n = X$, then

$$p\lim Y_n = X .$$

Proof. Given $\epsilon > 0$ and $\delta > 0$, there exists, by hypothesis, an N such that for $n > N$,

$$P\{|Y_n - X_n| \geq 0.5\epsilon\} \leq 0.5\delta \quad \text{and} \quad P\{|X_n - X| \geq 0.5\epsilon\} \leq 0.5\delta .$$

Applying Lemma 5.1.2, for $n > N$,

$$P\{|Y_n - X| \geq \epsilon\} \leq \delta . \qquad \blacktriangle$$

Definition 5.1.8. For $r \geq 1$, the sequence of random variables $\{X_n\}$ *converges in rth mean* to the random variable X if $E\{|X_n|^r\} < \infty$ for all n and

$$E\{|X_n - X|^r\} \to 0$$

as $n \to \infty$. We denote convergence in rth mean by writing $X_n \xrightarrow{r} X$.

We note that if $E\{|X_n - X_m|^r\} \to 0$ as $n \to \infty$ and $m \to \infty$, then there exists a random variable X such that $X_n \xrightarrow{r} X$. Using Chebyshev's inequality it is easy to demonstrate that convergence in rth mean implies convergence in probability.

Theorem 5.1.3. Let $\{X_n\}$ be a sequence of random variables with finite rth moments. If there exists a random variable X such that $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{P} X$.

Proof. Given $\epsilon > 0$,

$$P\{|X_n - X| > \epsilon\} \leq \epsilon^{-r} E\{|X_n - X|^r\}$$

by Chebyshev's inequality. For $\delta > 0$ there is, by hypothesis, an integer $N = N(\epsilon, \delta)$ such that for all $n > N$,

$$E\{|X_n - X|^r\} < \delta\epsilon^r$$

and therefore, for $n > N$,

$$P\{|X_n - X| > \epsilon\} < \delta . \qquad \blacktriangle$$

One useful consequence is Corollary 5.1.3.1, which can be paraphrased as follows. If the sequence of differences of two sequences of random variables converges in squared mean to zero, then the two sequences of random variables have a common probability limit if the limit exists.

Corollary 5.1.3.1. Let $\{X_n\}$ and $\{Y_n\}$ be sequences of [illegible]ndom variables such that

$$\lim_{n\to\infty} E\{(X_n - Y_n)^2\} = 0 .$$

If there exists a random variable X such that $p\lim X_n = X$, then $p\lim Y_n = X$.

Proof. By Theorem 5.1.3, we have that $p\lim(X_n - Y_n) = 0$. The conclusion follows by Theorem 5.1.2. ▲

Corollary 5.1.3.2. If the sequence of random variables $\{Y_n\}$ is such that

$$\lim_{n\to\infty} E\{Y_n\} = \mu$$

and

$$\lim_{n\to\infty} E\{(Y_n - E\{Y_n\})^2\} = 0 ,$$

then $p\lim Y_n = \mu$.

Proof. The proof follows directly by letting the constants μ and $\{E\{Y_n\}\}$ be, respectively, the X and $\{X_n\}$ of Corollary 5.1.3.1. ▲

Since we often work with functions of sequences of random variables, the following theorem is very important. The theorem states that if the function $\mathbf{g}(\mathbf{x})$ is continuous, then "the probability limit of the function is the function of the probability limit."

Theorem 5.1.4. Let $\{\mathbf{X}_n\}$ be a sequence of real valued k-dimensional random variables such that $p\lim \mathbf{X}_n = \mathbf{X}$. Let $\mathbf{g}(\mathbf{x})$ be a function mapping the real k-dimensional vector $\mathbf{x}$ into a real p-dimensional space. Let $\mathbf{g}(\mathbf{x})$ be continuous. Then $p\lim \mathbf{g}(\mathbf{X}_n) = \mathbf{g}(\mathbf{X})$.

Proof. Given $\epsilon > 0$ and $\delta > 0$, let A be a closed and bounded k-dimensional set such that

$$P\{\mathbf{X} \in A\} \geq 1 - 0.5\delta .$$

Since $\mathbf{g}(\mathbf{x})$ is continuous, it is uniformly continuous on A, and there exists a δ_ϵ such that

$$|\mathbf{g}(\mathbf{x}_1) - \mathbf{g}(\mathbf{x}_2)| < \epsilon$$

if $|\mathbf{x}_1 - \mathbf{x}_2| < \delta_\epsilon$ and $\mathbf{x}_1$ is in A. Since $p\lim \mathbf{X}_n = \mathbf{X}$, there exists an N such that for $n > N$,

$$P\{|\mathbf{X}_n - \mathbf{X}| > \delta_\epsilon\} < \tfrac{1}{2}\delta .$$

Therefore, for $n > N$,

$$\begin{aligned} P\{|\mathbf{g}(\mathbf{X}_n) - \mathbf{g}(\mathbf{X})| > \epsilon\} &= P\{|\mathbf{g}(\mathbf{X}_n) - \mathbf{g}(\mathbf{X})| > \epsilon | \mathbf{X} \notin A\} P\{\mathbf{X} \notin A\} \\ &\quad + P\{|\mathbf{g}(\mathbf{X}_n) - \mathbf{g}(\mathbf{X})| > \epsilon | \mathbf{X} \in A\} P\{\mathbf{X} \in A\} \\ &\leq P\{\mathbf{X} \notin A\} + P\{|\mathbf{X}_n - \mathbf{X}| > \delta_\epsilon | \mathbf{X} \in A\} P\{\mathbf{X} \in A\} \\ &\leq P\{\mathbf{X} \notin A\} + P\{|\mathbf{X}_n - \mathbf{X}| > \delta_\epsilon\} \\ &< \delta . \end{aligned}$$ ▲

Theorem 5.1.4 can be extended to functions that are continuous except on a set D where $P\{\mathbf{X} \in D\} = 0$. See, for example, Tucker (1967, p. 104).

Mann and Wald (1943b) demonstrated that the algebra of the common order relationships holds for order in probability. The following two theorems are similar to a paraphrase of Mann and Wald's result given by Pratt (1959). The proof follows more closely that of Mann and Wald, however.

Theorem 5.1.5. Let $\{\mathbf{X}_n\}$ be a sequence of k-dimensional random variables with elements $\{X_{jn}: j = 1, 2, \ldots, k\}$, and let $\{\mathbf{r}_n\}$ be a sequence of k-dimensional vectors with positive real elements $\{r_{jn}: j = 1, 2, \ldots, k\}$ such that

$$\begin{aligned} X_{jn} &= O_p(r_{jn}) , \qquad j = 1, 2, \ldots, t , \\ X_{jn} &= o_p(r_{jn}) , \qquad j = t+1, t+2, \ldots, k . \end{aligned}$$

Let $g_n(\mathbf{x})$ be a sequence of real valued (Borel measurable) functions defined on k-dimensional Euclidian space, and let $\{s_n\}$ be a sequence of positive real numbers. Let $\{\mathbf{a}_n\}$ be a nonrandom sequence of k-dimensional vectors. If

$$g_n(\mathbf{a}_n) = O(s_n)$$

for all sequences $\{\mathbf{a}_n\}$ such that

$$\begin{aligned} a_{jn} &= O(r_{jn}) , \qquad j = 1, 2, \ldots, t , \\ a_{jn} &= o(r_{jn}) , \qquad j = t+1, t+2, \ldots, k , \end{aligned}$$

then

$$g_n(\mathbf{X}_n) = O_p(s_n) .$$

Proof. Set $\epsilon > 0$. By assumption there exist real numbers $M_1, M_2, \ldots, M_t$ and sequences $\{M_{jn}\}$, $j = t+1, t+2, \ldots, k$, such that $\lim_{n\to\infty} M_{jn} = 0$ and

$$\begin{aligned} P\{|X_{jn}| \geq M_j r_{jn}\} &< \frac{\epsilon}{k} , \qquad j = 1, 2, \ldots, t , \\ P\{|X_{jn}| \geq M_{jn} r_{jn}\} &< \frac{\epsilon}{k} , \qquad j = t+1, t+2, \ldots, k , \end{aligned}$$

for all n. Let $\{A_n\}$ be a sequence of k-dimensional sets defined by

$$A_n = \{(y_1, y_2, \ldots, y_k): -M_j r_{jn} \leqslant y_j \leqslant M_j r_{jn} \text{ for } 1 \leqslant j \leqslant t$$
$$\text{and } -M_{jn} r_{jn} \leqslant y_j \leqslant M_{jn} r_{jn} \text{ for } t+1 \leqslant j \leqslant k\}.$$

Then, for $\mathbf{a} \in A_n$, there exists an M such that $|g_n(\mathbf{a})| < Ms_n$. Hence,

$$|g_n(\mathbf{X}_n)| < Ms_n$$

for all $\mathbf{X}_n$ contained in A_n, and the result follows, since the A_n were constructed so that $P\{\mathbf{X}_n \in A_n\} > 1 - \epsilon$ for all n. ▲

Theorem 5.1.6. If we replace

$$g_n(\mathbf{a}_n) = O(s_n)$$

by

$$g_n(\mathbf{a}_n) = o(s_n)$$

in the hypothesis of Theorem 5.1.5, we may replace

$$g_n(\mathbf{X}_n) = O_p(s_n)$$

by

$$g_n(\mathbf{X}_n) = o_p(s_n)$$

in the conclusion.

Proof. The set A_n is constructed exactly as in the proof of Theorem 5.1.5. There then exists a sequence $\{b_n\}$ such that $\lim_{n\to\infty} b_n = 0$ and

$$|g_n(\mathbf{a})| < b_n s_n$$

for $\mathbf{a}$ contained in A_n. Therefore, $|g_n(\mathbf{X}_n)| < b_n s_n$ for all $\mathbf{X}_n$ contained in A_n, and the result follows from the construction of the A_n. ▲

Corollary 5.1.5. Let $\{X_n\}$ be a sequence of scalar random variables such that

$$X_n = a + O_p(r_n),$$

where $r_n \to 0$ as $n \to \infty$. If $g(x)$ is a function with s continuous derivatives at $x = a$,

then

$$g(X_n) = g(a) + g^{(1)}(a)(X_n - a) + \cdots + \frac{1}{(s-1)!} g^{(s-1)}(a)(X_n - a)^{s-1} + O_p(r_n^s),$$

where $g^{(j)}(a)$ is the jth derivative of $g(x)$ evaluated at $x = a$.

Proof. Since the statement holds for a sequence of real numbers, the result follows from Theorem 5.1.5. A direct proof can be obtained by expanding $g(x)$ in a Taylor series with remainder

$$\frac{g^{(s)}(b)(X_n - a)^s}{s!},$$

where b is between X_n and a. Since $g^{(s)}(x)$ is continuous at a, for n sufficiently large, $g^{(s)}(b)$ is bounded in probability [i.e., $g^{(s)}(b) = O_p(1)$]. Therefore,

$$\frac{g^{(s)}(b)(X_n - a)^s}{s!} = O_p(r_n^s). \qquad \blacktriangle$$

Corollary 5.1.6. If

$$X_n = a + O_p(r_n)$$

in the hypothesis of Corollary 5.1.5 is replaced by

$$X_n = a + o_p(r_n),$$

then the remainder $O_p(r_n^s)$ is replaced by $o_p(r_n^s)$.

Proof. The proof is nearly identical to that of Corollary 5.1.5. ▲

Note that the condition on the derivative defining the remainder can be weakened. As we saw in the proof of Corollary 5.1.5, we need only that $g^{(s)}(b)$ is bounded in probability. Also see Exercise 5.30.

The corollaries generalize immediately to vector random variables. For example, let

$$\mathbf{X}_n = \mathbf{a} + o_p(r_n),$$

where $\mathbf{X}_n = (X_{1n}, X_{2n}, \ldots, X_{kn})'$, $\mathbf{a} = (a_1, a_2, \ldots, a_k)'$, and $r_n \to 0$ as $n \to \infty$. Let $g(\mathbf{x})$ be a real valued function defined on k-dimensional Euclidean space with continuous partial derivatives of order three at $\mathbf{a}$; then, for example,

$$g(\mathbf{X}_n) = g(\mathbf{a}) + \sum_{j=1}^{k} \frac{\partial g(\mathbf{a})}{\partial x_j}(X_{jn} - a_j)$$
$$+ \sum_{j=1}^{k}\sum_{i=1}^{k} \frac{1}{2!}\frac{\partial^2 g(\mathbf{a})}{\partial x_j\,\partial x_i}(X_{jn} - a_j)(X_{in} - a_i) + o_p(r_n^3),$$

where $\partial g(\mathbf{a})/\partial x_j$ is the partial derivative of $g(\mathbf{x})$ with respect to x_j evaluated at $\mathbf{x} = \mathbf{a}$ and $\partial^2 g(\mathbf{a})/\partial x_j\,\partial x_i$ is the second partial derivative of $g(\mathbf{x})$ with respect to x_j and x_i evaluated at $\mathbf{x} = \mathbf{a}$.

The Taylor expansions of Corollaries 5.1.5 and 5.1.6 are about a fixed point $\mathbf{a}$. Expansions about a random point are also valid under certain conditions.

Theorem 5.1.7. Let $\{\mathbf{X}_n\}$, $\{\mathbf{W}_n\}$ be two sequences of k-dimensional vector random variables defined on the same probability space. Let $l(\mathbf{X}_n, \mathbf{W}_n)$ be the line segment joining $\mathbf{X}_n$ and $\mathbf{W}_n$. For every $\epsilon > 0$, suppose there is an N_ϵ and a set B_ϵ in k-dimensional Euclidean space, such that

$$P\{l(\mathbf{X}_n, \mathbf{W}_n) \in B_\epsilon\} \geq 1 - \epsilon$$

for all $n > N_\epsilon$.

Let $g(\mathbf{x})$ be a real valued function that is continuous with continuous partial derivatives through order s on B_ϵ for every $\epsilon > 0$. Suppose the absolute values of the sth order partial derivatives are bounded on B_ϵ and that

$$\mathbf{X}_n - \mathbf{W}_n = O_p(r_n),$$

where $r_n \to 0$ as $n \to \infty$. Then

$$g(\mathbf{X}_n) = g(\mathbf{W}_n) + \sum_{i=1}^{k} g^{(i)}(\mathbf{W}_n)(X_{in} - W_{in})$$
$$+ 2^{-1}\sum_{i=1}^{k}\sum_{j=1}^{k} g^{(ij)}(\mathbf{W}_n)(X_{in} - W_{in})(X_{jn} - W_{jn}) + \cdots$$
$$+ [(s-1)!]^{-1}\sum_{i=1}^{k}\cdots\sum_{t=1}^{k} g^{(i\ldots t)}(\mathbf{W}_n)(X_{in} - W_{in})\ldots(X_{tn} - W_{tn})$$
$$+ O_p(r_n^s),$$

where $g^{(i)}(\mathbf{W}_n)$ is the first partial derivative of $g(\mathbf{x})$ with respect to x_i evaluated at $\mathbf{x} = \mathbf{W}_n$, and $g^{(i\ldots t)}(\mathbf{W}_n)$ is the sth partial derivative of $g(\mathbf{x})$ with respect to the elements of $\mathbf{x}$ whose indexes are $i, \ldots, t$, evaluated at $\mathbf{x} = \mathbf{W}_n$.

Proof. Let $\epsilon > 0$ be given, and let B_ϵ be the associated set. Then $g(\mathbf{x})$ and the first s derivatives of $g(\mathbf{x})$ are continuous on B_ϵ. For $l(\mathbf{X}_n, \mathbf{W}_n) \in B_\epsilon$, one can expand the function in a Taylor series with exact remainder. For scalar X_n, the expansion is

$$g(X_n) = g(W_n) + g^{(1)}(W_n)(X_n - W_n) + \cdots$$
$$+ [(s-1)!]^{-1} g^{(s-1)}(W_n)(X_n - W_n)^{s-1} + (s!)^{-1} g^{(s)}(W_n^\dagger)(X_n - W_n)^s ,$$

where $W_n^\dagger$ is on the line segment joining X_n and W_n, and for scalar arguments, $g^{(j)}(W_n)$ denotes the jth derivative. Now $g^{(s)}(W_n^\dagger)$ is bounded on B_ϵ, and $X_n - W_n = O_p(r_n)$. It follows that

$$g^{(s)}(W_n^\dagger)(X_n - W_n)^s = O_p(r_n^s) .$$

The arguments are analogous for the vector case, and the result is established. ▲

Example 5.1.1. Let $W_n = W$, where $W \sim N(0, 1)$, and let

$$X_n = W + Z_n ,$$

where $Z_n \sim U(-n^{-1/2}, n^{-1/2})$, independent of W. Let $g(x) = x^{-2}$. Let $1 > \epsilon > 0$ be given, and set

$$B_\epsilon = \{x: |x| > 2^{-1}\epsilon\} ,$$
$$B_{2\epsilon} = \{x: |x| > \epsilon\} .$$

The ordinate of the standard normal distribution is about 0.40 at zero. Therefore, $P\{W \in B_\epsilon\} \geq 1 - 0.4\epsilon$ and $P\{W \in B_{2\epsilon}\} \geq 1 - 0.8\epsilon$. If $W \in B_{2\epsilon}$ and $n > 4\epsilon^{-2}$, then $X_n \in B_\epsilon$. Hence, if $n > 4\epsilon^{-2}$ the probability is greater than $1 - \epsilon$ that $l(X_n, W_n)$ is in B_ϵ. The function x^{-2} is continuous with continuous derivatives on B_ϵ. The absolute value of the first derivative and of the second derivative are bounded by finite multiples of ϵ^{-3} and ϵ^{-4}, respectively, for $x \in B_\epsilon$. Therefore, the conditions of Theorem 5.1.7 are met and

$$X_n^{-2} = W^{-2} - 2W^{-3}(X_n - W) + O_p(n^{-1})$$
$$= W^{-2} + O_p(n^{-1/2}) .$$

Because W^2 is a chi-square random variable, $g(X_n)$ converges in probability to the reciprocal of a chi-square random variable, by Theorem 5.1.4. Theorem 5.1.7 enables us to establish the order of the remainder in the approximation. ▲▲

5.2. CONVERGENCE IN DISTRIBUTION

In the preceding section we discussed conditions under which a sequence of random variables converges in probability to a limit random variable. A second type of convergence important in statistics is the convergence of a sequence of distribution functions to a limit function. The classical example of such convergence is given by the central limit theorem wherein the sequence of distribution functions converges pointwise to the normal distribution function.

Definition 5.2.1. If $\{X_n\}$ is a sequence of random variables with distribution functions $\{F_{X_n}(x)\}$, then $\{X_n\}$ is said to converge in distribution (or in law) to the random variable X with distribution function $F_X(x)$, and we write $X_n \xrightarrow{\mathscr{L}} X$, if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all x for which $F_X(x)$ is continuous.

Note that the sequence of distribution functions is converging to a function that is itself a distribution function. Some authors define this type of convergence by saying the sequence $\{F_{X_n}(x)\}$ converges *completely* to $F_X(x)$. Thus our symbolism

$$X_n \xrightarrow{\mathscr{L}} X$$

is understood to mean that $F_{X_n}(x)$ converges to the distribution function of the random variable X. The notation

$$F_{X_n}(x) \xrightarrow{c} F_X(x)$$

is also used.

Theorem 5.2.1. Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables such that

$$p\lim(X_n - Y_n) = 0 .$$

If there exists a random variable X such that

$$X_n \xrightarrow{\mathscr{L}} X ,$$

then

$$Y_n \xrightarrow{\mathscr{L}} X .$$

Proof. Let W and Z be random variables with distribution functions $F_W(w)$ and $F_Z(z)$, respectively, and fix $\epsilon > 0$ and $\delta > 0$. We first show that

$$P\{|Z - W| > \epsilon\} \leq \delta$$

implies that

$$F_Z(z - \epsilon) - \delta \leq F_W(z) \leq F_Z(z + \epsilon) + \delta$$

for all z. This result holds because

$$\begin{aligned} F_Z(z-\epsilon) - F_W(z) &= P\{Z \le z-\epsilon\} - P\{W \le z\} \\ &\le P\{Z \le z-\epsilon\} - P\{Z \le z-\epsilon \text{ and } W \le z\} \\ &= P\{Z \le z-\epsilon \text{ and } W > z\} \\ &\le P\{(W-Z) > \epsilon\} \\ &\le P\{|W-Z| > \epsilon\} \le \delta\,, \end{aligned}$$

and, in a similar manner,

$$\begin{aligned} F_W(z) - F_Z(z+\epsilon) &= P\{W \le z\} - P\{Z \le z+\epsilon\} \\ &\le P\{W \le z\} - P\{Z \le z+\epsilon \text{ and } W \le z\} \\ &= P\{W \le z \text{ and } Z > z+\epsilon\} \\ &\le P\{|Z-W| > \epsilon\} \le \delta\,. \end{aligned}$$

Let x_0 be a continuity point for $F_X(x)$. Then, given $\delta > 0$, there is an $\eta > 0$ such that $|F_X(x) - F_X(x_0)| < \frac{1}{4}\delta$ for $|x - x_0| \le \eta$ and $F_X(x)$ is continuous at $x_0 - \eta$ and at $x_0 + \eta$. Furthermore, for this δ and η, there is an N_1 such that, for $n > N_1$,

$$P\{|X_n - Y_n| > \eta\} < \tfrac{1}{2}\delta$$

and therefore

$$F_{X_n}(x-\eta) - \tfrac{1}{2}\delta \le F_{Y_n}(x) \le F_{X_n}(x+\eta) + \tfrac{1}{2}\delta$$

for all x. Also, there is an N_2 such that, for $n > N_2$,

$$|F_{X_n}(x_0 - \eta) - F_X(x_0 - \eta)| < \tfrac{1}{4}\delta$$

and

$$|F_{X_n}(x_0 + \eta) - F_X(x_0 + \eta)| < \tfrac{1}{4}\delta\,.$$

Therefore, given the continuity point x_0 and $\delta > 0$, there is an $\eta > 0$ and an $N = \max(N_1, N_2)$ such that for $n > N$,

$$\begin{aligned} F_X(x_0) - \delta &< F_X(x_0 - \eta) - \tfrac{3}{4}\delta < F_{X_n}(x_0 - \eta) - \tfrac{1}{2}\delta \\ &\le F_{Y_n}(x_0) \le F_{X_n}(x_0 + \eta) + \tfrac{1}{2}\delta < F_X(x_0 + \eta) + \tfrac{3}{4}\delta \\ &< F_X(x_0) + \delta\,. \end{aligned}$$

▲

As a corollary we have the result that convergence in probability implies convergence in law.

Corollary 5.2.1.1. Let $\{X_n\}$ be a sequence of random variables. If there exists a random variable X such that $p\lim X_n = X$, then $X_n \xrightarrow{\mathscr{L}} X$.

Corollary 5.2.1.2. Let $\{X_n\}$ and X be random variables such that $p\lim X_n = X$. If $g(x)$ is a continuous function, then the distribution of $g(X_n)$ converges to the distribution of $g(X)$.

Proof. This follows immediately, since by Theorem 5.1.4,

$$p\lim g(X_n) = g(X) . \qquad \blacktriangle$$

We state the following two important theorems without proof.

Theorem 5.2.2 (Helly–Bray). If $\{F_n(\mathbf{x})\}$ is a sequence of distribution functions over k-dimensional Euclidean space $\mathscr{R}^{(k)}$ such that $F_n(\mathbf{x}) \xrightarrow{C} F(\mathbf{x})$, then

$$\int g(\mathbf{x})\, dF_n(\mathbf{x}) \rightarrow \int g(\mathbf{x})\, dF(\mathbf{x}) \qquad \text{as} \quad n \rightarrow \infty$$

for every bounded continuous function $g(\mathbf{x})$.

Theorem 5.2.3. Let $\{F_n(\mathbf{x})\}$ be a sequence of distribution functions over $\mathscr{R}^{(k)}$ with corresponding characteristic functions $\{\varphi_n(\mathbf{u})\}$.

(i) If $F_n(\mathbf{x}) \xrightarrow{C} F(\mathbf{x})$, then $\varphi_n(\mathbf{u}) \rightarrow \varphi(\mathbf{u})$ at all $\mathbf{u} \in \mathscr{R}^{(k)}$, where $\varphi(\mathbf{u})$ is the characteristic function associated with $F(\mathbf{x})$.

(ii) Continuity theorem. If $\varphi_n(\mathbf{u})$ converges pointwise to a function $\varphi(\mathbf{u})$ that is continuous at $(0, 0, \ldots, 0) \in \mathscr{R}^{(k)}$, then $\varphi(\mathbf{u})$ is the characteristic function of a distribution function $F(\mathbf{x})$ and $F_n(\mathbf{x}) \xrightarrow{C} F(\mathbf{x})$.

Theorem 5.2.4. Let $\{\mathbf{X}_n\}$ be a sequence of k-dimensional random variables with distribution functions $\{F_{\mathbf{X}_n}(\mathbf{x})\}$ such that $F_{\mathbf{X}_n}(\mathbf{x}) \xrightarrow{C} F_{\mathbf{X}}(\mathbf{x})$, and let T be a continuous mapping from $\mathscr{R}^{(k)}$ to $\mathscr{R}^{(p)}$. Then

$$F_{T(\mathbf{X}_n)}(\mathbf{y}) \xrightarrow{C} F_{T(\mathbf{X})}(\mathbf{y}) .$$

Proof. By the Helly–Bray theorem, the characteristic function of $T(\mathbf{X}_n)$ converges to the characteristic function of $T(\mathbf{X})$, and the result follows. $\blacktriangle$

Theorem 5.2.5. Let $\{\mathbf{X}_n\}$ and $\{\mathbf{Y}_n\}$ be two sequences of k-dimensional random variables such that $\mathbf{X}_n$ is independent of $\mathbf{Y}_n$ for all n. If there exist random variables $\mathbf{X}$ and $\mathbf{Y}$ such that $F_{\mathbf{X}_n}(\mathbf{x}) \xrightarrow{C} F_{\mathbf{X}}(\mathbf{x})$ and $F_{\mathbf{Y}_n}(\mathbf{y}) \xrightarrow{C} F_{\mathbf{Y}}(\mathbf{y})$, then

$$F_{\mathbf{X}_n \mathbf{Y}_n}(\mathbf{x}, \mathbf{y}) \xrightarrow{C} F_{\mathbf{X}}(\mathbf{x}) F_{\mathbf{Y}}(\mathbf{y}) .$$

Proof. The characteristic function of $(\mathbf{X}_n', \mathbf{Y}_n')'$ is given by

$$\int\int e^{i\mathbf{u}'\mathbf{x}} e^{i\mathbf{v}'\mathbf{y}}\, dF_{\mathbf{X}_n\mathbf{Y}_n}(\mathbf{x}, \mathbf{y}) = \int e^{i\mathbf{u}'\mathbf{x}}\, dF_{\mathbf{X}_n}(\mathbf{x}) \int e^{i\mathbf{v}'\mathbf{y}}\, dF_{\mathbf{Y}_n}(\mathbf{y})$$
$$= \varphi_{\mathbf{X}_n}(\mathbf{u})\varphi_{\mathbf{Y}_n}(\mathbf{v}) .$$

Now, by the Helly–Bray theorem, $\varphi_{\mathbf{X}_n}(\mathbf{u}) \rightarrow \varphi_{\mathbf{X}}(\mathbf{u})$ and $\varphi_{\mathbf{Y}_n}(\mathbf{v}) \rightarrow \varphi_{\mathbf{Y}}(\mathbf{v})$. Therefore,

$$\varphi_{\mathbf{XY}}(\mathbf{u}, \mathbf{v}) = \lim_{n\rightarrow\infty} \varphi_{\mathbf{X}_n}(\mathbf{u})\varphi_{\mathbf{Y}_n}(\mathbf{v}) = \varphi_{\mathbf{X}}(\mathbf{u})\varphi_{y}(\mathbf{v}) .$$

By the continuity theorem, this implies that $F_{\mathbf{X}_n\mathbf{Y}_n}(\mathbf{x}, \mathbf{y}) \xrightarrow{C} F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$, where $F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$ is the distribution function of independent random variables associated with the characteristic function $\varphi_{\mathbf{XY}}(\mathbf{u}, \mathbf{v})$. ▲

From Corollary 5.2.1.1, we know that convergence in probability implies convergence in law. For the special case wherein a sequence of random variables converges in law to a constant random variable, the converse is also true.

Lemma 5.2.1. Let $\{\mathbf{Y}_n\}$ be a sequence of p-dimensional random variables with corresponding distribution functions $\{F_{\mathbf{Y}_n}(\mathbf{y})\}$. Let $\mathbf{Y}$ be a p-dimensional random variable with distribution function $F_{\mathbf{Y}}(\mathbf{y})$ such that $P\{\mathbf{Y} = \mathbf{b}\} = 1$, $\mathbf{b}$ is a constant vector, and

$$F_{\mathbf{Y}_n}(\mathbf{y}) \xrightarrow{\ C\ } F_{\mathbf{Y}}(\mathbf{y}) .$$

Then, given $\epsilon > 0$, there exists an N such that, for $n > N$,

$$P\{|\mathbf{Y}_n - \mathbf{b}| \geq \epsilon\} < \epsilon .$$

Proof. Let $B = \{\mathbf{y}\colon\ y_1 > b_1 - \epsilon/p,\ y_2 > b_2 - \epsilon/p, \ldots, y_p > b_p - \epsilon/p\}$. Then $F_{\mathbf{Y}}(\mathbf{y}) = 0$ on the complement of B. Fix $\epsilon > 0$. As $F_{\mathbf{Y}_n}(\mathbf{y}) \xrightarrow{C} F_{\mathbf{Y}}(\mathbf{y})$, there exists an N_0 such that, for $n > N_0$,

$$F_{\mathbf{Y}_n}(\mathbf{g}_1) = F_{\mathbf{Y}_n}\left(b_1 - \frac{\epsilon}{p}, b_2 + \frac{\epsilon}{p}, \ldots, b_p + \frac{\epsilon}{p}\right) < \frac{\epsilon}{2p} ,$$
$$F_{\mathbf{Y}_n}(\mathbf{g}_2) = F_{\mathbf{Y}_n}\left(b_1 + \frac{\epsilon}{p}, b_2 - \frac{\epsilon}{p}, \ldots, b_p + \frac{\epsilon}{p}\right) < \frac{\epsilon}{2p} ,$$
$$\vdots$$
$$F_{\mathbf{Y}_n}(\mathbf{g}_p) = F_{\mathbf{Y}_n}\left(b_1 + \frac{\epsilon}{p}, b_2 + \frac{\epsilon}{p}, \ldots, b_p - \frac{\epsilon}{p}\right) < \frac{\epsilon}{2p} .$$

There also exists an N_1 such that, for $n > N_1$, $1 - F_{\mathbf{Y}_n}(b_1 + \epsilon/p, b_2 + \epsilon/p, \ldots, b_p +$

$\epsilon/p) < \epsilon/2$. Therefore, for $n > \max(N_0, N_1)$,

$$P\left\{b_1 - \frac{\epsilon}{p} \le y_1 \le b_1 + \frac{\epsilon}{p}, b_2 - \frac{\epsilon}{p} \le y_2 \le b_2 + \frac{\epsilon}{p}, \ldots, b_p - \frac{\epsilon}{p} \le y_p \le b_p + \frac{\epsilon}{p}\right\}$$

$$\ge F_{\mathbf{Y}_n}\left(b_1 + \frac{\epsilon}{p}, b_2 + \frac{\epsilon}{p}, \ldots, b_p + \frac{\epsilon}{p}\right) - \sum_{i=1}^{p} F_{\mathbf{Y}_n}(\mathbf{g}_i) \ge 1 - \epsilon,$$

and it follows that

$$P\{|\mathbf{Y}_n - \mathbf{b}| \ge \epsilon\} < \epsilon. \qquad \blacktriangle$$

Theorem 5.2.6. Let $\{(\mathbf{X}_n', \mathbf{Y}_n')'\}$ be a sequence of $(k+p)$-dimensional random variables where $\mathbf{X}_n$ is k-dimensional. Let the sequence of joint distribution functions be denoted by $\{F_{\mathbf{X}_n\mathbf{Y}_n}(\mathbf{x}, \mathbf{y})\}$ and the sequences of marginal distribution functions by $\{F_{\mathbf{X}_n}(\mathbf{x})\}$ and $\{F_{\mathbf{Y}_n}(\mathbf{y})\}$. If there exists a k-dimensional random variable $\mathbf{X}$ and a p-dimensional random variable $\mathbf{Y}$ such that $F_{\mathbf{X}_n}(\mathbf{x}) \xrightarrow{C} F_{\mathbf{X}}(\mathbf{x})$ and $F_{\mathbf{Y}_n}(\mathbf{y}) \xrightarrow{C} F_{\mathbf{Y}}(\mathbf{y})$, where $P\{\mathbf{Y} = \mathbf{b}\} = 1$ and $\mathbf{b} = (b_1, b_2, \ldots, b_p)'$ is a constant vector, then

$$F_{\mathbf{X}_n\mathbf{Y}_n}(\mathbf{x}, \mathbf{y}) \xrightarrow{C} F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}).$$

Proof. Now $P\{\mathbf{Y} = \mathbf{b}\} = 1$ implies that

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{XY}}(\mathbf{x}, \mathbf{b}),$$

that $F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = 0$ if any element of $\mathbf{y}$ is less than the corresponding element of $\mathbf{b}$, and that $F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = F_{\mathbf{X}}(\mathbf{x})$ if every element of $\mathbf{y}$ is greater than or equal to the corresponding element of $\mathbf{b}$. Fix $\epsilon > 0$, and consider a point $(\mathbf{x}_0, \mathbf{y}_0)$ where at least one element of $\mathbf{y}_0$ is less than the corresponding element of $\mathbf{b}$ by an amount ϵ. Then $F_{\mathbf{XY}}(\mathbf{x}_0, \mathbf{y}_0) = 0$. However, there is an N_0 such that, for $n > N_0$,

$$F_{\mathbf{X}_n\mathbf{Y}_n}(\mathbf{x}_0, \mathbf{y}_0) < \frac{\epsilon}{2},$$

by Lemma 5.2.1. Let $(\mathbf{x}_0, \mathbf{y}_1)$ be a continuity point of $F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$ where every element of $\mathbf{y}_1$ exceeds the corresponding element of $\mathbf{b}$ by $\epsilon/p^{1/2} > 0$. Because $F_{\mathbf{X}_n}(\mathbf{x}) \xrightarrow{C} F_{\mathbf{X}}(\mathbf{x})$ and $F_{\mathbf{Y}_n}(\mathbf{y}) \xrightarrow{C} F_{\mathbf{Y}}(\mathbf{y})$, we can choose N_1 such that, for $n \ge N_1$, $|F_{\mathbf{X}_n}(\mathbf{x}_0) - F_{\mathbf{X}}(\mathbf{x}_0)| < \epsilon/2$ and $|F_{\mathbf{Y}_n}(\mathbf{y}_1) - F_{\mathbf{Y}}(\mathbf{y}_1)| < \epsilon/2$. Hence,

$$\begin{aligned} |F_{\mathbf{X}_n\mathbf{Y}_n}(\mathbf{x}_0, \mathbf{y}_1) - F_{\mathbf{XY}}(\mathbf{x}_0, \mathbf{y}_1)| &\le |F_{\mathbf{X}_n\mathbf{Y}_n}(\mathbf{x}_0, \mathbf{y}_1) - F_{\mathbf{X}_n}(\mathbf{x}_0)| \\ &\quad + |F_{\mathbf{X}_n}(\mathbf{x}_0) - F_{\mathbf{X}}(\mathbf{x}_0)| \\ &< 1 - F_{\mathbf{Y}_n}(\mathbf{y}_1) + \frac{\epsilon}{2} < \epsilon. \end{aligned} \qquad \blacktriangle$$

Utilizing Theorems 5.2.4 and 5.2.6, we obtain the following corollary.

Corollary 5.2.6.1. Let $\{\mathbf{X}_n\}$ and $\{\mathbf{Y}_n\}$ be two sequences of k-dimensional random variables. If there exists a k-dimensional random variable $\mathbf{Y}$ and a fixed vector $\mathbf{b}$ such that $\mathbf{Y}_n \xrightarrow{\mathscr{L}} \mathbf{Y}$ and $\mathbf{X}_n \xrightarrow{\mathscr{L}} \mathbf{b}$, then

(i) $\mathbf{X}_n + \mathbf{Y}_n \xrightarrow{\mathscr{L}} \mathbf{b} + \mathbf{Y}$,
(ii) $\mathbf{X}_n'\mathbf{Y}_n \xrightarrow{\mathscr{L}} \mathbf{b}'\mathbf{Y}$.

Corollary 5.2.6.2. Let $\{\mathbf{Y}_n\}$ be a sequence of k-dimensional random variables, and let $\{\mathbf{A}_n\}$ be a sequence of $k \times k$ random matrices. If there exists a random vector $\mathbf{Y}$ and a fixed nonsingular matrix $\mathbf{A}$ such that $\mathbf{Y}_n \xrightarrow{\mathscr{L}} \mathbf{Y}$, $\mathbf{A}_n \xrightarrow{\mathscr{L}} \mathbf{A}$, then

$$\mathbf{A}_n^{-1}\mathbf{Y}_n \xrightarrow{\mathscr{L}} \mathbf{A}^{-1}\mathbf{Y}.$$

5.3. CENTRAL LIMIT THEOREMS

The exact distributions of many statistics encountered in practice have not been obtained. Fortunately, many statistics in the class of continuous functions of means or of sample moments converge in distribution to normal random variables. We give without proof the following central limit theorem.

Theorem 5.3.1 (Lindeberg central limit theorem). Let $\{Z_t: t = 1, 2, \ldots\}$ be a sequence of independent random variables with distribution functions $\{F_t(z)\}$. Let $E\{Z_t\} = \mu_t$, $E\{(Z_t - \mu_t)^2\} = \sigma_t^2$, and assume

$$\lim_{n\to\infty} V_n^{-1} \sum_{t=1}^{n} \int_{|z-\mu_t|>\epsilon V_n^{1/2}} (z - \mu_t)^2 \, dF_t(z) = 0 \tag{5.3.1}$$

for all $\epsilon > 0$, where $V_n = \Sigma_{t=1}^n \sigma_t^2$. Then,

$$V_n^{-1/2} \sum_{t=1}^{n} (Z_t - \mu_t) \xrightarrow{\mathscr{L}} N(0, 1),$$

where $N(0, 1)$ denotes the normal distribution with mean zero and variance one.

A form of the central limit theorem whose conditions are often more easily verified is the following theorem.

Theorem 5.3.2 (Liapounov central limit theorem). Let $\{Z_t: t = 1, 2, \ldots\}$ be a sequence of independent random variables with distribution functions $\{F_t(z)\}$. Let

$E\{Z_t\} = \mu_t$, $E\{(Z_t - \mu_t)^2\} = \sigma_t^2$, and $V_n = \Sigma_{t=1}^n \sigma_t^2$. If

$$\lim_{n\to\infty} \frac{\Sigma_{t=1}^n \int |z - \mu_t|^{2+\delta}\, dF_t(z)}{V_n^{1+\delta/2}} = 0$$

for some $\delta > 0$, then

$$V_n^{-1/2} \sum_{t=1}^n (Z_t - \mu_t) \xrightarrow{\mathscr{L}} N(0, 1)\,.$$

Proof. Let $\delta > 0$ and $\epsilon > 0$, and define the set A_ϵ by

$$A_\epsilon = \{z : |z - \mu_t| > \epsilon V_n^{1/2}\}\,.$$

Then

$$\begin{aligned} \frac{1}{V_n} \sum_{t=1}^n \int_{A_\epsilon} (z - \mu_t)^2\, dF_t(z) &\le \frac{1}{V_n(\epsilon V_n^{1/2})^\delta} \sum_{t=1}^n \int_{A_\epsilon} |z - \mu_t|^{2+\delta}\, dF_t(z) \\ &\le \frac{1}{V_n^{1+\delta/2}\epsilon^\delta} \sum_{t=1}^n \int |z - \mu_t|^{2+\delta}\, dF_t(z)\,, \end{aligned}$$

which, by hypothesis, goes to zero as $n \to \infty$. Therefore, the condition on the $2 + \delta$ moment implies the condition of the Lindeberg theorem. ▲

The reader is referred to the texts of Tucker (1967), Gnedenko (1967), and Loève (1963) for discussions of these theorems.

For a proof of the following extension of the central limit theorems to the multivariate case, see Varadarajan (1958).

Theorem 5.3.3. Let $\{\mathbf{Z}_n : n = 1, 2, \ldots\}$ be a sequence of k-dimensional random variables with distribution functions $\{F_{\mathbf{Z}_n}(\mathbf{z})\}$. Let $F_{\boldsymbol{\lambda},n}(x)$ be the distribution function of $X_n = \boldsymbol{\lambda}'\mathbf{Z}_n$, where $\boldsymbol{\lambda}$ is a fixed vector. A necessary and sufficient condition for $F_{\mathbf{Z}_n}(\mathbf{z})$ to converge to the k-variate distribution function $F(\mathbf{z})$ is that $F_{\boldsymbol{\lambda},n}(\mathbf{x})$ converges to a limit for each $\boldsymbol{\lambda}$.

In most of our applications of Theorem 5.3.3, each $F_{\boldsymbol{\lambda},n}(\mathbf{x})$ will be converging to a normal distribution function and hence the vector random variable $\mathbf{Z}_n$ will converge in distribution to a multivariate normal.

The Lindeberg and Liapounov central limit theorems are for independent random variables. It is possible to obtain the limiting normal distribution for sequences that satisfy weaker conditions. One type of sequence that has been studied is the *martingale* process.

Definition 5.3.1. Let $\{X_t\}_{t=1}^\infty$ be a sequence of random variables defined on the space $\{\Omega, \mathscr{A}, P\}$. The sequence is a martingale if, for all t,

$$E\{|X_t|\} < \infty$$

and

$$E\{X_t \mid \mathscr{A}_{t-1}\} = X_{t-1} \qquad \text{a.s.},$$

where $\mathscr{A}_t$ is the sigma-field generated by $\{X_1, X_2, \ldots, X_t\}$.

It is an immediate consequence of the definition that for $s \leq t-1$,

$$E\{X_t \mid \mathscr{A}_S\} = X_S \qquad \text{a.s.}$$

Definition 5.3.2. Let $\{X_t\}_{t=1}^{\infty}$ be a martingale sequence defined on $\{\Omega, \mathscr{A}, P\}$. Let $Z_t = X_t - X_{t-1}$. Then the sequence $\{Z_t\}_{t=1}^{\infty}$ is called a sequence of *martingale differences*.

We give a central limit theorem for martingale differences. Theorem 5.3.4 is due to Brown (1971). Related results have been obtained by Dvoretsky (1972), Scott (1973), and McLeish (1974). Also see Hall and Heyde (1980) and Pollard (1984).

Theorem 5.3.4. Let $\{Z_{tn}: 1 \leq t \leq n, n \geq 1\}$ denote a triangular array of random variables defined on the probability space $(\Omega, \mathscr{A}, P)$, and let $\{\mathscr{A}_{tn}: 0 \leq t \leq n, n \geq 1\}$ be any triangular array of sub-sigma-fields of $\mathscr{A}$ such that for each n and $1 \leq t \leq n$, Z_{tn} is $\mathscr{A}_{tn}$-measurable and $\mathscr{A}_{t-1,n}$ is contained in $\mathscr{A}_{tn}$. For $1 \leq k \leq n$, $1 \leq j \leq n$, and $n \geq 1$, let

$$S_{kn} = \sum_{t=1}^{k} Z_{tn},$$
$$\delta_{tn}^2 = E\{Z_{tn}^2 \mid \mathscr{A}_{t-1,n}\},$$
$$V_{jn}^2 = \sum_{t=1}^{j} \delta_{tn}^2,$$

and

$$s_{nn}^2 = E\{V_{nn}^2\}.$$

Assume

(i) $E(Z_{tn} \mid \mathscr{A}_{t-1,n}) = 0 \qquad$ a.s. for $1 \leq t \leq n$,

(ii) $V_{nn}^2 s_{nn}^{-2} \xrightarrow{P} 1$,

(iii) $\lim_{n\to\infty} s_{nn}^{-2} \sum_{j=1}^{n} E[Z_{jn}^2 I(|Z_{jn}| \geq \epsilon s_{nn}) \mid \mathscr{A}_{t-1,n}] = 0$ for all $\epsilon > 0$, where $I(A)$ denotes the indicator function of a set A. Then, as $n \to \infty$,

$$s_{nn}^{-1} S_{nn} \xrightarrow{\mathscr{L}} N(0, 1).$$

Proof. Omitted. ▲

Corollary 5.3.4. Let $\{e_{tn}\}$ be a triangular array satisfying

$$E\{(e_{tn}, e_{tn}^2) | \mathscr{A}_{t-1,n}\} = (0, \sigma^2) \qquad \text{a.s.},$$

$$E\{|e_{tn}|^{2+\delta} | \mathscr{A}_{t-1,n}\} < M < \infty \qquad \text{a.s.}$$

for some $\delta > 0$, where $\mathscr{A}_{t,n}$ is the sigma-field generated by $\{e_{jn}: j \le t\}$. Let $\{w_{tn}: 1 \le t \le n, n \ge 1\}$ be a triangular array of constants, $\sum_{t=1}^{n} w_{tn}^2 \neq 0$ for all n, satisfying

$$\lim_{n\to\infty} \sup_{1 \le j \le n} \left(\sum_{t=1}^{n} w_{tn}^2 \right)^{-1} w_{jn}^2 = 0.$$

Then

$$\left(\sum_{t=1}^{n} w_{tn}^2 \right)^{-1/2} \sum_{t=1}^{n} w_{tn} e_{tn} \xrightarrow{\mathscr{L}} N(0, \sigma^2).$$

Proof. Reserved for the reader. ▲

We now give a functional central limit theorem which is an extension of Theorem 5.3.4. For related results, we refer the reader to Billingsley (1968). In the following, we let $D = D[0, 1]$ denote the space of functions $f(u)$ on $[0, 1]$ that are right continuous and have left limits. Let $(\Omega, \mathscr{A}, P)$ be a probability space, and let $X(u, \omega)$ be a random function on D defined by an element ω of Ω. We will often abbreviate $X(u, \omega)$ to $X(u)$ or to X.

An important random function is the *Wiener process*. If W is a standard Wiener process, then $W(u)$ is normally distributed for every u in $[0, 1]$, where $W(u) \sim N(0, u)$. Furthermore, $W(u)$ has normal independent increments. A process on $[0, 1]$ has independent increments if, for

$$0 \le u_1 \le u_2 \le \cdots \le u_k \le 1,$$

the random variables

$$W(u_2) - W(u_1), \qquad W(u_3) - W(u_2), \ldots, \qquad W(u_k) - W(u_{k-1})$$

are independent. The weak convergence or convergence in distribution of a sequence of random elements X_n in D to a random element X in D will be denoted by $X_n \Rightarrow X$ or by $X_n \xrightarrow{\mathscr{L}} X$.

Theorem 5.3.5. Let $\{e_t\}_{t=1}^{\infty}$ be a sequence of random variables, and let $\mathscr{A}_{t-1}$

be the sigma-field generated by $\{e_1, e_2, \ldots, e_{t-1}\}$. Assume that

$$E\{(e_t, e_t^2) \mid \mathscr{A}_{t-1}\} = (0, \sigma^2) \qquad \text{a.s.}$$

and

$$E[|e_t|^{2+\delta} | \mathscr{A}_{t-1}] < M < \infty \qquad \text{a.s.}$$

for some $\delta > 0$. Let

$$S_n(u) = \sigma^{-1} n^{-1/2} \sum_{j=1}^{[nu]} e_j\,, \qquad 0 \leq u \leq 1\,,$$

where $[nu]$ denotes the integer part of nu and the sum is zero if $[nu] = 0$. Then

$$S_n \Rightarrow W\,,$$

where W is a standard Wiener process.

Proof. See Billinglsey (1968) or Chan and Wei (1988). ▲

The conclusion of Theorem 5.3.5 holds for e_t that are iid$(0, \sigma^2)$ random variables. That form of the theorem is known as Donsker's theorem. See Billingsley (1968, p. 68).

We now give a result for the limiting distribution of continuous functionals of S_n. The result follows from the continuous mapping theorem. See Theorem 5.1 of Billingsley (1968).

Theorem 5.3.6. Let S_n and S be random elements in D. For $0 \leq s \leq 1$, let

$$Z_{in}(s) = \int_0^s u^{k_i} f_i[S_n(u)]\, du$$

and

$$Z_i(s) = \int_0^s u^{k_i} f_i[S(u)]\, du\,, \qquad i = 1, \ldots, m\,,$$

where $f_1, \ldots, f_m$ are real valued continuous functions on the real line and $k_i \geq 0$. If $S_n \Rightarrow S$, then

$$(S_n, Z_{1n}, \ldots, Z_{mn}) \Rightarrow (S, Z_1, \ldots, Z_m)\,.$$

Proof. Omitted. ▲

The following corollary is useful in deriving the limiting distribution of estimators of the parameters of nonstationary autoregressive processes.

Corollary 5.3.6. Let $\{e_t\}$ be a sequence satisfying the conditions of Theorem 5.3.5. Define

$$Y_t = Y_{t-1} + e_t = \sum_{i=1}^{t} e_i ,$$

for $t \geq 1$ with $Y_0 = 0$. Then

$$\left(n^{-1/2}Y_n,\ n^{-3/2}\sum_{t=2}^{n} Y_{t-1},\ n^{-2}\sum_{t=2}^{n} Y_{t-1}^2,\ n^{-5/2}\sum_{t=2}^{n} tY_{t-1}\right)$$
$$\xrightarrow{\mathscr{L}} \left(\sigma W(1),\ \sigma\int_0^1 W(t)\,dt,\ \sigma^2\int_0^1 W^2(t)\,dt,\ \sigma\int_0^1 tW(t)\,dt\right).$$

Proof. From Theorem 5.3.5, we have that

$$S_n(u) = n^{-1/2}\sum_{j=1}^{[nu]} e_j$$

converges weakly to $\sigma W(u)$, where $W(u)$ is the Wiener process. If we let

$$f_1(x) = f_3(x) = x$$

and $f_2(x) = x^2$, then

$$n^{-3/2}\sum_{t=2}^{n} Y_{t-1} = \int_0^1 f_1[S_n(u)]\,du ,$$
$$n^{-2}\sum_{t=2}^{n} Y_{t-1}^2 = \int_0^1 f_2[S_n(u)]\,du ,$$

and

$$n^{-5/2}\sum_{t=2}^{n} tY_{t-1} = \int_0^1 uf_3[S_n(u)]\,du ,$$

The result follows from Theorem 5.3.6. ▲

Theorem 5.3.7 is the vector version of the functional central limit theorem. We define a standard vector Wiener process of dimension k to be a process such that the elements of $\mathbf{W}(u)$ are independent standard Wiener processes.

Theorem 5.3.7. Let $\{\mathbf{e}_t\}_{t=1}^{\infty}$ be a sequence of random vectors, and let $\mathscr{A}_{t-1}$ be the sigma-field generated by $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{t-1}\}$. Assume

$$E\{(\mathbf{e}_t, \mathbf{e}_t\mathbf{e}_t') \mid \mathscr{A}_{t-1}\} = (\mathbf{0}, \mathbf{\Sigma}_{ee}) \qquad \text{a.s.}$$

and

$$E\{|\mathbf{e}_t|^{2+\delta}|\mathscr{A}_{t-1}\}<M<\infty \qquad \text{a.s.}$$

for some $\delta > 0$, or assume the e_t are iid$(0, \mathbf{\Sigma}_{ee})$, where $\mathbf{\Sigma}_{ee}$ is positive definite. Let

$$\mathbf{Y}_t = \sum_{j=1}^{t} \mathbf{e}_j$$

and

$$\mathbf{S}_n(u) = n^{-1/2}\mathbf{\Sigma}_{ee}^{-1/2}\mathbf{Y}_{[nu]}, \qquad 0 \le u \le 1,$$

where $[nu]$ is the integer part of nu, and $\mathbf{\Sigma}_{ee}^{1/2}$ is the symmetric square root of $\mathbf{\Sigma}_{ee}$. Then

$$\mathbf{S}_n(u) \Rightarrow \mathbf{W},$$

where $\mathbf{W}$ is a standard vector Wiener process. Also,

$$n^{-3/2}\mathbf{\Sigma}_{ee}^{-1/2}\sum_{t=1}^{n}\mathbf{Y}_t \Rightarrow \int_0^1 \mathbf{W}(u)\,du \stackrel{\text{def}}{=} \boldsymbol{\zeta},$$

$$n^{-2}\mathbf{\Sigma}_{ee}^{-1/2}\sum_{t=1}^{n}\mathbf{Y}_t\mathbf{Y}_t'\mathbf{\Sigma}_{ee}^{-1/2} \Rightarrow \int_0^1 \mathbf{W}(u)\mathbf{W}'(u)\,du \stackrel{\text{def}}{=} \mathbf{G},$$

$$n^{-2}\mathbf{\Sigma}_{ee}^{-1/2}\sum_{t=1}^{n}(\mathbf{Y}_t-\bar{\mathbf{y}})(\mathbf{Y}_t-\bar{\mathbf{y}})'\mathbf{\Sigma}_{ee}^{-1/2} \Rightarrow \mathbf{G}-\boldsymbol{\zeta}\boldsymbol{\zeta}',$$

$$n^{-1}\mathbf{\Sigma}_{ee}^{-1/2}\sum_{t=1}^{n}\mathbf{Y}_{t-1}\mathbf{e}_t'\mathbf{\Sigma}_{ee}^{-1/2} \Rightarrow \int_0^1 \mathbf{W}(u)\,d\mathbf{W}'(u) \stackrel{\text{def}}{=} \mathbf{Y},$$

and

$$n^{-1}\mathbf{\Sigma}_{ee}^{-1/2}\sum_{t=1}^{n}(\mathbf{Y}_{t-1}-\bar{\mathbf{y}})\mathbf{e}_t'\mathbf{\Sigma}_{ee}^{-1/2} = \mathbf{Y}-\boldsymbol{\zeta}\mathbf{W}'(1).$$

Proof. See Phillips and Durlauf (1986). ▲

We shall have use for a strong consistency result for martingales.

Theorem 5.3.8. Let $\{S_n = \sum_{t=1}^{n} X_t,\ \mathscr{A}_n,\ n \ge 1\}$ be a martingale, and let $\{M_n, n \ge 1\}$ be a nondecreasing sequence of positive random variables where M_n is $\mathscr{A}_{n-1}$ measurable for each n. Let

$$\lim_{n\to\infty} M_n = \infty \qquad \text{a.s.}$$

and

$$\sum_{t=1}^{\infty} M_t^{-p}E\{|X_t|^p \mid \mathscr{A}_{t-1}\} < \infty \qquad \text{a.s.}$$

for some $1 \leq p \leq 2$. Then

$$\lim_{n\to\infty} M_n^{-1} S_n = 0 \qquad \text{a.s.}$$

Proof. See Hall and Heyde (1980, p. 36). ▲

Corollary 5.3.8. Let $\{e_t\}_{t=1}^{\infty}$ be a sequence of random variables, and let $\mathscr{A}_{t-1}$ be the sigma-field generated by $\{e_1, e_2, \ldots, e_{t-1}\}$. Assume

$$E\{e_t \mid \mathscr{A}_{t-1}\} = 0\,, \quad E\{|e_t|^{1+\delta} \mid \mathscr{A}_{t-1}\} < K_F < \infty \qquad \text{a.s.,}$$

where K_F is a fixed constant and $\delta > 0$. Then

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} e_t = 0 \qquad \text{a.s.}$$

Proof. By assumption

$$\sum_{t=1}^{n} t^{-(1+\delta)} E\{|e_t|^{1+\delta} \mid \mathscr{A}_{t-1}\} < \sum_{t=1}^{n} t^{-(1+\delta)} K_F < \infty\,.$$

The conclusion follows by letting $M_t = t$ in Theorem 5.3.8. ▲

5.4. APPROXIMATING A SEQUENCE OF EXPECTATIONS

Taylor expansions and the order in probability concepts introduced in Section 5.1 are very important in investigating the limiting behavior of sample statistics. Care must be taken, however, in understanding the meaning of statements about such behavior. To illustrate, consider the sequence $\{\bar{x}_n^{-1}\}$, where $\bar{x}_n$ is the mean of n normal independent $(\mu, 1)$ random variables, $\mu \neq 0$. By Corollary 5.1.5, we may write

$$\bar{x}_n^{-1} = \mu^{-1} - \mu^{-2}(\bar{x}_n - \mu) + \mu^{-3}(\bar{x}_n - \mu)^2 - \mu^{-4}(\bar{x}_n - \mu)^3 + O_p(n^{-2}) \tag{5.4.1}$$

and

$$n^{1/2}(\bar{x}_n^{-1} - \mu^{-1}) = -n^{1/2}\mu^{-2}(\bar{x}_n - \mu) + O_p(n^{-1/2})\,.$$

It follows that

$$p\lim n^{1/2}[(\bar{x}_n^{-1} - \mu^{-1}) + \mu^{-2}(\bar{x}_n - \mu)] = 0\,.$$

Therefore, by Theorem 5.2.1, the limiting distribution of $n^{1/2}(\bar{x}_n^{-1} - \mu^{-1})$ is the same as the limiting distribution of $-n^{1/2}\mu^{-2}(\bar{x}_n - \mu)$. The distribution of $-n^{1/2}\mu^{-2}(\bar{x}_n - \mu)$ is $N(0, \mu^{-4})$ for all n, and it follows that the limiting distribution of $n^{1/2}(\bar{x}_n^{-1} - \mu^{-1})$ is $N(0, \mu^{-4})$.

On the other hand, it can be demonstrated that $E\{\bar{x}_n^{-1}\}$ exists for *no* finite n. Since the expectation of $(\bar{x}_n^{-1} - \mu^{-1})$ is not defined, it is clear that one cannot speak of the sequence of expectations $\{E\{\bar{x}_n^{-1}\}\}$, and it is incorrect to say that $E\{\bar{x}_n^{-1}\}$ converges to μ^{-1}.

The example illustrates that a random variable Y_n may converge in probability and hence in distribution to a random variable Y that possesses finite moments even though $E\{Y_n\}$ is not defined. If we know that Y_n has finite moments of order $r > 1$, we may be able to determine that the sequence of expectations differs from a given sequence by an amount of specified order. The conditions required to permit such statements are typically more stringent than those required to obtain convergence in distribution. In this section we investigate these conditions and develop approximations to the expectation of functions of mean or "meanlike" statistics. In preparation for that investigation we consider the expectations of integer powers of sample means of random variables with zero population means. Let $(\bar{x}_n, \bar{y}_n, \bar{z}_n, \bar{w}_n)'$ be a vector of sample means computed from a random sample selected from a distribution function with zero mean vector and finite fourth moments. Then

$$\begin{aligned}
E\{\bar{x}_n\bar{y}_n\bar{z}_n\} &= \frac{1}{n^3} E\left\{\sum_i \sum_j \sum_k X_i Y_j Z_k\right\} \\
&= \frac{1}{n^3} E\left\{\sum_i X_i Y_i Z_i + \sum_{i\neq j} X_i Y_j Z_j + \sum_{i\neq j} X_j Y_i Z_j \right. \\
&\quad \left. + \sum_{i\neq j} X_j Y_j Z_i + \sum_{i\neq j\neq k} X_i Y_j Z_k\right\} \\
&= \frac{1}{n^2} E\{XYZ\},
\end{aligned}$$

where $E\{XYZ\}$ is the expectation of the product of the original random variables. Similarly,

$$\begin{aligned}
E\{\bar{x}_n\bar{y}_n\bar{z}_n\bar{w}_n\} &= \frac{1}{n^4} E\left\{\sum_i \sum_j \sum_k \sum_m X_i Y_j Z_k W_m\right\} \\
&= \frac{1}{n^4} E\left\{\sum_i X_i Y_i Z_i W_i \right. \\
&\quad + \sum_{i\neq j} (X_i Y_j Z_j W_j + X_j Y_i Z_j W_j + X_j Y_j Z_i W_j + X_j Y_j Z_j W_i) \\
&\quad + \sum_{i\neq j} (X_i Y_i Z_j W_j + X_i Y_j Z_i W_j + X_i Y_j Z_j W_i) \\
&\quad + \sum_{i\neq j\neq k} (X_i Y_i Z_j W_k + X_i Y_j Z_i W_k + X_i Y_j Z_k W_i + X_j Y_i Z_i W_k
\end{aligned}$$

$$+ X_jY_iZ_kW_i + X_jY_kZ_iW_i) + \sum_{i\neq j\neq k\neq m} X_iY_jZ_kW_m\Big\}$$

$$= \frac{1}{n^3} E\{XYZW\} + \frac{n-1}{n^3}(\sigma_{xy}\sigma_{zw} + \sigma_{xz}\sigma_{yw} + \sigma_{xw}\sigma_{yz}),$$

where σ_{xy} is the covariance between X and Y, σ_{zw} is the covariance between Z and W, and so forth.

We note that the expectation of a product of either three or four means is $O(n^{-2})$. This is an example of a general result that we state as a theorem. Our proof follows closely that of Hansen, Hurwitz, and Madow (1953).

Theorem 5.4.1. Let $\bar{\mathbf{x}}_n = (\bar{x}_{1n}, \bar{x}_{2n}, \ldots, \bar{x}_{mn})'$ be the mean of a random sample of n vector random variables selected from a distribution function with mean vector zero and finite Bth moment. Consider the sequence $\{\bar{\mathbf{x}}_n\}_{n=1}^{\infty}$, and let $b_1, b_2, \ldots, b_m$ be nonnegative integers such that $B = \sum_{i=1}^{m} b_i$. Then

$$E\{\bar{x}_{1n}^{b_1}\bar{x}_{2n}^{b_2}\cdots\bar{x}_{mn}^{b_m}\} = \begin{cases} O(n^{-B/2}) & \text{if } B \text{ is even},\\ O(n^{-(B+1)/2}) & \text{if } B \text{ is odd}.\end{cases}$$

Proof. Now $E\{\bar{x}_{1n}^{b_1}\bar{x}_{2n}^{b_2}\cdots\bar{x}_{mn}^{b_m}\}$ can be expanded into a sum of terms such as

$$n^{-B}X_{1i_1}X_{1i_2}\cdots X_{1i_{b_1}}X_{2i_{b_1+1}}X_{2i_{b_1+2}}\cdots X_{2i_{b_1+b_2}}\cdots X_{mi_{B-b_m+1}}X_{mi_{B-b_m+2}}\cdots X_{mi_B}.$$

If there is a subscript matched by no other subscript, the expected value of the product is zero. (Recall that in the four-variable case this included terms of the form $X_iY_jZ_jW_j$, $X_iY_jZ_kZ_k$, and $X_iY_jZ_kZ_r$.) If every subscript agrees with at least one other subscript, we group the terms with common subscripts to obtain, say, H groups. The expected value is then the product of the H expected values. The sum contains $n(n-1)\cdots(n-H+1)$ terms for a particular configuration of H different subscripts. The order of $n^{-B}[n(n-1)\cdots(n-H+1)]$ is n^{-B+H} and will be maximized if we choose H as large as possible. If B is even, the largest H that gives a nonzero expectation is $B/2$, in which case we have $B/2$ groups, each containing two indexes. If B is odd, the largest H that gives a nonzero expectation is $(B-1)/2$, in which case we have $(B-1)/2-1$ groups of two and one group of three. ▲

The following lemma is used in the proof of the principal results of this section.

Lemma 5.4.1. Let $\{\mathbf{X}_n\}$ be a sequence of k-dimensional random variables with corresponding distribution functions $\{F_n(\mathbf{x})\}$ such that

$$\int |x_i - \mu_i|^s \, dF_n(\mathbf{x}) = O(a_n^s), \qquad i = 1, 2, \ldots, k,$$

where the integral is over $\mathscr{R}^{(k)}$, $a_n > 0$, s is a positive integer, $E\{\mathbf{X}_n\} = \boldsymbol{\mu} =$

$(\mu_1, \mu_2, \ldots, \mu_k)'$, $|\mathbf{x}| = [\sum_{i=1}^{k} x_i^2]^{1/2}$ is the Euclidean norm, and

$$\lim_{n\to\infty} a_n = 0 .$$

Then

$$\int |x_1 - \mu_1|^{p_1} |x_2 - \mu_2|^{p_2} \cdots |x_k - \mu_k|^{p_k} \, dF_n(\mathbf{x}) = O(1) ,$$

where the p_i, $i = 1, 2, \ldots, k$, are nonnegative real numbers satisfying

$$\sum_{i=1}^{k} p_i \leq s .$$

Proof. Without loss of generality, set all $\mu_i = 0$. Define $A = [-1, 1]$, and let $I_A(x)$ be the indicator function with value one for $x \in A$ and zero for $x \notin A$. Then, for $0 \leq q \leq s$,

$$|x_i|^q \leq I_A(x_i) + |x_i|^s ,$$

so that

$$\int |x_i|^q \, dF_n(\mathbf{x}) \leq \int [I_A(x_i) + |x_i|^s] \, dF_n(\mathbf{x}) \leq 1 + O(a_n^s) = O(1) ,$$

where the integrals are over $\mathscr{R}^{(k)}$. By the Hölder inequality,

$$\begin{aligned}
&\int |x_1|^{p_1} |x_2|^{p_2} \cdots |x_k|^{p_k} \, dF_n(\mathbf{x}) \\
&\leq \left[\int |x_1|^r \, dF_n(\mathbf{x}) \right]^{p_1/r} \left[\int |x_2|^r \, dF_n(\mathbf{x}) \right]^{p_2/r} \cdots \left[\int |x_k|^r \, dF_n(\mathbf{x}) \right]^{p_k/r} \\
&= O(1) ,
\end{aligned}$$

where

$$r = \sum_{i=1}^{k} p_i . \qquad \blacktriangle$$

Theorem 5.4.2. Let $\{X_n\}$ be a sequence of real valued random variables with corresponding distribution functions $\{F_n(x)\}$, and let $\{f_n(x)\}$ be a sequence of real valued functions. Assume that for some positive integers s and N_0:

(i) $\int |x - \mu|^{2s} \, dF_n(x) = a_n^{2s}$, where $a_n \to 0$ as $n \to \infty$.
(ii) $\int |f_n(x)|^2 \, dF_n(x) = O(1)$.
(iii) $f_n^{(s)}(x)$ is continuous in x over a closed and bounded interval S for n greater than N_0, where $f_n^{(j)}(x)$ denotes the jth derivative of $f_n(x)$ evaluated at x and $f_n^{(0)}(x) \equiv f_n(x)$.

(iv) μ is an interior point of S.

(v) There is a K such that, for $n > N_0$,

$$|f_n^{(s)}(x)| \leq K \qquad \text{for all} \quad x \in S$$

and

$$|f_n^{(r)}(\mu)| \leq K \qquad \text{for} \quad r = 0, 1, \ldots, s-1 .$$

Then

$$\int f_n(x)\, dF_n(x) = f_n(\mu) + Q(n, s) + O(a_n^2),$$

where

$$Q(n, s) = \begin{cases} 0, & s = 1, \\ \sum_{j=1}^{s-1} (1/j!) f_n^{(j)}(\mu) \int (x-\mu)^j\, dF_n(x), & s > 1. \end{cases}$$

Proof. See the proof of Theorem 5.4.3. ▲

Theorem 5.4.3. Let $\{\mathbf{X}_n\}$ be a sequence of k-dimensional random variables with corresponding distribution functions $\{F_n(\mathbf{x})\}$, and let $\{f_n(\mathbf{x})\}$ be a sequence of functions mapping $\mathscr{R}^{(k)}$ into $\mathscr{R}$. Let $\delta \in (0, \infty)$, and define $\alpha = \delta^{-1}(1 + \delta)$. Assume that for some positive integers s and N_0:

(i) $\int |\mathbf{x} - \boldsymbol{\mu}|^{\alpha s}\, dF_n(\mathbf{x}) = a_n^{\alpha s}$, where $a_n \to 0$ as $n \to \infty$.

(ii) $\int |f_n(\mathbf{x})|^{1+\delta}\, dF_n(\mathbf{x}) = O(1)$.

(iii) $f_n^{(i_1, \ldots, i_s)}(\mathbf{x})$ is continuous in $\mathbf{x}$ over a closed and bounded sphere S for all n greater than N_0, where

$$f_n^{(i_1, \ldots, i_r)}(\mathbf{x}_0) = \frac{\partial^r}{\partial x_{i_1} \cdots \partial x_{i_r}} f_n(\mathbf{x})\big|_{\mathbf{x}=\mathbf{x}_0} .$$

(iv) $\boldsymbol{\mu}$ is an interior point of S.

(v) There is a finite number K such that, for $n > N_0$,

$$|f_n^{(i_1, \ldots, i_s)}(\mathbf{x})| \leq K \qquad \text{for all} \quad \mathbf{x} \in S,$$
$$|f_n^{(i_1, \ldots, i_r)}(\boldsymbol{\mu})| \leq K \qquad \text{for} \quad r = 1, 2, \ldots, s-1,$$

and

$$|f_n(\boldsymbol{\mu})| \leq K .$$

Then

$$\int f_n(\mathbf{x})\, dF_n(\mathbf{x}) = f_n(\boldsymbol{\mu}) + \sum_{j=1}^{s-1} \frac{1}{j!} \int D^j f_n(\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^j\, dF_n(\mathbf{x}) + O(a_n^s)\,,$$

where

$$D^r f_n(\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^r = \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \cdots \sum_{i_r=1}^{k} f_n^{(i_1,\ldots,i_r)}(\boldsymbol{\mu}) \prod_{j=1}^{r} (x_{i_j} - \mu_{i_j})$$

and, for $s = 1$, it is understood that

$$\int f_n(\mathbf{x})\, dF_n(\mathbf{x}) = f_n(\boldsymbol{\mu}) + O(a_n)\,.$$

The result also holds if we replace (ii) with the condition that the $f_n(\mathbf{x})$ are uniformly bounded for n sufficiently large and assume that (i), (iii), (iv), and (v) hold for $\alpha = 1$.

Proof. We consider only those n greater than N_0. By Taylor's theorem there is a sequence of functions $\{\mathbf{Y}_n\}$ mapping S into S such that

$$f_n(\mathbf{x}) = f_n(\boldsymbol{\mu}) + I_S(\mathbf{x}) \sum_{j=1}^{s-1} \frac{1}{j!} D^j f_n(\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^j + R_n(\mathbf{x})\,,$$

where

$$I_S(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in S\,, \\ 0 & \text{otherwise} \end{cases}$$

and

$$R_n(\mathbf{x}) = \begin{cases} (s!)^{-1} D^s f_n(\mathbf{Y}_n(\mathbf{x}))(\mathbf{x}-\boldsymbol{\mu})^s & \text{if } \mathbf{x} \in S\,, \\ f_n(\mathbf{x}) - f_n(\boldsymbol{\mu}) & \text{otherwise}\,. \end{cases}$$

For $\mathbf{x} \in S$ we have $\mathbf{Y}_n(\mathbf{x}) \in S$, so that

$$\begin{aligned} |R_n(\mathbf{x})| &\le (s!)^{-1} \sum_{i_1=1}^{k} \cdots \sum_{i_s=1}^{k} \{|f_n^{(i_1,\ldots,i_s)}(\mathbf{Y}_n(\mathbf{x}))|\, |x_{i_1} - \mu_{i_1}| \cdots |x_{i_s} - \mu_{i_s}|\} \\ &\le (s!)^{-1} \sum_{i_1=1}^{k} \cdots \sum_{i_s=1}^{k} K|\mathbf{x}-\boldsymbol{\mu}|^s \\ &= (s!)^{-1} K k^s |\mathbf{x}-\boldsymbol{\mu}|^s\,. \end{aligned}$$

Thus,

$$\int_S |R_n(\mathbf{x})|\, dF_n(\mathbf{x}) \le (s!)^{-1} K k^s \int_S |\mathbf{x} - \boldsymbol{\mu}|^s \, dF_n(\mathbf{x})$$
$$= O(a_n^s)\,.$$

Now, for finite δ, and letting $\bar{S}$ denote the complement of S,

$$\begin{aligned}
&\int_{\bar{S}} |R_n(\mathbf{x})|\, dF_n(\mathbf{x}) \\
&\quad = \int I_{\bar{S}}(\mathbf{x}) |f_n(\mathbf{x}) - f_n(\boldsymbol{\mu})|\, dF_n(\mathbf{x}) \\
&\quad \le \left[\int |f_n(\mathbf{x}) - f_n(\boldsymbol{\mu})|^{1+\delta} \, dF_n(\mathbf{x}) \right]^{1/(1+\delta)} \left[\int I_{\bar{S}}(\mathbf{x})\, dF_n(\mathbf{x}) \right]^{\delta/(1+\delta)}
\end{aligned}$$

by Hölder's inequality. By Theorem 5.1.1 (Chebyshev's inequality),

$$\int I_{\bar{S}}(\mathbf{x})\, dF_n(\mathbf{x}) \le M \int |\mathbf{x} - \boldsymbol{\mu}|^{\alpha s} \, dF_n(\mathbf{x}) = O(a_n^{\alpha s}) \qquad \text{for some} \quad M > 0\,.$$

Therefore, for $\delta \in (0, \infty)$,

$$\int |R_n(\mathbf{x})|\, dF_n(\mathbf{x}) = O(a_n^s)\,.$$

This result also holds for $\alpha = 1$ and $f_n(\mathbf{x})$ uniformly bounded (by $K^\dagger$, say), because then

$$\int I_{\bar{S}}(\mathbf{x}) |f_n(\mathbf{x}) - f_n(\boldsymbol{\mu})|\, dF_n(\mathbf{x}) \le (2K^\dagger) \int I_{\bar{S}}(\mathbf{x})\, dF_n(\mathbf{x}) = O(a_n^s)\,.$$

We now have that, for $s > 1$,

$$\begin{aligned}
\int f_n(\mathbf{x})\, dF_n(\mathbf{x}) = f_n(\boldsymbol{\mu}) &+ \int \sum_{j=1}^{s-1} (j!)^{-1} D^j f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^j \, dF_n(\mathbf{x}) \\
&- \int_{\bar{S}} \sum_{j=1}^{s-1} (j!)^{-1} D^j f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^j \, dF_n(\mathbf{x}) + O(a_n^s)\,.
\end{aligned}$$

However,

$$\begin{aligned}
&\int_{\bar{S}} \sum_{j=1}^{s-1} (j!)^{-1} D^j f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^j \, dF_n(\mathbf{x}) \\
&\quad = \int_{\bar{S}} \sum_{j=1}^{s-1} (j!)^{-1} \sum_{i_1=1}^{k} \cdots \sum_{i_j=1}^{k} f_n^{(i_1,\ldots,i_j)}(\boldsymbol{\mu}) \prod_{r=1}^{j} (x_{i_r} - \mu_{i_r})\, dF_n(\mathbf{x})
\end{aligned}$$

$$\begin{aligned}&\leq \int_S Kk^s[|\mathbf{x}-\boldsymbol{\mu}|^s+1]\,dF_n(\mathbf{x})\\&= O(a_n^s)\,.\end{aligned}$$ ▲

We now give an extension of Theorem 5.4.3 to a product, where one factor is a function of the type studied in Theorem 5.4.3 and the second need not converge to zero.

Corollary 5.4.3. Let the assumptions of Theorem 5.4.3 hold. Let Z_n be a sequence of random variables defined on the same space as $\mathbf{X}_n$. Let $F_n(\mathbf{x}, z)$ be the distribution function of $(\mathbf{X}_n, Z_n)$. Assume

(vi) $\int |z|^{\alpha} |(\mathbf{x}-\boldsymbol{\mu})|^{\alpha s}\, dF_n(\mathbf{x}, z) = O(a_n^{\alpha s})$,
(vii) $E\{|Z_n|\} = O(1)$.

Then

$$\int z f_n(\mathbf{x})\, dF_n(\mathbf{x}, z) = f_n(\boldsymbol{\mu}) E\{Z_n\} + \sum_{j=1}^{s-1} (j!)^{-1} \int D^j f_n(\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^j z\, dF_n(\mathbf{x}, z) + O(a_n^s)\,.$$

Proof. Omitted. ▲

Theorems 5.4.2 and 5.4.3 require that $\int |f_n(\mathbf{x})|^{1+\delta}\, dF_n(\mathbf{x})$ be bounded for all n sufficiently large. The following theorem gives sufficient conditions for a sequence of integrals $\int |f_n(\mathbf{x})|\, dF_n(\mathbf{x})$ to be $O(1)$.

Theorem 5.4.4. Let $\{f_n(\mathbf{x})\}$ be a sequence of real valued (measurable) functions, and let $\{\mathbf{X}_n\}$ be a sequence of k-dimensional random variables with corresponding distribution functions $\{F_n(\mathbf{x})\}$. Assume that:

(i) $|f_n(\mathbf{x})| \leq K_1$ for $\mathbf{x} \in \bar{S}$, where S is a bounded open set containing $\boldsymbol{\mu}$, $\bar{S}$ is the closure of S, and K_1 is a finite constant.
(ii) $|f_n(\mathbf{x})| \leq Y(\mathbf{x}) n^p$ for some $p > 0$ and for a function $Y(\cdot)$ such that $\int |Y(\mathbf{x})|^{\gamma}\, dF_n(\mathbf{x}) = O(1)$ for some γ, $1 < \gamma < \infty$.
(iii) $\int |\mathbf{x}-\boldsymbol{\mu}|^r\, dF_n(\mathbf{x}) = O(n^{-\eta p})$ for a positive integer r and an η such that $1/\eta + 1/\gamma = 1$.

Then

$$\int |f_n(\mathbf{x})|\, dF_n(\mathbf{x}) = O(1)\,.$$

The result also holds for $\eta = 1$ given that (i), (iii), and $|f_n(\mathbf{x})| \leq K_2 n^p$ hold for all $\mathbf{x}$, some $p > 0$, and K_2 a finite constant.

Proof. Let $\delta > 0$ be such that

$$A = \{\mathbf{x}: |\mathbf{x} - \boldsymbol{\mu}| \le \delta\} \subset S .$$

By Chebyshev's inequality,

$$\begin{aligned} P\{\mathbf{X}_n \in \bar{A}\} &= P\{|\mathbf{X}_n - \boldsymbol{\mu}| > \delta\} \\ &\le \delta^{-r} \int |\mathbf{x} - \boldsymbol{\mu}|^r \, dF_n(\mathbf{x}) \\ &= O(n^{-\eta p}) . \end{aligned}$$

For $1 < \gamma < \infty$,

$$\begin{aligned} \int |f_n(\mathbf{x})| \, dF_n(\mathbf{x}) &= \int_A |f_n(\mathbf{x})| \, dF_n(\mathbf{x}) + \int_{\bar{A}} |f_n(\mathbf{x})| \, dF_n(\mathbf{x}) \\ &\le \int_A K_1 \, dF_n(\mathbf{x}) + n^p \int_{\bar{A}} |Y(\mathbf{x})| \, dF_n(\mathbf{x}) \\ &\le K_1 + n^p \int I_{\bar{A}}(\mathbf{x}) |Y(\mathbf{x})| \, dF_n(\mathbf{x}) \\ &\le K_1 + n_p \left[\int I_{\bar{A}}(\mathbf{x}) \, dF_n(\mathbf{x}) \right]^{1/\eta} \left[\int |Y(\mathbf{x})|^\gamma \, dF_n(\mathbf{x}) \right]^{1/\gamma} \\ &= K_1 + n^p [P\{\mathbf{X}_n \in \bar{A}\}]^{1/\eta} [O(1)] \\ &= O(1) . \end{aligned}$$

For $|f_n(\mathbf{x})| \le K_2 n^p$ and $\eta = 1$, we have

$$\begin{aligned} \int |f_n(\mathbf{x})| \, dF_n(\mathbf{x}) &\le K_1 + K_2 n^p \int I_{\bar{A}}(\mathbf{x}) \, dF_n(\mathbf{x}) \\ &= O(1) . \end{aligned}$$

▲

Example 5.4.1. To illustrate some of the ideas of this section, we consider the problem of estimating $\log \mu$, where $\mu > 0$ is the mean of a random variable with finite sixth moment. Let $\bar{x}_n$ be the mean of a simple random sample of n such observations. Since $\log x$ is not defined for $x \le 0$, we consider the estimator

$$f_n(\bar{x}_n) = \begin{cases} \log \bar{x}_n & \text{if } \bar{x}_n > 1/n , \\ -\log n & \text{if } \bar{x}_n \le 1/n . \end{cases}$$

Suppose that we desire an approximation to the sequence of expectations of $f_n(\bar{x}_n)$ to order n^{-1}. We first apply Theorem 5.4.2 with $a_n = O(n^{-1/2})$ and $s = 3$. By Theorem 5.4.1, we have that

$$E\{(\bar{x}_n - \mu)^{2s}\} = O(n^{-s}) , \qquad s = 1, 2, 3 , \tag{5.4.2}$$

so that condition i of Theorem 5.4.2 is met. To establish condition ii, we demonstrate that $|f_n(x)|^2$ satisfies the conditions of Theorem 5.4.4. Since $f_n(x)$ is continuous for $x > 0$, condition i of Theorem 5.4.4 is satisfied. For $p = 1$, $\gamma = \eta = 2$, and

$$Y(x) = \begin{cases} 1, & x < 1, \\ x & \text{otherwise}, \end{cases}$$

we have that $|f_n(x)|^2 < nY(x)$ and $\int |Y(x)|^2 \, dF_n(x) = O(1)$. This is condition ii of Theorem 5.4.4; from equation (5.4.2) we see that condition iii also holds for $r = 4$, 5, or 6. Therefore, the conditions of Theorem 5.4.4 are satisfied and the integral of $|f_n(x)|^2$ is order one.

Since the third derivative of $\log x$ is continuous for positive x, we may choose an $N_0 > \mu^{-1}$ and an interval containing μ such that $f_n^{(3)}(x)$ is continuous on that interval for all $n > N_0$. To illustrate, let $\mu = 0.1$, and take the interval S to be $[0.05, 0.15]$. Then, for $n > 20$, $f_n^{(3)}(x)$ is continuous on S. Therefore, conditions iii, iv, and v of Theorem 5.4.2 are also met, and we may write

$$\begin{aligned} E\{f_n(\bar{x}_n)\} &= \log \mu + \frac{1}{\mu} E\{\bar{x}_n - \mu\} \\ &\quad - \frac{1}{2\mu^2} E\{(\bar{x}_n - \mu)^2\} + O(n^{-3/2}) \\ &= \log \mu - \frac{\sigma^2}{2n\mu^2} + O(n^{-3/2}), \end{aligned}$$

where σ^2 is the variance of the observations. To order n^{-1}, the bias in $f_n(\bar{x}_n)$ as an estimator of $\log \mu$ is $-(2n\mu^2)^{-1}\sigma^2$.

Since $\bar{x}_n$ possesses finite fifth moments, we can decrease the above remainder term to $O(n^{-2})$ by carrying the expansion to one more term and using the more general form of Theorem 5.4.3. By equation (5.4.2), condition i of Theorem 5.4.3 holds for $s = 4$, $\alpha = \frac{3}{2}$, and $a_n = O(n^{-1/2})$, so that $\delta = 2$ by definition. Condition ii is established by demonstrating that $|f_n(x)|^3$ satisfies the conditions of Theorem 5.4.4. Defining

$$Y(x) = \begin{cases} 1, & x < 1, \\ x^{3/2}, & x \geq 1, \end{cases}$$

and letting $p = \frac{3}{2}$, we see that $|f_n(x)|^3 \leq Y(x)n^p$ and

$$\int |Y(x)|^2 \, dF_n(x) \leq \int |1 + x^{3/2}|^2 \, dF_n(x) = O(1),$$

as $\bar{x}_n$ has finite third moments. Therefore, we are using $\gamma = \eta = 2$ and $\eta p = 3$. Since $\int |x - \mu|^5 \, dF_n(x) = O(n^{-3})$ and $|f_n(x)|^3$ is continuous, the conditions of Theorem 5.4.4 are met. The sixth derivative of $\log x$ is continuous for positive x, and we can find an N_0 and a set S such that conditions iii, iv, and v of Theorem 5.4.3 are also met. Because $f_n^{(3)}(\mu) = O(1)$ and $E\{(\bar{x}_n - \mu)^3\} = O(n^{-2})$, we have

$$\begin{aligned} E\{f_n(\bar{x}_n)\} &= \log \mu - (2n\mu^2)^{-1}\sigma^2 \\ &\quad + \frac{1}{3!} f_n^{(3)}(\mu) E\{(\bar{x}_n - \mu)^3\} + O(n^{-2}) \\ &= \log \mu - (2n\mu^2)^{-1}\sigma^2 + O(n^{-2}) . \end{aligned}$$

▲▲

5.5. ESTIMATION FOR NONLINEAR MODELS

5.5.1. Estimators That Minimize an Objective Function

The expected value of a time series is sometimes expressible as a nonlinear function of unknown parameters and observable functions of time. For example, we might have for $\{Y_t : t \in (0, 1, 2, \ldots)\}$

$$E\{Y_t\} = \sum_{j=0}^{t} \lambda^j x_{t-j} ,$$

where $\{x_t\}$ is a sequence of known constants and λ is unknown. The estimation of a parameter such as λ is considerably more complicated than the estimation of a parameter that enters the expected value function in a linear manner.

We assume the model

$$Y_t = f_t(\boldsymbol{\theta}^0) + e_t , \tag{5.5.1}$$

for $t = 1, 2, \ldots$, where the random variables are defined on a complete probability space $(\Omega, \mathcal{A}, P)$, the vector $\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_k^0)'$ of unknown parameters lies in a compact subset Θ of k-dimensional Euclidean space $\mathcal{R}^k$, the function f_t is defined on $\Omega \times \Theta$ with the form of f_t known, and $f_t(\omega, \cdot)$ is continuous on Θ almost surely, for all t. The expression $f_t(\omega, \cdot)$ represents the function for a particular ω in Ω. The dependence of the function on ω is generally suppressed for notational convenience. The function f_t may depend on an input vector $\mathbf{x}_t$, and sometimes we may write the model as

$$Y_t = f(\mathbf{x}_t; \boldsymbol{\theta}^0) + e_t .$$

The dimension of $\mathbf{x}_t$ need not be fixed, and it may increase with t. Some elements of $\mathbf{x}_t$ may be fixed and some may be random.

The derivatives of the function $f_t(\boldsymbol{\theta})$ are very important in the treatment of this problem, and we introduce a shorthand notation for them. Let $f_t^{(j)}(\hat{\boldsymbol{\theta}})$ denote the first partial derivative of $f_t(\boldsymbol{\theta})$ with respect to the jth element of $\boldsymbol{\theta}$ evaluated at the point $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Likewise, let $f_t^{(js)}(\hat{\boldsymbol{\theta}})$ denote the second partial derivative of $f_t(\boldsymbol{\theta})$ with respect to the jth and sth elements of $\boldsymbol{\theta}$ evaluated at the point $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The third partial derivative, $f_t^{(jrs)}(\hat{\boldsymbol{\theta}})$, is defined in a similar manner. Let

$$\mathbf{F}_t(\boldsymbol{\theta}) = [f_t^{(1)}(\boldsymbol{\theta}), f_t^{(2)}(\boldsymbol{\theta}), \ldots, f_t^{(k)}(\boldsymbol{\theta})] \tag{5.5.2}$$

be the vector of first derivatives, and let

$$\mathbf{F}_{nk}(\boldsymbol{\theta}) = [\mathbf{F}_1'(\boldsymbol{\theta}), \mathbf{F}_2'(\boldsymbol{\theta}), \ldots, \mathbf{F}_n'(\boldsymbol{\theta})]'$$

be the $n \times k$ matrix of first derivatives.

Let an estimator of the unknown $\boldsymbol{\theta}^0$ be the $\boldsymbol{\theta}$ that minimizes $Q_n(\boldsymbol{\theta})$, where $Q_n(\boldsymbol{\theta})$ is a function of the observations and $\boldsymbol{\theta}$. The least squares estimator of $\boldsymbol{\theta}^0$ is the $\boldsymbol{\theta}$ that minimizes

$$Q_{nl}(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^{n} [Y_t - f_t(\boldsymbol{\theta})]^2 . \tag{5.5.3}$$

For least squares, the function evaluated at $\boldsymbol{\theta}^0$ is

$$Q_{nl}(\boldsymbol{\theta}^0) = n^{-1} \sum_{t=1}^{n} e_t^2 .$$

The function $Q_n(\boldsymbol{\theta})$ is the negative of the logarithm of the likelihood in the case of maximum likelihood estimation for the model (5.5.1) with normal e_t.

In order to obtain a limiting distribution for the estimator, the sequence $\{\mathbf{x}_t, e_t\}$, the function $f_t(\boldsymbol{\theta})$, and the parameter space $\boldsymbol{\Theta}$ must satisfy certain conditions. Stronger conditions in one area will permit weaker conditions in another. The random variables and the function must be such that the estimator is uniquely defined, at least for large n. Furthermore, the function being minimized must have a unique minimum "close to" the true parameter value, and the distance must decrease as the sample size increases.

These ideas are made more precise in the following theorems. We begin with results for the general problem and then consider special cases. Let $\hat{\boldsymbol{\theta}}_n$ in $\boldsymbol{\Theta}$ be a measurable function that minimizes $Q_n(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}$ almost surely. Lemma 5.5.1 contains sufficient conditions for the consistency of $\hat{\boldsymbol{\theta}}_n$. Lemma 5.5.1 is given in Wu (1981) and is based on the idea used by Wald (1949) in a proof of the strong consistency of the maximum likelihood estimator.

Lemma 5.5.1. Let $\hat{\boldsymbol{\theta}}_n$ in $\boldsymbol{\Theta}$ be a measurable function that minimizes an objective function $Q_n(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}$ almost surely.

(i) Suppose, for any $\eta > 0$,

$$\lim_{n\to\infty} \inf_{|\theta - \theta^0| \geq \eta} [Q_n(\theta) - Q_n(\theta^0)] > 0 \tag{5.5.4}$$

almost surely. Then $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}^0$ almost surely as $n \to \infty$.

(ii) Suppose, for any $\eta > 0$,

$$\lim_{n\to\infty} P\left\{\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}^0|\geq\eta} [Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}^0)] > 0\right\} = 1\,. \tag{5.5.5}$$

Then $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}^0$ in probability as $n \to \infty$.

Proof. A proof of part i is given by Wu (1981). We prove part ii. Suppose $\hat{\boldsymbol{\theta}}_n$ does not converge to $\boldsymbol{\theta}^0$ in probability as $n \to \infty$. Then there exist $\eta > 0$, $1 \geq \epsilon_\eta > 0$, and a subsequence $\{n_k\}$ such that

$$P\{|\hat{\boldsymbol{\theta}}_{n_k} - \boldsymbol{\theta}^0| \geq \eta\} > \epsilon_\eta \qquad \text{for every } n_k\,,$$

implying

$$P\left\{\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}^0|\geq\eta} [Q_{n_k}(\boldsymbol{\theta}) - Q_{n_k}(\boldsymbol{\theta}^0)] \leq 0\right\} > \epsilon_\eta \qquad \text{for every } n_k\,.$$

Therefore,

$$\limsup_{n\to\infty} P\left\{\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}^0|\geq\eta} [Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}^0)] \leq 0\right\} > \epsilon_\eta\,.$$

That is,

$$\liminf_{n\to\infty} P\left\{\inf_{|\boldsymbol{\theta}-\boldsymbol{\theta}^0|\geq\eta} [Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}^0)] > 0\right\} \leq 1 - \epsilon_\eta < 1\,,$$

which is a contradiction of the condition (5.5.5). Hence, if (5.5.5) holds, then $\hat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}^0$ in probability as $n \to \infty$. ▲

The conditions in Lemma 5.5.1 are rather general. Gallant and White (1988, pp. 18–19) give a set of sufficient conditions for strong consistency. We give a form of their result in Lemma 5.5.2. The condition (5.5.6) of the lemma is a uniform law of large numbers for the function $Q_n(\boldsymbol{\theta})$. This condition guarantees that the sample function will be close to the limit function as the sample size increases. The condition (5.5.7) insures that the limit of $Q_n(\boldsymbol{\theta})$ has a minimum at $\boldsymbol{\theta}^0$. It is sometimes called the identification condition.

Lemma 5.5.2. Given $(\Omega, \mathscr{A}, P)$ and a compact set Θ that is a subset of $\mathscr{R}^k$, let $Q_n : \Omega \times \Theta \to \mathscr{R}$ be a random function continuous on Θ a.s., for $n = 1, 2, \ldots$. Let $\hat{\boldsymbol{\theta}}_n$ be as defined in Lemma 5.5.1. Suppose there exists a function $\bar{Q}_n : \Theta \to \mathscr{R}$ such that

$$Q_n(\boldsymbol{\theta}) - \bar{Q}_n(\boldsymbol{\theta}) \to 0 \qquad \text{a.s.} \tag{5.5.6}$$

uniformly on Θ. Assume that for any $\eta > 0$

$$\liminf_{n\to\infty} \left\{ \inf_{|\theta-\theta^0|\geq\eta} [\bar{Q}_n(\theta) - \bar{Q}_n(\theta^0)] \right\} > 0 . \tag{5.5.7}$$

Then $\hat{\theta}_n - \theta^0 \to 0$ a.s. as $n \to \infty$.

Proof. Now

$$Q_n(\theta) - Q_n(\theta^0) = Q_n(\theta) - \bar{Q}_n(\theta) + \bar{Q}_n(\theta) - \bar{Q}_n(\theta^0) + \bar{Q}_n(\theta^0) - Q_n(\theta^0) .$$

Therefore,

$$\begin{aligned}&\inf_{\Theta} [Q_n(\theta) - Q_n(\theta^0)] \\ &\qquad \geq \inf_{\Theta} [Q_n(\theta) - \bar{Q}_n(\theta)] + \inf_{\Theta} [\bar{Q}_n(\theta) - \bar{Q}_n(\theta^0)] + \bar{Q}_n(\theta^0) - Q_n(\theta^0) .\end{aligned}$$

The first term on the right of the inequality converges to zero as $n \to \infty$ because

$$\left|\inf_{\Theta} [Q_n(\theta) - \bar{Q}_n(\theta)]\right| \leq \sup_{\Theta} |Q_n(\theta) - \bar{Q}_n(\theta)|$$

and by the assumption (5.5.6), there exists a B in $\mathscr{A}$ with $P(B) = 1$ such that, given any $\epsilon > 0$, for each ω in B there exists an integer $N(\omega, \epsilon) < \infty$ such that for all $n > N(\omega, \epsilon)$,

$$\sup_{\Theta} |Q_n(\omega, \theta) - \bar{Q}_n(\theta)| < \epsilon .$$

The difference $\bar{Q}_n(\theta^0) - Q_n(\theta^0)$ also converges to zero as $n \to \infty$ by (5.5.6). The conclusion follows from the assumption (5.5.7) and Lemma 5.5.1. ▲

A critical assumption of Lemma 5.5.2 is (5.5.6), which guarantees that the random functions $Q_n(\theta)$ behave like nonrandom functions $\bar{Q}_n(\theta)$ for large n. If $Q_n(\theta)$ can be written as

$$Q_n(\theta) = n^{-1} \sum_{t=1}^{n} q_t(\theta) , \tag{5.5.8}$$

then a natural choice for $\bar{Q}_n(\theta)$ is

$$\bar{Q}_n(\theta) = n^{-1} \sum_{t=1}^{n} E\{q_t(\theta)\} . \tag{5.5.9}$$

Gallant and White (1988, p. 23) give a uniform law of large numbers for a function of the form (5.5.8) which is a slight modification of the result of Andrews (1986). We give the result as Lemma 5.5.3.

Lemma 5.5.3. Let $q_t(\boldsymbol{\theta})$ be a function from $\Omega \times \Theta$ to $\mathscr{R}$, $t = 1, 2, \ldots$, where $(\Omega, \mathscr{A}, P)$ is the underlying probability space, and Θ is a compact subset of $\mathscr{R}^k$. Assume $q_t(\omega, \cdot)$ is continuous in $\boldsymbol{\theta}$ almost surely. Assume $q_t(\cdot, \boldsymbol{\theta}^0)$ is $\mathscr{A}$-measurable for each $\boldsymbol{\theta}^0$ in Θ and for $t = 1, 2, \ldots$. For given $\boldsymbol{\theta}^0$ in Θ and $\delta > 0$, let

$$A(\delta) = \{\boldsymbol{\theta} \in \Theta: |\boldsymbol{\theta} - \boldsymbol{\theta}^0| < \delta\}.$$

Assume that for each $\boldsymbol{\theta}^0$ in Θ there exists a constant $\delta^0 > 0$ and positive random variables L_t^0 such that

$$n^{-1} \sum_{t=1}^{n} E\{L_t^0\} = O(1)$$

and

$$|q_t(\boldsymbol{\theta}) - q_t(\boldsymbol{\theta}^0)| \le L_t^0 |\boldsymbol{\theta} - \boldsymbol{\theta}^0|, \quad t = 1, 2, \ldots, \qquad \text{a.s.}$$

for all $\boldsymbol{\theta}$ in the closure of $A(\delta)$. Also assume that for all $0 < \delta < \delta^0$, where δ^0 may depend on $\boldsymbol{\theta}^0$,

$$\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} [l_t(\delta) - E\{l_t(\delta)\}] = 0 \quad \text{a.s.}$$

and

$$\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} [s_t(\delta) - E\{s_t(\delta)\}] = 0 \qquad \text{a.s.,}$$

where

$$l_t(\delta) = \sup_{A(\delta)} q_t(\boldsymbol{\theta}),$$

and

$$s_t(\delta) = \inf_{A(\delta)} q_t(\boldsymbol{\theta}).$$

Then $\bar{Q}_n(\boldsymbol{\theta})$ defined in (5.5.9) is continuous on Θ uniformly in n, and

$$\lim_{n \to \infty} \sup_{\Theta} |Q_n(\boldsymbol{\theta}) - \bar{Q}_n(\boldsymbol{\theta})| = 0.$$

Proof. See Gallant and White (1988, p. 38). ▲

It may be easier to obtain a uniform law of large numbers for the function $Q_n(\theta)$ directly than for $\{q_t(\boldsymbol{\theta}), t = 1, 2, \ldots, n\}$. A uniform strong law of large numbers for $Q_n(\boldsymbol{\theta})$ is given in Lemma 5.5.4, and a uniform weak law of large numbers is given in Lemma 5.5.5.

Lemma 5.5.4. Let Θ be a convex, compact subset of k-dimensional Euclidean space, $\mathscr{R}^k$, and let $\{Q_n(\theta),\ n = 1, 2, \ldots\}$ be a sequence of random functions. Assume

(i) For each $\boldsymbol{\theta}_1$ in Θ

$$\lim_{n\to\infty} [Q_n(\boldsymbol{\theta}_1) - E\{Q_n(\boldsymbol{\theta}_1)\}] = 0 \qquad \text{a.s.}$$

(ii) There exist random variables L_n such that

(a) $\displaystyle \max_{1\le i\le k} \left\{ \sup_{\Theta} \left| \frac{\partial Q_n(\boldsymbol{\theta})}{\partial \theta_i} \right| \right\} \le L_n ,$

(b) $\displaystyle \limsup_{n\to\infty} L_n < \infty$ a.s.,

(c) $\displaystyle \sup_n E\{L_n\} < \infty .$

Then

$$\lim_{n\to\infty} \left\{ \sup_{\Theta} |Q_n(\boldsymbol{\theta}) - E\{Q_n(\boldsymbol{\theta})\}| \right\} = 0 \qquad \text{a.s.}$$

Proof. Omitted. ▲

Lemma 5.5.5. Let Θ be a convex, compact subset of k-dimensional Euclidean space, and let $\{Q_n(\boldsymbol{\theta}): n = 1, 2, \ldots\}$ be a sequence of random functions. Assume:

(i) For each θ_1 in Θ,

$$\operatorname*{plim}_{n\to\infty} [Q_n(\boldsymbol{\theta}_1) - Q(\boldsymbol{\theta}_1)] = 0$$

for some $Q(\boldsymbol{\theta})$.

(ii) There exists a sequence of positive random variables $\{L_n\}$ and an L such that for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in Θ,
  (a) $|Q_n(\boldsymbol{\theta}_1) - Q_n(\boldsymbol{\theta}_2)| \le L_n |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|$,
  (b) $|Q(\boldsymbol{\theta}_1) - Q(\boldsymbol{\theta}_2)| \le L |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|$,
  (c) $L_n = O_p(1)$ and $L = O_p(1)$.

Then

$$\operatorname*{plim}_{n\to\infty} Q_n(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}) = 0$$

uniformly in $\boldsymbol{\theta}$ in Θ.

Proof. Given $\eta > 0$, consider the open set $G_{\boldsymbol{\theta}_i} = \{\boldsymbol{\theta}: |\boldsymbol{\theta} - \boldsymbol{\theta}_i| < \eta\}$ defined by $\boldsymbol{\theta}_i$ and η. Since Θ is compact, there exists a $k = k(\eta)$ and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_k$ such that

$\Theta = \bigcup_{i=1}^{k} G_i$, where $G_i = G_{\theta_i}$. Now,

$$\begin{aligned}\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| &= \max_{1\le i\le k} \sup_{\theta\in G_i} |Q_n(\theta) - Q(\theta)| \\ &\le \max_{1\le i\le k} \sup_{\theta\in G_i} |Q_n(\theta) - Q_n(\theta_i)| + \max_{1\le i\le k} |Q_n(\theta_i) - Q(\theta_i)| \\ &+ \max_{1\le i\le k} \sup_{\theta\in G_i} |Q(\theta_i) - Q(\theta)| .\end{aligned}$$

Given $\epsilon > 0$ and $\tau > 0$, by (i) there exists an $N_i = N_i(\epsilon, \tau)$ such that for $n > N_i$,

$$P[|Q_n(\theta_i) - Q(\theta_i)| > 3^{-1}\epsilon] < \tau .$$

Thus, for $n > N = \max(N_1, \dots, N_k)$,

$$P[\max_{1\le i\le k} |Q_n(\theta_i) - Q(\theta_i)| \ge 3^{-1}\epsilon] \le \sum_{i=1}^{k} P[|Q_n(\theta_i) - Q(\theta_i)| \ge 3^{-1}\epsilon] \le k\tau .$$

By (ii)(c), for given $\delta > 0$ there exists $M > 0$ such that

$$\sup_n P[L_n > M] < 3^{-1}\delta$$

and $P[L > M] < 3^{-1}\delta$. Thus, given $\epsilon > 0$ and $\delta > 0$, choose $\eta = (3M)^{-1}\epsilon$. Then,

$$P\left[\max_{1\le i\le k} \sup_{\theta\in G_i} |Q_n(\theta) - Q_n(\theta_i)| > 3^{-1}\epsilon\right] \le P[\eta L_n > 3^{-1}\epsilon] \le P[L_n > M] \le 3^{-1}\delta .$$

Similarly,

$$P\left[\max_{1\le i\le k} \sup_{\theta\in G_i} |Q(\theta) - Q(\theta_i)| > 3^{-1}\epsilon\right] \le 3^{-1}\delta .$$

Choosing $\tau = (3k)^{-1}\delta$ gives the result. ▲

Given conditions for consistency of $\hat{\theta}_n$, we investigate the limiting distribution of the estimator. We begin by giving conditions under which the least squares estimator converges in law to a normal random variable. In the theorem statement we assume the estimator to be consistent. The other assumptions of the theorem are sufficient for the existence of a consistent sequence of estimators (see Theorem 5.5.3), but are not sufficient to rule out multiple local minima. The theorem is general enough to cover models that contain lagged values of Y_t in the explanatory vector $\mathbf{x}_t$. To permit this, the sigma-field $\mathscr{A}_{t-1}$ of the theorem is generated by $\mathbf{x}_1$ through $\mathbf{x}_t$ and e_1 through e_{t-1}. For a vector $\mathbf{a}$, let $|\mathbf{a}| = (\mathbf{a}'\mathbf{a})^{1/2}$.

Theorem 5.5.1. Let the model (5.5.1) hold, and let $\hat{\theta}_n$ be a weakly consistent estimator of θ^0 that minimizes $Q_{nl}(\theta)$ of (5.5.3). Let $\mathscr{A}_{t-1}$ be the sigma-field generated by

$$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t, e_1, e_2, \ldots, e_{t-1}\}.$$

Assume:

(i) There exists a convex, compact neighborhood S of the true parameter vector $\boldsymbol{\theta}^0$ such that $f_t(\cdot, \boldsymbol{\theta})$ is measurable for each $\boldsymbol{\theta}$ in S, $f_t(\omega, \cdot)$ is almost surely twice continuously differentiable with respect to $\boldsymbol{\theta}$ on S, S is in Θ and $\boldsymbol{\theta}^0$ is an interior point of S.

(ii) The vector sequence $\{[\mathbf{F}_t(\boldsymbol{\theta}^0)e_t, e_t]\}$ satisfies

$$E\{[\mathbf{F}_t(\boldsymbol{\theta}^0)e_t, e_t, e_t^2] | \mathscr{A}_{t-1}\} = [\mathbf{0}, 0, \sigma^2],$$
$$E\{\|[\mathbf{F}_t(\boldsymbol{\theta}^0)e_t, e_t]\|^{2+\delta} | \mathscr{A}_{t-1}\} < M_F < \infty,$$
$$E\{\mathbf{F}_t'(\boldsymbol{\theta}^0)\mathbf{F}_t(\boldsymbol{\theta}^0)e_t^2 | \mathscr{A}_{t-1}\} = \mathbf{F}_t'(\boldsymbol{\theta}^0)\mathbf{F}_t(\boldsymbol{\theta}^0)\sigma^2$$

a.s., for all t, where $\delta > 0$, $\mathbf{F}_t(\boldsymbol{\theta})$ is defined in (5.5.2), and M_F is a positive constant.

(iii) There exists a $k \times k$ fixed matrix $\mathbf{B}(\boldsymbol{\theta})$ such that $\mathbf{B}(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}^0$,

$$\operatorname*{plim}_{n\to\infty} \left[0.5 \frac{\partial^2 Q_{nl}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} - \mathbf{B}(\boldsymbol{\theta})\right] = \mathbf{0} \qquad \text{uniformly on } S,$$
$$\operatorname*{plim}_{n\to\infty} n^{-1}\mathbf{F}_{nk}'(\boldsymbol{\theta}^0)\mathbf{F}_{nk}(\boldsymbol{\theta}^0) = \lim_{n\to\infty} n^{-1}E\{\mathbf{F}_{nk}'(\boldsymbol{\theta}^0)\mathbf{F}_{nk}(\boldsymbol{\theta}^0)\} = \mathbf{B}(\boldsymbol{\theta}^0),$$

and $\mathbf{B}(\boldsymbol{\theta}^0)$ is positive definite.

Let

$$s^2 = (n-k)^{-1} \sum_{t=1}^{n} [Y_t - f(\mathbf{x}_t; \hat{\boldsymbol{\theta}}_n)]^2. \tag{5.5.10}$$

Then

(a) $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{\mathscr{L}} N[\mathbf{0}, \mathbf{B}^{-1}(\boldsymbol{\theta}^0)\sigma^2]$ (5.5.11)

and

(b) s^2 converges to σ^2 in probability.

Proof. Because the estimators $\hat{\boldsymbol{\theta}}_n$ lie inside the convex, compact neighborhood S of $\boldsymbol{\theta}^0$ with probability approaching one as n increases, $\hat{\boldsymbol{\theta}}_n$ satisfies

$$\frac{\partial Q_{nl}(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} = 0$$

with probability approaching one as n increases. We can expand the first derivative in a Taylor series about $\boldsymbol{\theta}^0$ to obtain

$$\mathbf{0} = n^{1/2} \frac{\partial Q_{nl}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 Q_{nl}(\ddot{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \tag{5.5.12}$$

for $\hat{\boldsymbol{\theta}}_n$ in S, where $\ddot{\boldsymbol{\theta}}_n$ is on the line segment joining $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}^0$. By assumption iii and the consistency of $\hat{\boldsymbol{\theta}}_n$,

$$\frac{1}{2} \frac{\partial^2 Q_{nl}(\ddot{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} \xrightarrow{P} \mathbf{B}(\boldsymbol{\theta}^0)\,. \tag{5.5.13}$$

Now

$$\frac{\partial Q_{nl}(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} = -2n^{-1}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)\mathbf{e}_n\,, \tag{5.5.14}$$

where $\mathbf{e}_n = (e_1, e_2, \ldots, e_n)'$. Therefore

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = \mathbf{B}^{-1}(\boldsymbol{\theta}^0)n^{-1}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)\mathbf{e}_n + \mathbf{r}_n\,, \tag{5.5.15}$$

where $\mathbf{r}_n$ is of smaller order in probability than $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0$.

To prove that

$$n^{-1/2}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)\mathbf{e}_n \xrightarrow{\mathscr{L}} N[\mathbf{0}, \mathbf{B}(\boldsymbol{\theta}^0)\sigma^2]$$

it is enough to show that for any k-dimensional vector $\boldsymbol{\lambda}$, $\boldsymbol{\lambda} \neq 0$,

$$\boldsymbol{\lambda}'[n^{-1/2}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)\mathbf{e}_n] = \sum_{t=1}^{n} n^{-1/2}\left[\sum_{j=1}^{k} \lambda_j f_t^{(j)}(\boldsymbol{\theta}^0)e_t\right]$$

converges in law to a univariate normal distribution. We prove this by showing that the conditions of Theorem 5.3.4 hold. Let

$$Z_{tn} = n^{-1/2}\left[\sum_{j=1}^{k} \lambda_j f_t^{(j)}(\boldsymbol{\theta}^0)e_t\right],$$

and let $\mathscr{A}_{t,n}$ be $\mathscr{A}_t$, $0 \leq t \leq n$, $n \geq 1$. Then by assumption ii

$$E\{Z_{tn} \mid \mathscr{A}_{t-1,n}\} = 0 \qquad \text{a.s.}$$

and

$$V_{nn}^2 = \sum_{t=1}^{n} E\{Z_{tn}^2 \mid \mathscr{A}_{t-1,n}\} = \sigma^2\boldsymbol{\lambda}'[n^{-1}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)\mathbf{F}_{nk}(\boldsymbol{\theta}^0)]\boldsymbol{\lambda}\,.$$

Note that

$$V\{n^{-1/2}\boldsymbol{\lambda}'\mathbf{F}'_{nk}(\boldsymbol{\theta}_n^0)\mathbf{e}\} = \sigma^2\boldsymbol{\lambda}'\{E[n^{-1}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)\mathbf{F}_{nk}(\boldsymbol{\theta}^0)]\}\boldsymbol{\lambda}$$

and

$$s_{nn}^2 = E\{V_{nn}^2\} = \sigma^2 \boldsymbol{\lambda}'[E\{n^{-1}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)\mathbf{F}_{nk}(\boldsymbol{\theta}^0)\}]\boldsymbol{\lambda}.$$

By assumption iii

$$\begin{aligned}\sigma^{-2}(V_{nn}^2 - s_{nn}^2) = \boldsymbol{\lambda}'[n^{-1}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)\mathbf{F}_{nk}(\boldsymbol{\theta}^0)]\boldsymbol{\lambda} \\ - \boldsymbol{\lambda}'[E\{n^{-1}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)\mathbf{F}_{nk}(\boldsymbol{\theta}^0)\}]\boldsymbol{\lambda}\end{aligned}$$

converges to zero in probability,

$$s_{nn}^2 \to \sigma^2 \boldsymbol{\lambda}'\mathbf{B}(\boldsymbol{\theta}^0)\boldsymbol{\lambda} > 0,$$

and

$$(V_{nn}^2 - s_{nn}^2)s_{nn}^{-2} \xrightarrow{P} 0 \qquad \text{as} \quad n \to \infty.$$

Hence, condition i of Theorem 5.3.4 is satisfied. Now for any arbitrary $\epsilon > 0$,

$$\begin{aligned}
&s_{nn}^{-2} \sum_{j=1}^{n} E\{Z_{jn}^2 I(|Z_{jn}| \geq \epsilon s_{nn})\} \\
&\quad \leq s_{nn}^{-2} \sum_{j=1}^{n} (\epsilon s_{nn})^{-\delta} E\{|Z_{jn}|^{2+\delta} I(|Z_{jn}| \geq \epsilon s_{nn})\} \\
&\quad \leq \epsilon^{-\delta} s_{nn}^{-(2+\delta)} n^{-(1+\delta/2)} \\
&\qquad \times \sum_{t=1}^{n} E\left\{ \left| \sum_{j=1}^{k} \lambda_j f_t^{(j)}(\boldsymbol{\theta}^0) e_t \right|^{2+\delta} I(|Z_{tn}| \geq \epsilon s_{nn}) \right\} \\
&\quad \leq \epsilon^{-\delta} s_{nn}^{-(2+\delta)} n^{-(1+\delta/2)} \sum_{t=1}^{n} E\left\{ \left| \sum_{j=1}^{k} \lambda_j f_t^{(j)}(\boldsymbol{\theta}^0) e_t \right|^{2+\delta} \right\} \\
&\quad \leq \epsilon^{-\delta} s_{nn}^{-(2+\delta)} n^{-(\delta/2)} M_F,
\end{aligned}$$

where $I(A)$ is the indicator function for the set A. Hence, condition ii of Theorem 5.3.4 is satisfied and

$$n^{-1/2}\mathbf{F}'_{nk}(\boldsymbol{\theta}^0)e_n \xrightarrow{\mathscr{L}} N[\mathbf{0}, \mathbf{B}(\boldsymbol{\theta}^0)\sigma^2]. \tag{5.5.16}$$

By (5.5.12)–(5.5.14), using Corollary 5.2.6.2, we have

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{\mathscr{L}} N[\mathbf{0}, \mathbf{B}^{-1}(\boldsymbol{\theta}^0)\sigma^2].$$

To obtain the limit of s^2, we note that

$$\begin{aligned}
Q_n(\hat{\boldsymbol{\theta}}) &= Q_n(\boldsymbol{\theta}^0) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)' \frac{\partial Q_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + \tfrac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)' \frac{\partial^2 Q_n(\ddot{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\
&= Q_n(\boldsymbol{\theta}^0) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)' B(\boldsymbol{\theta}^0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) + O_p(n^{-1}) \\
&= Q_n(\boldsymbol{\theta}^0) + O_p(n^{-1}),
\end{aligned}$$

where $\ddot{\boldsymbol{\theta}}$ is on the line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$ and we have used (5.5.15). Therefore

$$s^2 = n(n-k)^{-1}Q_n(\hat{\boldsymbol{\theta}}) = (n-k)^{-1}\sum_{t=1}^{n} e_t^2 + O_p(n^{-1}). \tag{5.5.17}$$

By assumption ii, the e_t have bounded $2+\delta$ conditional moment. Therefore, by Corollary 5.3.5,

$$\lim_{n\to\infty}(n-k)^{-1}\sum_{t=1}^{n} e_t^2 = \sigma^2 \qquad \text{almost surely.}$$

Conclusion b follows by (5.5.17). ▲

Because s^2 is converging to σ^2 and $n^{-1}\mathbf{F}'_{nk}(\hat{\boldsymbol{\theta}})\mathbf{F}_{nk}(\hat{\boldsymbol{\theta}})$ is converging to $\mathbf{B}(\boldsymbol{\theta}^0)$, we can use

$$\hat{V}\{\hat{\boldsymbol{\theta}}\} = [\mathbf{F}'_{nk}(\hat{\boldsymbol{\theta}})\mathbf{F}_{nk}(\hat{\boldsymbol{\theta}})]^{-1}s^2$$

as an estimator of the variance of $\hat{\boldsymbol{\theta}}$. The statistics computed as the usual "t-statistics" of regression will be distributed as $N(0, 1)$ random variables in the limit.

There are situations in which the estimator $\hat{\boldsymbol{\theta}}_n$ may have a limiting distribution that is not normal. Also, for some functions $Q_n(\boldsymbol{\theta})$, the covariance matrix of the limiting distribution may not take the simple form of (5.5.11). Some of these situations are covered by Theorem 5.5.2.

Theorem 5.5.2. Suppose that:

(i) The assumptions (5.5.6) and (5.5.7) of Lemma 5.5.2 hold.

(ii) The first and second derivatives of $Q_n(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ are continuous for all n on a convex, compact neighborhood S containing $\boldsymbol{\theta}^0$ as an interior point.

(iii) $n^{1/2}\,\partial Q_n(\boldsymbol{\theta}^0)/\partial\boldsymbol{\theta} \xrightarrow{\mathscr{L}} 2\mathbf{X}$, where $\mathbf{X}$ represents the limiting distribution.

(iv) There exists a $k \times k$ fixed matrix function $\mathbf{B}(\boldsymbol{\theta})$ such that $\mathbf{B}(\boldsymbol{\theta})$ is continuous on S, $\mathbf{B}(\boldsymbol{\theta}^0)$ is a symmetric, positive definite matrix, and

$$\lim_{n\to\infty}\left[\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} - 2\mathbf{B}(\boldsymbol{\theta})\right] = \mathbf{0}$$

almost surely, uniformly on S.

Let $\hat{\boldsymbol{\theta}}_n$ be a measurable function that minimizes the objective function $Q_n(\boldsymbol{\theta})$. Then

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{\mathscr{L}} -\mathbf{B}^{-1}(\boldsymbol{\theta}^0)\mathbf{X}.$$

Proof. By assumption i, the estimator $\hat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}^0$ almost surely, and $\hat{\boldsymbol{\theta}}_n$ lies inside the convex, compact neighborhood S of $\boldsymbol{\theta}^0$ for large n, almost surely. Therefore, we can expand the partial derivative of $Q_n(\hat{\boldsymbol{\theta}}_n)$ with respect to $\boldsymbol{\theta}$ in a Taylor series about $\boldsymbol{\theta}^0$, for large n, almost surely. For $\ddot{\boldsymbol{\theta}}_n$ on the line segment joining $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}^0$ we have

$$\begin{aligned} \mathbf{0} &= n^{1/2} \frac{\partial Q_n(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \\ &= n^{1/2} \frac{\partial Q_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 Q_n(\ddot{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \end{aligned} \tag{5.5.18}$$

for large n, almost surely. By assumption iv, the second partial derivative of $Q_n(\boldsymbol{\theta})$ evaluated at $\ddot{\boldsymbol{\theta}}_n$ is converging to $2\mathbf{B}(\boldsymbol{\theta}^0)$ a.s. Therefore, using assumption iii, (5.5.18), and Corollary 5.2.6.2, we have the conclusion. ▲

If, in Theorem 5.5.2, $\mathbf{X} \sim N[\mathbf{0}, \boldsymbol{\Sigma}_{XX}]$, then

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{\mathscr{L}} N[\mathbf{0}, \boldsymbol{B}^{-1}(\boldsymbol{\theta}^0)\boldsymbol{\Sigma}_{XX}\mathbf{B}^{-1}(\boldsymbol{\theta}^0)]\,.$$

If $\boldsymbol{\Sigma}_{XX} = \mathbf{B}(\boldsymbol{\theta}^0)\sigma^2$, we obtain the result of Theorem 5.5.1 that the limiting distribution is $N[\mathbf{0}, \mathbf{B}^{-1}(\boldsymbol{\theta}^0)\sigma^2]$.

In Theorem 5.5.1 and Theorem 5.5.2, it is $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)$ that converges in distribution to a nondegenerate random variable. In assumption iii of Theorem 5.5.1, the matrix of derivatives divided by n is assumed to converge to a nonsingular matrix. The normalization of the sums of squares by n and the corresponding normalization of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0$ with $n^{1/2}$ is not always appropriate.

In time series analyses, it sometimes happens that the sum of squares of the derivative with respect to one parameter increases at a faster rate than the derivative with respect to a second parameter. Consider the simple model

$$Y_t = \beta_0 + \beta_1 t + e_t\,,$$

with vector of derivatives $\mathbf{F}(\mathbf{x}_t; \boldsymbol{\beta}) = [1, t]$. It follows that the sum of squares of the derivative with respect to β_0 increases at the rate n, while the sum of squares of the derivative with respect to β_1 increases at the rate n^3. Therefore, Theorem 5.5.1 could not be used to obtain the limiting distribution of the least squares estimator of the parameter vector. Following Sarkar (1990), we now give a theorem, Theorem 5.5.3, that is applicable to situations in which the sums of squares of the first derivatives increase at different rates. Standard results are obtained by setting $\mathbf{M}_n$ and a_n of the theorem equal to $n^{1/2}\mathbf{I}$ and $n^{1/2}$, respectively.

The assumptions of Theorem 5.5.3 are local in that assumption d is a local property of $Q_n(\boldsymbol{\theta})$ in the vicinity of $\boldsymbol{\theta}^0$. This assumption permits the existence of other local minima. Thus, the conclusion of Theorem 5.5.3 is local in that the sequence of estimators associated with the local region of assumption d are consistent for $\boldsymbol{\theta}^0$.

We assume that an estimator of $\boldsymbol{\theta}$ is constructed by minimizing the function $Q_n(\boldsymbol{\theta})$. Let the vector of first partial derivatives be

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2\mathbf{U}_n(\boldsymbol{\theta})$$

and the matrix of second partial derivatives be

$$\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} = 2\mathbf{B}_n(\boldsymbol{\theta}).$$

If the minimum of $Q_n(\boldsymbol{\theta})$ occurs in the interior of the parameter space, then the $\boldsymbol{\theta}$ that minimizes $Q_n(\boldsymbol{\theta})$ is a solution of the system

$$\mathbf{U}_n(\boldsymbol{\theta}) = \mathbf{0}. \tag{5.5.19}$$

Theorem 5.5.3. Suppose the true value $\boldsymbol{\theta}^0$ is an interior point of the parameter space $\boldsymbol{\Theta}$. Assume that there is a sequence of square matrices $\{\mathbf{M}_n\}$ and a sequence of real numbers $\{a_n\}$ such that

(a) $\lim_{n\to\infty} \mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta}^0)\mathbf{M}_n^{-1\prime} = \mathbf{B}$ a.s., where $\mathbf{B}$ is a $k \times k$ symmetric, positive definite matrix, a.s.
(b) $\lim_{n\to\infty} a_n^{-1}\mathbf{M}_n^{-1}\mathbf{U}_n(\boldsymbol{\theta}^0) = 0$ a.s.
(c) $\lim_{n\to\infty} a_n = \infty$.
(d) There exists a $\delta > 0$ and random variables L_n such that for all n, all $\boldsymbol{\theta}$ in $S_{n\delta_0}$, and all δ_0, $0 < \delta_0 < \delta$,

$$\|\mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta})\mathbf{M}_n^{-1\prime} - \mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta}^0)\mathbf{M}_n^{-1\prime}\| \le L_n\delta_0 \qquad \text{a.s.},$$

and

$$\limsup_{n\to\infty} L_n < \infty \qquad \text{a.s.},$$

where

$$S_{n\delta_0} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta}: \|a_n^{-1}\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\| < \delta_0\}$$

and for any $k \times r$ matrix $\mathbf{C}$, $\|\mathbf{C}\|^2 = \operatorname{tr} \mathbf{C}\mathbf{C}'$ denotes the sum of squares of the elements of $\mathbf{C}$.

Then there exists a sequence of roots of (5.5.19), denoted by $\{\tilde{\boldsymbol{\theta}}_n\}$, such that

$$\lim_{n\to\infty} \|a_n^{-1}\mathbf{M}_n'(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)\| = 0 \qquad \text{a.s.}$$

Proof. Given $\epsilon > 0$ and $0 < \delta_0 \le \delta$, it is possible to define a set A, an n_0, and positive constants K_1, K_2, and K_3 such that $P\{A\} \ge 1 - \epsilon$ and for all ω in A and all

$n > n_0$,

$$a_n > K_1\,, \tag{5.5.20}$$

$$\gamma_{\min}[\mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta}^0)\mathbf{M}_n^{-1\prime}] \geq 3K_2\,, \tag{5.5.21}$$

$$\|\mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta}^0)\mathbf{M}_n^{-1\prime} - \mathbf{B}\|^2 < K_2\,, \tag{5.5.22}$$

$$\|M_n^{-1}[\mathbf{B}_n(\boldsymbol{\theta}) - \mathbf{B}_n(\boldsymbol{\theta}^0)]\mathbf{M}_n^{-1\prime}\| < K_3\delta_0\,,$$

and $L_n(\omega) \leq K_3$ for all $\boldsymbol{\theta}$ in $S_{n\delta_0}$, by assumptions c, a, a, b, and d, respectively, where $\gamma_{\min}(\mathbf{C})$ is the smallest root of the matrix $\mathbf{C}$. Let $\delta_0^* = \min(\delta_0, K_3^{-1}K_2)$. Then, for ω in A, $\boldsymbol{\theta}$ in $S_{n\delta_0^*}$, and $n > n_0$, we have

$$\|\mathbf{d}_n\| < K_2\delta_0^* \tag{5.5.23}$$

and

$$\|\mathbf{M}_n^{-1}[\mathbf{B}_n(\boldsymbol{\theta}) - \mathbf{B}_n(\boldsymbol{\theta}^0)]\mathbf{M}_n^{-1\prime}\| < K_2\,, \tag{5.5.24}$$

where $\|\mathbf{d}_n\| = \|0.5a_n^{-1}\mathbf{M}_n^{-1}\mathbf{U}_n(\boldsymbol{\theta}^0)\|$.

Let the Taylor expansion of $Q_n(\boldsymbol{\theta})$ about $\boldsymbol{\theta}^0$ be

$$\begin{aligned} Q_n(\boldsymbol{\theta}) &= Q_n(\boldsymbol{\theta}^0) + 0.5(\boldsymbol{\theta} - \boldsymbol{\theta}^0)'\mathbf{U}_n(\boldsymbol{\theta}^0) + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)'\mathbf{B}_n(\overset{*}{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \boldsymbol{\theta}^0) \\ &= Q_n(\boldsymbol{\theta}^0) + 0.5(\boldsymbol{\theta} - \boldsymbol{\theta}^0)'\mathbf{U}_n(\boldsymbol{\theta}^0) + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)'[\mathbf{B}_n(\overset{*}{\boldsymbol{\theta}}_n) - \mathbf{B}_n(\boldsymbol{\theta}^0)](\boldsymbol{\theta} - \boldsymbol{\theta}^0) \\ &\quad + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)'\mathbf{B}_n(\boldsymbol{\theta}^0)(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\,, \end{aligned} \tag{5.5.25}$$

where $\overset{*}{\boldsymbol{\theta}}_n$ lies on the line segment joining $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^0$. Using (5.5.24),

$$(\boldsymbol{\theta} - \boldsymbol{\theta}^0)'\mathbf{M}_n\{\mathbf{M}_n^{-1}[\mathbf{B}_n(\overset{*}{\boldsymbol{\theta}}_n) - \mathbf{B}_n(\boldsymbol{\theta}^0)]\mathbf{M}_n^{-1\prime}\}\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0) \leq \|\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2 K_2 \tag{5.5.26}$$

for all $\boldsymbol{\theta}$ in $S_{n\delta_0^*}$. By (5.5.21),

$$(\boldsymbol{\theta} - \boldsymbol{\theta}^0)'\mathbf{B}_n(\boldsymbol{\theta}^0)(\boldsymbol{\theta} - \boldsymbol{\theta}^0) \geq 3K_2\|\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2\,. \tag{5.5.27}$$

Let $R(S_{n\delta_0^*})$ be the boundary of $S_{n\delta_0^*}$,

$$R(S_{n\delta_0^*}) = \{\boldsymbol{\theta} \in \boldsymbol{\Theta}: \|a_n^{-1}\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2 = \delta_0^{*2}\}\,.$$

From the expansion (5.5.25), using (5.5.26), (5.5.27), and (5.5.20),

$$Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}^0) \geq \|\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2[2K_2 + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)'\|\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^{-2}a_n\mathbf{M}_n\mathbf{d}_n] \tag{5.5.28}$$

and

$$\|\mathbf{M}_n'(\boldsymbol{\theta}-\boldsymbol{\theta}^0)\|^2 = \|a_n^{-1}\mathbf{M}_n'(\boldsymbol{\theta}-\boldsymbol{\theta}^0)\|^2 a_n^2 \geq \delta_0^{*2}K_1^2 \tag{5.5.29}$$

for all ω in A, all $n > n_0$, and all $\boldsymbol{\theta}$ in $R(S_{n\delta_0^*})$. Then, by (5.5.23) and (5.5.29),

$$\|\mathbf{M}_n'(\boldsymbol{\theta}-\boldsymbol{\theta}^0)\|^{-2}\|a_n(\boldsymbol{\theta}-\boldsymbol{\theta}^0)'\mathbf{M}_n\mathbf{d}_n\| \leq a_n\|\mathbf{M}_n'(\boldsymbol{\theta}-\boldsymbol{\theta}^0)\|^{-1}\|\mathbf{d}_n\| \leq \delta_0^{*-1}(K_2\delta_0^*)$$

and

$$Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}^0) \geq \delta_0^{*2}K_1K_2$$

for all ω in A, all $n > n_0$, and all θ in $R(S_{n\delta_0^*})$. It follows that $Q_n(\boldsymbol{\theta})$ must attain a minimum in the interior of $S_{n\delta_0^*}$, at which point the system (5.5.19) is satisfied. Therefore,

$$P\{\|a_n^{-1}\mathbf{M}_n'(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)\| < \delta_0^*\} \geq 1 - \epsilon$$

for $n > n_0$.

Now, let A_j, n_j, and δ_j^* denote A, n_0, and δ_0^* associated with $\epsilon_j = 2^{-j}$. Since δ_0 is arbitrary, we can choose the δ_j^* to be a sequence decreasing to zero and take n_j to be an increasing sequence. Define $A = \liminf A_j$. Then $P(A) = 1$. Given $\eta > 0$ and a fixed ω in A, there exists a j_0 such that ω is in A_j for all $j > j_0$. Choose $j_1 > j_0$ such that $\delta_{j_1} < \eta$. Then, for $j > j_1$ and $n > n_j$, we have

$$\|a_n^{-1}\mathbf{M}_n'(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)\| < \delta_j < \eta\,. \qquad \blacktriangle$$

Corollary 5.5.3.1. Let assumptions a through d of Theorem 5.5.3 hold in probability instead of almost surely. Then there exists a sequence of roots of (5.5.19), denoted by $\{\tilde{\boldsymbol{\theta}}_n\}$, such that

$$\operatorname*{plim}_{n\to\infty} \|a_n^{-1}\mathbf{M}_n'(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)\| = 0\,.$$

Proof. Omitted. ▲

If the roots of $a_n^{-2}\mathbf{M}_n\mathbf{M}_n'$ are bounded away from zero, then $\tilde{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}^0$. If $\tilde{\boldsymbol{\theta}}_n$ is consistent, and if the properly normalized first and second derivatives converge in distribution, then the estimator of $\boldsymbol{\theta}$ converges in distribution.

Corollary 5.5.3.2. Let assumptions a through d of Theorem 5.5.3 hold in probability. In addition, assume $\tilde{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}^0$ and

(e) $\{\mathbf{M}_n^{-1}\mathbf{U}_n(\boldsymbol{\theta}^0),\ \mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta}^0)\mathbf{M}_n^{-1\prime}\} \xrightarrow{\mathscr{L}} (\mathbf{X}, \mathbf{B})$ as $n \to \infty$.

Then

$$\mathbf{M}_n'(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{\mathscr{L}} -\mathbf{B}^{-1}\mathbf{X} \qquad \text{as} \quad n \to \infty\,,$$

where $\mathbf{B}$ is defined in assumption a and $\mathbf{U}_n(\boldsymbol{\theta}^0)$ is the vector of first partial derivatives evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$.

Proof. Because $\tilde{\boldsymbol{\theta}}_n$ is converging to $\boldsymbol{\theta}^0$ in probability, there is a compact set S containing $\boldsymbol{\theta}^0$ as an interior point such that, given $\epsilon > 0$, the probability that $\tilde{\boldsymbol{\theta}}_n$ is in S is greater than $1 - \epsilon$ for n greater than some N_ϵ. For $\tilde{\boldsymbol{\theta}}_n$ in S, we expand $\mathbf{U}_n(\boldsymbol{\theta})$ in a first order Taylor series about $\boldsymbol{\theta}^0$ and multiply the Taylor series expansion by $\mathbf{M}_n^{-1}$ to obtain

$$0 = \mathbf{M}_n^{-1}\mathbf{U}_n(\tilde{\boldsymbol{\theta}}_n) = \mathbf{M}_n^{-1}\mathbf{U}_n(\boldsymbol{\theta}^0) + \mathbf{M}_n^{-1}\mathbf{B}_n(\ddot{\boldsymbol{\theta}}_n)\mathbf{M}_n^{-1\prime}\mathbf{M}_n'(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0),$$

where $\ddot{\boldsymbol{\theta}}_n$ is an intermediate point between $\tilde{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}^0$. Note that

$$\mathbf{M}_n^{-1}\mathbf{B}_n(\ddot{\boldsymbol{\theta}}_n)\mathbf{M}_n^{-1\prime} = \mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta}^0)\mathbf{M}_n^{-1\prime} + \mathbf{M}_n^{-1}[\mathbf{B}_n(\ddot{\boldsymbol{\theta}}_n) - \mathbf{B}_n(\boldsymbol{\theta}^0)]\mathbf{M}_n^{-1\prime}$$

and the conclusion follows by assumptions a, d, and e. ▲

By Theorem 5.5.3, there is a sequence of roots of (5.5.19) that converges to $\boldsymbol{\theta}^0$. Hence, the assumption of the corollary that $\tilde{\boldsymbol{\theta}}$ is consistent is equivalent to assuming that we are considering the consistent sequence of roots.

Example 5.5.1. Consider the model

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 x_{t1} + \beta_0\beta_1 t + e_t \\ &= f[(1, x_{t1}, t), (\beta_0, \beta_1)] + e_t, \end{aligned} \tag{5.5.30}$$

where $e_t \sim \mathrm{NI}(0, \sigma^2)$, $\beta_1 \neq 0$, and

$$x_{t1} = \begin{cases} 1 & \text{if } t \text{ is odd}, \\ -1 & \text{if } t \text{ is even}. \end{cases}$$

Assume that a sample of size n is available and that an estimator of $(\beta_0, \beta_1)' = \boldsymbol{\beta}$ is constructed as the (β_0, β_1) that minimizes

$$Q_n(\boldsymbol{\beta}) = n^{-1} \sum_{t=1}^{n} (Y_t - \beta_0 - \beta_1 x_{t1} - \beta_0\beta_1 t)^2 .$$

The vector of first derivatives of $f[(1, x_{t1}, t), (\beta_0, \beta_1)]$ is

$$\mathbf{F}_t(\boldsymbol{\beta}) = (1 + \beta_1 t, x_{t1} + \beta_0 t) .$$

We note that $n^{-1} \sum_{t=1}^n t^2$ increases at the rate n^2. Therefore, if $\beta_0^0\beta_1^0 \neq 0$, $Q_n(\boldsymbol{\beta})$ will not converge for any (β_0, β_1) for which $\beta_0\beta_1 \neq \beta_0^0\beta_1^0$. Also,

$$\lim_{n\to\infty} n^{-3} \sum_{t=1}^{n} \mathbf{F}_t'(\boldsymbol{\beta})\mathbf{F}_t(\boldsymbol{\beta})$$

is singular if $\beta_0 \neq 0$ and $\beta_1 \neq 0$.

We now show how we can apply Theorem 5.5.3 to this problem. Let

$$\mathbf{M}_n^{-1\prime} = \begin{pmatrix} 1 & -\beta_0^0 \\ 0 & \beta_1^0 \end{pmatrix} \begin{pmatrix} n^2 & 0 \\ 0 & 1 \end{pmatrix}^{-1/2}.$$

Then

$$n^{-1/2}(1+\beta_1^0 t, x_{t1} + \beta_0^0 t)\mathbf{M}_n^{-1\prime} = \{n^{-3/2}(1+\beta_1^0 t), n^{-1/2}(\beta_1^0 x_{t1} - \beta_0^0)\}$$

and

$$\lim_{n\to\infty} \mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta}^0)\mathbf{M}_n^{-1\prime} = \begin{pmatrix} 3^{-1}(\beta_1^0)^2 & -2^{-1}\beta_0^0\beta_1^0 \\ -2^{-1}\beta_0^0\beta_1^0 & (\beta_0^0)^2 + (\beta_1^0)^2 \end{pmatrix}. \qquad (5.5.31)$$

Setting $a_n = n^{1/2}$, we have

$$n^{-1/2}\sum_{t=1}^{n} [n^{-3/2}(1+\beta_1^0 t)e_t, n^{-1/2}(\beta_1^0 x_{t1} - \beta_0^0)e_t] \to (0,0) \qquad \text{a.s.}$$

The second derivatives of $Q_n(\boldsymbol{\beta})$ are continuous with continuous derivatives, and hence condition d of Theorem 5.5.3 also holds. It follows that the estimator of $\boldsymbol{\beta}$ is consistent for $\boldsymbol{\beta}$. Using Theorem 5.3.4, one can show that

$$\mathbf{M}_n^{-1}\mathbf{U}_n(\boldsymbol{\beta}^0) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{V}_{\beta\beta}),$$

where $\mathbf{V}_{\beta\beta}$ is the matrix of (5.5.31) multiplied by σ^2.

An alternative method of defining a sequence of estimators with a nonsingular covariance matrix is to reparametrize the model. Assume that $\beta_1 \neq 0$. Let

$$(\alpha_0, \alpha_1) = (\beta_0\beta_1, \beta_1),$$

from which

$$(\beta_0, \beta_1) = (\alpha_1^{-1}\alpha_0, \alpha_1),$$

and

$$Y_t = \alpha_1^{-1}\alpha_0 + \alpha_1 x_{t1} + \alpha_0 t + e_t.$$

The vector of first derivatives of the $\boldsymbol{\alpha}$-form of the model is

$$\mathbf{G}_t(\boldsymbol{\alpha}) = (\alpha_1^{-1} + t, -\alpha_1^{-2}\alpha_0 + x_{t1}).$$

It follows that

$$\lim_{n\to\infty} \mathbf{H}_n^{-1/2}\sum_{t=1}^{n} \mathbf{G}_t'(\boldsymbol{\alpha})\mathbf{G}_t(\boldsymbol{\alpha})\mathbf{H}_n^{-1/2} = \mathbf{A},$$

where

$$\mathbf{A} = \begin{bmatrix} \frac{1}{3} & -0.5a_1^{-2}\alpha_0 \\ -0.5\alpha_1^{-2}\alpha_0 & 1 + \alpha_0^2\alpha_1^{-4} \end{bmatrix}$$

and $\mathbf{H}_n = \text{diag}(n^3, n)$. It can be shown that

$$\mathbf{H}_n^{-1/2} \sum_{t=1}^{n} \mathbf{G}_t(\boldsymbol{\alpha})e_t \xrightarrow{\mathscr{L}} N[\mathbf{0}, \mathbf{A}\sigma^2] .$$

Thus, the least squares estimator of $\boldsymbol{\beta}$ is consistent and there is a normalizer that gives a limiting normal distribution. However, the two normalizers that we used are functions of the true, unknown parameters. The practitioner prefers a distribution based on a normalizer that is a function of the sample.

It is natural to consider a normalizer derived from the derivatives evaluated at the least squares estimator. For the model in the $\boldsymbol{\alpha}$-form, let

$$\mathbf{K}_n = \sum_{t=1}^{n} \mathbf{G}_t'(\hat{\boldsymbol{\alpha}})\mathbf{G}_t(\hat{\boldsymbol{\alpha}}) .$$

One can show that

$$\mathbf{H}_n^{-1/2}\mathbf{K}_n\mathbf{H}_n^{-1/2} \xrightarrow{P} \mathbf{A}$$

and hence

$$\mathbf{K}_n^{-1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{I}\sigma^2) .$$

It follows that one can use the derivatives output by most nonlinear least squares programs to construct approximate tests and confidence intervals for the parameters of the model. ▲▲

Example 5.5.2. As an example where Theorem 5.5.3 does not hold, consider the model

$$Y_t = \beta + \beta^2 t + e_t ,$$

where $F(\mathbf{x}_t; \beta) = (1 + 2\beta t)$ and $e_t \sim \text{NI}(0, \sigma^2)$. Assume that the model is to be estimated by least squares by choosing β to minimize

$$Q_n = n^{-1} \sum_{t=1}^{n} (Y_t - \beta - \beta^2 t)^2 .$$

Then

$$U_n(\beta^0) = -n^{-1} \sum_{t=1}^{n} e_t(1 + 2\beta^0 t)$$

and

$$B_n(\beta^0) = n^{-1} \sum_{t=1}^{n} [(1 + 2\beta^0 t)^2 + 2te_t] .$$

If $\beta^0 = 0$,

$$B_n(\beta^0) = 1 + 2n^{-1} \sum_{t=1}^{n} te_t .$$

The quantity $n^{-1} \sum_{t=1}^{n} te_t$ does not converge to zero in probability. If one divides $B_n(\beta^0)$ by $n^{1/2}$, the normalized quantity converges in distribution, but not in probability. If one divides $B_n(\beta^0)$ by a quantity increasing at a rate faster than $n^{1/2}$, the limit is zero. Therefore, the model with $\beta^0 = 0$ does not satisfy assumption a of Theorem 5.5.3. However, one can use Lemma 5.5.1 to demonstrate that the least squares estimator is a consistent estimator of β^0 when $\beta^0 = 0$. ▲▲

5.5.2. One-Step Estimation

An important special case of nonlinear estimation occurs when one is able to obtain an initial consistent estimator of the parameter vector $\boldsymbol{\theta}^0$. In such cases, asymptotic results can be obtained for a single step or for a finite number of steps of a sequential minimization procedure.

We retain the model (5.5.1), which we write as

$$Y_t = f(\mathbf{x}_t; \boldsymbol{\theta}^0) + e_t , \qquad t = 1, 2, \ldots , \tag{5.5.32}$$

where the e_t are iid$(0, \sigma^2)$ random variables or are $(0, \sigma^2)$ martingale differences satisfying condition ii of Theorem 5.5.1. Let $\hat{\boldsymbol{\theta}}$ be an initial consistent estimator of the unknown true value $\boldsymbol{\theta}^0$. A Taylor's series expansion of $f(\mathbf{x}_t; \boldsymbol{\theta}^0)$ about $\hat{\boldsymbol{\theta}}$ gives

$$f(\mathbf{x}_t; \boldsymbol{\theta}^0) = f(\mathbf{x}_t; \hat{\boldsymbol{\theta}}) + \sum_{j=1}^{k} f^{(j)}(\mathbf{x}_t; \hat{\boldsymbol{\theta}})(\theta_j^0 - \hat{\theta}_j) + d_t(\mathbf{x}_t; \hat{\boldsymbol{\theta}}) , \tag{5.5.33}$$

where θ_j^0 and $\hat{\theta}_j$ are the jth elements of $\boldsymbol{\theta}^0$ and $\hat{\boldsymbol{\theta}}$, respectively,

$$d_t(\mathbf{x}_t; \hat{\boldsymbol{\theta}}) = \frac{1}{2} \sum_{r=1}^{k} \sum_{j=1}^{k} f^{(jr)}(\mathbf{x}_t; \overset{*}{\boldsymbol{\theta}}_t)(\theta_j^0 - \hat{\theta}_j)(\theta_r^0 - \hat{\theta}_r) ,$$

and $\overset{*}{\boldsymbol{\theta}}_t$ is on the line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$. On the basis of equation (5.5.33), we consider the modified sum of squares

$$\begin{aligned} \hat{Q}(\boldsymbol{\delta}) &= n^{-1} \sum_{t=1}^{n} \left\{ Y_t - f(\mathbf{x}_t; \hat{\boldsymbol{\theta}}) - \sum_{j=1}^{k} f^{(j)}(\mathbf{x}_t; \hat{\boldsymbol{\theta}})[\theta_j - \hat{\theta}_j] \right\}^2 \\ &= n^{-1}[\mathbf{w} - \mathbf{F}(\hat{\boldsymbol{\theta}})\boldsymbol{\delta}]'[\mathbf{w} - \mathbf{F}(\hat{\boldsymbol{\theta}})\boldsymbol{\delta}] , \end{aligned}$$

where $\boldsymbol{\delta} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, $\mathbf{w}$ is the $n \times 1$ vector with tth element given by

$$w_t = Y_t - f(\mathbf{x}_t; \hat{\boldsymbol{\theta}}),$$

and $\mathbf{F}(\boldsymbol{\theta})$ is the $n \times k$ matrix with tth row $F_t(\boldsymbol{\theta})$ defined in (5.5.2). Minimizing $\hat{Q}(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ leads to

$$\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} + \tilde{\boldsymbol{\delta}} \tag{5.5.34}$$

as an estimator of $\boldsymbol{\theta}^0$, where

$$\tilde{\boldsymbol{\delta}} = [\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1}\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{w}.$$

We call $\tilde{\boldsymbol{\theta}}$ the one-step Gauss–Newton estimator.

Theorem 5.5.4. Assume that the nonlinear model (5.5.32) holds. Let $\tilde{\boldsymbol{\theta}}$ be the one-step Gauss–Newton estimator defined in (5.5.34). Assume:

(1) There is an open set S such that S is in Θ, $\boldsymbol{\theta}^0 \in S$, and

$$\operatorname*{plim}_{n\to\infty} n^{-1}\mathbf{F}'(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta})$$

is nonsingular for all $\boldsymbol{\theta}$ in S, where $\mathbf{F}(\boldsymbol{\theta})$ is the $n \times k$ matrix with tjth element given by $f^{(j)}(\mathbf{x}_t; \boldsymbol{\theta})$.

(2) $\operatorname{plim}_{n\to\infty} n^{-1}\mathbf{G}'(\boldsymbol{\theta})\mathbf{G}(\boldsymbol{\theta}) = \mathbf{L}(\boldsymbol{\theta})$ uniformly in $\boldsymbol{\theta}$ on the closure $\bar{S}$ of S, where the elements of $\mathbf{L}(\boldsymbol{\theta})$ are continuous functions of $\boldsymbol{\theta}$ on $\bar{S}$, and $\mathbf{G}(\boldsymbol{\theta})$ is an $n \times (1 + k + k^2 + k^3)$ matrix with tth row given by

$$\begin{aligned}&[f(\mathbf{x}_t; \boldsymbol{\theta}), f^{(1)}(\mathbf{x}_t; \boldsymbol{\theta}), f^{(2)}(\mathbf{x}_t; \boldsymbol{\theta}), \ldots, f^{(k)}(\mathbf{x}_t; \boldsymbol{\theta}),\\ &f^{(11)}(\mathbf{x}_t; \boldsymbol{\theta}), f^{(12)}(\mathbf{x}_t; \boldsymbol{\theta}), \ldots, f^{(1k)}(\mathbf{x}_t; \boldsymbol{\theta}), f^{(21)}(\mathbf{x}_t; \boldsymbol{\theta}), \ldots, f^{(kk)}(\mathbf{x}_t; \boldsymbol{\theta}),\\ &f^{(111)}(\mathbf{x}_t; \boldsymbol{\theta}), \ldots, f^{(kkk)}(\mathbf{x}_t; \boldsymbol{\theta})].\end{aligned}$$

(3) The initial estimator of $\boldsymbol{\theta}^0$, say $\hat{\boldsymbol{\theta}}$, satisfies $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) = O_p(a_n)$, where $\lim_{n\to\infty} a_n = 0$.

Then

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = [\mathbf{F}'(\boldsymbol{\theta}^0)\mathbf{F}(\boldsymbol{\theta}^0)]^{-1}\mathbf{F}'(\boldsymbol{\theta}^0)\mathbf{e} + O_p(\max\{a_n^2, a_n n^{-1/2}\}).$$

Furthermore, if $n^{-1/2}\sum_{t=1}^n \mathbf{F}_t'(\boldsymbol{\theta}^0)e_t \xrightarrow{\mathscr{L}} N[\mathbf{0}, \boldsymbol{B}(\boldsymbol{\theta}^0)\sigma^2]$ and $a_n^2 = o(n^{-1/2})$, then

$$n^{1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{\mathscr{L}} N[\mathbf{0}, \mathbf{B}^{-1}(\boldsymbol{\theta}^0)\sigma^2].$$

Proof. Because $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0$ is $O_p(a_n)$, given $\epsilon_0 > 0$, one can choose N_0 such that the probability is greater than $1-\epsilon_0$ that $\hat{\boldsymbol{\theta}}$ is in S for all $n > N_0$. On the basis of a Taylor expansion, we can write, for $\hat{\boldsymbol{\theta}}$ in S,

$$\begin{aligned}\tilde{\boldsymbol{\delta}} &= [\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1}\mathbf{F}'(\hat{\boldsymbol{\theta}})[\mathbf{f}(\boldsymbol{\theta}^0) - \mathbf{f}(\hat{\boldsymbol{\theta}}) + \mathbf{e}] \\ &= [\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1}\mathbf{F}'(\hat{\boldsymbol{\theta}})[\mathbf{F}(\hat{\boldsymbol{\theta}})\boldsymbol{\delta}^0 + \mathbf{e}] \\ &\quad + [\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1}\mathbf{R}(\hat{\boldsymbol{\theta}}),\end{aligned}$$

where $\mathbf{f}(\boldsymbol{\theta})$ is the $n \times 1$ vector with tth element $f(\mathbf{x}_t; \boldsymbol{\theta})$, $\mathbf{e}$ is the $n \times 1$ vector with tth element e_t, $\boldsymbol{\delta}^0 = \boldsymbol{\theta}^0 - \hat{\boldsymbol{\theta}}$, and the jth element of $n^{-1}\mathbf{R}(\hat{\boldsymbol{\theta}})$ is

$$(2n)^{-1} \sum_{r=1}^{k} \sum_{s=1}^{k} \sum_{t=1}^{n} f^{(j)}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}) f^{(rs)}(\mathbf{x}_t; \ddot{\boldsymbol{\theta}})[\theta_r^0 - \hat{\theta}_r][\theta_s^0 - \hat{\theta}_s],$$

in which $\ddot{\boldsymbol{\theta}}$ is on the line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$. The elements of $\mathbf{L}(\boldsymbol{\theta})$ are bounded on $\bar{S}$, and, given $\epsilon_1 > 0$, there exists an N_1 such that the probability is greater than $1 - \epsilon_1$ that the elements of

$$n^{-1}\left\{\sum_{t=1}^{n} [f^{(j)}(\mathbf{x}_t; \boldsymbol{\theta})]^2, \sum_{t=1}^{n} [f^{(rs)}(\mathbf{x}_t; \boldsymbol{\theta})]^2\right\}$$

differ from the respective elements of $\mathbf{L}(\boldsymbol{\theta})$ by less than ϵ_1 for all $\boldsymbol{\theta} \in \bar{S}$ and all $n > N_1$. Therefore,

$$(2n)^{-1} \sum_{t=1}^{n} f^{(j)}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}) f^{(rs)}(\mathbf{x}_t; \ddot{\boldsymbol{\theta}}) = O_p(1), \qquad j, r, s = 1, 2, \ldots, k,$$

and

$$n^{-1}\mathbf{R}(\hat{\boldsymbol{\theta}}) = O_p(a_n^2). \tag{5.5.35}$$

Because $\boldsymbol{B}(\boldsymbol{\theta})$ is nonsingular on S, there is an N_2 such that $\mathbf{F}'(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta})$ is nonsingular with probability greater than $1 - \epsilon_2$ for all $\boldsymbol{\theta}$ in S and all $n > N_2$. It follows that

$$n^{-1}\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}}) = n^{-1}\mathbf{F}'(\boldsymbol{\theta}^0)\mathbf{F}(\boldsymbol{\theta}^0) + O_p(a_n)$$

and

$$p\lim[n^{-1}\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1} = \mathbf{B}^{-1}(\boldsymbol{\theta}^0),$$

where the jsth element of $\mathbf{B}(\boldsymbol{\theta}^0)$ is

$$\{\mathbf{B}(\boldsymbol{\theta}^0)\}_{js} = \underset{n\to\infty}{p\lim}\left[n^{-1} \sum_{t=1}^{n} f^{(j)}(\mathbf{x}_t; \boldsymbol{\theta}^0) f^{(s)}(\mathbf{x}_t; \boldsymbol{\theta}^0)\right].$$

Therefore,

$$[\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1}\mathbf{R}(\hat{\boldsymbol{\theta}}) = O_p(a_n^2)\,.$$

The jth element of $n^{-1}\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{e}$ is

$$\begin{aligned} n^{-1}\sum_{t=1}^{n} f^{(j)}(\mathbf{x}_t;\hat{\boldsymbol{\theta}})e_t &= n^{-1}\sum_{t=1}^{n} f^{(j)}(\mathbf{x}_t;\boldsymbol{\theta}^0)e_t \\ &\quad + n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{k} f^{(js)}(\mathbf{x}_t;\boldsymbol{\theta}^0)[\hat{\theta}_s-\theta_s^0]e_t \\ &\quad + (2n)^{-1}\sum_{t=1}^{n}\sum_{s=1}^{k}\sum_{r=1}^{k} f^{(jsr)}(\mathbf{x}_t;\boldsymbol{\theta}^\dagger)[\hat{\theta}_s-\theta_s^0][\hat{\theta}_r-\theta_r^0]e_t\,, \end{aligned}$$

where $\boldsymbol{\theta}^\dagger$ is on the line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$. By assumption 2,

$$n^{-1}\sum_{t=1}^{n}[f^{(jsr)}(\mathbf{x}_t;\boldsymbol{\theta}^\dagger)-f^{(jsr)}(\mathbf{x}_t;\boldsymbol{\theta}^0)]^2 \xrightarrow{P} 0$$

as $n\to\infty$. Therefore,

$$n^{-1}\sum_{t=1}^{n}[f^{(jsr)}(\mathbf{x}_t;\boldsymbol{\theta}^\dagger)-f^{(jsr)}(\mathbf{x}_t;\boldsymbol{\theta}^0)]e_t = o_p(1)$$

and

$$n^{-1}\sum_{t=1}^{n} f^{(j)}(\mathbf{x}_t;\hat{\boldsymbol{\theta}})e_t = n^{-1}\sum_{t=1}^{n} f^{(j)}(\mathbf{x}_t;\boldsymbol{\theta}^0)e_t + O_p(\max\{a_n n^{-1/2}\})\,. \tag{5.5.36}$$

The results follow from (5.5.35) and (5.5.36). ▲

To estimate the variance of the limiting distribution of $n^{1/2}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)$, we must estimate $\mathbf{B}^{-1}(\boldsymbol{\theta}^0)$ and σ^2. By the assumptions, $[n^{-1}\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1}$ and $[n^{-1}\mathbf{F}'(\tilde{\boldsymbol{\theta}})\mathbf{F}(\tilde{\boldsymbol{\theta}})]^{-1}$ are consistent estimators for $\mathbf{B}^{-1}(\boldsymbol{\theta}^0)$. Also

$$s^2 = (n-k)^{-1}\sum_{t=1}^{n}[Y_t - f(\mathbf{x}_t;\tilde{\boldsymbol{\theta}})]^2 \tag{5.5.37}$$

is a consistent estimator of σ^2. Hence, using the matrix $[\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1}$ and the mean square s^2, all of the standard linear regression theory holds approximately for $\tilde{\boldsymbol{\theta}}$.

The theorem demonstrates that the one-step estimator is asymptotically unchanged by additional iteration if $\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0 = o_p(n^{-1/4})$. For small samples, we may choose to iterate the procedure. For a particular sample we are not guaranteed that iteration of (5.5.34), using $\tilde{\boldsymbol{\theta}}$ of the previous step as the initial estimator, will converge. Therefore, if one iterates, the estimator $\tilde{\boldsymbol{\theta}}$ should be replaced at each step by

$$\tilde{\boldsymbol{\theta}}_\nu = \hat{\boldsymbol{\theta}} + \nu\tilde{\boldsymbol{\delta}}\,,$$

where $\hat{\boldsymbol{\theta}}$ is the estimator of the previous step and $\nu \in (0, 1]$ is chosen so that $n^{-1}\sum_{t=1}^{n}[Y_t - f(\mathbf{x}_t; \tilde{\boldsymbol{\theta}}_\nu)]^2$ is less than $n^{-1}\sum_{t=1}^{n}[Y_t - f(\mathbf{x}_t; \hat{\boldsymbol{\theta}})]^2$, and so that $\tilde{\boldsymbol{\theta}}_\nu \in \boldsymbol{\Theta}$. This iteration furnishes a method of obtaining the least squares estimator that minimizes (5.5.3).

Example 5.5.3. To illustrate the Gauss–Newton procedure, consider the model

$$Y_t = \theta_0 + \theta_1 x_{t1} + \theta_1^2 x_{t2} + e_t\,, \tag{5.5.38}$$

where the e_t are normal independent $(0, \sigma^2)$ random variables. While the superscript 0 was used on $\boldsymbol{\theta}$ in our derivation to identify the true parameter value, it is not a common practice to use that notation when discussing a particular application of the procedure. Observations generated by this model are given in Table 5.5.1. To obtain initial estimators of θ_0 and θ_1, we ignore the nonlinear restriction and regress Y_t on $x_{t0} \equiv 1$, x_{t1}, and x_{t2}. This gives the regression equation

$$\hat{Y}_t = 0.877 + 1.262x_{t1} + 1.150x_{t2}\,.$$

Assuming that $n^{-1}\sum x_{t1}^2$, $n^{-1}\sum x_{t2}^2$, and $n^{-1}\sum x_{t1}x_{t2}$ converge to form a positive definite matrix, the coefficients for x_{t0} and x_{t1} are estimators for θ_0 and θ_1, respectively, with errors $O_p(n^{-1/2})$. We note that $(1.150)^{1/2}$ is also a consistent estimate of θ_1. Using 0.877 and 1.262 as initial estimators, we compute

$$w_t = Y_t - 0.877 - 1.262x_{t1} - 1.593x_{t2}\,.$$

The rows of $\mathbf{F}(\hat{\boldsymbol{\theta}})$ are given by

$$\mathbf{F}_{t.}(\hat{\boldsymbol{\theta}}) = (1, x_{t1} + 2\hat{\theta}_1 x_{t2})\,, \qquad t = 1, 2, \ldots, 6\,.$$

Table 5.5.1. Data and Regression Variables Used in the Estimation of the Parameters of the Model (5.5.38)

t	Y_t	x_{t0}	x_{t1}	x_{t2}	w_t	$x_{t1} + 2\hat{\theta}_1 x_{t2}$
1	9	1	2	4	−0.777	12.100
2	19	1	7	8	−3.464	27.199
3	11	1	1	9	−5.484	23.724
4	14	1	3	7	−1.821	20.674
5	9	1	7	0	−0.714	7.000
6	3	1	0	2	−1.065	5.050

Regressing w_t on x_{t0} and $x_{t1}+2\hat{\theta}_1 x_{t2}$, we obtain

$$\tilde{\boldsymbol{\delta}}=\begin{pmatrix}0.386\\-0.163\end{pmatrix},$$

$$\tilde{\boldsymbol{\theta}}=\begin{pmatrix}0.877\\1.262\end{pmatrix}+\begin{pmatrix}0.386\\-0.163\end{pmatrix}=\begin{pmatrix}1.263\\1.099\end{pmatrix}$$

as the one-step Gauss–Newton estimate.

Our computations illustrate the nature of the derivatives that enter the covariance matrix of the asymptotic distribution. In practice, we would use one of the several nonlinear least squares programs to estimate the parameters. These programs often have the option of specifying the initial values for the iteration or permitting the program to use a search technique. In this example, we have excellent start values. If we use the start values (0.877, 1.262), the nonlinear least squares estimate is $\hat{\boldsymbol{\theta}}=(1.2413, 1.0913)'$, the estimated covariance matrix is

$$\hat{V}\{\hat{\boldsymbol{\theta}}\}=\begin{pmatrix}1.3756 & -0.0763\\-0.0763 & 0.00054\end{pmatrix},$$

and the residual mean square is 1.7346. We used procedure NLIN of SAS® [SAS (1989)] for the computations. ▲▲

5.6. INSTRUMENTAL VARIABLES

In many applications, estimators of the parameters of an equation of the regression type are desired, but the classical assumption that the matrix of explanatory variables is fixed is violated. Some of the columns of the matrix of explanatory variables may be measured with error and (or) may be generated by a stochastic mechanism such that the assumption that the error in the equation is independent of the explanatory variables becomes suspect.

Let us assume that we have the model

$$\mathbf{y}=\boldsymbol{\Phi}\boldsymbol{\beta}+\mathbf{X}\boldsymbol{\lambda}+\mathbf{z}, \tag{5.6.1}$$

where $\boldsymbol{\beta}$ is a $k_1 \times 1$ vector of unknown parameters, $\boldsymbol{\lambda}$ is a $k_2 \times 1$ vector of unknown parameters, y is an $n \times 1$ vector, $\boldsymbol{\Phi}$ is an $n \times k_1$ matrix, $\mathbf{X}$ is an $n \times k_2$ matrix, and $\mathbf{z}$ is an $n \times 1$ vector of unknown random variables with zero mean. The matrix $\boldsymbol{\Phi}$ is fixed, but the elements of $\mathbf{X}$ may contain a random component that is correlated with $\mathbf{z}$.

Estimators obtained by ordinary least squares may be seriously biased because of the correlation between $\mathbf{z}$ and $\mathbf{X}$. If information is available on variables that do not enter the equation, it may be possible to use these variables to obtain consistent estimators of the parameters. Such variables are called *instrumental variables*. The instrumental variables must be correlated with the variables entering the matrix $\mathbf{X}$ but not with the random components of the model. Assume that we have observations on k_3 instrumental variables, $k_3 \geq k_2$. We denote the $n \times k_3$ matrix of observations on the instrumental variables by $\boldsymbol{\psi}$ and assume the elements of $\boldsymbol{\psi}$ are fixed. We express $\mathbf{X}$ as a linear combination of $\boldsymbol{\Phi}$ and $\boldsymbol{\psi}$:

$$\begin{aligned}\mathbf{X} &= \boldsymbol{\Phi}\boldsymbol{\delta}_1 + \boldsymbol{\psi}\boldsymbol{\delta}_2 + \mathbf{w} \\ &= (\boldsymbol{\Phi} : \boldsymbol{\psi})\boldsymbol{\delta} + \mathbf{w}\,, \end{aligned} \tag{5.6.2}$$

where

$$\boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}'\boldsymbol{\Phi} & \boldsymbol{\Phi}'\boldsymbol{\psi} \\ \boldsymbol{\psi}'\boldsymbol{\Phi} & \boldsymbol{\psi}'\boldsymbol{\psi} \end{pmatrix}^{-1} \begin{pmatrix} E\{\boldsymbol{\Phi}'\mathbf{X}\} \\ E\{\boldsymbol{\psi}'\mathbf{X}\} \end{pmatrix}.$$

Note that the residuals $\mathbf{w}$ follow from the definition of $\boldsymbol{\delta}$. Therefore, $\mathbf{w}$ may be a sum of random and fixed components, but the fixed component is orthogonal to $\boldsymbol{\Phi}$ and $\boldsymbol{\psi}$ by construction.

The instrumental variable estimators we consider are obtained by regressing $\mathbf{X}$ on $\boldsymbol{\Phi}$ and $\boldsymbol{\psi}$, computing the estimated values $\hat{\mathbf{X}}$ from this regression, and then replacing $\mathbf{X}$ by $\hat{\mathbf{X}}$ in the regression equation

$$\mathbf{y} = \boldsymbol{\Phi}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\lambda} + \mathbf{z}\,.$$

The instrumental variable estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are given by the regression of $\mathbf{y}$ on $\boldsymbol{\Phi}$ and $\hat{\mathbf{X}}$:

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}'\boldsymbol{\Phi} & \boldsymbol{\Phi}'\hat{\mathbf{X}} \\ \hat{\mathbf{X}}'\boldsymbol{\Phi} & \hat{\mathbf{X}}'\hat{\mathbf{X}} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\Phi}'\mathbf{y} \\ \hat{\mathbf{X}}'\mathbf{y} \end{pmatrix}, \tag{5.6.3}$$

where

$$\hat{\mathbf{X}} = (\boldsymbol{\Phi} : \boldsymbol{\psi})[(\boldsymbol{\Phi} : \boldsymbol{\psi})'(\boldsymbol{\Phi} : \boldsymbol{\psi})]^{-1}(\boldsymbol{\Phi} : \boldsymbol{\psi})'\mathbf{X}\,.$$

These estimators are called two-stage least squares estimators in the econometrics literature. An alternative instrumental variable estimator is the limited information maximum likelihood estimator. See Johnston (1984) and Fuller (1987).

To investigate the properties of estimator (5.6.3), we assume:

1. $\mathbf{Q}_n = \mathbf{D}_{1n}^{-1}(\boldsymbol{\Phi} : \boldsymbol{\psi})'(\boldsymbol{\Phi} : \boldsymbol{\psi})\mathbf{D}_{1n}^{-1}$ is a nonsingular matrix for all $n > k_1 + k_3$, and $\lim_{n\to\infty} \mathbf{Q}_n = \mathbf{Q}$, where $\mathbf{Q}$ is nonsingular and $\mathbf{D}_{1n}$ is a diagonal matrix whose elements are the square roots of the diagonal elements of $(\boldsymbol{\Phi} : \boldsymbol{\psi})'(\boldsymbol{\Phi} : \boldsymbol{\psi})$.

2. $\mathbf{M}_n$ is nonsingular for all $n > k_1 + k_3$, and

$$\lim_{n\to\infty} \mathbf{M}_n = \mathbf{M}\,,$$

where $\mathbf{M}$ is a positive definite matrix,

$$\mathbf{M}_n = \mathbf{D}_{2n}^{-1}(\mathbf{\Phi}:\bar{\mathbf{X}})'(\mathbf{\Phi}:\bar{\mathbf{X}})\mathbf{D}_{2n}^{-1},$$
$$\bar{\mathbf{X}} = (\mathbf{\Phi}:\boldsymbol{\psi})\boldsymbol{\delta},$$

and $\mathbf{D}_{2n}$ is a diagonal matrix whose elements are the square roots of the diagonal elements of $(\mathbf{\Phi}:\bar{\mathbf{X}})'(\mathbf{\Phi}:\bar{\mathbf{X}})$.

3. $\lim_{n\to\infty} \mathbf{R}_n = \mathbf{R}$, where $\mathbf{R}$ is finite and

$$\mathbf{R}_n = E\{\mathbf{D}_{2n}^{-1}[\mathbf{\Phi}:\bar{\mathbf{X}}]'\mathbf{z}\mathbf{z}'[\mathbf{\Phi}:\bar{\mathbf{X}}]\mathbf{D}_{2n}^{-1}\}.$$

4. (a) $\lim_{n\to\infty} \mathbf{B}_n = \mathbf{B}$, where $\mathbf{B}$ is finite and

$$\mathbf{B}_n = E\{\mathbf{D}_{1n}^{-1}(\mathbf{\Phi}:\boldsymbol{\psi})'\mathbf{z}\mathbf{z}'(\mathbf{\Phi}:\boldsymbol{\psi})\mathbf{D}_{1n}^{-1}\}.$$

(b) $\lim_{n\to\infty} \mathbf{G}_{nij} = \mathbf{G}_{ij}$, $i, j = 1, 2, \ldots, k_1 + k_2$, where $\mathbf{G}_{ij}$ is finite,

$$\mathbf{G}_{nij} = E\{\mathbf{D}_{1n}^{-1}[\mathbf{\Phi}:\boldsymbol{\psi}]'\mathbf{w}_{.i}\mathbf{w}_{.j}'[\mathbf{\Phi}:\boldsymbol{\psi}]\mathbf{D}_{1n}^{-1}\},$$

and $\mathbf{w}_{.i}$ is the ith column of the matrix $\mathbf{w}$.

5. (a) $\lim_{n\to\infty} d_{jnii} = \infty$, $j = 1, 2$, $i = 1, 2, \ldots, k_1 + k_{4-j}$, where d_{jnii} is the ith diagonal element of $\mathbf{D}_{jn}$.

(b) $$\lim_{n\to\infty} \left(\sum_{t=1}^{n} \varphi_{ti}^2\right)^{-1} \varphi_{ni}^2 = 0, \qquad i = 1, 2, \ldots, k_1,$$

$$\lim_{n\to\infty} \left(\sum_{t=1}^{n} \psi_{tj}^2\right)^{-1} \psi_{nj}^2 = 0, \qquad j = 1, 2, \ldots, k_3,$$

where φ_{ti} is the tith element of $\mathbf{\Phi}$ and ψ_{tj} is the tjth element of $\boldsymbol{\psi}$.

Theorem 5.6.1. Let the model of (5.6.1) to (5.6.2) and assumptions 1 through 4 and 5(a) hold. Then

$$\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{M}_n^{-1}\mathbf{D}_{2n}^{-1}(\mathbf{\Phi}:\bar{\mathbf{X}})'\mathbf{z} + o_p(1),$$

where

$$\hat{\boldsymbol{\theta}} = [(\mathbf{\Phi}:\hat{\mathbf{X}})'(\mathbf{\Phi}:\hat{\mathbf{X}})]^{-1}(\mathbf{\Phi}:\hat{\mathbf{X}})'\mathbf{y},$$
$$\hat{\mathbf{X}} = (\mathbf{\Phi}:\boldsymbol{\psi})\hat{\boldsymbol{\delta}},$$

and

$$\hat{\boldsymbol{\delta}} = [(\mathbf{\Phi}:\boldsymbol{\psi})'(\mathbf{\Phi}:\boldsymbol{\psi})]^{-1}(\mathbf{\Phi}:\boldsymbol{\psi})'\mathbf{X}.$$

Proof. Define $\hat{\mathbf{M}}_n$ to be the matrix $\mathbf{M}_n$ with $\bar{\mathbf{X}}$ replaced by $\hat{\mathbf{X}}$. By assumption 4, the variance of $\mathbf{D}_{1n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})$ is of order one, and by assumption 3, the variance of $\mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \bar{\mathbf{X}})'\mathbf{z}$ is of order one. Thus,

$$
\begin{aligned}
\hat{\mathbf{M}}_n - \mathbf{M}_n &= \mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \hat{\mathbf{X}})'(\boldsymbol{\Phi} : \hat{\mathbf{X}})\mathbf{D}_{2n}^{-1} - \mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \bar{\mathbf{X}})'(\boldsymbol{\Phi} : \bar{\mathbf{X}})\mathbf{D}_{2n}^{-1} \\
&= \mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \bar{\mathbf{X}})'[\mathbf{0} : (\boldsymbol{\Phi} : \boldsymbol{\psi})\mathbf{D}_{1n}^{-1}\mathbf{D}_{1n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})]\mathbf{D}_{2n}^{-1} \\
&\quad + \mathbf{D}_{2n}^{-1}[\mathbf{0} : (\boldsymbol{\Phi} : \boldsymbol{\psi})\mathbf{D}_{1n}^{-1}\mathbf{D}_{1n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})]'(\boldsymbol{\Phi} : \bar{\mathbf{X}})\mathbf{D}_{2n}^{-1} \\
&\quad + \mathbf{D}_{2n}^{-1}[\mathbf{0} : (\boldsymbol{\Phi} : \boldsymbol{\psi})\mathbf{D}_{1n}^{-1}\mathbf{D}_{1n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})]'[\mathbf{0} : (\boldsymbol{\Phi} : \boldsymbol{\psi})\mathbf{D}_{1n}^{-1}\mathbf{D}_{1n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})]\mathbf{D}_{2n}^{-1} \\
&= o_p(1) .
\end{aligned}
$$

Similarly,

$$\mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \hat{\mathbf{X}})'\mathbf{z} - \mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \bar{\mathbf{X}})'\mathbf{z} = o_p(1) .$$

Using

$$
\begin{aligned}
\mathbf{y} &= (\boldsymbol{\Phi} : \mathbf{X})\boldsymbol{\theta} + \mathbf{z} \\
&= \boldsymbol{\Phi}\boldsymbol{\beta} + (\hat{\mathbf{X}} + \hat{\mathbf{w}})\boldsymbol{\lambda} + \mathbf{z}
\end{aligned}
$$

and

$$(\boldsymbol{\Phi} : \hat{\mathbf{X}})'\hat{\mathbf{w}} = \mathbf{0} ,$$

we have

$$
\begin{aligned}
\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &= \hat{\mathbf{M}}_n^{-1}\mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \hat{\mathbf{X}})'\mathbf{z} = [\mathbf{M}_n^{-1} + o_p(1)]\mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \bar{\mathbf{X}})'\mathbf{z} + o_p(1) \\
&= \mathbf{M}_n^{-1}\mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \bar{\mathbf{X}})'\mathbf{z} + o_p(1) .
\end{aligned}
$$
▲

Since $\lim_{n\to\infty} \mathbf{M}_n = \mathbf{M}$, we can also write

$$\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{M}^{-1}\mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \bar{\mathbf{X}})'\mathbf{z} + o_p(1) . \tag{5.6.4}$$

In many applications $n^{-1/2}\mathbf{D}_{1n}$ and $n^{-1/2}\mathbf{D}_{2n}$ have finite nonsingular limits. In these cases $n^{1/2}$ can be used as the normalizing factor, and the remainder in (5.6.4) is $O_p(n^{-1/2})$.

It follows from Theorem 5.6.1 that the sampling behavior of $\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is approximately that of $\mathbf{M}_n^{-1}\mathbf{D}_{2n}^{-1}(\boldsymbol{\Phi} : \bar{\mathbf{X}})'\mathbf{z}$, which has variance $\mathbf{M}_n^{-1}\mathbf{R}_n\mathbf{M}_n^{-1}$, where $\mathbf{R}_n$ was defined in assumption 3. In some situations it is reasonable to assume the elements of $\mathbf{z}$ are independent $(0, \sigma_z^2)$ random variables.

Corollary 5.6.1. Let the model (5.6.1) to (5.6.2) and assumptions 1 to 5 hold with the elements of $\mathbf{z}$ independently distributed $(0, \sigma_z^2)$ random variables with

$E\{Z_t^4\} = \eta\sigma_z^4$. Then

$$\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{M}\sigma_z^2)\,.$$

Furthermore, a consistent estimator for σ_z^2 is

$$s_z^2 = \frac{1}{n - k_1 - k_2}\hat{\mathbf{z}}'\hat{\mathbf{z}}\,, \tag{5.6.5}$$

where $\hat{\mathbf{z}} = \mathbf{y} - (\boldsymbol{\Phi} : \mathbf{X})\hat{\boldsymbol{\theta}}$.

Proof. A proof is not presented here. The normality result is a special case of Theorem 6.3.4 of Chapter 6. That s_z^2 is a consistent estimator of σ_z^2 can be demonstrated by the arguments of Theorem 9.8.3. ▲

Our analysis treated $\boldsymbol{\Phi}$ and $\boldsymbol{\psi}$ as fixed. Theorem 5.6.1 will hold for $\boldsymbol{\Phi}$ and (or) $\boldsymbol{\psi}$ random, provided the second moments exist, the probability limits analogous to the limits of assumptions 1, 2, and 5 exist, and the error terms in $\mathbf{y}$ and $\mathbf{X}$ are independent of $\boldsymbol{\Phi}$ and $\boldsymbol{\psi}$.

A discussion of instrumental variables particularly applicable when k_3 is larger than k_2 is given in Sargan (1958). A model where the method of instrumental variables is appropriate will be discussed in Chapter 9.

Example 5.6.1. To illustrate the method of instrumental variables, we use some data collected in an animal feeding experiment. Twenty-four lots of pigs were fed six different rations characterized by the percentage of protein in the ration. The remainder of the ration was primarily carbohydrate from corn. We simplify by calling this remainder corn. Twelve of the lots were weighed after two weeks, and twelve after four weeks. The logarithms of the gain and of the feed consumed are given in Table 5.6.1. We consider the model

$$G_i = \beta_0 + \beta_1 P_i + \beta_2 C_i + Z_i\,, \qquad i = 1, 2, \ldots, 24\,,$$

where G_i is the logarithm of gain, C_i is the logarithm of corn consumed, P_i is the logarithm of protein consumed, and Z_i is the random error for the ith lot.

It is clear that neither corn nor protein, but instead their ratio, is fixed by the experimental design. The observations on corn and protein for a particular ration are constrained to lie on a ray through the origin with slope corresponding to the ratio of the percentages of the two items in the ration. The logarithms of these observations will lie on parallel lines. Since $C_i - P_i$ is fixed, we rewrite the model as

$$G_i = \beta_0 + (\beta_1 + \beta_2)P_i + \beta_2(C_i - P_i) + Z_i\,.$$

In terms of the notation of (5.6.1), $G_i = Y_i$ and $P_i = X_i$. Candidates for ψ_i are functions of $C_i - P_i$ and of time on feed. As one simple model for the protein

Table 5.6.1. Gain and Feed Consumed by 24 Lots of Pigs

Lot i	Time on Feed Weeks	Log Gain G	Log Corn C	Log Protein P	$\hat{P}$	$\hat{Z}$
1	2	4.477	5.366	3.465	3.601	−0.008
2	2	4.564	5.488	3.587	3.601	−0.042
3	2	4.673	5.462	3.647	3.682	0.035
4	2	4.736	5.598	3.783	3.682	−0.036
5	2	4.718	5.521	3.787	3.757	−0.033
6	2	4.868	5.580	3.846	3.757	0.059
7	2	4.754	5.516	3.858	3.828	−0.043
8	2	4.844	5.556	3.898	3.828	0.007
9	2	4.836	5.470	3.884	3.895	0.035
10	2	4.828	5.463	3.877	3.895	0.034
11	2	4.745	5.392	3.876	3.961	−0.026
12	2	4.852	5.457	3.941	3.961	0.017
13	4	5.384	6.300	4.399	4.453	−0.021
14	4	5.493	6.386	4.485	4.453	0.003
15	4	5.513	6.350	4.535	4.537	0.001
16	4	5.583	6.380	4.565	4.537	0.041
17	4	5.545	6.314	4.580	4.616	0.013
18	4	5.613	6.368	4.634	4.616	0.028
19	4	5.687	6.391	4.733	4.690	0.028
20	4	5.591	6.356	4.698	4.690	−0.033
21	4	5.591	6.288	4.702	4.760	−0.015
22	4	5.700	6.368	4.782	4.760	0.015
23	4	5.700	6.355	4.839	4.828	−0.019
24	4	5.656	6.332	4.816	4.828	−0.041

Source: Data courtesy of Research Department, Moorman Manufacturing Company. The data are a portion of a larger experiment conducted by the Moorman Manufacturing Company in 1974.

consumption we suggest

$$P_i = \delta_0 + \delta_1(C_i - P_i) + \delta_2 t_i + \delta_3(t_i - 3)(C_i - P_i) + W_i\,,$$

where t_i is the time on feed of the ith lot. The ordinary least squares estimate of the equation is

$$\underset{}{\hat{P}_i =} \underset{(0.48)}{4.45} - \underset{(0.09)}{0.95}\,(C_i - P_i) + \underset{(0.15)}{0.46}\,t_i - \underset{(0.09)}{0.02}\,(t_i - 3)(C_i - P_i)\,,$$

where the numbers in parentheses are the estimated standard errors of the regression coefficients. If the W_i are independent $(0, \sigma^2)$ random variables, the usual regression assumptions are satisfied. The interaction term contributes very little to the regression, but the time coefficient is highly significant. This supports assumption 2 because it suggests that the partial correlation between P_i and t_i after adjusting for $C_i - P_i$ is not zero. This, in turn, implies that the matrix M_n is nonsingular. The $\hat{P}$-values for this regression are given in Table 5.6.1. Regressing G_i on $\hat{P}_i$ and $C_i - P_i$, we obtain

$$\hat{G}_i = 0.49 + 0.98P_i + 0.31(C_i - P_i) .$$

In this problem it is reasonable to treat the Z's as independent random variables. We also assume that they have common variance. The estimated residuals are shown in the last column of Table 5.6.1. These must be computed directly as

$$G_i - 0.49 - 0.98P_i - 0.31(C_i - P_i) .$$

The residuals obtained in the second round regression computations are $G_i - 0.49 - 0.98\hat{P}_i - 0.31(C_i - P_i)$ and are inappropriate for the construction of variance estimates. From the $\hat{Z}$'s we obtain

$$s_z^2 = (21)^{-1} \sum_{i=1}^{24} \hat{Z}_i^2 = 0.0010 .$$

The inverse of the matrix used in computing the estimates is

$$\begin{pmatrix} 14.73 & -1.32 & -5.37 \\ -1.32 & 0.23 & 0.21 \\ -5.37 & 0.21 & 2.62 \end{pmatrix},$$

and it follows that the estimated standard errors of the estimates are (0.121), (0.015), and (0.051), respectively. ▲▲

5.7. ESTIMATED GENERALIZED LEAST SQUARES

In this section, we investigate estimation for models in which the covariance matrix of the error is estimated. Our treatment is restricted to linear models, but many of the results extend to nonlinear models.

Consider the linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} , \tag{5.7.1}$$

where $\mathbf{Y}$ is an $n \times 1$ vector, $\mathbf{X}$ is an $n \times k$ matrix, and $\boldsymbol{\beta}$ is the $k \times 1$ vector of unknown parameters. We assume

$$E\{(\mathbf{u}, \mathbf{uu}') \mid \mathbf{X}\} = (\mathbf{0}, \mathbf{V}_{uu}) , \tag{5.7.2}$$

where $\mathbf{V}_{uu}$ is positive definite. In many of our applications, $\mathbf{u}' = (u_1, u_2, \ldots, u_n)$ will be a portion of a realization of a time series. For example, the time series may be a pth order autoregressive process.

For known $\mathbf{V}_{uu}$, the generalized least squares estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{Y}, \tag{5.7.3}$$

where we have assumed $\mathbf{V}_{uu}$ and $\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X}$ to be nonsingular. The conditional variance of the estimator is

$$V\{\hat{\boldsymbol{\beta}} \mid \mathbf{X}\} = (\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X})^{-1}. \tag{5.7.4}$$

Often $\mathbf{V}_{uu}$ is not known. We are interested in the situation where an estimator of $\mathbf{V}_{uu}$, denoted by $\hat{\mathbf{V}}_{uu}$, is used to construct an estimator of $\boldsymbol{\beta}$. We define the estimated generalized least squares estimator by

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{Y}, \tag{5.7.5}$$

where we assume $\hat{\mathbf{V}}_{uu}$ and $\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X}$ are nonsingular. The estimated generalized least squares estimator of $\boldsymbol{\beta}$ is consistent for $\boldsymbol{\beta}$ under mild conditions. In the theorem, we use a general normalizer matrix, $\mathbf{M}_n'$. A natural choice for $\mathbf{M}_n'$ is $\mathbf{G}_n^{1/2} = (\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X})^{1/2}$.

Theorem 5.7.1. Let the model (5.7.1) and (5.7.2) hold. Assume there exists a sequence of estimators $\hat{\mathbf{V}}_{uu}$ and a sequence of nonsingular matrices $\{\mathbf{M}_n\}$ such that

$$\mathbf{M}_n'(\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X})^{-1}\mathbf{M}_n = O_p(1), \tag{5.7.6}$$

$$\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{u} = O_p(1), \tag{5.7.7}$$

$$\mathbf{M}_n^{-1}\mathbf{X}'(\hat{\mathbf{V}}_{uu}^{-1} - \mathbf{V}_{uu}^{-1})\mathbf{X}\mathbf{M}_n^{-1\prime} = O_p(\xi_n), \tag{5.7.8}$$

$$\mathbf{M}_n^{-1}\mathbf{X}'(\hat{\mathbf{V}}_{uu}^{-1} - \mathbf{V}_{uu}^{-1})\mathbf{u} = O_p(\xi_n), \tag{5.7.9}$$

where $\xi_n \to 0$ as $n \to \infty$. Then

$$\mathbf{M}_n'(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) = O_p(\xi_n).$$

Proof. By the assumption (5.7.8),

$$\mathbf{M}_n'\hat{\mathbf{G}}_n^{-1}\mathbf{M}_n = \mathbf{M}_n'\mathbf{G}_n^{-1}\mathbf{M}_n + O_p(\xi_n), \tag{5.7.10}$$

where

$$(\mathbf{G}_n, \hat{\mathbf{G}}_n) = (\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X}, \mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X}). \tag{5.7.11}$$

Therefore

$$\begin{aligned}\mathbf{M}_n'(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) &= \mathbf{M}_n'(\hat{\mathbf{G}}_n^{-1} - \mathbf{G}_n^{-1})\mathbf{M}_n\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{u} \\ &\quad + \mathbf{M}_n'\hat{\mathbf{G}}_n^{-1}\mathbf{M}_n\mathbf{M}_n^{-1}\mathbf{X}'(\hat{\mathbf{V}}_{uu}^{-1} - \mathbf{V}_{uu}^{-1})\mathbf{u} \\ &= O_p(\xi_n). \end{aligned}$$ ▲

In working with models such as (5.7.1)–(5.7.2), it is often assumed that

$$\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X}\mathbf{M}_n^{-1\prime} \xrightarrow{P} \mathbf{A}_0, \tag{5.7.12}$$

where $\mathbf{A}_0$ is a fixed positive definite matrix. This assumption is sufficient for the condition (5.7.6). If there exists a sequence $\{\mathbf{M}_n\}$ that satisfies the assumptions of Theorem 5.7.1, then any nonsingular $\mathbf{H}_n$ such that $\mathbf{G}_n = \mathbf{H}_n\mathbf{H}_n'$ also will satisfy the assumptions because (5.7.6) implies that $\mathbf{H}_n^{-1}\mathbf{M}_n = O_p(1)$.

Theorem 5.7.1 is given for a general normalizing matrix $\mathbf{M}_n$. If $\mathbf{M}_n$ is chosen to be $\mathbf{G}_n^{1/2}$, where $\mathbf{G}_n$ is defined in (5.7.11), then (5.7.6) and (5.7.7) follow directly.

Under the assumptions of Theorem 5.7.1, the limiting distribution of the normalized estimated generalized least squares estimator is the same as the limiting distribution of the normalized generalized least squares estimator constructed with known $\mathbf{V}_{uu}$, provided the limiting distribution exists. Note that the estimator $\hat{\mathbf{V}}_{uu}$ can converge to $\mathbf{V}_{uu}$ rather slowly.

Corollary 5.7.1.1. Let the assumptions (5.7.10)–(5.7.12) of Theorem 5.7.2 hold. In addition assume

$$\mathbf{M}_n^{-1}\mathbf{G}_n\mathbf{M}_n^{-1\prime} \xrightarrow{P} A_0, \tag{5.7.13}$$

where $\mathbf{A}_0$ is a fixed positive definite matrix, and

$$\mathbf{M}_n'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{A}_0^{-1}). \tag{5.7.14}$$

Then

$$\mathbf{M}_n'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{A}_0^{-1}), \tag{5.7.15}$$

and

$$(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\hat{\mathbf{G}}_n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} \chi^2(k), \tag{5.7.16}$$

where $\hat{\mathbf{G}}_n$ is defined in (5.7.11) and $\chi^2(k)$ is a chi-square random variable with k degrees of freedom.

Proof. The assumption (5.7.13) implies the assumption (5.7.6) of Theorem 5.7.1. Hence, (5.7.15) follows from the variance of the generalized least squares estimator and (5.7.14). Also, (5.7.16) follows from

$$\begin{aligned}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})'\hat{\mathbf{G}}_n(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}) &= (\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{M}_n[\mathbf{M}_n^{-1}\hat{\mathbf{G}}_n\mathbf{M}_n^{-1\prime}]\mathbf{M}_n'(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\ &= [\mathbf{M}_n'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]'\mathbf{A}_0[\mathbf{M}_n'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})] + o_p(1),\end{aligned}$$

where we have used (5.7.13), (5.7.4), and the assumption (5.7.8). ▲

The conditions (5.7.6) and (5.7.7) of Theorem 5.7.1 hold if $\mathbf{M}_n'$ is chosen equal to $\mathbf{G}_n^{1/2}$. However, in practice, one wishes a matrix $\mathbf{M}_n'$ that is a function of the data. In some situations, there is a known transformation such that the estimated parameters of the transformed model can be normalized with a diagonal matrix. The transformation can be a function of n. In Corollary 5.7.1.2, we demonstrate how the estimator in such a case can be normalized using sample statistics.

Corollary 5.7.1.2. Let the assumptions of Theorem 5.7.1 hold with $\mathbf{M}_n = \mathbf{D}_n^{1/2}$, where $\mathbf{D}_n = \operatorname{diag}\{\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X}\}$. Also assume (5.7.14) holds with $\mathbf{M}_n = \mathbf{D}_n^{1/2}$. Then

$$\hat{\mathbf{D}}_n^{1/2}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{A}_0^{-1}),$$

where $\hat{\mathbf{D}}_n = \operatorname{diag}\{\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X}\}$, and

$$\hat{\mathbf{A}}^{1/2}\hat{\mathbf{D}}_n^{1/2}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{I}), \tag{5.7.17}$$

where $\hat{\mathbf{A}} = \hat{\mathbf{D}}_n^{-1/2}\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X}\hat{\mathbf{D}}_n^{-1/2}$.

Proof. By assumption (5.7.8) of Theorem 5.7.1,

$$\operatorname{diag}\{\mathbf{D}_n^{-1/2}\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X}\mathbf{D}_n^{-1/2}\} = \mathbf{D}_n^{-1}\hat{\mathbf{D}}_n = \mathbf{I} + O_p(\xi_n). \tag{5.7.18}$$

Thus,

$$\hat{\mathbf{D}}_n^{1/2}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \mathbf{D}_n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + O_p(\xi_n)$$

and the distribution result is established. By (5.7.18) and the assumptions (5.7.6) and (5.7.8) of Theorem 5.7.1,

$$\hat{\mathbf{A}} = \hat{\mathbf{D}}_n^{-1/2}\mathbf{D}_n^{1/2}\mathbf{D}_n^{-1/2}\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X}\mathbf{D}_n^{-1/2}\mathbf{D}_n^{1/2}\hat{\mathbf{D}}_n^{-1/2} \xrightarrow{P} \mathbf{A}_0$$

and $\hat{\mathbf{A}}^{1/2} \xrightarrow{P} \mathbf{A}_0^{1/2}$. ▲

Given the conditions of Theorem 5.7.1 and Corollary 5.7.1.1, the estimated variance of the estimated generalized least squares estimator can be used to construct pivotal statistics. We state the result in Corollary 5.7.1.3.

Corollary 5.7.1.3. Let the assumptions of Theorem 5.7.1 and Corollary 5.7.1.1 hold, where $\{\mathbf{M}_n\}$ is a sequence of fixed matrices. Then

$$\hat{\sigma}_{n\lambda}^{-1}\boldsymbol{\lambda}_n'(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})\xrightarrow{\mathscr{L}} N(0,1),$$

where $\hat{\sigma}_{n\lambda}^2=\boldsymbol{\lambda}_n'\hat{\mathbf{G}}_n^{-1}\boldsymbol{\lambda}_n$, $\hat{\mathbf{G}}_n$ is defined in (5.7.12), and $\boldsymbol{\lambda}_n$ is a fixed nonzero vector that can depend on n.

Proof. Under the assumptions, we show that

$$\sigma_{n\lambda}^{-1}\boldsymbol{\lambda}_n'(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})\xrightarrow{\mathscr{L}} N(0,1),$$

where

$$\sigma_{n\lambda}^2=\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{A}_0^{-1}\mathbf{M}_n^{-1}\boldsymbol{\lambda}_n.$$

By Corollary 5.7.1.1, we have

$$\mathbf{L}_n=\mathbf{M}_n'(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})\xrightarrow{\mathscr{L}} N(\mathbf{0},\mathbf{A}_0^{-1}).$$

By Skorohod's theorem [see Theorem 2.9.6 of Billingsley (1979)], there exist random vectors $\mathbf{Z}_n$ and $\mathbf{Z}$ on a common probability space such that $\mathbf{Z}_n$ has the same distribution as $\mathbf{L}_n$, $\mathbf{Z}\sim N(\mathbf{0},\mathbf{A}_0^{-1})$, and $\lim\mathbf{Z}_n=\mathbf{Z}$ a.s. Therefore,

$$\sigma_{n\lambda}^{-1}\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{Z}_n=\sigma_{n\lambda}^{-1}\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{Z}+\sigma_{n\lambda}^{-1}\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}(\mathbf{Z}_n-\mathbf{Z}).$$

Since $\sigma_{n\lambda}^{-2}\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{A}_0^{-1}\mathbf{M}_n^{-1}\boldsymbol{\lambda}_n=1$, we have

$$\sigma_{n\lambda}^{-1}\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{Z}\sim N(0,1)$$

and

$$\begin{aligned}|\sigma_{n\lambda}^{-1}\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{A}_0^{-1/2}\mathbf{A}_0^{1/2}(\mathbf{Z}_n-\mathbf{Z})|^2&\le(\mathbf{Z}_n-\mathbf{Z})'\mathbf{A}_0(\mathbf{Z}_n-\mathbf{Z})\\&\le\zeta_{\max}(\mathbf{A}_0)(\mathbf{Z}_n-\mathbf{Z})'(\mathbf{Z}_n-\mathbf{Z})\\&\to 0\qquad\text{a.s.},\end{aligned}$$

where $\zeta_{\max}(\mathbf{A}_0)$ is the maximum eigenvalue of $\mathbf{A}_0$. Therefore, $\sigma_{n\lambda}^{-1}\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{Z}_n$ and hence $\sigma_{n\lambda}^{-1}\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{L}_n$ converges in distribution to the standard normal distribution. See Sanger (1992, Section 7.1.3).

From the assumption (5.7.8) of Theorem 5.7.1 and the assumption (5.7.13) of Corollary 5.7.1.1,

$$\begin{aligned}\mathbf{M}_n^{-1}\hat{\mathbf{G}}_n\mathbf{M}_n^{-1\prime}&=\mathbf{M}_n^{-1}\mathbf{G}_n\mathbf{M}_n^{-1\prime}+O_p(\xi_n)\\&=\mathbf{A}_0+o_p(1),\end{aligned}$$

and $\mathbf{M}_n'\hat{\mathbf{G}}_n^{-1}\mathbf{M}_n = \mathbf{A}_0^{-1} + \mathbf{\Delta}_n$, where $\mathbf{\Delta}_n = o_p(1)$. Let $\boldsymbol{\delta}_n = \mathbf{M}_n^{-1}\boldsymbol{\lambda}_n$. Then

$$\begin{aligned}\left|\frac{\hat{\sigma}_{n\lambda}^2}{\sigma_{n\lambda}^2} - 1\right| &= \left|\frac{\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{M}_n'\hat{\mathbf{G}}_n^{-1}\mathbf{M}_n\mathbf{M}_n^{-1}\boldsymbol{\lambda}_n}{\boldsymbol{\lambda}_n'\mathbf{M}_n^{-1\prime}\mathbf{A}_0^{-1}\mathbf{M}_n^{-1}\boldsymbol{\lambda}_n} - 1\right| \\ &= \left|(\boldsymbol{\delta}_n'\mathbf{A}_0^{-1}\boldsymbol{\delta}_n)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{k}\delta_{ni}\delta_{nj}\Delta_{nij}\right| \\ &\le [\boldsymbol{\delta}_n'\boldsymbol{\delta}_n\zeta_{\max}^{-1}(\mathbf{A}_0)]^{-1}\delta_{n,\max}^2\sum_{i=1}^{k}\sum_{j=1}^{k}|\Delta_{nij}| \\ &= o_p(1),\end{aligned}$$

where $\delta_{n,\max}$ is the maximum element of $\boldsymbol{\delta}_n$. Therefore $\hat{\sigma}_{n\lambda}^{-1}\sigma_{n\lambda}$ converges to one in probability, and the result follows. ▲

Under the conditions of Theorem 5.7.1, the asymptotic distribution of the estimated generalized least squares estimator is the same as that of the generalized least squares estimator, provided the limiting distribution of the generalized least squares estimator exists. The following theorem gives conditions such that the generalized least squares estimator is asymptotically normal. To obtain a limiting normal distribution, the transformed errors must satisfy a central limit theorem. In Theorem 5.7.2, we assume the transformed errors to be independent. For example, if $\{u_t\}$ is an autoregressive process with independent increments, then the assumption is satisfied.

Theorem 5.7.2. Assume the model (5.7.1) and (5.7.2) with fixed $\mathbf{X}$. Assume there exists a sequence of fixed, nonsingular matrices $\{\mathbf{M}_n\}$ such that

$$\lim_{n\to\infty} \mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X}\mathbf{M}_n^{-1\prime} = \mathbf{A}_0\,, \tag{5.7.19}$$

where $\mathbf{A}_0$ is positive definite. Assume there exists a sequence of nonsingular transformations $\{\mathbf{T}_n\}$ such that the elements of $\mathbf{e} = \mathbf{T}_n\mathbf{u}$ are independent with zero expectation, variance one, and

$$E\{|e_{tn}|^{2+\delta}\} < K$$

for some $\delta > 0$ and finite K. Also assume that

$$\lim_{n\to\infty}\sup_{\substack{1\le j\le n\\ 1\le i\le p}} |(\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{T}_n')_{ij}| = 0\,. \tag{5.7.20}$$

Then

$$\mathbf{M}_n'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{A}_0^{-1})\,,$$

where $\hat{\boldsymbol{\beta}}$ is the generalized least squares estimator defined in (5.7.3).

Proof. Consider the linear combination

$$\boldsymbol{\lambda}'\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{T}_n'\mathbf{e} = \sum_{t=1}^{n} c_{tn}e_{tn},$$

where

$$c_{tn} = \boldsymbol{\lambda}'\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{T}_{nt.}',$$

$\boldsymbol{\lambda}$ is an arbitrary k-dimensional vector with $0 < |\boldsymbol{\lambda}| < \infty$, and $\mathbf{T}_{nt.}$ is the tth row of $\mathbf{T}_n$. By construction, the variance of $\Sigma_{t=1}^{n} c_{tn}e_{tn}$, denoted by V_n, is

$$V_n \equiv \sum_{t=1}^{n} c_{tn}^2 V\{e_{tn}\} = \sum_{t=1}^{n} c_{tn}^2 .$$

Hence, $V_n^{-1}\Sigma_{t=1}^{n} c_{tn}^2$ is bounded. By the assumption (5.7.19),

$$\begin{aligned}\lim_{n\to\infty} \sum_{t=1}^{n} c_{tn}^2 &= \lim_{n\to\infty} \boldsymbol{\lambda}'\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{T}_n'\mathbf{T}_n\mathbf{X}\mathbf{M}_n^{-1\prime}\boldsymbol{\lambda} \\ &= \boldsymbol{\lambda}'\mathbf{A}_0\boldsymbol{\lambda} > 0 .\end{aligned} \tag{5.7.21}$$

Let b_{ij} be the ijth element of $\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{T}_n'$. Thus,

$$c_{tn}^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} \lambda_i\lambda_j b_{it}b_{jt} \le k^2 \max_{1\le i\le k} \lambda_i^2 \max_{1\le i\le k} b_{it}^2 ,$$

which implies that

$$\sup_{1\le t\le n} c_{tn}^2 \le k^2 \max_{1\le i\le k} \lambda_i^2 \sup_{\substack{1\le j\le n\\ 1\le i\le k}} b_{ij}^2 .$$

By the assumption (5.7.20) and (5.7.21),

$$\lim_{n\to\infty} V_n^{-1} \sup_{1\le t\le n} c_{tn}^2 = 0 .$$

Hence, by Corollary 5.3.4,

$$V_n^{-1/2} \sum_{t=1}^{n} c_{tn}e_{tn} \xrightarrow{\mathscr{L}} N(0, 1) .$$

Since $\boldsymbol{\lambda}$ is arbitrary, by the assumption (5.7.19) we have

$$\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{u} \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{A}_0) ,$$

and we obtain the conclusion. ▲

To this point, no structure has been imposed on $\mathbf{V}_{uu}$ other than that it must be positive definite. The number of unknown parameters in $\mathbf{V}_{uu}$ could conceivably grow with the sample size n. A model of practical importance is

$$
\begin{aligned}
E\{\mathbf{u} \mid \mathbf{X}\} &= \mathbf{0} \\
E\{\mathbf{uu}' \mid \mathbf{X}\} &= \mathbf{V}_{uu} = \mathbf{V}_{uu}(\boldsymbol{\theta}^0),
\end{aligned}
\tag{5.7.22}
$$

where $\boldsymbol{\theta}$ is an $l \times 1$ vector of unknown parameters, l is fixed, and $\boldsymbol{\theta}^0$ is the true value. The parameter space for $\boldsymbol{\theta}$ is $\boldsymbol{\Theta}$. It is assumed that the form of the function $\mathbf{V}_{uu}(\boldsymbol{\theta})$ is known and that $\mathbf{V}_{uu}(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$. If, for example, the time series is known to be a pth order autoregressive process, the vector $\boldsymbol{\theta}$ will contain the parameters of the autoregressive process. We are interested in the situation in which an estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, is used to construct an estimator of $\mathbf{V}_{uu}(\boldsymbol{\theta})$, denoted by $\hat{\mathbf{V}}_{uu} = \mathbf{V}_{uu}(\hat{\boldsymbol{\theta}})$.

The following theorem gives sufficient conditions for the estimated generalized least squares estimator to have the same asymptotic distribution as the generalized least squares estimator for general normalizing matrices under the model (5.7.1) and (5.7.22).

Theorem 5.7.3. Let the model (5.7.1) and (5.7.22) hold. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_l)'$ and let

$$
\mathbf{B}_{ni}(\boldsymbol{\theta}) = \frac{\partial \mathbf{V}_{uu}^{-1}(\boldsymbol{\theta})}{\partial \theta_i}, \qquad i = 1, 2, \ldots, l,
$$

be continuous in $\boldsymbol{\theta}$. Let $\tilde{\boldsymbol{\beta}}$ be defined by (5.7.5). Assume there exists a sequence of nonsingular matrices $\{\mathbf{M}_n\}$ such that

$$
\mathbf{M}_n'\mathbf{G}_n^{-1}\mathbf{M}_n = O_p(1), \tag{5.7.23}
$$

$$
\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}(\boldsymbol{\theta}^0)\mathbf{u} = O_p(1), \tag{5.7.24}
$$

where $\mathbf{G}_n = \mathbf{X}'\mathbf{V}_{uu}^{-1}(\boldsymbol{\theta}^0)\mathbf{X}$. Also assume

$$
\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{B}_{ni}(\boldsymbol{\theta})\mathbf{X}\mathbf{M}_n^{-1\prime} = O_p(1), \tag{5.7.25}
$$

$$
\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{B}_{ni}(\boldsymbol{\theta})\mathbf{u} = O_p(1), \tag{5.7.26}
$$

for $i = 1, 2, \ldots, l$, uniformly in an open neighborhood of $\boldsymbol{\theta}^0$, denoted by $C(\boldsymbol{\theta}^0)$. Let an estimator of $\boldsymbol{\theta}^0$, denoted by $\hat{\boldsymbol{\theta}}$, be available, and assume

$$
\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(\xi_n), \tag{5.7.27}
$$

where $\xi_n \to 0$ as $n \to \infty$. Then

$$
\mathbf{M}_n'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{M}_n'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + O_p(\xi_n).
$$

Proof. By a Taylor expansion

$$\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{X}\mathbf{M}_n^{-1\prime} = \mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}(\boldsymbol{\theta}^0)\mathbf{X}\mathbf{M}_n^{-1\prime} + \sum_{i=1}^{l} \mathbf{M}_n^{-1}\mathbf{X}'\overset{*}{\mathbf{B}}_{ni}(\boldsymbol{\theta}^0, \hat{\boldsymbol{\theta}})\mathbf{X}\mathbf{M}_n^{-1\prime}(\hat{\theta}_i - \theta_i^0),$$

where $\boldsymbol{\theta}^0$ is the true value of $\boldsymbol{\theta}$, and $\overset{*}{\mathbf{B}}_{ni}(\boldsymbol{\theta}^0_*, \hat{\boldsymbol{\theta}})$ is a matrix whose js element is the js element of $\mathbf{B}_{ni}(\boldsymbol{\theta})$ evaluated at a point $\overset{*}{\boldsymbol{\theta}}_{js}$ on the line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^0$. Let the probability space be $(\Omega, \mathscr{A}, P)$. Let $\epsilon > 0$ be given. Choose δ such that $\{\boldsymbol{\theta}: |\boldsymbol{\theta}^0 - \boldsymbol{\theta}| < \delta\} \subset C(\boldsymbol{\theta}^0)$, where for any vector $\mathbf{v}$, $|\mathbf{v}|^2 = \mathbf{v}'\mathbf{v}$. Since $\hat{\boldsymbol{\theta}}$ converges in probability to $\boldsymbol{\theta}^0$, there exists an N_1 and a set $D_{1,n} \in \mathscr{A}$ such that $P(D_{1,n}) > 1 - \epsilon/2$ and

$$|\overset{*}{\boldsymbol{\theta}}_{js} - \boldsymbol{\theta}^0| < \delta$$

for $j = 1, 2, \ldots, n$, for $s = 1, 2, \ldots, n$, for all $n > N_1$, and for all $\omega \in D_{1,n}$. By the assumption (5.7.25), there exists a K, an N_2, and a set $D_{2,n} \in \mathscr{A}$ with $P(D_{2,n}) > 1 - \epsilon/2$ such that

$$\|\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{B}_{ni}(\boldsymbol{\theta})\mathbf{X}\mathbf{M}_n^{-1\prime}\| < K$$

for all $n > N_2$, all $\omega \in D_{2,n}$, and all $\boldsymbol{\theta} \in C(\boldsymbol{\theta}^0)$, where $\|\mathbf{H}\| = [\mathrm{tr}(\mathbf{H}'\mathbf{H})]^{1/2}$. Let $D_n = D_{1,n} \cap D_{2,n}$, and observe that $P(D_n) > 1 - \epsilon$. Let $N = \max(N_1, N_2)$. Therefore, for all $n > N$ and for all $\omega \in D_n$,

$$\|\mathbf{M}_n^{-1}\mathbf{X}'\overset{*}{\mathbf{B}}_{ni}(\boldsymbol{\theta}^0, \hat{\boldsymbol{\theta}})\mathbf{X}\mathbf{M}_n^{-1\prime}\| < K.$$

Hence,

$$\mathbf{M}_n^{-1}\mathbf{X}'\overset{*}{\mathbf{B}}_{ni}(\boldsymbol{\theta}^0, \hat{\boldsymbol{\theta}})\mathbf{X}\mathbf{M}_n^{-1\prime} = O_p(1), \tag{5.7.28}$$

which implies the assumption (5.7.28) of Theorem 5.7.1. By a similar expansion,

$$\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{u} = \mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}(\boldsymbol{\theta}^0)\mathbf{u} + \sum_{i=1}^{l} \mathbf{M}_n^{-1}\mathbf{X}'\ddot{\mathbf{B}}_{ni}(\boldsymbol{\theta}^0, \hat{\boldsymbol{\theta}})\mathbf{u}(\hat{\theta}_i - \theta_i^0),$$

where $\ddot{\mathbf{B}}_{ni}(\boldsymbol{\theta}^0, \hat{\boldsymbol{\theta}})$ is analogous to $\overset{*}{\mathbf{B}}_{ni}(\boldsymbol{\theta}^0, \hat{\boldsymbol{\theta}})$. Using the assumption (5.7.26) and an argument similar to the one used to show (5.7.28),

$$\mathbf{M}_n^{-1}\mathbf{X}'\ddot{\mathbf{B}}_{ni}(\boldsymbol{\theta}^0, \hat{\boldsymbol{\theta}})\mathbf{u} = O_p(1).$$

Thus the assumption (5.7.9) of Theorem 5.7.1 is satisfied. Because the assumptions (5.7.23) and (5.7.24) are the same as the assumptions (5.7.6) and (5.7.7) of Theorem 5.7.1, the result follows. ▲

A sufficient condition for (5.7.23) is

$$\mathbf{M}_n^{-1}\mathbf{G}_n\mathbf{M}_n^{-1\prime} \xrightarrow{P} \mathbf{A}_0, \tag{5.7.29}$$

where $\mathbf{A}_0$ is a fixed positive definite matrix, and a sufficient condition for (5.7.25) is

$$\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{B}_{ni}(\boldsymbol{\theta})\mathbf{X}\mathbf{M}_n^{-1\prime} \xrightarrow{P} \mathbf{A}_i(\boldsymbol{\theta}), \qquad i = 1, 2, \ldots, l,$$

uniformly in a neighborhood of $\boldsymbol{\theta}^0$ as $n \to \infty$, where the $\mathbf{A}_i(\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$.

Under the assumptions of Theorem 5.7.3, the limiting distribution of the normalized estimated generalized least squares estimator is the same as the limiting distribution of the normalized generalized least squares estimator constructed with known $\mathbf{V}_{uu}$, provided the limiting distribution exists. Note that the estimator of $\boldsymbol{\theta}$ can converge to $\boldsymbol{\theta}^0$ at a relatively slow rate.

A common procedure is to estimate $\boldsymbol{\beta}$ by the ordinary least squares estimator,

$$\ddot{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \tag{5.7.30}$$

compute residuals

$$\ddot{\mathbf{u}} = \mathbf{Y} - \mathbf{X}\ddot{\boldsymbol{\beta}}, \tag{5.7.31}$$

and use these residuals to estimate $\boldsymbol{\theta}$. Then the estimator of $\boldsymbol{\theta}$ is used in (5.7.5) to estimate $\boldsymbol{\beta}$. In order for this procedure to be effective, the estimator of $\boldsymbol{\theta}$ based on $\ddot{\mathbf{u}}$ must be a consistent estimator of $\boldsymbol{\theta}^0$, and this, in turn, requires $\ddot{\boldsymbol{\beta}}$ of (5.7.30) to be a consistent estimator of $\boldsymbol{\beta}$. The following lemma gives sufficient conditions for the estimator of $\boldsymbol{\theta}^0$ based on the ordinary least squares residuals to be consistent.

Lemma 5.7.1. Consider the model (5.7.1). Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{u})$ be an estimator of $\boldsymbol{\theta}^0$ based upon the true $\mathbf{u}$, and let

$$\overset{*}{\mathbf{u}} = [\mathbf{I} - c\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{u},$$

where $0 \le c \le 1$. Assume that $\hat{\boldsymbol{\theta}}(\mathbf{u})$ is a continuous function of $\mathbf{u}$ with a continuous first derivative, and that

$$\hat{\boldsymbol{\theta}}(\mathbf{u}) - \boldsymbol{\theta}^0 = O_p(\xi_n), \tag{5.7.32}$$

where $\xi_n \to 0$ as $n \to \infty$. Also assume

$$\sum_{i=1}^{n} \left(\frac{\partial \hat{\theta}_j(\overset{*}{\mathbf{u}})}{\partial u_i}\right)\mathbf{X}_i(\ddot{\boldsymbol{\beta}} - \boldsymbol{\beta}) = o_p(\xi_n), \tag{5.7.33}$$

for $j = 1, 2, \ldots, l$, uniformly in c for $0 \leq c \leq 1$, where $\ddot{\boldsymbol{\beta}}$ is defined in (5.7.30). Let $\hat{\boldsymbol{\theta}}(\ddot{\mathbf{u}})$ be an estimator of $\boldsymbol{\theta}$ constructed with the residuals defined in (5.7.31). Then

$$\hat{\boldsymbol{\theta}}(\ddot{\mathbf{u}}) = \hat{\boldsymbol{\theta}}(\mathbf{u}) + o_p(\xi_n) = \boldsymbol{\theta}^0 + O_p(\xi_n)\,.$$

Proof. By a Taylor expansion about $\mathbf{u}$, we have

$$\hat{\theta}_j(\ddot{\mathbf{u}}) = \hat{\theta}_j(\mathbf{u}) - \sum_{i=1}^{n} \frac{\partial \hat{\theta}_j(\overset{*}{\mathbf{u}})}{\partial u_i} \mathbf{X}_i(\ddot{\boldsymbol{\beta}} - \boldsymbol{\beta})\,, \tag{5.7.34}$$

where $\overset{*}{\mathbf{u}}$ is on the line segment joining $\mathbf{u}$ and $\ddot{\mathbf{u}}$, $\mathbf{X}_i$ is the ith row of $\mathbf{X}$, and $\ddot{u}_i - u_i = -\mathbf{X}_i(\ddot{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Because $\overset{*}{\mathbf{u}}$ is of the form given in the assumptions, it follows from (5.7.33) that

$$\hat{\theta}_j(\ddot{\mathbf{u}}) = \hat{\theta}_j(\mathbf{u}) + o_p(\xi_n) = \theta_j^0 + O_p(\xi_n)\,. \qquad \blacktriangle$$

Given that $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}^0$, the difference between the estimator of $\boldsymbol{\beta}$ based upon $\hat{\boldsymbol{\theta}}$ and the estimator of $\boldsymbol{\beta}$ based on $\boldsymbol{\theta}^0$ converges to zero.

Theorem 5.7.4. Let the assumptions (5.7.23)–(5.7.26) of Theorem 5.7.3 hold. In addition, assume (5.7.32),

$$\mathbf{M}_n'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{M}_n = O_p(1)\,, \tag{5.7.35}$$

$$E\{\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}(\boldsymbol{\theta}^0)\mathbf{X}\mathbf{M}_n^{-1\prime}\} = O(1)\,, \tag{5.7.36}$$

and

$$\sum_{i=1}^{n} \frac{\partial \hat{\theta}_j(\overset{*}{\mathbf{u}})}{\partial u_i} \mathbf{X}_i \mathbf{M}_n^{-1\prime} = o_p(\xi_n) \tag{5.7.37}$$

uniformly in c for $0 \leq c \leq 1$ and $j = 1, 2, \ldots, l$, where $\mathbf{X}_i$ is the ith row of $\mathbf{X}$, and $\overset{*}{\mathbf{u}}$ is defined in Lemma 5.7.1. Then

$$\mathbf{M}_n'[\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}(\ddot{u})) - \boldsymbol{\beta}] = \mathbf{M}_n'[\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}^0) - \boldsymbol{\beta}] + O_p(\xi_n)\,,$$

where

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}^0) = [X'\mathbf{V}_{uu}^{-1}(\boldsymbol{\theta}^0)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}(\boldsymbol{\theta}^0)\mathbf{Y}\,.$$

If it is further assumed that

$$\mathbf{M}_n'[\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}^0) - \boldsymbol{\beta}] \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{A}_0^{-1})\,, \tag{5.7.38}$$

then

$$\mathbf{M}_n'[\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}(\ddot{\mathbf{u}})) - \boldsymbol{\beta}] \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{A}_0^{-1}), \tag{5.7.39}$$

where $\hat{\boldsymbol{\theta}}(\ddot{\mathbf{u}})$ is defined in Lemma 5.7.1.

Proof. Under the assumptions (5.7.36) and (5.7.22),

$$E\{E[\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}\mathbf{M}_n^{-1\prime} \mid \mathbf{X}]\} = E\{\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}(\boldsymbol{\theta}^0)\mathbf{X}\mathbf{M}_n^{-1\prime}\} = O(1),$$

which implies that $\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{u} = O_p(1)$ and

$$\mathbf{M}_n'(\ddot{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{M}_n'(X'X)^{-1}\mathbf{M}_n\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{u} = O_p(1). \tag{5.7.40}$$

By (5.7.40) and the assumption (5.7.37),

$$\sum_{i=1}^{n} \frac{\partial \hat{\theta}_j(\overset{*}{\mathbf{u}})}{\partial u_i} \mathbf{X}_i(\ddot{\boldsymbol{\beta}} - \boldsymbol{\beta}) = o_p(\xi_n). \tag{5.7.41}$$

By (5.7.41), the assumption (5.7.32), and Lemma 5.7.1, we have $\hat{\boldsymbol{\theta}}(\ddot{\mathbf{u}}) = \boldsymbol{\theta}^0 + O_p(\xi_n)$, and the assumption (5.7.27) of Theorem 5.7.3 holds. Also, the other assumptions of Theorem 5.7.3 hold for $\hat{\boldsymbol{\theta}}(\ddot{\mathbf{u}})$. Therefore,

$$\mathbf{M}_n'[\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}(\ddot{\mathbf{u}})) - \boldsymbol{\beta}] = \mathbf{M}_n'[\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}^0) - \boldsymbol{\beta}] + O_p(\xi_n),$$

and (5.7.39) follows from the assumption (5.7.38). ▲

Under the conditions of Theorem 5.7.4, the procedure of using the ordinary least squares residuals to estimate the covariance matrix and then using the estimated covariance matrix in the estimated generalized least squares estimator produces an estimator of $\boldsymbol{\beta}$ with the same large sample properties as the estimator constructed with known $\boldsymbol{\theta}$. Also

$$\hat{\mathbf{V}}\{\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}(\ddot{\mathbf{u}}))\} = [\mathbf{X}'\mathbf{V}_{uu}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{X}]^{-1} \tag{5.7.42}$$

can be used as an estimator of the covariance matrix of the approximate distribution of $\hat{\boldsymbol{\beta}}$.

5.8. SEQUENCES OF ROOTS OF POLYNOMIALS

In Chapter 2, we saw that the roots of the characteristic polynomial were important in defining the behavior of autoregressive and autoregressive moving average time series. Hence, the relationships between the coefficients of a polynomial and the roots are of interest. We write the pth order polynomial as

$$g(m) = m^p + a_1 m^{p-1} + \cdots + a_p = (m - m_1)(m - m_2)\cdots(m - m_p), \tag{5.8.1}$$

where the m_i, $i = 1, 2, \ldots, p$, are the roots of $g(m) = 0$. Let $\bar{m}_1, \bar{m}_2, \ldots, \bar{m}_q$ denote the distinct roots with multiplicities $r_1, r_2, \ldots, r_q$, where $\Sigma_{i=1}^{q} r_i = p$. If $r_i = 1$, we say that $\bar{m}_i = m_i$ is a simple root. The coefficients can be expressed in terms of the roots by

$$\begin{aligned} a_1 &= -\sum_{j=1}^{p} m_j, \qquad a_2 = \sum_{i=1}^{p} \sum_{j=i+1}^{p} m_i m_j, \\ a_3 &= -\sum_{i=1}^{p} \sum_{j=i+1}^{p} \sum_{l=j+1}^{p} m_i m_j m_l, \ldots, \qquad a_p = (-1)^p \prod_{i=1}^{p} m_i. \end{aligned} \tag{5.8.2}$$

Thus, the coefficients are continuous differentiable functions of the roots. The functions (5.8.2) define a mapping from a p-dimensional space to a p-dimensional space. The differentials are

$$\begin{aligned} da_1 &= -\sum_{j=1}^{p} dm_j, \\ da_2 &= -a_1 \sum_{j=1}^{p} dm_j - \sum_{j=1}^{p} m_j\, dm_j, \\ da_3 &= -a_2 \sum_{j=1}^{p} dm_j - a_1 \sum_{j=1}^{p} m_j\, dm_j - \sum_{j=1}^{p} m_j^2\, dm_j, \\ &\ \vdots \\ da_p &= -a_{p-1} \sum_{j=1}^{p} dm_j - a_{p-2} \sum_{j=1}^{p} m_j\, dm_j - \cdots - \sum_{j=1}^{p} m_j^{p-1}\, dm_j. \end{aligned} \tag{5.8.3}$$

Thus

$$d\mathbf{a} = \mathbf{BH}\, d\mathbf{m}, \tag{5.8.4}$$

where $d\mathbf{a} = (da_1, da_2, \ldots, da_p)'$, $d\mathbf{m} = (dm_1, dm_2, \ldots, dm_p)'$,

$$\mathbf{B} = \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ -a_1 & -1 & 0 & \cdots & 0 \\ -a_2 & -a_1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -a_{p-1} & -a_{p-2} & -a_{p-3} & \cdots & -1 \end{pmatrix}, \tag{5.8.5}$$

and

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ m_1 & m_2 & \cdots & m_p \\ m_1^2 & m_2^2 & \cdots & m_p^2 \\ \vdots & \vdots & & \vdots \\ m_1^{p-1} & m_2^{p-1} & \cdots & m_p^{p-1} \end{pmatrix} . \tag{5.8.6}$$

The matrix $\mathbf{H}$ is nonsingular if all roots are simple. Because $\mathbf{B}$ is nonsingular, it is possible to solve for the differentials of m_i as functions of the differentials of a_j when all roots are simple.

If the coefficients of the polynomial are real, the complex roots are complex conjugate pairs. In that case, we can use a transformation to define real quantities

$$c_1 = m_1 + m_2 , \tag{5.8.7}$$

$$c_2^2 = (m_1 - m_2)^2 \tag{5.8.8}$$

for every complex conjugate pair. The inverse transformation is

$$m_1 = 0.5(c_1 + c_2) ,$$

$$m_2 = 0.5(c_1 - c_2) , \tag{5.8.9}$$

where we adopt the convention that m_1 is defined with the positive coefficient on the radical and m_2 is defined with the negative coefficient. The inverse transformation has continuous derivatives except at $c_2 = 0$, the point where $m_1 = m_2$. The derivatives of the coefficients with respect to c_1 and with respect to c_2^2 are derivatives of real valued functions with respect to real valued arguments.

From (5.8.6) we see that repeated roots will produce a singularity in the Jacobian of the transformation. However, even if there are repeated roots, the roots of a polynomial are continuous functions of the coefficients of the polynomial.

Lemma 5.8.1. Let $g(m)$ of (5.8.1) have distinct roots $\bar{m}_1, \bar{m}_2, \ldots, \bar{m}_q$ with multiplicities $r_1, r_2, \ldots, r_q$, where $a_1, a_2, \ldots, a_p$ are complex numbers, and $\sum_{j=1}^{q} r_j = p$. Let

$$G(m) = m^p + (a_1 + \delta_1)m^{p-1} + \cdots + (a_p + \delta_p) ,$$

where $\delta_1, \ldots, \delta_p$ are complex numbers. For $i = 1, 2, \ldots, q$, let ϵ_i be an arbitrary real number satisfying

$$0 < \epsilon_i < \min_{j \neq i} |\bar{m}_j - \bar{m}_i| .$$

Then there exists a $\delta > 0$ such that, if $|\delta_i| \leq \delta$ for $i = 1, 2, \ldots, p$, then $G(m)$ has precisely r_i roots in the circle ϕ_i with center at $\bar{m}_i$ and radius ϵ_i, for $i = 1, 2, \ldots, q$.

Proof. Omitted. See Marden (1966, p. 3). ▲

To extend the results of Lemma 5.8.1 to sequences of polynomials, let

$$g_n(m) = m^p + a_{1n}m^{p-1} + \cdots + a_{pn}, \tag{5.8.10}$$

where a_{in}, $i = 1, 2, \ldots, p$, are sequences of real numbers such that $a_{in} \to a_i$, $i = 1, 2, \ldots, p$, as $n \to \infty$, where the a_i are the coefficients of (5.8.1). Let $\bar{m}_1, \bar{m}_2, \ldots, \bar{m}_q$ be the distinct roots of (5.8.1) with multiplicities $r_1, r_2, \ldots, r_q$, and let ϵ be the arbitrary real number defined in Lemma 5.8.1. Then, by Lemma 5.8.1, there exists an N such that, for $n \geq N$, $g_n(m)$ has precisely r_i roots in the sphere ϕ_i with center $\bar{m}_i$ and radius ϵ, for $i = 1, \ldots, q$.

We now investigate the special case where a root has multiplicity two. To study the relation between the coefficients and the roots in the neighborhood of $m_1 = m_2$, we use the transformation introduced in (5.8.7) and (5.8.8) and assume $m_3, m_4, \ldots, m_p$ are simple roots. Then

$$\begin{aligned}
a_1 &= -c_1 - \sum_{j=3}^{p} m_j\,, \\
a_2 &= m_1 m_2 + c_1 \sum_{j=3}^{p} m_j + \sum_{j=3}^{p} \sum_{i=j+1}^{p} m_i m_j \\
&= 0.25(c_1^2 - c_2^2) + c_1 \sum_{j=3}^{p} m_j + \sum_{j=3}^{p} \sum_{i=j+1}^{p} m_i m_j\,, \\
a_3 &= -0.25(c_1^2 - c_2^2) \sum_{j=3}^{p} m_j - c_1 \sum_{j=3}^{p} \sum_{i=j+1}^{p} m_i m_j \\
&\quad + \sum_{i=3}^{p} \sum_{j=i+1}^{p} \sum_{k=j+1}^{p} m_i m_j m_k \\
&\;\;\vdots \\
a_p &= 0.25(c_1^2 - c_2^2)(-1)^p \prod_{j=3}^{p} m_j\,.
\end{aligned} \tag{5.8.11}$$

We observe that the partial derivatives of the coefficients with respect to c_1 evaluated at $m_1 = m_2$, are equal to the partial derivatives of the functions (5.8.2) with respect to m_1. Also, c_2 enters the expressions for the coefficients only as a square. In the proof of Lemma 5.8.2, we show that $c_1, c_3, m_3, \ldots, m_p$, where $c_3 = c_2^2$ and $m_3, \ldots, m_p$ are simple roots, can be expressed as locally continuous differentiable functions of the coefficients. Thus, Lemma 5.8.2 extends the result of Lemma 5.8.1 to the case in which some of the roots are of multiplicity two. The bound on the difference between two equal roots is the square root of the bound on the error in a simple root.

Lemma 5.8.2. Let $\mathbf{a}_n = (a_{1n}, a_{2n}, \ldots, a_{pn})$, $n = 1, 2, \ldots$, be a sequence of real vectors such that

$$|\mathbf{a}_n - \mathbf{a}| \le M_n ,$$

where $M_n \to 0$ as $n \to \infty$ and $\mathbf{a} = (a_1, a_2, \ldots, a_p)$ is the vector of real coefficients of (5.8.1). Let m_{in}, $i = 1, 2, \ldots, p$, be the roots of

$$m^p + a_{1n}m^{p-1} + a_{2n}m^{p-2} + \cdots + a_{pn} = 0 ,$$

and let m_i, $i = 1, 2, \ldots, p$, be the roots of (5.8.1). Assume the roots of (5.8.1) are either simple roots or roots of multiplicity two. If m_i is a simple root of (5.8.1),

$$|m_{in} - m_i| \le KM_n$$

for some $K < \infty$, where it is understood that for n sufficiently large, K can be chosen such that there is exactly one m_{in} satisfying the inequality for each simple m_i.

If $m_1 = m_2$,

$$|m_{1n} + m_{2n} - 2m_1| \le KM_n ,$$
$$|m_{1n} - m_{2n}|^2 \le KM_n$$

for some $K < \infty$, where, for n sufficiently large, K can be chosen such that exactly one pair of roots satisfies the inequalities.

Proof. Let the roots, denoted by $\mathbf{m} = (m_1, m_2, \ldots, m_p)$, be arranged so that the repeated roots occur as pairs in the first part of the vector. Let (m_1, m_2), $(m_3, m_4), \ldots, (m_{2k-1}, m_{2k})$ be the pairs of equal roots. Let

$$c_{1i} = m_{2i-1} + m_{2i} \quad \text{and} \quad c_{3i} = (m_{2i-1} - m_{2i})^2$$

and

$$m_{2i-1} = 0.5(c_{1i} + c_{3i}^{1/2}) \quad \text{and} \quad m_{2i} = 0.5(c_{1i} - c_{3i}^{1/2})$$

for $i = 1, 2, \ldots, k$. The total differential of $\mathbf{a}$ with respect to

$$\mathbf{u} = (c_{11}, c_{31}, c_{12}, c_{32}, \ldots, c_{1k}, c_{3k}, m_{2k+1}, \ldots, m_p) ,$$

analogous to (5.8.4), is

$$d\mathbf{a} = \mathbf{BH}_1\, d\mathbf{u} , \tag{5.8.12}$$

where $\mathbf{H}_1 = \mathbf{HJ}$,

$$\mathbf{J} = \text{block diag}(\mathbf{J}_{11}, \mathbf{J}_{22}, \ldots, \mathbf{J}_{kk}, \mathbf{I}),$$

$\mathbf{B}$ is defined in (5.8.5), $\mathbf{H}$ is defined in (5.8.6), $\mathbf{J}$ is the Jacobian of the transformation of $\mathbf{u}$ into $\mathbf{m}$, and

$$\mathbf{J}_{ii} = \begin{pmatrix} 0.5 & 0.25c_{3i}^{-1/2} \\ 0.5 & -0.25c_{3i}^{-1/2} \end{pmatrix}.$$

The matrix $\mathbf{H}$ is a Vandermonde matrix with determinant

$$|\mathbf{H}| = \prod_{1 \le i < j \le p} (m_j - m_i),$$

and the determinant of $\mathbf{H}_1$ is

$$|\mathbf{H}_1| = (0.25)^k \prod_{i=1}^{k} (m_{2i-1} - m_{2i})^{-1} |\mathbf{H}|.$$

The determinant $|\mathbf{H}_1|$ is not zero at $m_{2i-1} = m_{2i}$ because every zero product in the determinant of $\mathbf{H}$ is removed by the transformation. Hence, the differentials of $\mathbf{u}$ are locally continuous differentiable functions of $\mathbf{a}$, and the results follow. ▲

It is important that the bound given for the difference in Lemma 5.8.2 can be achieved. Consider, for example, the sequence of polynomials

$$g_n(m) = m^2 - (2 + 2n^{-1})m + 1 + n^{-1} + n^{-2}.$$

The roots of $g_n(m) = 0$ are $1 + n^{-1} + n^{-1/2}$ and $1 + n^{-1} - n^{-1/2}$. Thus, the difference between the two roots is $O(n^{-1/2})$. The conclusion of Lemma 5.8.2 can also be obtained from the following theorem.

Theorem 5.8.1. Let $\mathbf{a}_n = (a_{1n}, a_{2n}, \ldots, a_{pn})$, $n = 1, 2, \ldots$, be a sequence of (possibly complex valued) vectors, and let $\mathbf{a} = (a_1, a_2, \ldots, a_p)$ be the vector of (possibly complex valued) coefficients of (5.8.1). Assume that

$$|\mathbf{a}_n - \mathbf{a}| = O(\kappa_n), \tag{5.8.13}$$

where $\kappa_n \to 0$ as $n \to \infty$. Let $\bar{m}_1, \bar{m}_2, \ldots, \bar{m}_q$ denote the distinct roots of $g(m) = 0$ with multiplicities $r_1, r_2, \ldots, r_q$, where $\sum_{i=1}^{q} r_i = p$. Let ϵ be an arbitrary real number satisfying

$$0 < \epsilon < 0.5 \min_{2 \le j \le q} |\bar{m}_1 - \bar{m}_j|,$$

and let $r_1 = s$. Then there exists an N such that if $n \ge N$, the equation

$$m^p + a_{n1}m^{p-1} + \cdots + a_{np} = 0 \tag{5.8.14}$$

has exactly s roots in the circle ϕ_1 with center $\bar{m}_1$ and radius ϵ and

$$\begin{aligned}
&\sum_{i=1}^{s} \xi_{ni} = O(\kappa_n) \\
&\sum_{\substack{i=1 \\ i \neq j}}^{s} \sum_{j=1}^{s} \xi_{ni}\xi_{nj} = O(\kappa_n) \\
&\qquad \vdots \\
&\prod_{i=1}^{s} \xi_{ni} = O(\kappa_n),
\end{aligned} \tag{5.8.15}$$

where $\xi_{ni} = m_{ni} - \bar{m}_1$, $i = 1, 2, \ldots, s$, in which m_{ni}, $i = 1, 2, \ldots, s$, are the s roots in ϕ_1.

Proof. That s roots will fall in ϕ_1 when $n \geq N$ follows from Lemma 5.8.1. Therefore, we only demonstrate the order results.

If $\bar{m}_1 \neq 0$, we can replace m in (5.8.1) with $m - \bar{m}_1$ to obtain a polynomial with s zero roots. Therefore, without loss of generality, we assume $\bar{m}_1 = 0$. Then

$$a_p = a_{p-1} = \cdots = a_{p-s+1} = 0$$

and

$$a_{np} = O(\kappa_n), \qquad a_{n,p-1} = O(\kappa_n), \ldots, \qquad a_{n,p-s+1} = O(\kappa_n).$$

The polynomial (5.8.1) can be written as

$$(m - \xi_{n1})(m - \xi_{n2}) \cdots (m - \xi_{ns})(m^{p-s} + b_{n1}m^{p-s-1} + \cdots + b_{n,p-s}), \tag{5.8.16}$$

where

$$\begin{aligned}
(-1)^s \prod_{i=1}^{s} \xi_{ni} b_{n,p-s} &= a_{np} = O(\kappa_n), \\
(-1)^{s-1} \sum_{j=1}^{s} \prod_{\substack{i=1 \\ i \neq j}}^{s} \xi_{ni} b_{n,p-s} + (-1)^s \prod_{i=1}^{s} \xi_{ni} b_{n,p-s-1} &= a_{n,p-1} = O(\kappa_n), \\
(-1)^{s-2} \sum_{\substack{j=1 \\ j \neq k}}^{s} \sum_{k=1}^{s} \prod_{\substack{i=1 \\ i \neq j \\ i \neq k}}^{s} \xi_{ni} b_{n,p-s} + (-1)^{s-1} \sum_{j=1}^{s} \prod_{\substack{i=1 \\ i \neq j}}^{s} \xi_{ni} b_{n,p-s-1} & \\
+ (-1)^s \prod_{i=1}^{n} \xi_{ni} b_{n,p-s-2} &= a_{n,p-2} = O(\kappa_n),
\end{aligned}$$

$$\vdots \tag{5.8.17}$$

$$(-1)\sum_{i=1}^{s} \xi_{ni} b_{n,p-s} + (-1)^2 \sum_{\substack{i=1 \\ i \neq j}}^{s} \sum_{j=1}^{s} \xi_{ni}\xi_{nj} b_{n,p-s-1} + \cdots + (-1)^s \prod_{i=1}^{s} \xi_{ni} b_{n,p-2s+1} = a_{n,p-s+1} = O(\kappa_n)\,.$$

Because the a_{ni}, $i = 1, 2, \ldots, p$, converge, the b_{nj}, $j = 1, 2, \ldots, p-s$, also converge, and we write

$$(b_{n1}, b_{n2}, \ldots, b_{n,p-s}) \to (b_1, b_2, \ldots, b_{p-s})\,.$$

Since the multiplicity of $\bar{m}_1 = 0$ is s, we have $b_{p-s} \neq 0$. Thus, for example,

$$\prod_{i=1}^{s} \xi_{ni} = O(\kappa_n)$$

from the first equation of (5.8.17). ▲

Corollary 5.8.1.1. If the assumption (5.8.13) is replaced with

$$|\mathbf{a}_n - \mathbf{a}| = O_p(\kappa_n)\,,$$

then $O(\kappa_n)$ of (5.8.15) is replaced with $O_p(\kappa_n)$.

Proof. Omitted. ▲

Corollary 5.8.1.2. Let the assumptions of Theorem 5.8.1 hold with $s = 2$. Then, for $n \geq N$,

$$(m_{n1} - m_{n2})^2 = O(\kappa_n)\,,$$

where m_{n1} and m_{n2} are the two roots in ϕ_1.

Proof. We have, for m_{n1} and m_{n2} in ϕ_1,

$$\begin{aligned}(m_{n1} - m_{n2})^2 &= m_{n1}^2 + m_{n2}^2 - 2m_{n1}m_{n2} \\ &= O(\kappa_n)\end{aligned}$$

by Lemma 5.8.1 and by Theorem 5.8.1. ▲

We now give a result on the roots and vectors of square matrices. Let $\mathbf{A}$ be any $p \times p$ matrix with possibly complex elements. The roots of the determinantal equation

$$|\mathbf{A} - m\mathbf{I}| = 0 \tag{5.8.18}$$

are called the characteristic roots of $\mathbf{A}$, the eigenvalues of $\mathbf{A}$, the eigenroots of $\mathbf{A}$, or simply the roots of $\mathbf{A}$. The vector $\mathbf{u}_i$ that satisfies the equation

$$(\mathbf{A} - m_i\mathbf{I})\mathbf{u}_i = \mathbf{0}\,, \tag{5.8.19}$$

where m_i is a root of (5.8.18), is called a characteristic vector of $\mathbf{A}$ or an eigenvector of $\mathbf{A}$. The following theorem, taken from Magnus and Neudecker (1988), demonstrates that the simple roots of $\mathbf{A}$ are continuous differentiable functions of the elements of $\mathbf{A}$. The theorem gives the differentials of the roots and of the vectors.

Theorem 5.8.2. Let m_1 be a simple root of (5.8.18) for $\mathbf{A}_1$, where $\mathbf{A}_1$ is any (possibly complex) $p \times p$ matrix. Let $\mathbf{u}_1$ be the associated eigenvector. Then a complex valued function $m(\mathbf{A})$ and a complex vector valued function $\mathbf{u}(\mathbf{A})$ are defined for all $\mathbf{A}$ in some neighborhood S of $\mathbf{A}_1$, such that

$$m(\mathbf{A}_1) = m_1\,, \qquad \mathbf{u}(\mathbf{A}_1) = \mathbf{u}_1\,, \tag{5.8.20}$$

$$\mathbf{A}\mathbf{u}(\mathbf{A}) = m(\mathbf{A})\mathbf{u}(\mathbf{A})\,, \quad \text{and} \quad \mathbf{u}_1^*(\mathbf{A})\mathbf{u}(\mathbf{A}) = 1 \tag{5.8.21}$$

for all $\mathbf{A}$ in S, where $\mathbf{u}_1^*(\mathbf{A})$ is the complex conjugate of $\mathbf{u}_1(\mathbf{A})$. Furthermore, the functions are differentiable any number of times on S, and the differentials are

$$d\mathbf{m} = (\mathbf{v}_1^*\mathbf{u}_1)^{-1}\mathbf{v}_1^*(d\mathbf{A})\mathbf{u}_1 \tag{5.8.22}$$

and

$$d\mathbf{u} = (\mathbf{A}_1 - m_1\mathbf{I})^{\dagger}[\mathbf{I} - (\mathbf{v}_1^*\mathbf{u}_1)^{-1}\mathbf{u}_1\mathbf{v}_1^*](d\mathbf{A})\mathbf{u}_1\,, \tag{5.8.23}$$

where $\mathbf{B}^{\dagger}$ is the Moore–Penrose generalized inverse of $\mathbf{B}$, $\mathbf{v}_1$ is the eigenvector associated with the eigenvalue m_1^* of $\mathbf{A}_1^*$, satisfying

$$\mathbf{A}_1^*\mathbf{v}_1 = m_1^*\mathbf{v}_1\,,$$

and $\mathbf{B}^*$ is the complex conjugate of $\mathbf{B}$.

Proof. Omitted. See Magnus and Neudecker (1988, p. 161). ▲

If $\mathbf{A}_1$ is a real symmetric matrix, then the roots are real, and the expressions (5.8.20) and (5.8.21) simplify to

$$\mathrm{d}m = \mathbf{u}_1'(d\mathbf{A})\mathbf{u}_1\,, \tag{5.8.24}$$

$$d\mathbf{u} = (\mathbf{A}_1 - m_1\mathbf{I})^{\dagger}(d\mathbf{A})\mathbf{u}_1\,. \tag{5.8.25}$$

It follows from Theorem 5.8.1 that if the coefficients of a matrix are converging to limiting values at the rate M_n, then a simple root m_i is converging to a limiting value at the same rate.

The results for the roots of a matrix are applicable to the roots of a polynomial written in the form (5.8.1). Let the $p \times p$ matrix $\mathbf{A}$ be defined by

$$\mathbf{A} = \begin{pmatrix} -a_1 & -a_2 & -a_3 & \cdots & -a_{p-1} & a_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \tag{5.8.26}$$

where the a_i are the coefficients of (5.8.1). Then the roots of the determinantal equation

$$|\mathbf{A} - m\mathbf{I}| = 0 \tag{5.8.27}$$

are the same as the roots of (5.8.1).

REFERENCES

Section 5.1. Chernoff (1956), Chung (1968), Mann and Wald (1943b), Pratt (1959), Tucker (1967).

Sections 5.2, 5.3. Billingsley (1968), Brown (1971), Chan and Wei (1988), Chung (1968), Cramér (1946), Dvoretsky (1972), Gnedenko (1967), Hall and Heyde (1980), Loeve (1963), Pollard (1984), Rao (1973), Scott (1973), Tucker (1967), Varadarajan (1958), Wilks (1962).

Section 5.4. Cramér (1946), DeGracie and Fuller (1972), Hansen, Hurwitz, and Madow (1953), Hatanaka (1973).

Section 5.5. Andrews (1987), Gallant (1971, 1987), Gallant and White (1988), Gumpertz and Pantula (1992), Hartley (1961), Hartley and Booker (1965), Jennrich (1969), Klimko and Nelson (1978), Malinvaud (1970b), Potscher and Prucha (1991a,b), Sarkar (1990), Wald (1949), Wu (1981).

Section 5.6. Basmann (1957), Fuller (1987), Sargan (1958, 1959), Theil (1961).

Section 5.7. Carroll and Ruppert (1988), Jobson and Fuller (1980), Rothenberg (1984), Sanger (1992), Toyooka (1982, 1990), Van Der Genugten (1983).

Section 5.8. Amemiya (1990), Magnus and Neudecker (1988), Marden (1966).

EXERCISES

1. Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers, and let $\{f_n\}$ and $\{g_n\}$ be sequences of positive real numbers such that $a_n = O(f_n)$ and $b_n = O(g_n)$. Show that:
   (a) $|a_n|^s = O(f_n^s)$, $s > 0$.

(b) $a_n b_n = O(f_n g_n)$.

2. Let $\{f_n\}$, $\{g_n\}$, and $\{r_n\}$ be sequences of positive real numbers, and let $\{X_n\}$, $\{Y_n\}$, and $\{Z_n\}$ be sequences of random variables such that $X_n = O_p(f_n)$, $Y_n = O_p(g_n)$, and $Z_n = o_p(r_n)$. Without recourse to Theorems 5.1.5 and 5.1.6, show that:
   (a) $|X_n|^s = O_p(f_n^s)$, $s > 0$.
   (b) $X_n Y_n = O_p(f_n g_n)$.
   (c) $X_n + Y_n = O_p(\max\{f_n, g_n\})$.
   (d) $X_n Z_n = o_p(f_n r_n)$.
   (e) If $g_n / r_n = o(1)$, then $Y_n + Z_n = o_p(r_n)$.

3. Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ be two independent random samples, each of size n. Let the X_i be $N(0, 4)$ and the Y_i be $N(2, 9)$. Find the order in probability of the following statistics.
   (a) $\bar{X}_n$. (d) $\bar{Y}_n^2$.
   (b) $\bar{Y}_n$. (e) $\bar{X}_n \bar{Y}_n$.
   (c) $\bar{X}_n^3$. (f) $(\bar{X}_n + 1)$.
   Find the most meaningful expression (i.e., the smallest quantity for which the order in probability statement is true).

4. Prove that $p\lim_{n\to\infty} \hat{\theta} = \theta^0$ implies that there exists a sequence of positive real numbers $\{a_n\}$ such that $\lim_{n\to\infty} a_n = 0$ and $\hat{\theta} - \theta^0 = O_p(a_n)$.

5. Let $\{a_t\}$ be a sequence of constants satisfying $\sum_{t=1}^{\infty} |a_t| < \infty$; also let $\{X_{tn}: t = 1, 2, \ldots, n; n = 1, 2, \ldots\}$ be a triangular array of random variables such that $E\{X_{tn}^2\} = O(b_n^2)$, $t = 1, 2, \ldots, n$, where $\lim_{n\to\infty} b_n = 0$. Prove

$$\sum_{t=1}^{n} a_t X_{tn} = O_p(b_n).$$

6. Let $\bar{x}_n$ be distributed as a normal $(\mu, \sigma^2/n)$ random variable with $\mu \neq 0$, and define $Y_n = \bar{x}_n^3 - \bar{x}_n^{-2}$. Expand Y_n in a Taylor's series through terms of $O_p(n^{-1})$. Find the expectation of these terms.

7. Let $F_1, F_2, \ldots, F_M$ be a finite collection of distribution functions with finite variances and common mean μ. A sequence of random variables $\{X_t: t \in (1, 2, \ldots)\}$ is created by randomly choosing, for each t, one of the distribution functions and then making a random selection from the chosen distribution. Show that, as $n \to \infty$,

$$\left(\sum_{t=1}^{n} \sigma_t^2\right)^{-1/2} \sum_{t=1}^{n} (X_t - \mu) \xrightarrow{\mathscr{L}} N(0, 1),$$

where σ_t^2 is the variance of the distribution chosen for index t. Does

$$n^{1/2}\left(M^{-1}\sum_{j=1}^{M}\sigma_j^2\right)^{-1/2}\left(n^{-1}\sum_{t=1}^{n}X_t-\mu\right)\xrightarrow{\mathscr{L}} N(0,1)\,,$$

where σ_j^2 is the variance of the jth distribution, $j=1,2,\ldots,M$?

8. Let X_t be normal independent (μ, σ^2) random variables, and let $\bar{x}_n = n^{-1}\sum_{t=1}^{n} X_t$.

(a) It is known that $\mu \neq 0$. How would you approximate the distribution of $\bar{x}_n^2$ in large samples?

(b) It is known that $\mu = 0$. How would you approximate the distribution of $\bar{x}_n^2$ in large samples?

Explain and justify your answers, giving the normalizing constants necessary to produce nondegenerate limiting distributions.

9. Let $\bar{x}_n$ be distributed as a normal $(\mu, \sigma^2/n)$ random variable, $\mu > 0$.

(a) Find $E\{Y_n\}$ to order $(1/n)$, where

$$Y_n = \operatorname{sgn}\bar{x}_n\,|\bar{x}_n|^{1/2}$$

and $\operatorname{sgn}\bar{x}_n$ denotes the sign of $\bar{x}_n$.

(b) Find $E\{(Y_n - \mu^{1/2})^2\}$ to order $(1/n)$.

(c) What is the limiting distribution of $n^{1/2}(Y_n - \mu^{1/2})$?

10. Let $\bar{x}_n$ be distributed as a normal $(\mu, \sigma^2/n)$ random variable, $\mu \neq 0$, and define

$$Z_n = \frac{n\bar{x}_n}{n\bar{x}_n^2 + \sigma^2}\,.$$

Show that $E\{\bar{x}_n^{-1}\}$ is not defined but that $E\{Z_n\}$ is defined. Show further that Z_n satisfies the conditions of Theorem 5.4.3, and find the expectation of Z_n through terms of $O(1/n)$.

11. Let $(X, Y)'$ be distributed as a bivariate normal random variable with mean $(\mu_x, \mu_y)'$ and covariance matrix

$$\begin{pmatrix}\sigma_x^2 & \sigma_{xy}\\ \sigma_{xy} & \sigma_y^2\end{pmatrix}.$$

Given a sample of size n from this bivariate population, derive an estimator for the product $\mu_x\mu_y$ that is unbiased to $O(1/n)$.

12. Assume the model

$$Y_t = \alpha + \lambda e^{-\beta x_t} + u_t ,$$

where the u_t are distributed as normal independent $(0, \sigma^2)$ random variables. We have available the following data:

t	Y_t	x_t	t	Y_t	x_t
1	47.3	0.0	7	136.5	2.0
2	87.0	0.4	8	132.0	4.0
3	120.1	0.8	9	68.8	0.0
4	130.4	1.6	10	138.1	1.5
5	58.8	0.0	11	145.7	3.0
6	111.9	1.0	12	143.0	5.9

where Y is yield of corn and x is applied nitrogen. Given the initial values $\hat{\alpha} = 143$, $\hat{\lambda} = -85$, $\hat{\beta} = 1.20$, carry out two iterations of the Gauss–Newton procedure. Using the estimates obtained at the second iteration, estimate the covariance matrix of the estimator.

13. In the illustration of Section 5.5 only the coefficient of x_{t1} was used in constructing an initial estimate of θ_1. Identifying the original equation as

$$Y_t = \theta_0 + \theta_1 x_{t1} + \alpha^2 x_{t2} + e_t ,$$

construct an estimator for the covariance matrix of $(\hat{\theta}_0, \hat{\theta}_1, \hat{\alpha})$, where the coefficients $\hat{\theta}_0$, $\hat{\theta}_1$, and $\hat{\alpha}^2$ are the ordinary least squares estimates. Using this covariance matrix, find the λ that minimizes the estimated variance of

$$\lambda \hat{\theta}_1 + (1 - \lambda)\hat{\alpha}$$

as an estimator of θ_1. Use the estimated linear combination as an initial estimate in the Gauss–Newton procedure.

14. Assuming that the e_t are normal independent $(0, \sigma^2)$ random variables, obtain the likelihood function associated with the model (5.5.1). By evaluating the expectations of the second partial derivatives of the likelihood function with respect to the parameters, demonstrate that the asymptotic covariance matrix of the maximum likelihood estimator is the same as that given in Theorem 5.5.1.

15. Let Y_t satisfy the model

$$Y_t = \theta_0 + \theta_1 e^{-\theta_2 t} + e_t , \qquad t = 1, 2, \ldots ,$$

where $\theta_2 > 0$ and e_t is a sequence of normal independent $(0, \sigma^2)$ random variables. Does this model satisfy assumptions 1 and 2 of Theorem 5.4.4? Would the model with t in the exponent replaced by x_t, where $\{x_t\} = \{1, 2, 3, 4, 1, 2, 3, 4, \ldots\}$, satisfy the three assumptions?

16. An experiment is conducted to study the relationship between the phosphate content of the leaf and the yield of grain for the corn plant. In the experiment different levels of phosphate fertilizer were applied to the soil of 20 experimental plots. The leaf phosphate and the grain yield of the corn were recorded for each plot. Denoting yield by Y, applied phosphate by A, and leaf phosphate by P, the sums of squares and cross products matrix for A, P, and Y were computed as

$$\begin{pmatrix} \Sigma A^2 & \Sigma AP & \Sigma AY \\ \Sigma PA & \Sigma P^2 & \Sigma PY \\ \Sigma YA & \Sigma YP & \Sigma Y^2 \end{pmatrix} = \begin{pmatrix} 69{,}600 & 16{,}120 & 3{,}948 \\ 16{,}120 & 8{,}519 & 1{,}491 \\ 3{,}948 & 1{,}491 & 739 \end{pmatrix}.$$

The model

$$Y_t = \beta P_t + v_t\,, \qquad t = 1, 2, \ldots, 20\,,$$

where the v_t are normal independent $(0, \sigma^2)$ random variables, is postulated. Estimate β by the method of instrumental variables, using A_t as the instrumental variable. Estimate the standard error of your coefficients. Compare your estimate with the least squares estimate.

17. Show that if $X_n = O_p(a_n)$ and $a_n = o(b_n)$, then $X_n = o_p(b_n)$.

18. Prove the following corollary to Theorem 5.5.1.

Corollary 5.5.1. Let the assumptions of Theorem 5.5.1 hold. Also assume

$$E\{(e_t^2 - \sigma^2)^2 | \mathscr{A}_{t-1}\} = \kappa$$

and

$$E\{|e_t^2 - \sigma^2|^{2+\delta}\} < M_e\,.$$

Then

$$n^{1/2}(s^2 - \sigma^2) \xrightarrow{\mathscr{L}} N(0, \kappa)\,.$$

19. Let s^2 be defined by (5.5.57). Show that

$$s^2 = (n-k)^{-1}\mathbf{e}'\{\mathbf{I} - \mathbf{F}(\boldsymbol{\theta}^0)[\mathbf{F}'(\boldsymbol{\theta}^0)\mathbf{F}(\boldsymbol{\theta}^0)]^{-1}\mathbf{F}'(\boldsymbol{\theta}^0)\}\mathbf{e} + O_p(\max\{a_n^3, n^{-1/2}a_n^2\})$$

under the assumptions of Theorem 5.5.4.

20. Show that if the assumptions (5.7.6) through (5.7.9) hold and if $\mathbf{M}_n^{-1} = o_p(1)$, then $\tilde{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}$ in probability, where $\tilde{\boldsymbol{\beta}}$ is defined in (5.7.5).

21. Let X_{tn}, $t = 1, 2, \ldots$, $n = 1, 2, \ldots$, be a triangular array of random variables. Let $\{g_t\}_{t=1}^{\infty}$ and $\{h_n\}_{n=1}^{\infty}$ be sequences of positive real numbers. Show that if

$$E\{|X_{tn}|\} = g_t h_n \,, \qquad t = 1, 2, \ldots, n \,, \quad n = 1, 2, \ldots,$$

then

$$\sum_{t=1}^{n} X_{tn} = O_p\left(h_n \sum_{t=1}^{n} g_t\right).$$

Show that $X_{tn} = O_p(g_t h_n)$ does not imply that

$$\sum_{t=1}^{n} X_{tn} = O_p\left(h_n \sum_{t=1}^{n} g_t\right).$$

22. Let the model (5.7.1)–(5.7.2) hold with $X_t = t$ and $\mathbf{V}_{uu} = \text{diag}\{e^{\theta}, e^{2\theta}, \ldots, e^{n\theta}\}$ for $\theta \in \Theta = (1, \infty)$. Let $M_n = G_n^{1/2}$. Suppose an estimator of θ^0, the true value, is available such that $\hat{\theta} - \theta^0 = O_p(n^{-1/2})$.

(a) Show that

$$\|\mathbf{V}_{uu}^{1/2}\hat{\mathbf{V}}_{uu}^{-1}\mathbf{V}_{uu}^{1/2} - \mathbf{I}_n\|_2$$

does not go to zero in probability, where $\|\mathbf{A}\|_2$ is the square root of the largest eigenvalue of $\mathbf{A}'\mathbf{A}$.

(b) Show that the assumption (5.7.8) holds for this model.

23. Let

$$f_n(z) = a_{0n} + a_{1n}z + \cdots + a_{kn}z^k \,,$$

and let

$$f(z) = a_0 + a_1 z + \cdots + a_k z^k = 0$$

have distinct roots $\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_q$ with multiplicities $r_1, r_2, \ldots, r_q$. Let

$$a_{in} - a_i = O_p(n^{-\xi})$$

for $i = 1, 2, \ldots, k$ and some $\xi > 0$. Prove that, given $\epsilon > 0$, there exists an N such that the probability is greater than $1 - \epsilon$ that exactly r_i of the roots of $f_n(z)$ are within $M_\epsilon n^{-\xi}$ of $\bar{z}_i$ for $i = 1, 2, \ldots, q$.

24. Consider the regression model

$$Y_t = X_{t1}\beta_1 + X_{t2}\beta_2 + e_t ,$$

where

$$(X_{t1}, X_{t2}) = (t + u_{t1}, t + u_{t2}) ,$$
$$\mathbf{u}_t = (u_{t1}, u_{t2})' \sim \mathrm{II}(\mathbf{0}, \sigma_u^2\mathbf{I}) ,$$

$e_t \sim \mathrm{II}(0, \sigma_e^2)$, and e_t is independent of $\mathbf{u}_j$ for all t and j. The least squares estimator of $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{n} \mathbf{X}_t'\mathbf{X}_t\right)^{-1} \sum_{t=1}^{n} \mathbf{X}_t'\mathbf{Y}_t .$$

The variance of $\hat{\boldsymbol{\beta}}$ conditional on $\mathbf{X}_t$, $t = 1, 2, \ldots, n$, is

$$V\{\hat{\boldsymbol{\beta}} \mid (\mathbf{X}_1, \ldots, \mathbf{X}_n)\} = \left(\sum_{t=1}^{n} \mathbf{X}_t'\mathbf{X}_t\right)^{-1} \sigma_e^2 .$$

Show that

$$\operatorname*{plim}_{n\to\infty} n^{-3} \sum_{t=1}^{n} \mathbf{X}_t'\mathbf{X}_t = 3^{-1}\mathbf{J}'\mathbf{J} ,$$

where $\mathbf{J} = (1, 1)$. Construct a sequence of matrices $\{\mathbf{M}_n\}$ such that

$$\operatorname*{plim}_{n\to\infty} \mathbf{M}_n^{-1} \sum_{t=1}^{n} \mathbf{X}_t'\mathbf{X}_t\mathbf{M}_n^{-1\prime} = \mathbf{I}$$

and such that

$$\mathbf{M}_n'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \longrightarrow N(\mathbf{0}, \mathbf{I}) .$$

25. Use Theorem 5.3.4 to prove the assertion made in Example 5.5.1 that

$$\mathbf{M}_n^{-1}\mathbf{U}_n(\boldsymbol{\beta}^0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{V}_{\beta\beta}) .$$

26. Let $(a_{1n}, a_{2n})' \sim N[(-1.40, 0.49)', n^{-1}\mathbf{I}]$. Let (m_{1n}, m_{2n}) be the roots of

$$m^2 + a_{1n}m + a_{2n} = 0 ,$$

where $|m_{1n}| \geq |m_{2n}|$. What can you say about the distribution of $n^{1/2}(m_{1n} - 0.7, m_{2n} - 0.7)$ as n increases?

27. Let $(a_{1n}, a_{2n})' \sim N[(-1.20, 0.61)', n^{-1}\mathbf{I}]$. Let (m_{1n}, m_{2n}) be the roots of

$$m^2 + a_{1n}m + a_{2n} = 0 .$$

What can you say about the distribution of $n^{1/2}(m_{1n} - m_1, m_{2n} - m_2)'$, where (m_1, m_2) are the roots of

$$m^2 - 1.20m + 0.61 = 0?$$

Hint: See Exercise 10 of Chapter 4.

28. Consider the sequence of polynomials

$$g_n(m) = m^2 - 3n^{-1}m + 2n^{-2}.$$

What is the order of the difference of the two roots? How do you explain the difference between this result and the example of the text? Use the proof of Theorem 5.8.1 to prove a general result for the order of the difference of the two roots of a second order polynomial that is converging to a polynomial with repeated roots.

29. Let $\{\mathbf{Y}_i\}_{i=1}^{\infty}$ be a sequence of k-dimensional random variables, where $\mathbf{Y}_i \sim \mathrm{NI}(\mu\mathbf{J}, \mathbf{\Sigma})$, and $\mathbf{J}$ is a k-dimensional column vector of ones. Find the limiting distribution, as $n \to \infty$, of

$$\hat{\mu}_g = (\mathbf{J}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{J})^{-1}\mathbf{J}'\hat{\mathbf{\Sigma}}^{-1}\bar{\mathbf{y}},$$

where $\bar{\mathbf{y}} = n^{-1}\sum_{i=1}^{n}\mathbf{Y}_i$ and

$$\hat{\mathbf{\Sigma}} = (n-1)^{-1}\sum_{i=1}^{n}(\mathbf{Y}_i - \bar{\mathbf{y}})(\mathbf{Y}_i - \bar{\mathbf{y}})'.$$

Compare the variance of the limiting distribution of $\hat{\mu}_g$ with that of $\bar{\mu} = k^{-1}\mathbf{J}'\bar{\mathbf{y}}$.

30. Prove the following.

Corollary 5.1.6.2. Let $\{X_n\}$ be a sequence of scalar random variables such that

$$X_n = a + O_p(r_n),$$

where $r_n \to 0$ as $n \to \infty$. If $g(x)$ and $g'(x)$ are continuous at a, then

$$g(X_n) = g(a) + g'(a)(X_n - a) + o_p(r_n).$$

31. Let

$$Y_t = e^{\beta t} + a_t, \qquad t = 0, 1, 2, \ldots,$$

where $\beta \in (-\infty, 0)$ and $a_t \sim \mathrm{NI}(0, \sigma^2)$. Let $\hat{\beta}$ be the value of β that minimizes

$$Q_n(\beta) = \sum_{t=0}^{n} (Y_t - e^{\beta t})^2$$

for $\beta \in (-\infty, 0)$. Are the conditions of Theorem 5.5.1 satisfied for $M_n = 1$ and $a_n = n^{\zeta}$ for some $\zeta > 0$? What do you conclude about the consistency of $\hat{\beta}$?