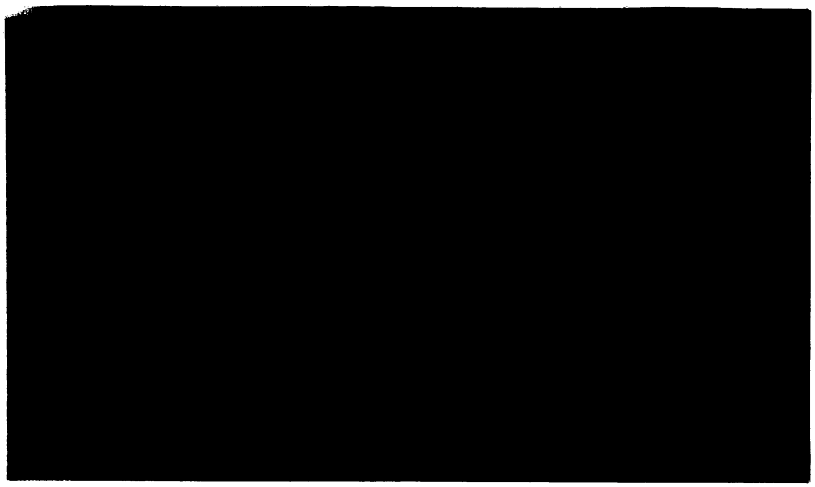
Center for Research on Economic and Social Theory
Research Seminar in Quantitative Economics

Discussion Paper

[redacted]

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, Michigan 48109



A CENTRAL LIMIT THEOREM

WITH APPLICATIONS TO ECONOMETRICS*

E. Philip Howrey
Saul H. Hymans
The University of Michigan

Revised July, 1983

R-110.83

*The authors wish to thank the members of the Statistics Seminar and the Econometric Theory Seminar, both of the University of Michigan, for valuable contributions made following presentations of the earliest versions of this paper. We are especially grateful to Professors Bruce M. Hill, Christopher Sims, Michael Woodroofe, and several anonymous referees, who read earlier versions of the paper with extreme care. Any remaining errors, of course, are ours. Research on this topic was supported, in part, by a grant from the National Science Foundation.

Abstract

This paper is concerned primarily with the asymptotic distribution of the least squares estimator in a linear equation with stochastic regressors. We prove a central limit theorem dealing with a sequence of products of random variables. The theorem is then applied to show asymptotic normality of the least squares estimator in a wide variety of cases, including: a) autoregressive regressors, b) moving average regressors, c) lagged dependent variable regressors. The results are generalized to handle Aitken estimation with stochastic regressors, and instrumental variable estimation in simultaneous equation models.

I. Motivation

This paper is concerned with the asymptotic distribution of the least squares estimator of β in the regression model

$$y_t = \beta x_t + \varepsilon_t \qquad (t = 1, 2, \ldots, T)$$

where $\{\varepsilon_t\}$ is a sequence of independent, identically distributed (i.i.d.) random variables and x_t is a scalar stochastic regressor.[1] In particular, the asymptotic distribution of the stabilized least squares estimator

$$\sqrt{T}\,(\hat{\beta} - \beta) = \frac{\sqrt{T}\,(T^{-1}\Sigma x_t \varepsilon_t)}{(T^{-1}\Sigma x_t^2)}$$

is derived under alternative assumptions about the stochastic process governing the generation of the regressor x_t. Provided that $T^{-1}\Sigma x_t^2$ has a finite, non-zero probability limit, it follows from the convergence theorem of Cramér [1946; p. 254] that $\sqrt{T}\,(\hat{\beta} - \beta)$ will be asymptotically normally distributed if $\sqrt{T}\,(T^{-1}\Sigma x_t \varepsilon_t)$ converges in distribution to normality. In Sections III and IV of the paper we state and prove a central limit theorem dealing with a stochastic sequence of the form $\{x_t \varepsilon_t\}$. Section V of the paper applies the general theorem to the regression model under alternative assumptions about the

---

[1] In the final section of the paper we extend our basic result to the case of a vector of regressors and also relax the assumption of an i.i.d. error structure. For expositional purposes, however, the bulk of the paper focuses on the simple regression model with i.i.d. errors.

generation of x_t. The paper concludes with Section VI which discusses extensions of the basic results. Before turning to the theorem itself, we present a brief review of the existing literature relating to the central question of the paper.

II. The Existing Literature

Most econometrics textbooks provide an explicit derivation of the asymptotic distribution of the least squares estimator only for the "fixed regressor" case. It is generally assumed that the regressor (or vector of regressors) is nonstochastic or, if stochastic, fully independent of the disturbance vector $\underline{\varepsilon}$ in which case the asymptotic distribution of $\sqrt{T}\,(\hat{\beta} - \beta)$ is obtained conditional on the observed values of the regressor. For example, Theil [1971, pp. 380-1] uses the familiar Lindeberg-Lévy central limit theorem to prove the asymptotic normality of the least squares estimator for the fixed regressor case. Hannan [1961] considers what amounts to a system of seemingly unrelated regressions. Using the Liapunov form of the central limit theorem, he proves asymptotic normality of the least squares estimator conditional on regressors which satisfy a form of strong law convergence.

In connection with models which contain lagged values of the dependent variable among the regressors, both Theil [1971, pp. 412-13] and Malinvaud [1966, p. 453], for example, state without proof theorems which assert that the least squares estimator is asymptotically normally distributed. They both cite Mann and Wald [1943] as the original reference for this result.

Other treatments of this problem include Koopmans, Rubin, and Leipnik [1950], Grenander and Rosenblatt [1957], Durbin [1960], and most recently Fuller, Hasza, and Goebel [1981], Hansen [1982], and Lai and Wei [1982]. Koopmans, Rubin and Leipnik were primarily concerned with the extension of the Mann and Wald result to the case where (nonstochastic) exogenous variables are present among the regressors. Moreover, as Durbin notes, the results given in Koopmans, Rubin, and Leipnik depend on a theorem attributed to Rubin [1948] the proof of which was never published. In their proof of the asymptotic distribution of the least squares estimator, Grenander and Rosenblatt refer to Diananda [1953] who in turn uses a result from Mann and Wald. A careful reading of the Durbin paper reveals that at a critical point in his proof, a result from Mann and Wald is again used.

The most recent work on this problem relies on martingale central limit theorems. Fuller, Hasza, and Goebel consider the limiting distribution of the least squares estimator of the parameters of a regression model with lagged values of the dependent variable and nonstochastic exogenous variables as regressors. The proof relies on a martingale central limit theorem given by Scott [1973]. The central limit theorem in Hansen [1982] appears to be based on Theorem 1 in Hannan [1973] which is concerned with the distribution of the least squares estimator in a model with stochastic regressors which are independent of the disturbances. Hannan's proof of this theorem again uses a result given in Scott. Finally, the paper by Lai and Wei includes a central limit theorem for the least squares estimator in the stochastic regressor model and uses a martingale central limit theorem of Dvoretzky [1972].

Thus, while there appear at first glance to be several complete discussions of the asymptotic properties of the least squares estimator in the case of stochastic regressors, the basic references are Mann and Wald, or the more recent treatments that rely on martingale central limit theorems. Mann and Wald maintain a level of generality which renders their notation and derivations cumbersome and difficult to follow. In addition, their primary focus on the presence of a lagged dependent variable in the single-equation model makes it somewhat difficult to see the generalization of their result to a stochastic regressor other than a lagged dependent variable. Martingale central limit theorems have not yet penetrated the traditional econometrics curriculum sufficiently to appeal to a wide audience. In view of the importance of the stochastic regressor case in econometrics, a uniform treatment which is fairly simple and sufficiently general to include the classic Mann and Wald result as well as other stochastic regressor cases seems to be highly desirable.

III. Statement of the Theorem

In the statement and proof of the theorem we use notation which translates naturally into the linear regression context in which the theorem is to be applied. Thus we are concerned with the expression $\sqrt{T}\ (T^{-1}\Sigma x_t\varepsilon_t)$ which is in turn constructed from the sequences $\{x_t\}$ and $\{\varepsilon_t\}$. The following four assumptions specify the properties of $\{x_t\}$ and $\{\varepsilon_t\}$.

A.1) The stochastic sequence $\{\varepsilon_t\}$ is i.i.d with mean zero and variance σ^2.

A.2) The stochastic sequence $\{v_t\}$ is i.i.d. with mean zero and variance δ^2.

A.3) The random variables ε_t and v_{t-j} are stochastically independent for $j \geq 0$ and $j < J < 0$.

A.4) The stochastic sequence $\{x_t\}$ is defined by $x_t = \sum_{j=0}^{\infty} \alpha_j v_{t-j}$, where the α_j (not all zero) are scalar constants which are absolutely summable, i.e., $\sum_{j=0}^{\infty} |\alpha_j|$ is finite.

<u>Theorem</u>. Assumptions (A.1) - (A.4) imply that as $T \to \infty$, $\sqrt{T}\,(T^{-1} \sum_{t=1}^{T} x_t \varepsilon_t)$ converges in distribution to a normal random variable with mean zero and variance $\sigma^2 \delta^2 A$ where $A = \sum_{j=0}^{\infty} \alpha_j^2$.

Before proving the theorem, we remark that it is easily shown that the sequence $\{x_t \varepsilon_t\}$ is uncorrelated, though not independent. It may be thought that uncorrelatedness (orthogonality) would be sufficient to establish the theorem. Unfortunately, this is not the case; there exists no general central limit theorem for uncorrelated random variables.[2]

A theorem similar to ours was proved by Moran [1947]. Our

---

[2]See Granander and Rosenblatt (1957, pp. 180-1) for several examples of uncorrelated random variables with stabilized means that are not asymptotically normal.

proof, unlike the proofs of Moran and Mann and Wald, does not rely on a form of the Liapunov central limit theorem for a doubly subscripted sequence of random variables. Instead, we use results that involve only a modest extension of the standard Lindeberg-Lévy central limit theorem. We first state a lemma that indicates the essential features of this extension.

Lemma. [Anderson (1971, p. 425)] Suppose the sequence $\{Y_T; T = 1, 2, \ldots\}$ of random variables can be expressed as

$$Y_T = S_{kT} + R_{kT}$$

for $T = 1, 2, \ldots$ and $k = 1, 2, \ldots$. If

(A) $\operatorname*{plim}_{k\to\infty} R_{kT} = 0$

uniformly in T,

(B) $S_{kT} \xrightarrow{D} Z_k$

as $T \to \infty$,[3] and

(C) $Z_k \xrightarrow{D} Z$

as $k \to \infty$, then

$Y_T \xrightarrow{D} Z$

as $T \to \infty$.

This lemma, which can be proved using elementary methods,

---

[3]The notation $\xrightarrow{D}$ means that the lefthand quantity has the same asymptotic distribution as the righthand quantity.

indicates that a "sequential" argument can be used to establish the limiting distribution of $\{Y_T\}$. In particular, if condition (A) is satisfied, Y_T and S_{kT} have the same distribution as $k \to \infty$. Thus it is sufficient to examine the doubly subscripted sequence S_{kT} first as $T \to \infty$ and then as $k \to \infty$ to determine the limiting distribution of Y_T. Thus, for example, if for each k S_{kT} converges in distribution to a normal $(0, \sigma_k^2)$ random variable as $T \to \infty$ and $\sigma_k^2 \to \sigma^2$ as $k \to \infty$, we conclude that Y_T converges in distribution to a normal $(0, \sigma^2)$ random variable.

IV. Proof of the Theorem

In order to prove the theorem stated at the beginning of Section III, let

$$x'_{kt} = \sum_{j=0}^{k} \alpha_j v_{t-j}$$

and

$$x''_{kt} = \sum_{j=k+1}^{\infty} \alpha_j v_{t-j}$$

so that

$$x_t = x'_{kt} + x''_{kt}$$

for k = 1, 2, ... The sequence

$$Y_T = \sqrt{T}(T^{-1} \sum_{t=1}^{T} x_t \varepsilon_t) = (1/\sqrt{T}) \sum_{t=1}^{T} x_t \varepsilon_t$$

can be written as

$$Y_T = S_{kT} + R_{kT}$$

where

$$S_{kT} = (1/\sqrt{T}) \sum_{t=1}^{T} x'_{kt} \varepsilon_t$$

and

$$R_{kT} = (1/\sqrt{T}) \sum_{t=1}^{T} x''_{kt} \varepsilon_t \ .$$

The desired result is established by verifying that

(A) $\underset{k\to\infty}{\text{plim}}\ R_{kT} = 0$ uniformly in T,

(B) $S_{kT} \xrightarrow{D} Z_k \sim N(0, \sigma_k^2)$ as $T \to \infty$,

and observing that

(C) $\lim_{k\to\infty} \sigma_k^2 = \sigma^2\delta^2 A$.

(A) We note that $\{x''_{kt}\varepsilon_t;\ t = 1, 2, \ldots\}$ is a sequence of uncorrelated random variables with mean zero and variance $\sigma^2\delta^2(A-A_k)$ where $A_k = \sum_{j=0}^{k} \alpha_j^2$. It follows that $E(R_{kT}) = 0$ and $\text{Var}(R_{kT}) = \sigma^2\delta^2(A-A_k)$. Thus R_{kT} converges in mean square and hence in probability to zero as $k \to \infty$. (The convergence is uniform in T since $Var(R_{kT})$ does not depend on T.)

(B)[4] For any given value of k, let $\chi_t^{(k)} = x_{kt}'\varepsilon_t$ and observe that $\{\chi_t^{(k)};\ t = 1, 2, \ldots\}$ is a sequence of uncorrelated $(0, \sigma_k^2)$ random variables where $\sigma_k^2 = \sigma^2\delta^2 A_k$. In addition, it follows from the definition of $\chi_t^{(k)}$ and the assumptions imposed on ε_t and v_t that $\chi_t^{(k)}$ and $\chi_{t+s}^{(k)}$ are independent for $|s| > k - J \equiv n$; i.e., $\chi_t^{(k)}$ is an n-dependent sequence.

For any integral value of $m > n$ let $M = [T/m]$ be the largest integer less than or equal to T/m so that $T = mM + r$ with $0 \le r < m$. For $m < T$, let[5]

$$U_{mT} = 1/\sqrt{T}\ \big((\chi_1 + \ldots + \chi_{m-n}) + (\chi_{m+1} + \ldots + \chi_{2m-n}) + \ldots + (\chi_{(M-1)m+1} + \ldots + \chi_{Mm-n}) \big)$$

$$V_{mT} = 1/\sqrt{T}\ \big((\chi_{m-n+1} + \ldots + \chi_m) + (\chi_{2m-n+1} + \ldots + \chi_{2m}) + \ldots + (\chi_{Mm-n+1} + \ldots + \chi_{Mm}) \big)$$

$$W_{mT} = 1/\sqrt{T}\ (\chi_{Mm+1} + \ldots + \chi_{Mm+r})\ ;$$

for $T \le m < T+n$, let

---

[4]The proof of this result is given in Anderson [1971, pp. 427-428]. It is summarized here to provide continuity and to make our proof of the theorem self contained.

[5]The explicit dependence of χ_t and hence U_{mT}, V_{mT}, and W_{mT} on k is supressed for notational convenience. This is permissible since the argument is concerned with the limiting distribution of S_{kT} for a given value of k.

$$U_{mT} = 1/\sqrt{T}\ (\ \chi_1 + \dots + \chi_{m-n}\ )$$

$$V_{mT} = 1/\sqrt{T}\ (\ \chi_{m-n+1} + \dots + \chi_T\ )$$

$$W_{mT} = 0;$$

and for $T+n \le m$, let

$$U_{mT} = S_{kT}$$
$$V_{mT} = 0$$
$$W_{mT} = 0\ .$$

Thus, for any given value of k

$$S_{kT} = U_{mT} + V_{mT} + W_{mT}$$

for $T = 1, 2, \dots$ and $m = n+1, n+2, \dots$.

We now use the lemma to obtain the limiting distribution of S_{kT} as $T \to \infty$ with k fixed. In particular, we first show that

(i) $\operatorname{plim}_{m\to\infty} V_{mT} = 0$ uniformly in T, so that only $U_{mT} + W_{mT}$ need be considered in determining the limiting distribution of S_{kT} as $T \to \infty$.

Then, we show that for any fixed value of $m > n$,

(ii) $U_{mT} + W_{mT} \overset{D}{\to} Z_m \sim N(0,\ (1-n/m)\sigma_k^2)$ as $T \to \infty$.

Thus we conclude from the lemma that

$$S_{kT} \overset{D}{\to} Z_k \sim N(0,\ \sigma_k^2) \text{ as } T \to \infty.$$

The proof proceeds as follows.

(i) Since χ_t is a sequence of uncorrelated $(0,\ \sigma_k^2)$ random variables, it follows that $E(V_{mT}) = 0$. For $n < m < T$,

$$\begin{aligned} \mathrm{Var}(V_{mT}) &= (nM/T)\sigma_k^2 \\ &= [nM/(mM+r)]\sigma_k^2 \\ &\leq (nM/mM)\sigma_k^2 \\ &\leq n\sigma_k^2/m; \end{aligned}$$

for $T \leq m < T+n$,

$$\begin{aligned} \mathrm{Var}(V_{mT}) &= [(T-m+n)/T]\sigma_k^2 \\ &\leq n\sigma_k^2/m \end{aligned}$$

since $(m-T)(m-n) \geq 0$ implies that $(T-m+n)/T \leq n/m$; and for $m \geq T+n$, $\mathrm{Var}(V_{mT}) = 0$. Thus

$$\mathrm{Var}(V_{mT}) \leq n\sigma_k^2/m$$

for all T (and $m > n$) so that V_{mT} converges in mean square and hence in probability to zero as $m \to \infty$. (The convergence is uniform in T since $\mathrm{Var}(V_{mT})$ is bounded by $n\sigma_k^2/m$ which is independent of T.)

(ii) Since $E(W_{mT}) = 0$ and

$$\begin{aligned} \mathrm{Var}(W_{mT}) &= (r/T)\sigma_k^2 \\ &\leq (m/T)\sigma_k^2 , \end{aligned}$$

it follows that for any fixed value of m, $\operatorname{plim}_{T\to\infty} W_{mT} = 0$ and $U_{mT} + W_{mT}$ has the same limiting distribution as U_{mT} as $T \to \infty$. Since for $m > n$, $\sqrt{T}U_{mT}$ is the sum of M

i.i.d. $(0,\ (m-n)\sigma_k^2)$ random variables, it follows from the Lindeberg-Lévy central limit theorem that

$$(\sqrt{T}/\sqrt{M})U_{mT} \overset{D}{\rightarrow} Z_m \sim N(0,\ (m-n)\sigma_k^2)$$

or

$$(\sqrt{T}/\sqrt{mM})U_{mT} \overset{D}{\rightarrow} (1/\sqrt{m})Z_m \sim N(0,\ (1-n/m)\sigma_k^2).$$

Since $\sqrt{T}/\sqrt{mM} = \sqrt{mM+r}/\sqrt{mM} = \sqrt{1+r/(mM)} \rightarrow 1$ as T and hence $M \rightarrow \infty$,

$$U_{mT} \overset{D}{\rightarrow} (1/\sqrt{m})Z_m \sim N(0,\ (1-n/m)\sigma_k^2).$$

Since the conditions of the lemma are satisfied: namely,

$$\operatorname*{plim}_{m\rightarrow\infty} V_{mT} = 0 \text{ uniformly in } T,$$

$$U_{mT} + W_{mT} \overset{D}{\rightarrow} Z_m \sim N(0,\ (1-n/m)\sigma_k^2) \text{ as } T \rightarrow \infty,$$

and

$$Z_m \overset{D}{\rightarrow} Z_k \sim N(0,\ \sigma_k^2) \text{ as } m \rightarrow \infty;$$

we conclude that

$$S_{kT} \overset{D}{\rightarrow} Z_k \sim N(0,\ \sigma_k^2)$$

as $T \rightarrow \infty$.

(C) The proof is completed by observing that $\sigma_k^2 \to \sigma^2\delta^2 A$ as $k \to \infty$ so that $(1/\sqrt{T})\Sigma x_t \varepsilon_t$ is indeed asymptotically normal $(0, \sigma^2\delta^2 A)$.

V. Econometric Applications

The central limit theorem of the preceding section is directly applicable to a number of specific models that are commonly encountered in econometrics. This section is devoted to a discussion of the following special cases: 1) the regressor x_t is generated by an autoregressive process, 2) x_t is generated by a finite moving average process, 3) x_t is an i.i.d. sequence, 4) x_t is a lagged dependent variable, 5) x_t is an endogenous variable in a Wold recursive system, and 6) x_t is an exogenous variable to be used as an instrument in a simultaneous equations model.

In each of the cases considered below the estimator to be examined is of the form

$$\sqrt{T}\ (\hat{\beta} - \beta) = D_T\ \sqrt{T}\ (T^{-1}\ \Sigma x_t \varepsilon_t)\ .$$

The asymptotic normality of $\sqrt{T}\ (\hat{\beta} - \beta)$ is obtained by applying the central limit theorem to $\sqrt{T}\ (T^{-1}\ \Sigma\ x_t \varepsilon_t)$ after observing that D_T converges in probability to a finite non-zero constant. In the first five cases that are considered, D_T is given by

$$D_T^{-1} = T^{-1} \Sigma x_t^2 .$$

With x_t defined by (A.2) and (A.4) it is not difficult to verify that

$$\text{plim } T^{-1} \Sigma x_t^2 = \delta^2 A$$

provided v_t has a finite fourth moment. This means that the sample second moment is a consistent estimator of the variance of x_t -- an assumption commonly made in the econometric literature. If $\{x_t\}$ is an i.i.d. sequence as in cases 3 and 5, second moment consistency follows immediately from the weak law of large numbers. If $\{x_t\}$ is a correlated sequence, second moment consistency is less obvious. However, it is true that the sample variance is a consistent estimator of the population variance if the fourth moment of v_t is finite.[6]

1) Autoregressive x_t. In this case the model is written as

$$y_t = \beta x_t + \varepsilon_t$$

where

---

[6]It is immediately apparent that $E(T^{-1}\Sigma x_t^2) = \delta^2 A$. A sufficient condition for $T^{-1}\Sigma x_t^2$ to converge in probability to $\delta^2 A$ is that $\lim_{T\to\infty} \text{Var}(T^{-1}\Sigma x_t^2) = 0$ or equivalently, for the case at hand, that $\lim_{T\to\infty} E[(T^{-1}\Sigma x_t^2)^2] = \delta^4 A^2$. If $(T^{-1}\Sigma x_t^2)$ is written in the terms of the generating process $x_t = \Sigma\alpha_j v_{t-j}$, an examination of the expectation of the resulting expression indicates that the limiting variance of $T^{-1}\Sigma x_t^2$ is zero if v_t has a finite fourth moment. A proof of this assertion is given by Fuller (1976), pp. 239-240.

i) $x_t = \rho x_{t-1} + v_t,\ |\rho| < 1$

ii) A.1, A.2, and A.3 are satisfied.

From i) it follows that the moving average representation of x_t is

$$x_t = \sum_{j=0}^{\infty} \rho^j v_{t-j}$$

so that $\alpha_j = \rho^j$ in (A.4). Since $|\rho| < 1$, the α_j are absolutely summable and $A = \Sigma\alpha_j^2 = 1/(1 - \rho^2)$. Thus the assumptions of the theorem are satisfied and we conclude that

$$\sqrt{T}\ (T^{-1}\ \Sigma\ x_t\varepsilon_t) \overset{D}{\rightarrow} N(0,\ \sigma^2\delta^2 A)$$

and, if v_t has a finite fourth moment

$$\text{plim}\ D_T^{-1} = \delta^2/(1 - \rho^2),$$

so that

$$\sqrt{T}\ (\hat{\beta} - \beta) \overset{D}{\rightarrow} N[0,\ (1 - \rho^2)\sigma^2/\delta^2].$$

We merely note that a similar result holds if x_t is generated by a stable autoregressive process of any finite order. The moving-average representation as well as the expression for the variance of x_t ($\delta^2 A$) are more complicated but no further difficulties are involved in the consideration of higher order autoregressive processes.

2) Finite Moving Average x_t. With a finite moving average

regressor the model is written as

$$y_t = \beta x_t + \varepsilon_t$$

where

i) $x_t = \sum_{j=0}^{q} \alpha_j v_{t-j}$

ii) A.1, A.2, and A.3 are satisfied.

Since x_t is already in moving average form, the central limit theorem applies directly. Hence

$$\sqrt{T}(T^{-1}\Sigma x_t \varepsilon_t) \overset{D}{\to} N(0,\ \sigma^2\delta^2 A)$$

where $A = \sum_{j=0}^{q} \alpha_j^2$. In addition, if v_t has a finite fourth moment,

$$\text{plim } D_T^{-1} = \delta^2 A$$

so that

$$\sqrt{T}(\hat{\beta} - \beta) \overset{D}{\to} N[0,\ \sigma^2/(\delta^2 A)].$$

3) I.i.d. x_t. This is a special case of moving average x_t where $q = 0$ and $\alpha_0 = 1$. Hence we conclude immediately that

$$\sqrt{T}\ (\hat{\beta} - \beta) \overset{D}{\to} N(0,\ \sigma^2/\delta^2).$$

4) $x_t = y_{t-1}$. In the lagged dependent variable case, the model is

$$y_t = \beta x_t + \varepsilon_t \qquad |\beta| < 1$$

with

$$\text{i)}\quad x_t = y_{t-1} = \sum_{j=1}^{\infty} \beta^{j-1}\varepsilon_{t-j} = \sum_{j=0}^{\infty} \beta^{j} v_{t-j}$$

$$\text{ii)}\quad v_t = \varepsilon_{t-1} \sim \text{i.i.d. } (0, \sigma^2).$$

It is clear from the definition of v_t that ε_t and v_{t-j} are independent for all $j \geq 0$ and $j < -1$. Under the restriction that $|\beta| < 1$, it follows that Var $x_t = \sigma^2/(1 - \beta^2)$ and if ε_t has a finite fourth moment,

$$\text{plim } D_T^{-1} = \sigma^2/(1 - \beta^2).$$

This case is thus similar to case 1 and we conclude that

$$\sqrt{T}\,(\hat{\beta} - \beta) \overset{D}{\rightarrow} N[0, (1 - \beta^2)].$$

Note that the model $y_t = \beta x_t + \varepsilon_t$ with $x_t = y_{t-k}$ can be handled in exactly the same fashion.

5) Wold Recursive System. Suppose that x_t is an endogenous variable in the recursive system

$$x_t = \gamma z_t + \eta_t$$
$$y_t = \beta x_t + \varepsilon_t$$

where

$$\text{i)}\quad z_t \sim \text{i.i.d. } (0, \sigma_z^2)$$

ii) $\eta_t \sim$ i.i.d. $(0, \sigma_\eta^2)$ and independent of $z_{t'}$ for all t'

iii) $\varepsilon_t \sim$ i.i.d. $(0, \sigma^2)$

iv) ε_t is independent of both $\eta_{t'}$ and $z_{t'}$ for all t'.

It follows that $v_t = \gamma z_t + \eta_t$ is i.i.d. $(0, \sigma_v^2)$ where $\sigma_v^2 = \gamma^2\sigma_z^2 + \sigma_\eta^2$ and ε_t is independent of $v_{t'}$ for all t'. This case is therefore equivalent to Case 3 and we conclude that

$$\sqrt{T}\,(\hat{\beta} - \beta) \overset{D}{\rightarrow} N(0,\ \sigma^2/\sigma_v^2).$$

We note in passing that if assumption i) is relaxed to allow the exogenous variable z_t to be generated by either an autoregressive or a moving average process, x_t is no longer of the form postulated in (A.4). For example, suppose z_t is generated by

$$z_t = \rho z_{t-1} + \xi_t$$

where $\xi_t \sim$ i.i.d. $(0, \sigma_\xi^2)$ and ε_t and $\xi_{t'}$ are independent for all t'. Then x_t becomes

$$x_t = \eta_t + \gamma \sum_{j=0}^{\infty} \rho^j \xi_{t-j}$$

which is not directly of the form $\Sigma\alpha_j v_{t-j}$. It would not be difficult, however, to modify our theorem to accommodate such a case.

6) Simultaneous Equations Model. Consider a single equation

$$y_t = \beta y_t^* + \varepsilon_t$$

embedded in a simultaneous system where y_t^* is also an endogenous variable. If x_t is an exogenous variable in some other equation in the model, an instrumental variable estimator of β is

$$\hat{\beta} = \beta + (\Sigma\, x_t y_t^*)^{-1}\, \Sigma\, x_t \varepsilon_t$$

and

$$\sqrt{T}\,(\hat{\beta} - \beta) = D_T\, \sqrt{T} \cdot (T^{-1}\, \Sigma x_t \varepsilon_t)$$

where

$$D_T^{-1} = T^{-1}\, \Sigma x_t y_t^* \,.$$

If $\varepsilon_t \sim$ i.i.d. $(0,\ \sigma^2)$ and x_t is a) i.i.d., b) stationary autoregressive, or c) finite moving average, the conditions of the central limit theorem will be satisfied. Therefore

$$\sqrt{T}\,(T^{-1}\, \Sigma x_t \varepsilon_t) \overset{D}{\to} N[0,\ \sigma^2\, \mathrm{Var}(x_t)] \,.$$

Provided that D_T converges in probability to a finite nonzero constant, say Q, we conclude that

$$\sqrt{T}\,(\hat{\beta} - \beta) \overset{D}{\to} N[0,\ \sigma^2\, \mathrm{Var}(x_t) Q^2] \,.$$

VI. Extensions and Conclusions

The central limit theorem of Section III is readily applied to establish asymptotic normality of the Aitken estimator corresponding to a regression equation with an autoregressive error term. We present the result for the case of first-order autoregression; the generalization to any finite order stable autoregressive process is immediately apparent. Suppose that

$$y_t^* = \beta x_t^* + u_t$$

where

i) $u_t = \rho u_{t-1} + \varepsilon_t \qquad |\rho| < 1$

ii) the stochastic sequence $\{x_t^*\}$ is defined by $x_t^* = \sum_{j=0}^{\infty} \alpha_j^* v_{t-j}$ where the α_j^* (not all zero) are scalar constants and absolutely summable, and

iii) A.1, A.2, and A.3 are satisfied[7]

Let

$$y_t = y_t^* - \rho y_{t-1}^*$$

$$x_t = x_t^* - \rho x_{t-1}^*$$

so that the original equation may be transformed to yield

$$y_t = \beta x_t + \varepsilon_t \ .$$

Now observe that

$$x_t = \sum_{j=0}^{\infty} \alpha_j^* v_{t-j} - \rho \sum_{j=1}^{\infty} \alpha_j^* v_{t-j}$$

[7]Assumptions A.1 and A.3 refer to $\{\varepsilon_t\}$, not $\{u_t\}$.

$$= \sum_{j=0}^{\infty} \alpha_j v_{t-j}$$

where

$$\alpha_0 = \alpha_0^*$$

$$\alpha_j = \alpha_j^* - \rho\alpha_{j-1}^*, \text{ for } j \geq 1.$$

Since

$$\sum_{j=0}^{\infty} |\alpha_j| = \left|\alpha_0^*\right| + \sum_{j=1}^{\infty} \left|\alpha_j^* - \rho\alpha_{j-1}^*\right|$$

$$\leq \left|\alpha_0^*\right| + \sum_{j=1}^{\infty} \left(\left|\alpha_j^*\right| + |\rho| \left|\alpha_{j-1}^*\right| \right)$$

$$\leq \sum_{j=0}^{\infty} \left|\alpha_j^*\right| + |\rho| \sum_{j=0}^{\infty} \left|\alpha_j^*\right|$$

we see that the α_j are absolutely summable. It follows that the transformed equation

$$y_t = \beta x_t + \varepsilon_t$$

satisfies all of the conditions for the central limit theorem of Section III to be applied and

$$\sqrt{T}(\hat{\beta} - \beta) = D_T \sqrt{T}(T^{-1} \Sigma x_t \varepsilon_t) \overset{D}{\to} N[0, \sigma^2/(\delta^2 A)]$$

where as before $D_T^{-1} = T^{-1}\Sigma x_t^2$ is assumed to converge in probability to $\delta^2 A$. But when expressed in terms of the original variables,

$$\sqrt{T}(\hat{\beta} - \beta) = \frac{1}{T^{-1}\Sigma(x_t^* - \rho x_{t-1}^*)^2} \sqrt{T}T^{-1} \Sigma(x_t^* - \rho x_{t-1}^*)(u_t - \rho u_{t-1})$$

so that $\hat{\beta}$ is the Aitken estimator of β.[8]

Up to now we have restricted attention to the case of a single explanatory variable. Our results, however, can be extended to the multiple regression model without difficulty. We first state the multivariate analogue of the central limit theorem of Section III and then show how the result would be used in practice.

The assumptions that underlie the multivariate central limit theorem are as follows.

A'.1) $\varepsilon_t \sim$ i.i.d. $(0, \sigma^2)$

A'.2) $V_t \sim$ i.i.d. $(0, \Delta)$ where V_t is a $(p \times 1)$ vector.

A'.3) ε_t is independent of (each element of) V_{t-j} for $j \geq 0$ and $j < J < 0$.

A'.4) The random vector X_t' is defined as

$$X_t' = \sum_{j=0}^{\infty} V_{t-j}' D(\underline{\alpha}_j)$$

where $D(\underline{\alpha}_j)$ denotes a diagonal matrix with elements of the vector $\underline{\alpha}_j' = (\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{pj})$ on the diagonal. The sequences $\{\alpha_{ij};\ j = 1, 2, \ldots\}$ are assumed to be

---

[8]Obviously, there is no need to be concerned with the so-called "first-observation problem" in this asymptotic context.

absolutely summable so that the sequence of vectors $\{\underline{\alpha}_j\}$ satisfies the condition

$$\sum_{j=0}^{\infty} \underline{\alpha}_j \underline{\alpha}_j' = A ,$$

where A is a non-null matrix of finite constants.

It is readily apparent that these assumptions are generalizations of those in Section III and guarantee that each element of the vector X_t satisfies the conditions which were previously postulated for the scalar x_t.

<u>Theorem</u>. Assumptions A'.1-A'.4 imply that as $T \to \infty$, $\sqrt{T}\ (T^{-1}\ \Sigma X_t' \varepsilon_t)$ converges in distribution to the p-variate normal with mean vector 0 and covariance matrix $\sigma^2(\Delta * A)$, where $\Delta * A$ denotes the element by element product of Δ and A (each of which is p x p).

A proof of this theorem is obtained by going through the steps of Section IV for the vector case. Rather than do this here, we simply show how the covariance matrix of $\sqrt{T}\ (T^{-1}\Sigma X_t' \varepsilon_t)$ is obtained. Since $E(X_t' \varepsilon_t) = 0$ it follows that the covariance matrix (denoted in general by Ω) of $X_t' \varepsilon_t$ is

$$\Omega_{X_t' \varepsilon_t} = E(\varepsilon_t^2\ X_t X_t') = \sigma^2\ \Omega_{X_t'} .$$

From the definition of X_t', the covariance matrix of X_t' is

$$\Omega_{X_t'} = \sum_{j=0}^{\infty} E\ \{D(\underline{\alpha}_j)\ V_{t-j}\ V_{t-j}'\ D(\underline{\alpha}_j)\}$$

$$= \sum_{j=0}^{\infty} D(\underline{\alpha}_j)\ \Delta\ D(\underline{\alpha}_j)\ .$$

Further, it can be shown that

$$\sum_{j=0}^{\infty} D(\underline{\alpha}_j)\ \Delta\ D(\underline{\alpha}_j) = [\delta_{i\ell}\ a_{i\ell}] \qquad i,\ \ell = 1,\ 2,\ \ldots,\ p$$

$$= \Delta * A$$

where $[a_{i\ell}] = A$ as defined in (A'.4). We conclude that

$$\Omega_{X_t' \varepsilon_t} = \sigma^2\ (\Delta * A)\ .$$

Further, since successive elements of the sum $\Sigma_t\ X_t' \varepsilon_t$ are uncorrelated,

$$\Omega_{(\sqrt{T}\ T^{-1}\ \Sigma\ X_t' \varepsilon_t)} = \sigma^2 (\Delta * A)\ .$$

Thus the assumption (A'.4) guarantees that $\sqrt{T}\ T^{-1}\ \Sigma\ X_t' \varepsilon_t$ has a finite covariance matrix.

As an illustration of the use of this theorem in practice, consider the multiple regression model

$$y_t = X_t' \beta + \varepsilon_t \qquad (t = 1, 2, \ldots, T)$$

where β is now a $(p \times 1)$ vector. The stabilized least squares estimator is given by

$$\sqrt{T}\,(\hat{\beta} - \beta) = (T^{-1} \Sigma X_t X_t')^{-1} \sqrt{T}\,(T^{-1} \Sigma X_t' \varepsilon_t) \ .$$

Provided that i) X_t' and ε_t satisfy (A'.1) - (A'.4), and ii) $T^{-1} \Sigma X_t X_t'$ converges in probability to $\Delta{*}A$, and iii) $\Delta{*}A$ is nonsingular, it follows that

$$\sqrt{T}\,(\hat{\beta} - \beta) \overset{D}{\rightarrow} N(0, \sigma^2 (\Delta{*}A)^{-1}) \ .$$

Clearly, the elements of X_t' can be any mixture of autoregressive, moving average, or lagged dependent variables which satisfy the assumptions (A'.2) - (A'.4).

As a final illustration we re-cast the preceding example in the matrix notation most used in the econometric literature. The matrix $\Delta{*}A$ is, of course, the population covariance matrix associated with the vector of regressors X_t'. The form $\Delta{*}A$ emphasizes the functional dependence of X_t' on the sequence $\{V_t\}$. Ignore this dependence and denote the matrix $\Delta{*}A$ by M_x. Write the multiple regression model in matrix form as

$$Y = X\beta + \varepsilon$$

where Y is $(T \times 1)$, X is $(T \times p)$, β is $(p \times 1)$, and ε is $(T \times 1)$.

Assume

$$\text{Plim } (T^{-1}X'X) = M_x$$

and M_x non-singular. Then if the rows of the matrix X satisfy the assumptions of the central limit theorem, it follows that

$$\sqrt{T}\,(\hat{\beta} - \beta) \overset{D}{\to} N(0, \sigma^2 M_x^{-1}) .$$

The extension of the multiple regression result to allow for Aitken estimation of the vector β when ε_t is a stable autoregressive process is entirely analogous to the extension already presented for the simple regression case.

In effect, this paper shows that under fairly general conditions, it is valid to assume asymptotic normality of the least squares (or Aitken) estimator in a multivariate, stochastic regressor, linear model -- just as most of us have done all along.

REFERENCES

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

Cramér, Harold (1946), Mathematical Methods of Statistics, Princeton: Princeton University Press.

Diananda, P.H. (1953), "Some Probability Limit Theorems with Statistical Applications," Proceedings of the Cambridge Philosophical Society, 49, 239-246.

Durbin, J. (1960), "Estimation of Parameters in Time-Series Regression Models," Journal of the Royal Statistical Society, Series B, 22, 139-153.

Dvoretzky, A. (1972), "Asymptotic Normality for Sums of Dependent Random Variables," Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley: University of California Press.

Feller, William (1966), An Introduction to Probability Theory and its Applications, Volume II, New York: Wiley.

Fuller, Wayne A. (1976), Introduction to Statistical Time Series, New York: Wiley.

Fuller, Wayne A., David P. Hasza and J. Jeffery Goebel (1981), "Estimation of the Parameters of Stochastic Difference Equations," The Annals of Statistics, 9, 531-543.

Grenander, Ulf and Murray Rosenblatt (1957), Statistical Analysis of Stationary Time Series, New York: Wiley.

Hannan, E.J. (1961), "A Central Limit Theorem for Systems of Regressions," Proceedings of the Cambridge Philosophical Society, 57, 583-588.

Hannan, E.J. (1973), "Central Limit Theorems for Time Series," Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gabiete, 26, 157-170.

Hansen, Lars Peter (1982), "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 50, 1029-1054.

Koopmans, T.C., H. Rubin, and R.B. Leipnik (1950), "Measuring the Equation Systems of Dynamic Economics," Chapter 2 of Statistical Inference in Dynamic Economic Models, edited by T.C. Koopmans, New York: Wiley.

Lai, Tze Leung and Ching Zong Wei (1982), "Least Squares Estimates in Stochastic Regression Models With Applications to Identification and Control of Dynamic Systems," The Annals of Statistics, 10, 154-166.

Malinvaud, E. (1966), Statistical Methods of Econometrics, Chicago: Rand McNally.

Mann, H.B. and A. Wald (1943), "On the Statistical Treatment of Linear Stochastic Difference Equations," Econometrica, 11, 173-220.

Moran, P.A.P. (1947), "Some Theorems on Time Series I," Biometrika, 34, 281-291.

Rubin, Herman (1948), "Some Results on the Asymptotic Distribution of Maximum - and Quasi-Maximum - Likelihood Estimates," Annals of Mathematical Statistics, 20, 598.

Scott, D.J. (1973), "Central Limit Theorems for Martingales and for Processes With Stationary Increments Using A Skorokhod Representation Approach," Advances in Applied Probability, 5, 119-137.

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