Proceedings of the 1986 Winter Simulation Conference
J. Wilson, J. Henriksen, S. Roberts (eds.)

BEHAVIOR OF SAMPLE MEANS OF NEARLY NONSTATIONARY TIME SERIES

Ruey S. Tsay
Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213, U.S.A.

ABSTRACT

Sample means are perhaps the most commonly used statistics in data analysis. When the data points are independent, behavior of sample means is governed by the Law of Large Numbers. When the data are serial correlated, theoretical results of stationary stochastic processes are often used to describe the behavior of sample means. How do the sample mean behave when data are nonstationary (or nearly so) is yet to be discussed. In this paper, I give the limiting distribution of sample means when the processes are either nonstationary or nearly nonstationary. I also discuss why the conventional formula fails to provide adequate inference for sample means when the processes are nearly nonstationary.

1. INTRODUCTION

For a given set of data $\{X_1, X_2, \ldots, X_n\}$, the sample mean $\bar{X}_n = n^{-1}\Sigma_{i=1}^n X_i$ is perhaps the most commonly used statistic in inference making. It is routinely reported in the summary table of any data analysis in every statistical package. It is also a well-studied statistic in the literature. For example, from the Law of Large Numbers we have the following results.

<u>Case 1: Independent and Identically Distributed</u>. Let X_i be a sequence of independent and identically distributed (IID) random variables.

- Khinchine's WLLN:

$$E(X_i) = \nu < \infty \Rightarrow \bar{X}_n \rightarrow_p \nu.$$

- Kolmogorov's SLLN:

$\bar{X}_n \rightarrow_{a.s.} \nu$ IFF $E(X_i)$ exists and is equal to ν.

<u>Case 2: Independent but Not Identically Distributed</u>. Let X_i be a sequence of independent variables such that $E(X_i) = \nu_i$ and $Var(X_i) = \sigma_i^2$. Define $\bar{\nu}_n = n^{-1}\Sigma_{i=1}^n \nu_i$.

- Chebyshev's WLLN:

$$\lim_{n\rightarrow\infty} n^{-2}\sum_{i=1}^{n} \sigma_i^2 = 0 \Rightarrow \bar{X}_n - \bar{\nu}_n \rightarrow_p 0.$$

- Kolmogorov's SLLN:

$$\sum_{i=1}^{\infty} \sigma_i^2 / i^2 < \infty \Rightarrow \bar{X}_n - \bar{\nu}_n \rightarrow_{a.s.} 0.$$

In the above, $\rightarrow_p$ denotes convergence in probability, $\rightarrow_{a.s.}$ convergence with probability 1, i.e., almost sure convergence, WLLN and SLLN denote weak and strong laws of large numbers, respectively.

When the independence condition is replaced by weak stationarity, that is $E(X_i) = \nu$, $Var(X_i) = \sigma_x^2$, and $Cov(X_i, X_{i+k}) = R(|k|)$, which does not depend on the time i. Let $\rho_x(k) = R(k)/R(0)$ be the autocorrelation of X_i at lag k. Then, it is easily seen that, (Priestley, 1980),

$$E(\bar{X}_n) = \nu$$

$$Var(\bar{X}_n) = n^{-1}\sigma_x^2 \sum_{k=-(n-1)}^{n-1} (1 - |k|/n)\rho_x(k). \tag{1}$$

Thus, $\bar{X}_n$ is an unbiased estimator of ν and it is consistent if the right hand side of (1) converges. In particular, if X_i has a purely continuous spectrum with (normalized) spectral density function,

$$f(\omega) = (2\pi)^{-1} \sum_{-\infty}^{\infty} \rho_x(k)\exp(-i\omega k) \quad \text{for } -\pi \le \omega \le \pi,$$

then we have

$$\lim_{n\rightarrow\infty} \sum_{k=-(n-1)}^{n-1} (1 - |k|/n)\rho_x(k) = \sum_{k=-\infty}^{\infty} \rho_k(k) = 2\pi f(0). \tag{2}$$

Therefore, for large n

$$Var(\bar{X}_n) \sim n^{-1} 2\pi f(0)\sigma_x^2. \tag{3}$$

To illustrate the use of (3), suppose that X_i follows a stationary autoregressive process of order 1, i.e., AR(1) model

$$X_i = \Phi X_{i-1} + \epsilon_i, \quad |\Phi| < 1$$

where ϵ_i is a sequence of IID Gaussian noises with mean zero and variance σ^2. Then, (3) reduces to

$$\mathrm{Var}(\bar{X}_n) \sim n \; [illegible] \; 1+\Phi)/(1-\Phi)]. \tag{4}$$

Compared with the IID case in which $\mathrm{Var}(\bar{X}_n) = n^{-1}\sigma_x^2$, it is clear that the effect of serial correlation is $(1+\Phi)/(1-\Phi)$. For this reason, $n(1-\Phi)/(1+\Phi)$ is often regarded as the *equivalent degree of freedom* of an AR(1) process.

In theory (3) applies to all stationary processes with purely continuous spectrum. However, the approximation could be very poor for the finite sample case when X_i is nearly nonstationary. This can easily be seen from the result of the next section.

2. NONSTATIONARY PROCESSES

For better understanding of the nonstationary case, I consider the simple random walk model in this section. X_t follows a random walk model if it satisfies

$$X_i = X_{i-1} + \epsilon_i \, . \tag{5}$$

where ϵ_i is defined as before in Section 1. I shall assume that the starting value X_0 is zero. This assumption has no effect in the limiting behavior of $\bar{X}_n$, see Tiao and Tsay (1986). For finite sample, the effect might not be negligible. However, the assumption can always be met in practice by treating X_1 as a fixed number and subjecting it from all the observations.

Since X_i is nonstationary, the asymptotic property of $\bar{X}_n$ is non-standard in the sense that it is different from those obtained by the central limit or ergodic theorem. The result involves stochastic integrals of Brownian motions, and I use the following notations: (a) D[0,1] is the space of functions g(t) on the unit interval [0,1] which are right continuous and have left-hand limits, see Billingsley (1968); (b) D[0,1] is equipped with the Skorohod topology; (c) $Y_i \to_d Y$ denotes that the sequence $\{Y_i\}$ converges in distribution to an element Y in D[0,1]; (d) For $0 \le s \le 1$, [ns] denotes the largest integer less than or equal to ns.

From (5) and under the assumption of zero starting value, we have

$$X_i = \sum_{t=1}^{i} \epsilon_t. \tag{6}$$

Let

$$P_i = \sum_{t=1}^{i} \epsilon_t$$

be the i-th partial sum of $\{ \epsilon_t \mid 0 < t < \infty \}$ and define the function $Z_n(s)$ on [0,1] as

$$Z_n(s) = (n^{1/2}\sigma_\epsilon)^{-1} P_{[ns]} \quad \text{for } 0 \le s \le 1 \tag{7}$$

where $Z_n(0) = 0$. Then, by Donsker's Theorem (Billingsley, 1968, pp. 137),

$$Z_n(s) \to_d W(s) \tag{8}$$

where W(s) is a standard Brownian motion.

<u>THEOREM 1</u>. Suppose that X_t follows the model (5) with $X_0 = 0$. Then,

$$n^{-3/2} \sum_{i=1}^{n} X_i \to_d \sigma_\epsilon \int_0^1 W(s)ds.$$

<u>PROOF</u>: From (6)

$$\begin{aligned} & n^{-3/2} \sum_{i=1}^{n} X_i \\ &= n^{-3/2} \sum_{i=1}^{n} (P_{i-1} + \epsilon_i) \\ &= n^{-1}\sigma_\epsilon \sum_{i=1}^{n} (n^{1/2}\sigma_\epsilon)^{-1} P_{i-1} + n^{-1/2}(n^{-1} \sum_{i=1}^{n} \epsilon_i) \\ &= \sigma_\epsilon \sum_{i=1}^{n} (n^{1/2}\sigma_\epsilon)^{-1} P_{i-1} [i/n - (i-1)/n] + o_p(1) \\ &= \sigma_\epsilon \sum_{i=1}^{n} \int_{(i-1)/n}^{i/n} (n^{1/2}\sigma_\epsilon)^{-1} P_{[ns]} \, ds + o_p(1) \\ &= \sigma_\epsilon \int_0^1 Z_n(s)ds + o_p(1). \end{aligned}$$

In the above, I have used the WLLN so that $n^{-1}\Sigma\epsilon_t$ converges to zero in probability. The theorem then follows from the continuous mapping theorem (Billingsley, 1968, Theorem 5.2) and (8). QED.

By Theorem 1, the "appropriate" normalization factor of the sum of observations is $n^{3/2}$ when X_t follows the random walk model. Therefore, $\bar{X}_n$ goes to infinity as the sample size n increases. In other words, in the nonstationary case the sample mean $\bar{X}_n$ does not have a well-defined limit.

Since the stationary AR(1) process can be regarded as the transition status between independence and the random walk model, one would expect the sample mean of a stationary AR(1) to play a transition role between the limits of the two extremes, independence and high correlation. (Note that the random walk case can be interpreted as highly correlated situation because the sample autocorrelations all approach unity, see Tiao and Tsay, 1983, Corollary 2.6.) However, the results of Section 1 and Theorem 1 clearly point out that this is not the case. The normalization factor *must* be modified. This explains the failure of formula (4) in providing aedquate inferences on the sample mean when X_i is nearly nonstationary.

3. NEARLY NONSTATIONARY PROCESSES

To derive a unified limiting distribution for sample means or their variants, I parameterize the AR(1) model in a triangular array setting. Consider

$$X_{n,i} = \beta_n X_{n,i-1} + \epsilon_i \tag{9}$$

with $\beta_n = 1 - \gamma/n$.

By (9), $X_{n,i}$ reduces to the random walk model (5) if $\gamma = 0$, and it approaches a stationary case when γ/n goes to a fixed limit between 0 and 1 Again, for simplicity I assume $X_{n,0} = 0$. From (9),

$$X_{n,i} = \sum_{t=0}^{i-1} \beta_n^t \epsilon_{i-t} \tag{10}$$

Let

$$Z_{n,i} = (n^{1/2}\sigma_\epsilon)^{-1} \beta_n^{-i} \epsilon_i$$

$$Z_n(s) = \sum_{i=0}^{[ns]} Z_{n,i} \quad \text{for } 0 \le s \le 1.$$

Lemma 1. Let $B_\gamma(t) = \exp(-2\gamma)[\exp(2\gamma t) - 1]/2\gamma$. Then,

$$Z_n(s) \to_d W(B_\gamma(s)) \quad \text{as } n \to \infty,$$

where W(s) is a standard Brownian motion.

PROOF: See Lemma 2.1 of Chan and Wei (1986). QED.

THEOREM 2: Suppose that $X_{n,i}$ follows the model (9) and $X_{n,0} = 0$. Then,

$$n^{-3/2} \sum_{i=1}^{n} X_{n,i} \to_d \sigma_\epsilon \int_0^1 \exp[-\gamma(s-1)]\, W(B_\gamma(s))ds.$$

PROOF: From (10),

$$n^{-3/2} \sum_{i=1}^{n} X_{n,i}$$

$$= n^{-1} \sigma_\epsilon \sum_{i=1}^{n} (n^{1/2}\sigma_\epsilon)^{-1} X_{n,i}$$

$$= \sigma_\epsilon \sum_{i=1}^{n} \beta_\gamma^{i-n} \sum_{t=1}^{i} Z_{n,t}\, n^{-1}$$

$$= \sigma_\epsilon \sum_{i=1}^{n} \beta_n^{i-n} Z_n(i/n)\, n^{-1}$$

$$= \sigma_\epsilon \sum_{i=1}^{n} \beta_n^{i-n} Z_n(i/n)\, n^{-1} - \sigma_\epsilon \int_0^1 \exp[-\gamma(s-1)] Z_n(s)ds + \sigma_\epsilon \int_0^1 \exp[-\gamma(s-1)]\, Z_n(s)ds$$

$$= \sigma_\epsilon \int_0^1 \exp[-\gamma(s-1)]\, Z_n(s)ds + \sigma_\epsilon H_n. \tag{11}$$

Next, I show that $|H_n| = o_p(1)$. First, claim that

$$T_n = \sup_{\{1 \le i \le n\}} \sup_{\{(i-1)/n \le s \le i/n\}} \left| \beta_n^{i-n} - \exp[-\gamma(s-1)] \right| = o(1). \tag{12}$$

If (12) holds, then

$$|H_n| = \left| \sum_{i=1}^{n} \beta_n^{i-n} Z_n(i/n)\, n^{-1} - \int_0^1 \exp[-\gamma(s-1)] Z_n(s)ds \right|$$

$$= \left| \sum_{i=1}^{n-1} \beta_n^{i-n} Z_n(i/n) n^{-1} + Z_n(1) n^{-1} - \sum_{i=0}^{n-1} \int_{i/n}^{(i+1)/n} \exp[-\gamma(s-1)] Z_n(i/n)ds \right|$$

$$= \left| \sum_{i=1}^{n-1} \int_{i/n}^{(i+1)/n} [\beta_\gamma^{i-n} - e^{-\gamma(s-1)}]\, Z_n(i/n)\, ds + Z_n(1) n^{-1} \right|$$

$$\le \left| \sum_{i=1}^{n-1} T_n \int_{i/n}^{(i+1)/n} Z_n(i/n)ds \right| + n^{-1} \left| Z_n(1) \right|$$

$$= T_n \left| \int_{1/n}^{1} Z_n(s)ds \right| + n^{-1} \left| Z_n(1) \right|$$

$$= o_p(1).$$

The theorem then follows from (11), Lemma 1, and the continuous mapping theorem. In the above, I have used $Z_n(0) = 0$. To show (12), let $d_n = -n\gamma^{-1}\log(1 - \gamma/n)$. Then, $d_n \to 1$ as $n \to \infty$ for fixed γ. Moreover, we may rewrite β_n^{i-n} as

$$\beta_n^{i-n} = \exp[\gamma(1 - i/n)d_n].$$

Therefore, we have

$$\left| \beta_n^{i-n} - e^{-\gamma(s-1)} \right|$$

$$= e^{\gamma(1 - i/n)} \left| \exp[\gamma(1 - i/n)(d_n - 1)] - \exp[\gamma(i/n - s)] \right|$$

$$\le \exp(|\gamma|) \left\{ \left| \exp[(d_n - 1)\gamma(1 - i/n)] - 1 \right| + \left| 1 - \exp[\gamma(i/n - s)] \right| \right\}.$$

Now,

$$\sup_{\{1 \le i \le n\}} \left| (d_n - 1)\gamma(1 - i/n) \right| \le \left| d_n - 1 \right| |\gamma| \to 0$$

as $n \to \infty$,

$$\sup_{\{(i-1)/n \le s \le i/n\}} \left| 1 - \exp[\gamma(i/n - s)] \right|$$

$$\le \max\{ 1 - \exp(-|\gamma|/n),\ \exp(|\gamma|/n) - 1 \} \to 0$$

as $n \to \infty$.

Therefore, (12) holds and the proof of the theorem is complete. QED.

Note that the above distribution can be evaluated by using a result of Kac (1980),

$$W(s) = \sum_{i=0}^{\infty} [(2i+1)\pi]^{-1} 2^{3/2} \sin[(i + 0.5)\pi s] Y_i$$

where Y_i are IID standard Gaussian variables. Furthermore, the value of γ may be estimated by $1 - \hat{\beta}$ with $\hat{\beta}$ being the least squares estimate of β in the usual AR(1) regression,

$$X_i = \beta X_{i-1} + \epsilon_i.$$

The inference of the sample mean of X_i can then be made by using the result of Theorem 2.

4. GENERAL MODELS WITH A SINGLE NEARLY NONSTATIONARY ROOT

The results of the preceeding two sections are relatively limited. For they only deals with AR(1) type of models. In practice, a stochastic process may not well follow such models and it is necessary to consider the general situation. In this section, we consider the model

$$\phi(B)(1 - \beta_n B)X_i = \theta(B)\epsilon \qquad (13)$$

where β_n and ϵ_i are defined as those of Section 3, B is the backshift operator such that $BX_i = X_{i-1}$, and $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ are two polynomials in B of degrees p and q, respectively. For model (13), I assume that $\phi(B)$ and $\theta(B)$ have no common factors and that all of the zeros of $\phi(B)$ are greater than unity in modulus. Moreover, all of the zeros of $\phi(B)$ should be far away from the unit circle as compared with β_n^{-1}. The latter condition implies that $(1 - \beta_n B)$ is the dominating factor of the autoregressive part of X_i. This condition is needed because the asymptotic behavior of X_i, hence that of $\bar{X}_n$, is determined by zeros of the autoregressive part which are on or close to the unit circle, see Tiao and Tsay (1983).

THEOREM 3: Suppose that X_i follows the model (13) and $X_i = 0$ for $i \leq 0$. Then,

$$n^{-3/2} \sum_{i=1}^{n} X_i \to_d$$

$$\Delta(\phi\ ,\ \theta)\ \sigma_\epsilon \int_0^1 \exp[-\gamma(s-1)]\ W(B_\gamma(s))ds,$$

where γ, $B_\gamma(s)$, and W(s) are defined in Theorem 2, and $\Delta(\phi, \theta) = (1 - \theta_1 - \dots - \theta_q)/(1 - \phi_1 - \dots - \phi_p)$.

PROOF: This theorem can be proved in three steps depending on the values of p and q. The first case is p = 0, the second q = 0, and the third $p \neq 0$ and $q \neq 0$. The details are along the same line as those of Theorems 1 - 3 of Tsay (1986) and are omitted. QED.

Note that (13) can be rewritten as

$$X_i = \beta_n X_{i-1} + V_i$$

where $V_i = [\phi(B)]^{-1}\theta(B)\epsilon_i$ is a stationary autoregressive moving average model, i.e., ARMA(p,q) model, see Box and Jenkins (1976). Thus Theorem 3 can be regarded as an extension of Theorem 2 by allowing for serial correlations in the disturbance term V_i.

In practice, the order (p, q) and the parameters ϕ_i and θ_i are unknown. They must be estimated from the data. The following procedure is suggested.

- Identify an overall ARMA model for X_i by using the Extended Sample Autocorrelation Functions (ESACF) of Tsay and Tiao (1984) or the Smallest Canonical (SCAN) Correlation approach of Tsay and Tiao (1985). The overall order of X_i of (13) is (p+1, q).

- Estimate the parameters ϕ_i, β_n, θ_i, and σ_ϵ simultaneously by maximum likelihood method. The Kalman filter recursion may be used here.

- Factor the fitted AR polynomial to locate the zeros. The one which is closest to the unit circle is treated as an estimate β_n and the rests are used to estimate the ϕ_i's.

- Make inferences conerning $\bar{X}_n$ by using Theorem 3.

This procedure is based on several considerations: First, I suggest the use of either ESACF or SCAN because both methods can handle nonstationary and stationary processes in the same manner. They do not require any differencing in handling nonstationary series. Secondly, if desired, one may replace the maximum likelihood estimation by least squares method via the use of iterated autoregressions of Tsay and Tiao (1984). The iterated autoregressions produce consistent least squares estimates of ARMA parameters which may not be efficient, compared with MLE, but are much easier computationally. Thirdly, Theorem 3 still holds when ϕ_i, θ_i, and σ_ϵ are replaced by consistent estimates because convergence in probability implies convergence in distribution.

REFERENCES

Billingsley, P. (1968). *Convergence of Probability Measure.* John Wiley, New York.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Chan, N. H. and Wei, C. Z. (1985). Asymptotic inference of nearly nonstationary AR(1) process. Technical Report, Department of Mathematics, University of Maryland - College Park.

Kac, M. (1980). *Integration in Function Spaces and Some of Its Applications*. Academia Nationale Dei Lincei, Pisa.

Priestley, M. B. (1980). *Spectral Analysis and Time Series*, Vol. I & II. Academic Press, New York.

Tiao, G. C. and Tsay, R. S. (1983). Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Annals of Statistics*, **11**, 856-871.

Tiao, G. C. and Tsay, R. S. (1986). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. Technical Report, Department of Statistics, Carnegie-Mellon Univeristy.

Tsay, R. S. (1986). Asymptotic inference of least squares estimates of nearly nonstationary time series. Technical Report 360, Department of Statistics, Carnegie-Mellon University.

Tsay, R. S. and Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation functions of nonstationary and stationary ARMA models. *Journal of the American Statistical Association*, **79**, 84-96.

Tsay, R. S. and Tiao, G. C. (1985). Use of canonical correlation in time series model identification. *Biometrika*, **72**; 299-314.

Author's Biography

RUEY S. TSAY is an Assistant Professor of Statistics in the Statistics Department at Carnegie Mellon. He received a B.S. in Mathematics from the National Tsing Hua University, Taiwan, R.O.C. in 1974, and a Ph.D. in Statistics from the University of Wisconsin in 1982. As a Research Assistant at the University of Wisconsin at Madison from 1980 to 1982, he worked on developing and modifying the Wisconsin Multiple time series package. At the Graduate School of Business, University of Chicago, he was a Research Associate in 1985 working on multiple time series modeling. He is a fellow in the Royal Statistical Society and a member of the American Statistical Association, Institute of Mathematical Statistics, Econometric Society and the Institute of Electrical and Electronic Engineers. His current research involves time series analysis and applications.

Ruey S. Tsay
Department of Statistics
Carnegie Mellon
Pittsburgh, PA 15213, U.S.A.
(412) 268-8874