A CENTRAL LIMIT THEOREM FOR SYSTEMS OF REGRESSIONS

by

E. J. Hannan

University of North Carolina

and

Canberra University College

This research was supported by the Office of Naval Research under Contract No. Nonr 855(09) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part for any purpose of the United States Government is permitted

Institute of Statistics
Mimeograph Series No. 247
March 1960

A CENTRAL LIMIT THEOREM FOR SYSTEMS OF REGRESSIONS[1]

E. J. Hannan
The University of North Carolina
and
Canberra University College

1. Introduction:

The theory of regression on fixed variables, when the residuals are generated by a stationary process, has been illuminated by the introduction of certain restrictions on the regressor vectors, of a very general form, by Grenander (see Grenander and Rosenblatt (1957)). These restrictions are designed to concentrate attention on cases where consistent estimates of the regression coefficients exist and the asymptotic properties of these estimates may be investigated by Fourier methods. In the case of a single regression on a single fixed variable, x_t, it is required that x_t be generated by some law such that $\Sigma x_t^2 \longrightarrow \infty$ as the number, n, of observations increases while the limit of $(\Sigma x_t^2)^{-1} \Sigma x_t x_{t+h}$ exists. In order to make this limit independent of end effects due to slight variations in the definition of the correlation it is further required that $x_n^2/\Sigma x_t^2$ should tend to zero. The restrictions will be satisfied in most cases of regression on functions of time as well as in cases where the x_t are generated by some stationary process for which the sample serial correlations converge with probably one.

1. This research was supported by the Office of Naval Research under Contract No. Nonr 855(09) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part for any purpose of the United States Government is permitted.

It is the purpose of this paper to show that these conditions are (together with certain restrictions on the nature of the process generating the residuals) sufficient to ensure that the estimates of the regression coefficients are asymptotically normal. We begin immedicately, however, with the case of a multiple system of regressions.

There is, of course, already in the literature a very considerable number of forms of central limit theorem, similar to that to be proved below. Many of these (Bernstein (1927), Diaranda (1953), Moran (1947)) refer only to estimation of the mean and they may also impose stronger conditions on the nature of the serial dependence of the residuals (balancing this with weaker conditions on their higher moments). In Koopmans (1950) multiple systems of regressions are considered (see also Mann and Wald (1943)) but the discussion given there overlaps only slightly with that to be given below mainly because the residual process is required to be one of independent random variables while lagged values of the dependent variables are included in the regressor set. The "independent" (exogenous) variables are required to be bounded and there are other important differences. The justification for the present extension in another direction lies in the importance and generality of the class of regressors considered, which was discussed in the first paragraph, together with the fact that the relatively simple conditions imposed on the regressors appear to be reasonably near to necessary conditions for the central limit property to hold, for a sufficiently general class of stationary residuals.

2. Conditions of the Problem and Notation

We consider a system of regressions

$$z_t = By_t + x_t \tag{1}$$

where z_t and x_t are vectors of p components, y_t is a vector of q components and B is a (pxq) matrix of regression coefficients, β_{ij}. We consider the case where

A. $$x_t = \sum_{-\infty}^{\infty} A_j \epsilon_{t-j}$$

where ϵ_t are mutually independent and identically distributed random vectors of r components with covariance matrix G and finite absolute moments.* We define the norms $\| A_j \|_p$, by

$$\| A_j \|_p = \sup \| x \|_p^{-1} \|A_j x\|_p$$

where the supremum is over all non-null vectors in the domain of A_j and $\| x \|_p = \left\{\Sigma |x_i|^p\right\}^{1/p}$.

We then require

B. $$\sum_{-\infty}^{\infty} \| A_j \|_2 < \infty$$

Then the matrix spectral density of the x_t process is

$$F(\lambda) = \frac{1}{2\pi} \left(\sum_{-\infty}^{\infty} A_j e^{ij\lambda}\right) G \left(\sum_{-\infty}^{\infty} A_j' e^{-ij\lambda}\right).$$

Following Grenander we require that the components, $y_{j,t}$ of y_t are generated by a process such that, with probability one, as n increases,

* Actually the finiteness of an absolute moment of order $2+\delta$, $\delta > 0$, is all that is required below.

C. $$d^2_{j,n} = \sum_1^n y^2_{j,t} \longrightarrow \infty \quad , \text{ all } j,$$

D. $$y^2_{j,n}/d^2_{j,n} \longrightarrow 0 \quad , \text{ all } j,$$

E $$\text{Lt} \sum_1^n y_{j,t} y_{k,t-h} \Big/ d_{j,n} d_{k,n} \text{ exists, all } j,k.$$

We call this limit $\rho_{j,k}(h)$.

We put

(2) $$R(h) = [\rho_{jk}(h)]$$

and assume that $R(o)$ is non singular.

We may write

(3) $$R(h) = \int_{\pi}^{\pi} e^{ih\lambda} dM(\lambda)$$

where $M(\lambda)$ is a matrix valued function whose increments are Hermitian non-negative (Grenander and Rosenblatt (1957)).

The least squares estimate of B is

(4) $$\hat{B} = [b_{ij}] = ZY' [YY']^{-1}$$

Where Y has y_t and Z has z_t as the t-th column.

It is preferable to express (4) in the form of a column vector b having b_{ij} in row $(p(i-1)+j$. The expectation of b we call β. Then

$$b = (I_p \otimes YY')^{-1} (\sum_t z_t \otimes y_t)$$

where $\otimes$ means the Kronecker product of two vectors, or of two matrices, and I_p is the p dimensional unit matrix. We introduce the diagonal matrix D_n having $d_{j,n}$ in the j-th place. Then Rosenblatt (1956) has shown that

$$\text{(5)} \quad \mathcal{E}\left\{(I_p \otimes D_n)(b-\beta)(b-\beta)'(I_p \otimes D_n)\right\} \longrightarrow (I_p \otimes R(0))^{-1} 2\pi \int_{-\pi}^{\pi} F(\lambda) \otimes dM(\lambda)(I_p \otimes R(0))^{-1}$$

The integral here is of course a matrix of integrals of the form

$$\int_{-\pi}^{\pi} f_{ij}(\lambda)\ dM_{k\ell}(\lambda)$$

where $f_{ij}(\lambda)$ and $M_{k\ell}(\lambda)$ are typical elements of $F(\lambda)$ and $M(\lambda)$ respectively.

In the derivation of (5) by Rosenblatt, which he expresses in a different form, it is assumed that $F(\lambda)$ is non singular for every λ in $[-\pi,\pi]$. However (5) may be extended to the general case by replacing x_t by $x_t + \alpha\eta_t$ where η_t is a process of independent random vectors, independent also of x_t with unit covariance matrix, and $\alpha > 0$. Then $F(\lambda)$ is replaced by $F(\lambda)+(\alpha/2\pi)\ I_p$ which is non singular. Allowing α to tend to zero (5) follows for the general case.

We make the final assumption:

F. The right hand member of (5) is not the null matrix. A word of explanation is called for in relation to this condition. A corresponding condition has not been inserted in Diananda (1953)(or Anderson (1953) Theorem 4.4) so that the central limit theorems must there be interpreted, in some cases, in a sense in which the central limit theorem would not usually be interpreted. For example, if we consider the case where $p = q = 1$, $\beta_{11} = 0$ and $y_{1,t} \equiv 1$ while $x_t = \epsilon_t - \epsilon_{t-1}$ we see that $d_{11}b_{11} = \Sigma x_t/\sqrt{n} = (\epsilon_n - \epsilon_o)/\sqrt{n}$ which converges in probability to zero. The difficulty arises in this case fundamentally because $M_{11}(\lambda)$ has a single point of increase, at $\lambda = 0$, at which $f_{11}(\lambda) = 0$. Since $M(\lambda)$ will have the origin as a single point of increase when the $y_{j,t}$ correspond to a polynomial regression $F(\lambda)$ must not be null at the origin in this case. Thus, we should not difference any initial system of the form (1) for example, (as is sometimes done) since this will make $F(\lambda)$ null to the origin and prevent the theorem stated below from being applicable. It is known, from the work of Grenander and Rosenblatt that, in the situation just described, differencing will also result in a loss of asymptotic efficiency.

We have allowed the right hand side of (5) to be singular. However, if it is not null we are reasonable in calling our theorem a central limit theorem since the dominant term in an asymptotic expansion for the distribution is, at least, not now a distribution concentrated at the origin.

3. The Central Limit Theorem.

We wish to prove that the distribution of $(I_p \quad D_n)(b-\beta)$ converges to a (possibly singular) normal distribution with covariance matrix given by (5). Let us write

$$(I_p \otimes D_n)(b-\beta) = c_1 + c_2$$

where for c_1 we replace x_t in $(I_p \otimes D_n(b-\beta)$ by

$$\hat{x}_t = \sum_{-N}^{N} A_j \epsilon_{t-j}, \text{ N fixed,}$$

while in c_2 we replace x_t by

$$\hat{\hat{x}}_t = \sum_{j>N} A_j \epsilon_{t-j}.$$

Evidently

$$(6) \qquad (c_i c_i') \longrightarrow \left\{I_p \otimes R(o)\right\}^{-1} 2\pi \int_{-\pi}^{\pi} F_i(\lambda) \quad dM(\lambda) \left\{I_p \otimes R(o)\right\}^{-1}, \ i=1,2,\ldots$$

where $F_i(\lambda)$ is obtained from $F(\lambda)$ by making the summation run over the appropriate set of indices.

Since

$$\|F_2(\lambda)\|_2 \leq \frac{1}{2\pi}\| G \|_2 \|\sum_{j>N} A_j e^{ij\lambda}\|_2^2$$

$$\leq \frac{1}{2\pi} \| G \|_2 \sum_{j>N} \|A_j\|_2{}^2$$

we may make this norm arbitrarily small, uniformly in λ, by choosing N sufficiently large, because of B. Then the same norm of the right hand member of (6), for $i = 2$, may also be made arbitrarily small by taking N sufficiently large. If now we can show that when n increases the distribution of c_1 converges to a normal distribution we shall have accomplished what we wish since by choosing first N large and then allowing n to increase we may take the distribution of $(I_p \otimes D_n)(b-\beta)$ differ by as little as we like from its limiting form; a normal distribution.

Now

$$c_1 = (I_p \otimes D_n)(I_p \otimes YY')^{-1}(I_p \otimes D_n).(I_p \otimes D_n)^{-1}(\sum_t \hat{x}_t \otimes y_t)$$

and the product of the first three factors converges to $\left\{ I_p \otimes R(o) \right\}^{-1}$. Thus we need to show that the product of the last two has a limiting normal distribution. However

$$\sum_t \hat{x}_t \otimes y_t = \sum_t \left(\sum_{-N}^{N} A_j \epsilon_{t-j} \right) \otimes y_t$$

$$= \sum_{s=-N+1}^{N} \left\{ \sum_{j=-s+1}^{N} A_j \epsilon_s \otimes y_{j+s} \right\} + \sum_{n-N+1}^{n+N} \left\{ \sum_{j=-N}^{n-s} A_j \epsilon_s \otimes y_{j+s} \right\} \tag{7}$$

$$+ \sum_{s=N+1}^{n-N} \left\{ \sum_{j=-N}^{N} A_j \, \epsilon_s \otimes y_{j+s} \right\}.$$

Since

$$\mathcal{E}\left\{\|(I_p\otimes D_n)^{-1}\sum_{-N}^{N} A_j\,\epsilon_s\otimes y_{j+s}\|_2^2\right\}$$

$$= (I_p\otimes D_n)^{-1}\sum\sum_{-N}^{N}\mathcal{E}\,[(A_j\otimes I_q)(\epsilon_s\epsilon_s'\otimes y_{j+s}y'_{k+s})(A_j'\otimes I_q)](I_p\otimes D_n)^{-1}$$

$$= (I_p\otimes D_n)^{-1}\sum\sum_{-N}^{N} A_jG\,A_k\otimes y_{j+s}y'_{k+s}\quad(I_p\otimes D_n)^{-1}$$

$$= \sum\sum_{-N}^{N}(A_jG\,A_k)\otimes(D_n^{-1}\,y_{j+s}y'_{k+s}D_n^{-1}),$$

which converges to a null matrix by D above, we see that the first two terms in (7) may be neglected. The expected value of the same (squared) norm of the last term in (7) converges to the right hand member of (6) (with i=1) which will not be null if N is large enough because of F. We shall have accomplished our purpose if we show that the sum, over s, of the third absolute moments* of any element of

$$(8)\qquad (I_p\otimes D_n)^{-1}\sum_{j=-N}^{N} A_j\,\epsilon_s\otimes y_{j+s}$$

converges to zero (see for example, Bernstein (1927) pp. 44-50.). The sum of the third absolute moments of the elements of (8) is however

$$\mathcal{E}\left\{\|(I_p\otimes D_n)^{-1}\sum_{j=-N}^{N} A_j\,\epsilon_s\otimes y_{j+s}\|_3^3\right\}$$

$$\leq \mathcal{E}\left[\left\{\sum_{N}^{N}\|A_j\epsilon_s\otimes D_n^{-1}\,y_{j+s}\|_3\right\}^3\right]$$

$$\leq \left[\sum\left\{\mathcal{E}\,(\|A_j\epsilon_s\otimes D_n^{-1}\,y_{j+s}\|_3^3)\right\}^{1/3}\right]^3$$

$$\leq \left[\sum\left\{\mathcal{E}\,(\|A_j\|_3^3\,\|\epsilon_s\|_3^3\,\|D_n^{-1}\,y_{j+s}\|_3^3)\right\}^{1/3}\right]^3$$

$$(9)\qquad \leq Mp\left\{\sum\|A_j\|_3\|D_n^{-1}y_{j+s}\|_3\right\}^3$$

* Again a consideration of a moment of order $2+\delta$, $\delta>0$, would, of course suffice.

where M is an upper bound to the third absolute moments of the elements of ϵ_s. Summing (9) over 5 from N+1 to n-N we find that the sum, over s, of the sum of the third absolute moments of the elements of (8), is dominated by

$$M_p(2N+1) \sum_j \| A_j\|_3^3 \quad \left\{ \sum_{s=1}^{n} \| D_n^{-1} y_s \|_3^3 \right\}$$

The term in braces is, however

$$\sum_{k=1}^{q} \sum_{s=1}^{n} \left| \frac{y_{k,s}}{d_{k,n}} \right|^3 \leq \sum_{k=1}^{\epsilon} \frac{\max_{s \leq n} |y_{k,s}|}{d_{k,n}}$$

which converges to zero because of D. Thus the sum over s of the third absolute moments of any element of (8) converges to zero also and the central limit theorem is proved.

References

Anderson T. W., "On asymptotic distributions of estimates of parameters of stochastic difference equations," Ann. Math. Stat. 30, (1959) 676-687.

Bernstein, S., "Sur l'extension du théorème limite du calcul des probabilites aux sommes de quantités dépendantes," Math. Annalen, 97 (1927), 1-59.

Diananda, P. H. "Some probability limit theorems with statistical applications," Proc. Camb. Phil. Soc., 49, (1953) 239-246.

Grenander, U. and Rosenblatt, M., "Statistical Analysis of Stationary Time Series." New York. John Wiley and Sons (1957).

Koopmans, T. J., (Editor) "Statistical Inference in Dynamic Economic Models," New York (1959).

Mann, H. B., and Wald, A. "On the statistical treatment of linear stochastic difference equations, Econometrica, 11 (1943) 173-220.

Moran, P. A., "Some theorems on time series," Biometrika, 34, (1947) 281-291.

Rosenblatt, M., "On the estimation of regression coefficients of a vector-valued time series with a stationary residual," Ann. Math. Stat., 27, (1956) 99-121.