AN INVESTIGATION OF THE TRANSIENT STAGE IN SIMULATION DATA BY SPECTRAL ANALYSIS

by

PETER SHIH-YU CHANG

B.A., CHUNG-YUANG COLLEGE OF SCIENCE AND ENGINEERING, 1969

A MASTER'S THESIS

SUMBITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1972

Approved by:

L. E. Grosh

Major Professor

Table of Contents

ACKNOWLEDGEMENT

I dedicate this thesis to the Chang's family whose happiness means to me more than any success. I express my deepest appreciation to Dr. L. E. Grosh for his help and valuable guidance throughout this work. I am very grateful to Mrs. Marie Jirak for her prompt and excellent typing work. I also thank Mr. Cheng Yung-Nan for his beautiful drawing.

CHAPTER 1

INTRODUCTION

1.1. Identify The Problem

Simulation is a technique used to study the dynamic properties of a system in which the elements of this system are represented by arithmetic and logical process. This technique may be used in the following instances:

(1) To investigate a model so complex that we can not obtain the answer any other way.

(2) To model a new previously non-existant system.

(3) To evaluate an alternative to an existing system.

By running a simulation, we can obtain an uninterrupted recording of a system's performance. This recording is a set of observations which is a segment of a run sufficient for estimating the value of each of the desired performance measures.

Two methods of estimating a simulation output can be distinguished by:

(1) The instantaneous value of a variable observed at a uniform time interval. For example consider the number of items waiting in a queue, the mean and variance of the variable can be derived as follow;

$$\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^{N} X_i \qquad (1.1.)$$

$$\hat{\sigma}_N^2 = \sum_{i=1}^{N} (X_i - \hat{\mu}_N)^2/(N-1) \qquad (1.2.)$$

where N is the number of observations, and X_i is the observation (number of items waiting in a queue at time $T_0 + i\Delta T$, T_0 is beginning record time, ΔT is the time interval).

(2) The time integrated average of the system performance is a continuous weighted measure with high autocorrelation. Measures of system performance are taken when the variable changes level, see Figure 1.1. This method could be used to represent the average number of items waiting in the system;

$$\hat{\mu}_{T_N} = \frac{1}{T_N - T_b} \sum_{i=1}^{N} X_i \, \Delta T_i, \qquad (1.3.)$$

$$\hat{\sigma}_{T_N}^2 = \sum_{i=1}^{N} X_i^2 \frac{\Delta T_i}{T_N - T_b} - \hat{\mu}_{T_N}^2, \qquad (1.4.)$$

$$T_N - T_b = \sum_{i=1}^{N} \Delta T_i, \qquad (1.5.)$$

where T_N is the present time, and T_b is the beginning time, ΔT_i = time increment between changes in the level of X_i and X_{i-1}. This method can not be applied to our technique.

According to Harris (12), the mean of the simulation output will

THIS BOOK CONTAINS NUMEROUS PAGES WITH DIAGRAMS THAT ARE CROOKED COMPARED TO THE REST OF THE INFORMATION ON THE PAGE.

THIS IS AS RECEIVED FROM CUSTOMER.

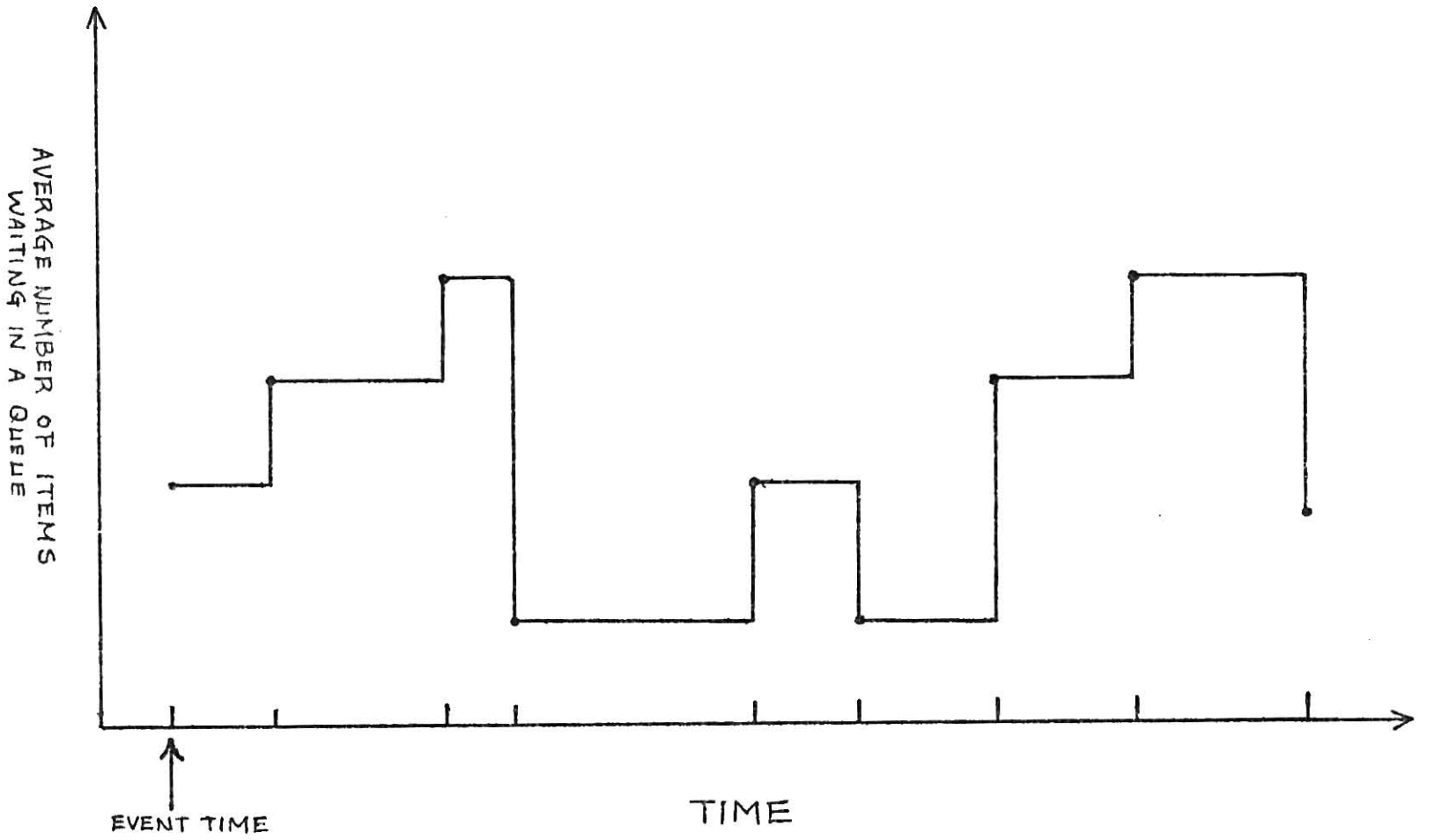

Figure 1.1 RESPONSE MEASUREMENT OF TIME INTEGRATED AVERAGE OF THE SYSTEM PERFORMANCE

converge first, then the variance. The reason can be summarized as

$$\hat{\mu}_N = \frac{1}{N} \sum_{i=1}^{N} X_i, \tag{1.6.}$$

where N is the number of observations, and

$$\hat{\sigma}^2_N = \frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{\mu}_N)^2, \tag{1.7.}$$

If the mean is biased (affected by the transient period), the variance is inflated because $E(X_i) \neq \mu$ for all i. When the biased mean gradually approaches steady state, the variance also converges toward the minimum sum of squares. It is more obvious if we decompose the variance estimate;

$$\hat{\mu}_N = \frac{N_T}{N} \hat{\mu}_{N_T} + (1 - \frac{N_T}{N}) \hat{\mu}_S, \tag{1.8.}$$

$$\hat{\sigma}^2_N = \frac{N_T}{N} \hat{\sigma}^2_{N_T} + (1 - \frac{N_T}{N}) \hat{\sigma}^2_S, \tag{1.9.}$$

where $\hat{\mu}_{N_T}$: the estimated mean in the transient stage,

$\hat{\mu}_S$: the estimated mean in steady state,

$\hat{\mu}_N$: the overall estimated mean,

$\hat{\sigma}^2_{N_T}$: the estimated variance in the transient stage,

$\hat{\sigma}_s^2$: the estimated variance in steady state,

$\hat{\sigma}_N^2$: the overall estimated variance,

N_T : the number of observations in the transient stage (fixed value)

N : the total number of observations.

For N sufficiently large,

$$\frac{N_T}{N} \hat{\sigma}_{N_T}^2 < \delta, \ (\delta \text{ is a small value}) \tag{1.10.}$$

then,

$$(1 - \frac{N_T}{N}) \hat{\sigma}_S^2 \xrightarrow{N \to \infty} \sigma_x^2, \quad (\sigma_x^2 \text{ is true variance}) \tag{1.11.}$$

At this stage, the variable becomes statistically stable, i.e., there is no change in expected value (mean) and variance. On the contrary, a system whose behavior does not satisfy the steady state conditions is usually described as in a transient stage, see Figure 1.2.

The transient properties of a system may be caused by:

(1) The atypical starting conditions used to initiate the model.

(2) A natural transient phenomenon that may occur in the system being simulated.

To the analyst, the information gathered during the transient stage is of little use if his main goal is to estimate a steady state level of performance. In the transient stage, the variance estimates are inflated and the mean estimates are biased. Therefore, in simulation

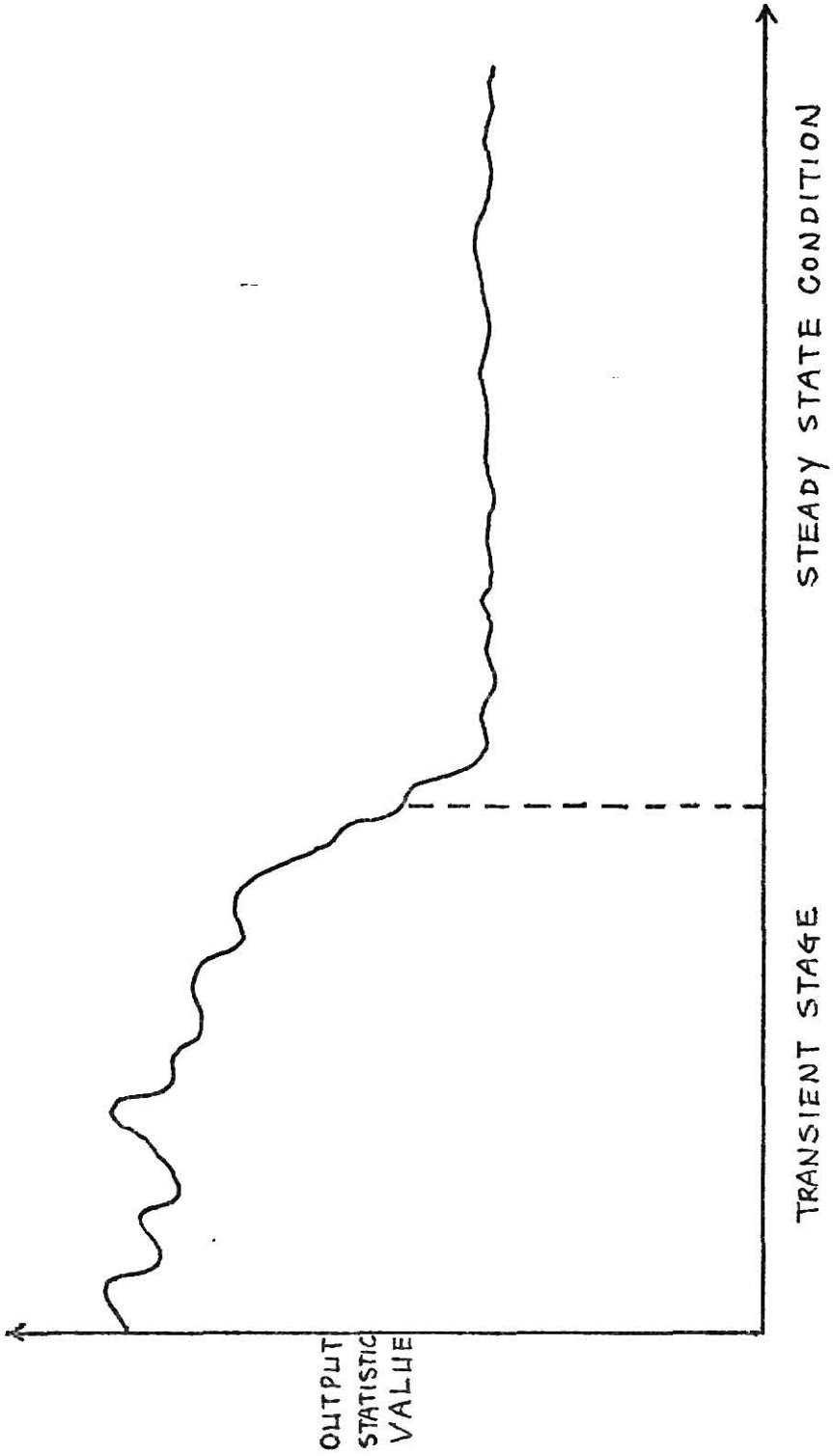

Figure 1.2 SIMULATION RUN

experimentation, the first and most important point is to remove this unwanted transient. Two methods are commonly used to remove the effects of transients. The first one is to use a long simulation run so that data being collected from the transient period becomes insignificant relative to the data in the steady state period. From the economic point of view, this method is not attractive because the cost of the computing time is excessive. The second method, is not to record the statistics until the simulation goes into a steady state condition. The condition at the end of the transient stage becomes an a priori estimate of the steady state condition and, in effect, is used to start a new run.

Now, a question is raised as to how one should decide the length of the transient stage. In other words, where is the end of the transient period. Generally speaking, there is no clear-cut point estimate to indicate the end of the transient period. A reasonable way to solve the transient period problem is to determine a better method of estimating the end point of the phase. Many techniques have been developed, and will be discussed in length in the next section.

1.2. Literature Review

Conway (4) suggested making several pilot runs and analyzing the results. This procedure is experimental, not statistical, because the investigator may either pre-load the system at the expected value of the parameters, or determine the average length of the transient stage and begin data collection at that point in the next simulation. It is not

a well-defined solution because it does not indicate the end of the transient stage while the simulation is in process, i.e., it is not a dynamic decision making procedure. In addition it requires long simulation runs which increases the computation cost. However, this method is very attractive to those people whose problem is not complicated.

Bueno (3), approached the problem by comparing sample means with a grand mean through a student's t-test. He mentioned that the individual stochastic processes of a simulation seems to stabilize before the entire system reaches equilibrium. The simulation is divided into N time periods. The process mean within the i^{th} period, $\bar{X}_i$ is compared to the grand process mean for period i+1 through N periods. Bueno suggested discarding the first period as we can assume it to contain transient biases. Bueno states that since the sample mean and grand mean are calculated from several single observations, they are normally distributed by virtue of the Central Limit Theorem. This technique is formulated below:

$$\text{Sample mean} \quad \bar{X}_j = \frac{1}{n} \sum_{i=(j-1)n+1}^{jn} X_i, \quad j = 1, \ldots, N \tag{1.12.}$$

where n is the number of observations in each time period, N is the number of time periods.

$$\text{Grand mean} \quad \bar{\bar{X}} = \frac{1}{N-1} \sum_{i=2}^{N} \bar{X}_i, \quad \text{since we ignore } \bar{X}_1, \tag{1.13.}$$

$$S^2_{\bar{x}} = \frac{1}{N-2} \sum_{i=2}^{N} (\bar{X}_i - \bar{\bar{X}})^2, \tag{1.14.}$$

$$U.C.L. = \bar{\bar{X}} + 2S_{\bar{x}} / \sqrt{N-1}, \tag{1.15.}$$

$$L.C.L. = \bar{\bar{X}} - 2S_{\bar{x}} / \sqrt{N-1} . \tag{1.16.}$$

Starting the search from $\bar{X}_2$, one finds the first $\bar{X}_m$ such that both $\bar{X}_m$ and $\bar{X}_{m+1}$ lie within the range bounded by UCL and LCL. The period 1 through m-1 are defined as the transient stage. This is a dynamic test, with the restriction that all observations ($\bar{X}_i$) are required to be independent. Unfortunately, Fishman (8) and Reese (20) both point out that successive estimates of the mean of simulation-generated time series are highly autocorrelated.

Reese (20) developed a technique to compare successive sample means by using a sequential t-test of observations made at m intervals. He suggested selecting a lag time of length m, which corresponds to the largest autocorrelation interval, then observing the sample separated only by m time units. The samples, if taken or used in this manner, form a normal distribution and are independent. Reese states a simple hypothesis, H_0: $\theta = \theta_0$ with the alternate H_1: $\theta \neq \theta_0$. The decision criterion is $B < \frac{\theta_0}{\theta} f(X,\mu,\sigma) < A$, where $f(X,\mu,\sigma)$ is the normally distributed density function of which μ and σ are parameters, and where $A = \alpha/1-\beta$ and $B = \beta/1-\alpha$. If the sample statistic is less than B, accept H_0; If the sample statistic is greater than A, reject H_0. If the sample statistic is between A and B, no decision can be made, i.e., the simulation still in the transient stage.

Dynamic testing can be made continuously in time using the sequential t-test. In this test we first give our sample lag associated with the process, m

$$\text{Grand mean} \qquad \bar{\bar{X}} = \frac{1}{N} \sum_{i=1}^{N} X_i, \tag{1.17.}$$

where N is the number of observations to the present.

$$\text{Sample mean} \qquad \bar{X}_j = \frac{1}{n} \sum_{i=(j-1)n+m+1}^{m+nj} X_i, \qquad j = 1, \ldots, J. \tag{1.18.}$$

where j is the number of sample group, J is the group number which the solution is obtained.

$$\text{sample variance} \qquad S^2_{x_j} = \frac{1}{n-1} \sum_{i=(j-1)n+m+1}^{m+nj} (X_i - \bar{X}_j)^2, \quad j = 1, \ldots, J \tag{1.19.}$$

$$\text{Sample autocovariance} \quad R_{\tau_j} = \frac{1}{n} \sum_{i=(j-1)n+m+1}^{m+jn} (X_i - \bar{X}_j)(X_{i-\tau} - \bar{X}_j), \tag{1.20.}$$

where $\tau < n$.

Variance, adjusted for autocovariance

$$S^2_{\bar{x}_j} = \frac{1}{n} \left(S^2_{\bar{x}_j} + 2 \sum_{\tau=1}^{m} R_{\tau_j}\right). \tag{1.21.}$$

Let the sample statistic be λ_n, the test is $B < \lambda_n < A$, where

$$\lambda_n = \frac{1}{2} \left\{ [(n + t_0^2)/(n + t_1^2)]^{1/2(n+1)} + [(n + t_0^2)/(n + t_2^2)]^{1/2(n+1)} \right\}, \tag{1.22.}$$

where

$$t_0 = \frac{\bar{X}_j - \bar{\bar{X}}}{S_{\bar{x}_j}}, \quad t_1 = \frac{\bar{X}_j - \bar{\bar{X}} - \delta}{S_{\bar{x}_j}}, \quad t_2 = \frac{\bar{X}_j - \bar{\bar{X}} + \delta}{S_{\bar{x}_j}},$$

δ is the longest correlated interval. $j = 1, 2, \ldots$.

This method is dynamic, thus no repetitive simulations are necessary. Reese admitted that this method is too conservative and forces the accumulation of a large number of samples. The grand mean is biased, and this bias diminishes only after the transient stage has been passed. We also noted that Reese adjusts the sample variance for autocorrelation. This removes information provided by the autocorrelation, thus decreasing comprehension of the system being studied.

Fishman and Kiviat (9) present a discussion of spectral analysis and show how it can be applied to simulation-generated time series. They advocate spectral analysis for two reasons. First, as mentioned before, simulation data is autocorrelated; thus, the researcher can not apply commonly used statistical tools. Second, they do not believe that the autocorrelation can be removed. If it is removed, the information which the researcher would be concerned with will be found to be altered. Fishman (8) suggested a technique to determine the sample size of autocorrelated data so that its mean can be estimated with a prespecified level of confidence. In other words, Fishman's technique estimates the sample mean such that the variance of this sample mean is within prespecified limits. Thus, if we can obtain the condition in a time series

where the sample mean and the variance of this mean become stable, it will indicate that the system has reached steady-state. This approach, is dynamic and the sample elements are not independent, but it is not easy to calculate and is disturbed by a cyclic behavior of the time series.

Kirtane (16) also, worked with Fishman's technique. He estimated the parameters of the linear autoregressive scheme that represents the autocorrelated data. His basic idea is that when the process reaches steady-state condition, the successive samples are expected to have the same relative size and the same order for the data, and proved that autocorrelation between observations does occur. However, it failed to give any conclusive results regarding the end of the transient stage in the simulation.

Harris (12), investigated the use of cross-spectral analysis. He reasoned that if two successive spectra were of different character, the transient stage will still be in process, or conversely, if two successive spectra appeared to be similar, then the transient phase will have terminated. He used cross-spectral analysis to illustrate this thought. The similarity of two spectra can be reflected in the coherence diagram which is a smoothed plot of the squared correlation against the frequency over all frequencies. If the two segments are from stationary time series, the coherence will be approximately one for all frequencies. Similarly, if the two segments are from dissimilar time series, the coherence diagram will approach zero for all frequencies.

This seems to be a fruitful approach but requires much computer time.

Harris calculated the spectrum and the cross spectrum for each of the two segments. From an economic point of view, if we could solve the problem by calculating only one spectrum, we could expect to reduce the cost to approximately a third of that of Harris' method. This approach would require the comparison of the sample spectrum with an idealized spectrum from a stationary time series. If the comparison is significant, then the process is still in the transient phase. Conversely, if the difference approaches zero, then the process has reached steady state.

Rao and Shapiro's paper (19) derived the confidence interval (control limit) of an idealized spectrum. They introduced a simple stationary time series,

$$X(t) = a + e(t), \quad \text{where } a \text{ is a constant,} \tag{1.23.}$$

and where

$$E[e(t)] = 0, \tag{1.24.}$$

thus

$$E[X(t)] = a, \quad \text{and} \tag{1.25.}$$

$$E[e(t)e(t+\tau)] = R(\tau), \quad \text{which is a function of } \tau \text{ only.} \tag{1.26.}$$

If the stationary time series is changed to a nonstationary time series, for example to a step pattern, the changing of the time series can be

indicated by observing when the spectral estimates exceed the control limit. They examine two successive segments of the series. If these two successive segments are dissimilar, the process has changed. If these two successive segments are similar there has been no change in the process. Expanding this idea, we shall inspect two successive segments of simulation output. We shall assume they are dissimilar at the beginning and indicate the end of the transient stage as soon as two successive segments become "similar". Hence, we can solve our problem by creating a process,

$$X(t) = T(t) + a + e(t), \qquad (1.27.)$$

where $T(t)$ is a trend function which can be interpreted as the transient stage in a simulation run. This trend function will approach zero as time goes to infinity, i.e., the process becomes a stationary time series. Theoretically at the beginning of the nonstationary time series, the spectral estimates should lie outside the control limit, and as the process becomes stationary, the spectral estimates will fall within the control limit. Therefore, from this observation we can conclude that the transient phase has been passed.

1.3. Proposal for identifying the end of a transient stage

(1) Expand upon Rao and Shapiro's control limit calculation to make it easy to identify the control limit for various parameter combinations.

(2) Program and validate the Rao and Shapiro's algorithm.

(3) Apply the algorithm to "pseudo" simulation data. A "pseudo" simulation is not a simulation but it has the character of a simulation. We use a pseudo simulation for ease of control and minimal computer expense. If the statistic works well on the "pseudo" data, then it should work for a simulation.

CHAPTER 2

THEORY AND PROCEDURE

2.1. Theory

The output of a simulation can be interpreted as a time series. In general it is a sequence, either discrete or continuous, of quantitative data observed at uniformly spaced points in time.

Time series can be categorized as stationary and nonstationary. A stationary time series is a series whose parameters do not change with time. Let X_t represent a stationary time series. Its mean, variance and covariance can be formulated as

$$E[X_t] = \mu, \text{ (constant)}, \tag{2.1.}$$

$$Var[X_t] = \sigma^2, \text{ (constant)}, \tag{2.2.}$$

and

$$Cov[X_t, X_{t+\tau}] = \gamma_\tau, \tag{2.3.}$$

where τ is an integral multiple of the time ("lags") between observations. Any series whose parameters exhibit a different behavior can be defined as a nonstationary time series. Box and Jenkins (1, pp. 35) state that for a stationary time series, the autocorrelation function dies out rapidly when the number of lags increase. The autocorrelation function of a stationary time series is sometimes difficult to interpret especially if more than one cyclic process is affecting the generation of the original time series. This means that the autocorrelation function may

be complex. Thus the spectrum, which is the Fourier transformation of the autocorrelation function of the time series, has been introduced as an analysis tool. It is used to decompose the process' variance into contributions for various frequency components. Jenkins and Watts (15) state that the spectrum is often preferable to the autocorrelation function in the analysis of a finite length of record. The estimates of the spectrum at neighboring frequencies are approximately independent, and the interpretation of the spectrum is usually easier than of the autocorrelation function.

Let a decomposable non-stationary series be

$$X_t = T(t) + C(t) + e(t), \tag{2.4.}$$

where T(t) is a trend term whose mean value is generally a monotonic function of time and C(t) is a cyclic function of time that for some period P

$$C(t) = C(t + P), \tag{2.5.}$$

$$\sum_{t=T+1}^{T+P} C(t) = 0. \tag{2.6.}$$

And e(t) is white noise which is an independent random series, such that

$$E[e(t)] = 0, \tag{2.7.}$$

$$\mathrm{Var}[e(t)] = \sigma^2, \qquad \text{(fixed value)}, \tag{2.8.}$$

$$E[e(t_i)\ e(t_i+\tau)] = 0, \quad \tau = 1, 2, \ldots \ . \tag{2.9.}$$

Hence, the mean, variance, and covariance of the noise process are not a function of time. Thus, white noise can be interpreted as a stationary time series which has a "flat" **stratum.**

Now, if we adjust the series, it becomes

$$X_t = T(t) - C(t) = e(t), \tag{2.10.}$$

which is a stationary time series. In the basic idea, T(t) represents the transient. The cyclic terms will contribute only at one frequency band and destroy the white noise terms. If

$$E[X_t - T(t) - C(t)] = 0, \tag{2.11.}$$

$$Var[X_t - T(t) - C(t)] = \sigma^2, \quad \text{(constant)} \tag{2.12.}$$

and

$$Cov[X_t - T(t) - C(t)] = \gamma_\tau, \quad \text{(constant)} . \tag{2.13.}$$

For a simulation, this means that the process reached steady state. On the other hand, if $E[X_t - T(t) - C(t)]$ is not close to zero, then the trend term is inadequate. The spectral estimates are inflated, especially at lower frequencies (first frequency band). This means the process still remains in the transient stage.

The estimated power spectrum (i.e., the Fourier transformation of the autocorrelation function) is defined as (19)

$$\hat{S}_x(h\Delta f) = [\hat{C}_x(0) + \sum_{P=1}^{m-1} \hat{C}_x(P) \cos(Ph\Delta f) +$$

$$\hat{C}_x(m) \cos(mh\Delta f)], \quad h = 0,1, \ldots, m. \tag{2.14.}$$

where $\Delta f = \frac{2\pi}{m}$, is the resolution bandwidth, and

$$\hat{C}_x(P) = \frac{1}{N-P} \sum_{i=1}^{N-P} X_i(t)\, X_{i+P}(t) -$$

$$\frac{1}{(N-P)^2} \sum_{i=1}^{N-P} X_i(t) \sum_{i=1}^{N-P} X_{i+P}(t), \quad P = 0,1, \ldots, m. \tag{2.15.}$$

Hannan (11) states that the estimate of the spectrum does not provide a smooth lag function. Thus, for ease of interpretation, it is necessary to add a weighting factor, a spectral window, in order to get a smoothed estimate of the spectrum. Rao and Shapiro in their example choose a so-called Hanning window, a simple moving average of the spectral estimates.

Thus

$$\hat{f}(0) = \frac{1}{2} \hat{S}_x(0) + \frac{1}{2} \hat{S}_x(\Delta f),$$

$$\hat{f}(h\Delta f) = \frac{1}{4} \hat{S}_x((h-1)\Delta f) + \frac{1}{2} \hat{S}_x(h\Delta f) + \frac{1}{4} \hat{S}_x((h+1)\Delta f),$$

$$h = 1, \ldots, m-1, \tag{2.16.}$$

and

$$\hat{f}(m\Delta f) = \frac{1}{2}\hat{S}_x((m-1)\Delta f) + \frac{1}{2}\hat{S}_x(m\Delta f).$$

A Hanning window is easy to program. It has a very small amount of leakage from one frequency band to another frequency band. The estimated spectra for non-adjacent frequency bands is effectively uncorrelated. For further discussion of weighting factors, see Granger (10), Hannan (11), and Jenkins and Watts (15).

Here, it should be indicated that in this estimate, the integer m, is called "the number of lags used". It represents the number of frequency bands for which the spectrum is estimated. In usage, the larger the variance of the estimate at each point and the smaller m is, the result will be a more accurate estimate. Granger (10) states, it is reasonable for m = N/3 if N is a large value, m = N/5 or N/6 for N is small.

2.2. Rao And Shapiro's Technique

Rao and Shapiro have introduced a technique using evolutionary spectra to inspect the structure changes in a time series. This is equivalent to watching the time series through a moving window of fixed length. Two contradictory conditions should be considered when choosing the window length:

(1) The length of the window must be long enough so that a stable estimate can be obtained for a reasonable number of spectral components (i.e., small standard error of estimation).

(2) The length of the window should be short so that the procedure is sensitive to small changes in the process mean.

Two parameters are to be chosen, sample length N and the maximum lag to be considered m. Rao and Shapiro require repeated estimates of m+1 spectral bands for overlapping samples. A typical situation is shown in Table 2.1.

TIME SERIES	SPECTRAL ESTIMATES
DATA	1 2 ..... n
$X_1, X_2, \ldots\ldots, X_N$	$\hat{F}_{1,0}, \hat{F}_{1,1}, \ldots\ldots, \hat{F}_{1,m}$
$X_2, X_3, \ldots\ldots, X_{N+1}$	$\hat{F}_{2,0}, \hat{F}_{2,1}, \ldots\ldots, \hat{F}_{2,m}$
$X_3, X_4, \ldots\ldots, X_{N+2}$	$\hat{F}_{3,0}, \hat{F}_{3,1}, \ldots\ldots, \hat{F}_{3,m}$

Table 2.1 THE EXHIBITION OF THE SPECTRAL

Where

$$\hat{F}_{i,h} = \text{Ln } \hat{f}_i(h\Delta f), \quad h = 0, 1, \ldots, m, \tag{2.17}$$

i, h denote the i^{th} spectrum for the h^{th} frequency band and $\hat{f}_i(h\Delta f)$ is defined by equation (2.16). Ln $\hat{f}(h\Delta f)$ is usually plotted rather than $\hat{f}(h\Delta f)$. Since the confidence interval of the spectrum is a constant when plotted on a logarithmic scale.

Since th $\hat{F}_{i,h}$ are subject to sampling variation, it is necessary to

introduce a moving average to smooth them before comparing the latest spectral estimates with previous estimates.

Thus

$$\delta_{i,h} = \frac{1}{3} (\hat{F}_{i-2,h} + \hat{F}_{i-1,h} + \hat{F}_{i,h}) - \hat{F}_{i,h}, \tag{2.18.}$$

and

$$\Delta_i = \max_h |\delta_{i,h}|. \tag{2.19.}$$

Now, following Rao and Shapiro, we can obtain the distribution and control limits of the Δ series.

$$\delta_{i,h} = \sum_{j=1}^{i} a_{ij} \hat{F}_{j,h} - \hat{F}_{i,h}, \tag{2.20.}$$

where

$$0 \leq a_{ij} \leq 1, \tag{2.21.}$$

and

$$\sum_{j=1}^{i} a_{ij} = 1.$$

Let F_h be the true value of the spectral estimates $\hat{F}_{i,h}$. Then,

$$E[\hat{F}_{i,h}] \doteq F_h, \tag{2.22.}$$

therefore

$$E[\delta_{i,h}] \doteq \left(\sum_{j=1}^{i} a_{ij} - 1 \right) F_h, \tag{2.23.}$$

from equation (2.21), thus

$$E[\delta_{i,h}] \doteq 0. \tag{2.24.}$$

The variance of $\delta_{i,h}$ has been defined as (19)

$$Var[\delta_{i,h}] \doteq \frac{Km}{N} (1 + \delta_n) \left\{ \sum_{j=1}^{i} (a'_{ij})^2 + \sum_{j \neq \ell = 1}^{i} \sum_{\ell=1}^{i} a'_{ij} \, a'_{i\ell} \, R_P^2 \right\}, \tag{2.25.}$$

where

$$a'_{ij} = a_{ij}, \qquad j = 1, \ldots, j-1,$$

$$a'_{ii} = a_{ii} - 1,$$

and K is a constant which depends on the spectral window being used (K = 0.75 for the Tukey and Hanning window) and

$$\delta_h = 1, \qquad h = 0, m,$$

$$\delta_h = 0, \qquad \text{otherwise.}$$

R_P^2 is the coherence between two segments of N points with P time periods apart. It has been formulated (19) as

$$R_P^2 = \left(\frac{1}{1 + \frac{P}{N-P}} \right)^2 = \left(\frac{N-P}{N} \right)^2 = \left(1 - \frac{P}{N} \right)^2. \tag{2.26.}$$

Rao and Shapiro state that the $\hat{F}_{i,h}$ are asymptotically normally distributed, and $\delta_{i,h}$ as well, because the $\delta_{i,h}$ are a linear combination of the $\hat{F}_{i,h}$. As known, the two spectral estimates are independent if they are not adjacent. Thus $\delta_{i,0}$, $\delta_{i,1}$, ..., $\delta_{i,m}$ is a 1-dependent sequence of random variables that are distributed asymptotically as the Chi-distribution with one degree of freedom (19). A sequence of random variables X_t has been called 1-dependent if $|i-j| > 1$ implies that X_i and X_j are independent (23), then from Watson (23),

$$\lim_{n\to\infty} P(\text{Max}_h \ |\delta_{i,h}| \leq C_n(\alpha)) = e^{-\alpha}, \tag{2.27.}$$

where

$$\alpha = nP(|\delta_{i,h}| > C_n(\alpha)), \quad \text{for fixed } \alpha,$$

$n=m+1$, is the number of spectral components, $C_n(\alpha)$ is the control limit. Thus,

$$P(\Delta_i < \chi) = e^{-nP(|\delta_{i,h}| > \chi)}. \tag{2.28.}$$

χ is the Chi value ($\sqrt{\chi^2}$). From equation (2.28), the control limit of the idealized spectral estimates can be built up for various confidence levels. For a detailed derivation of the control limit, see (19). As a numerical example, let N = 50, m = 10, therefore n = 11 and the desired confidence level be 0.95, then

$$P(\Delta_i < \chi) = e^{-nP(|\delta_{i,h}| > \chi)} = 0.95,$$

taking the natural logarithm

$$nP(|\delta_{i,h}| > \chi) = 0.051,$$

$$P(|\delta_{i,h}| > \chi) = 0.0046,$$

interpolating from Chi-Square Table,

$$\frac{\chi^2}{\sigma_{\delta_{i,h}}} = 9.151,$$

thus

$$\chi = \sigma_{\delta_{i,h}} \cdot 3.025.$$

Thus, the control limit can be computed (see Appendix B). The confidence levels (α) plotted are for 0.99, 0.95, and 0.90. For each confidence level, the control limit is plotted (see Fig. A, B, C in Appendix C) against sample size for various m = 3, 5, 6, 10. The control limit is a smooth surface, decreasing with an increase in sample size but increasing as the number of spectral bands increases.

Theoretically if the process is not a stationary time series, if for example it has a trend, then the Δ series should lie outside the control limit. However when the process reaches steady-state the Δ series should fall within the control limit. Thus, by noticing this phenomenon we can determine where the structure of the time series changes.

2.3. Computational Program

The computer program consisted of a main program, three subroutines and a random number generating function. The main program is used to generate a time series. The subroutines process the generated time series to obtain a series of Δ values. Subroutine FIRST computes the summations for the covariance function (Eq. 2.15.). Subroutine SECOND stores the generated observation and updates the summations recursively. Subroutine THIRD is used to calculate the Δ value and return this value to the main program. Everytime an observation, X, is generated, a Δ value is obtained. The programming list is shown in Appendix A. The three subroutines have been made general; the maximum number of observations (N) allowed is 200, and the maximum number of lags (m) allowed is 10. However, more observations or lags or both may be used by changing the proper dimension statement.

When studying simulation behavior we do not know the length of the transient stage, therefore, if we keep storing generated observations we will need an unlimited storage area. This may restrict the storage available for other purposes. We propose a recursive approach which does not store all the generated observations. Consider a summation term which is part of the covariance function (Eq. 2.15.)

$$S_j(N,P) = \sum_{i=j}^{j+N-P-1} X_i \, X_{i+j}, \qquad 0 \leq P \leq m, \qquad j = 1, 2, \ldots, \quad (2.29.)$$

or

$$S_j(N,P) = S_{j-1}(N,P) + X_{j+N-P-1} \, X_{j+N-1} - X_{j-1} \, X_{j+P-1}, \quad j > 1. \quad (2.30.)$$

For each time period we add the newly generated observation to the summation and subtract the oldest observation from the summation in order to update the summation. In this case, the newly generated observation X_{j+N-1} occupied the location of the old X_{j-2} and remains in storage until it becomes the oldest generated observation, i.e., N+1 time periods later. The location will be occupied by the newly generated observation X_{j+2N}. Therefore, the newly generated observation occupies the location of an old observation with a subscript difference of N+1, i.e., only N+1 spaces are necessary for storing the observations. Subroutine FIRST calculates the initial summations which are needed to initialize the recursive approach.

However, when applying the recursive idea to the computer program, a complication is encountered due to different subscript series, generated observations and storage spaces. Subroutine SECOND does the indexing job. Everytime an observation is generated, subroutine SECOND will compute its index and store the value in the proper cell. We use a modular statement to adjust the subscript of the generated observation. Suppose we want to index X_T, T is 1, 2, ..., and the total storage is S = N+1, then,

$$T \equiv C \text{ MOD } (S), \tag{2.31.}$$

where

$C = T$, if $T < S$,

$C = T - q.S$ if $T \geq S$, q is maximum integer such that $0 \leq C < S$.

If T is an integral multiple of S, C becomes 0. In this case, we set C equal to S (the last cell of the storage). The computer program is dynamic since these calculations may be made while the simulation is in progress. Everytime an observation is generated in the main program, subroutine SECOND adjusts its index and returns it to the main program. We can analyze the behavior of the Δ series while the process is still running.

2.4. Validation of the Δ computation

Rao and Shapiro illustrate their procedure by applying it to various perturbed time series, in their example the step process, a spike process and a ramp process. To test the validation of the Δ value, we shall reproduce the step process and compare these results with those of Rao and Shapiro. For testing, we generated a step process with a mean of 15.0 for the first 170 values, then imposed a jump to 16.5 for the next 170 data points and then drop back to 15.0 for the remaining data points. The generated time series is formulated as

$$X_t = 15.0 + Z_t\sigma, \quad 0 < t \leq 170,\ 340 < t \leq 511,$$

$$= 16.5 + Z_t\sigma, \quad 170 < t \leq 340,$$

where

$$Z \sim N(0,1),$$

$$\sigma = 0.15.$$

Figure 2.1 shows the behavior of the process. Part of the output (Δ series)

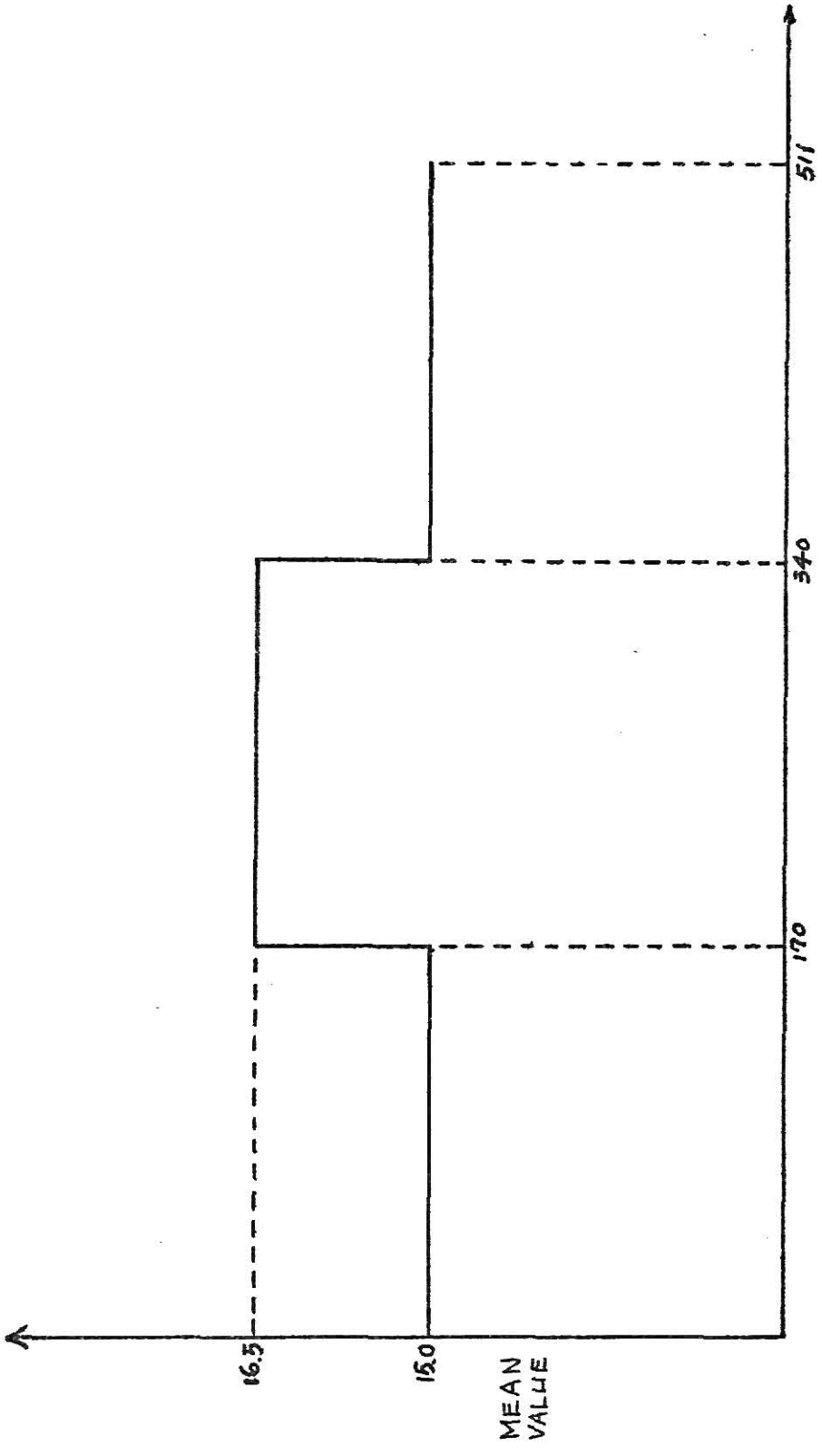

Figure 2.1 THE MEAN BEHAVIOR OF A STEP PROCESS

for the parameters N = 50, m = 5 is shown in Table 2.2. The Δ series recorded were from Δ_{101} to Δ_{140}. Let the confidence level of the control limit be 0.95, then from Appendix C, we know the control limit is 0.15490. Next we replace the rapidly fluctuating Δ series by a binary sequence such that each Δ less than the control limit is replaced by an I and each Δ out of control is replaced by an 0, see column 2, Table 2.2. Now we must construct a statistic from the binary sequence whose distribution is known. The crucial point in constructing such a statistic is the behavior of the original Δ values since all but one of the observations used to calculate Δ_i are used to calculate Δ_{i+1}. An investigation of the Δ behavior was made by computing the autocorrelation function of two time series. Figure 2.2a is the autocorrelation function for Δ from a constant process, (N = 50, m = 6). Thus it is not unreasonable to assume that the Δ series of a constant process will behave as if they had formed an independent series. Contrary, Figure 2.2b shows a similar plot for a decay process (non-stationary, N = 70, m = 5) and from it, it is obvious that independence is not a reasonable assumption here. Hence for the constant process we shall assume that the binary sequence is generated by a series of Bernoulli trials where the probability of exceeding the control limit is given by the confidence level of the control limit. Therefore, we divide the binary sequence into several sub-intervals and count the number outside the control limit for each sub-interval, i.e., $K(n,\alpha)$ in column 3, where n is the sub-interval size, and α is the confidence level of the control limit. This $K(n,\alpha)$ is binomially distributed, and can be formulated (17) as

$C_{50}(0.95) = 0.15490$		
Δ VALUE	IN OR OUT THE C.L.	K(10,0.95)
0.07073	I	
0.05733	I	
0.05644	I	
0.03592	I	
0.02853	I	0
0.05120	I	
0.04064	I	
0.03240	I	
0.04149	I	
0.04469	I	
0.04418	I	
0.02428	I	
0.11164	I	
0.06936	I	
0.06089	I	
0.05950	I	1
0.05236	I	
0.11991	I	
0.05281	I	
0.70678	O	
0.75035	O	
0.56565	O	
0.40642	O	
0.27141	O	
0.25026	O	8
0.11866	I	
0.12271	I	
0.21902	O	
0.20851	O	
0.19186	O	
0.07288	I	
0.06539	I	
0.05667	I	
0.23832	O	
0.10960	I	3
0.17478	O	
0.13528	I	
0.09786	I	
0.18697	O	
0.14419	I	

TABLE 2.2 NUMBER OF Δ VALUE OUTSIDE THE CONTROL LIMIT

FOR EVERY SET (10 OBSERVATIONS)

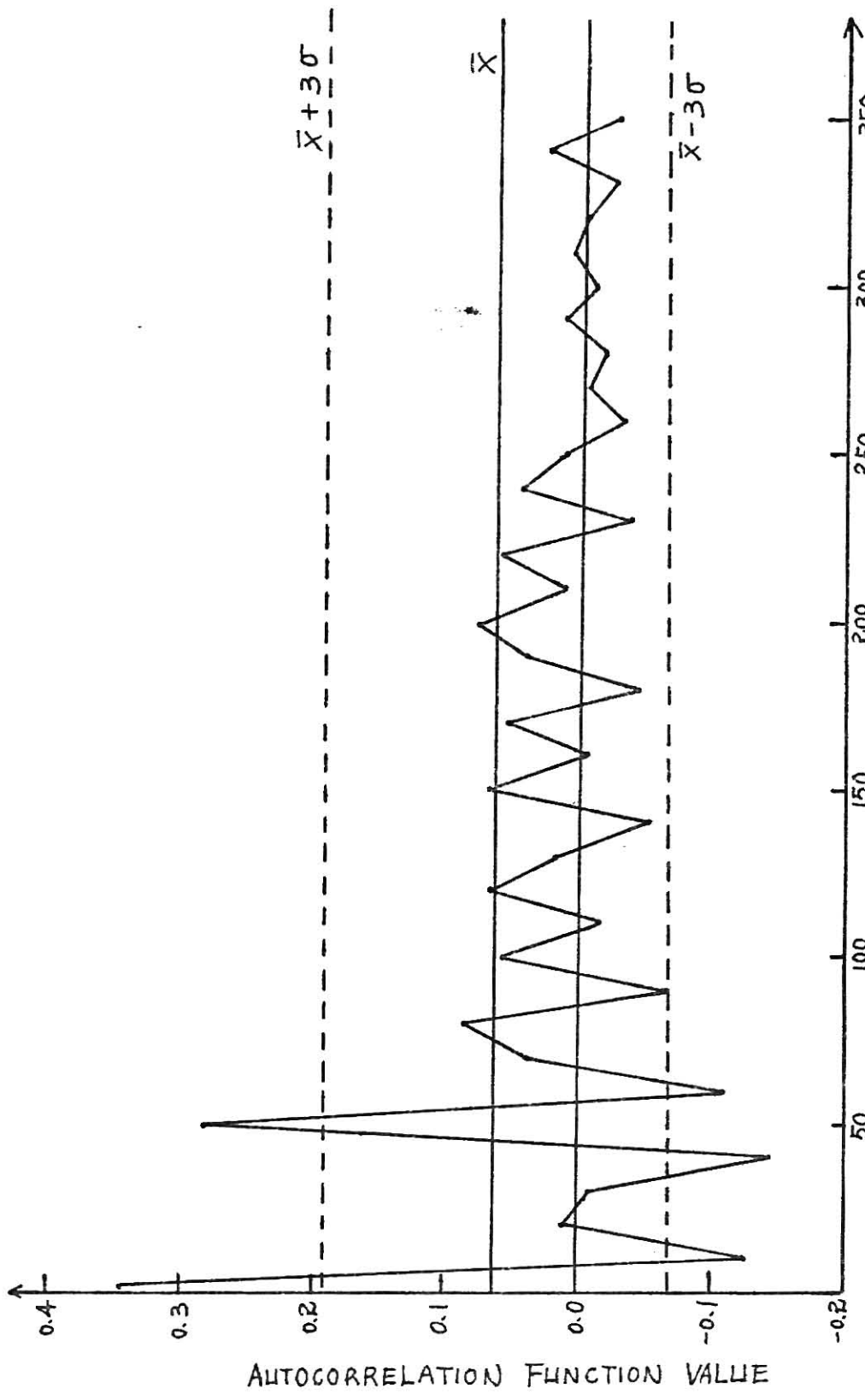

Figure 2.2a AUTOCORRELATION FUNCTION VALUE OF A CONSTANT PROCESS FOR THE LAG FROM 1 TO 350

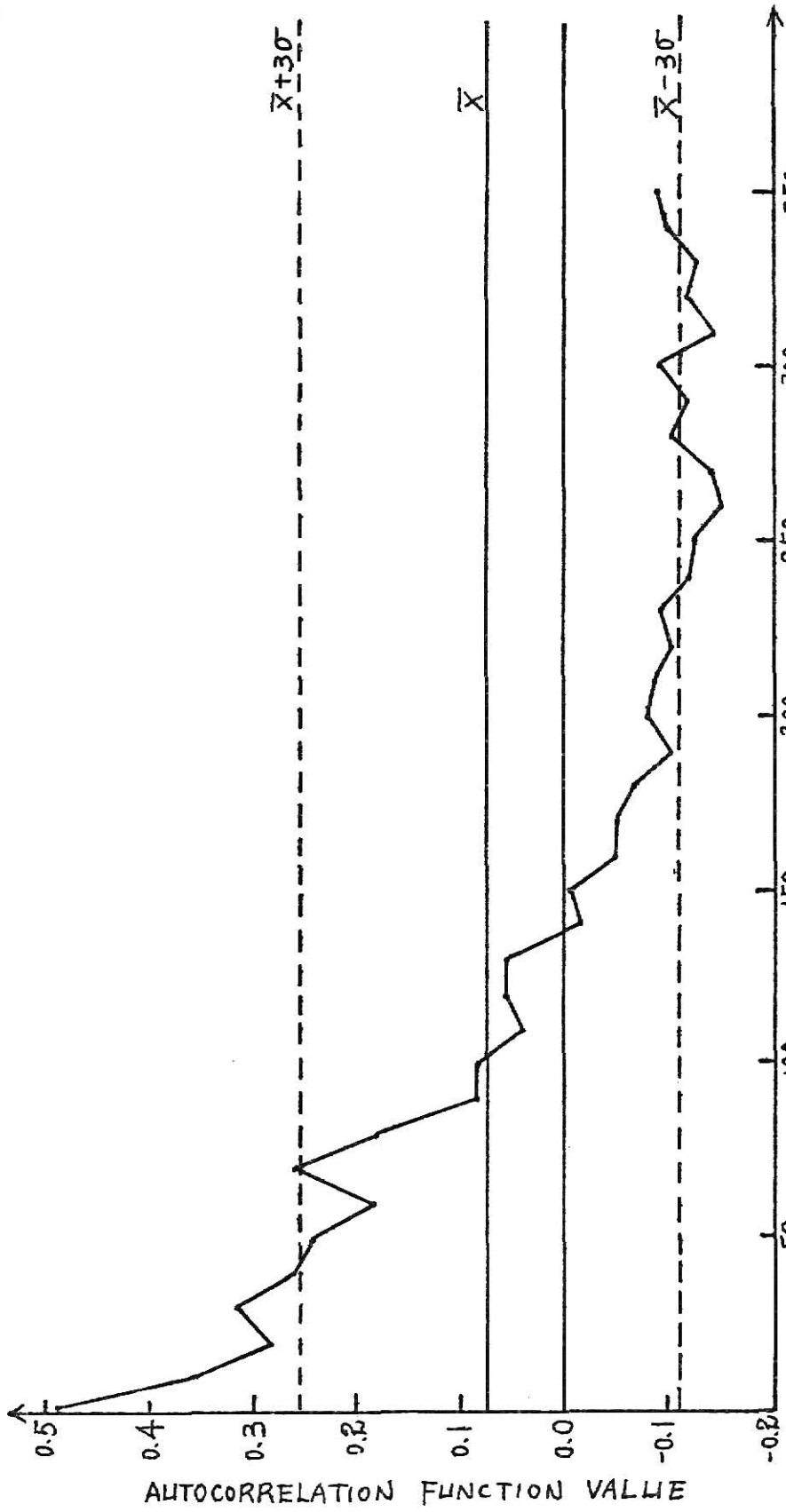

Figure 2.2b AUTOCORRELATION FUNCTION VALUE OF A DECAY PROCESS FOR THE LAG FROM 1 TO 350

$$P(S = r) = \frac{n!}{r!(n-r)!} p^r q^{n-r},$$

$$= \binom{n}{r} p^r q^{n-r}, \tag{2.32.}$$

$$P(S \geq r) = \sum_{S=r}^{n} \binom{n}{S} p^S q^{n-S} \tag{2.33.}$$

where $r = K(n,\alpha)$ with probability $p = 1 - \alpha$ in single trial, $q = \alpha$, and n is the total number of independent trials (sub-interval size). Therefore, a significance level ($SL(n,\alpha,\beta)$) for various confidence limits (β) on the sub-interval can be computed. For a numerical example, let $n = 12$, $q = \alpha = 0.95$, $\beta = 0.05$, then

$$P(S \geq r) = \sum_{S=r}^{n} \binom{n}{S} p^S q^{n-S} \leq 0.05.$$

from the Binomial Table,

$$P(S \geq 2) = 0.0861384 > 0.05,$$

$$P(S \geq 3) = 0.0115036 < 0.05.$$

Thus, the significance level $SL(n,\alpha,\beta) = 3$. Here, we make a hypothesis that the observations are generated from a stationary time series $X_t = A + E_t$. If a subinterval is significant (i.e. $K(n,\alpha) \geq SL(n,\alpha,\beta)$), then we reject the hypothesis and consider the observations which are

not generated from that time series. Otherwise, if it is not significant (i.e. $K(n,\alpha) < SL(N,\alpha,\beta)$), then we accept the hypothesis.

Three runs of the same step process have been tested, each one for 500 Δ observations for the parameters ($N = 50$, $m = 3$), ($N = 50$, $m = 5$), and ($N = 160$, $m = 10$). The results for $n = 10$, $\alpha = 0.95$, $p = 0.95$, $\beta = 0.05$ are shown in Figure 2.3a, 2.3b and 2.3c. A significant sub-interval appears N/n sub-intervals ahead of the step occurence because at that time we start to pick up the observations with a different mean. According to Rao and Shapiro, the significant sub-intervals should last for N/n sub-intervals, then the N observations are all taken from the same mean. This pattern is repeated twice in the experiment, once for each change in the mean. From Figure 2.3a, 2.3b and 2.3c, we know that when m is increasing, the method is more sensitive to detecting a change in the time series.

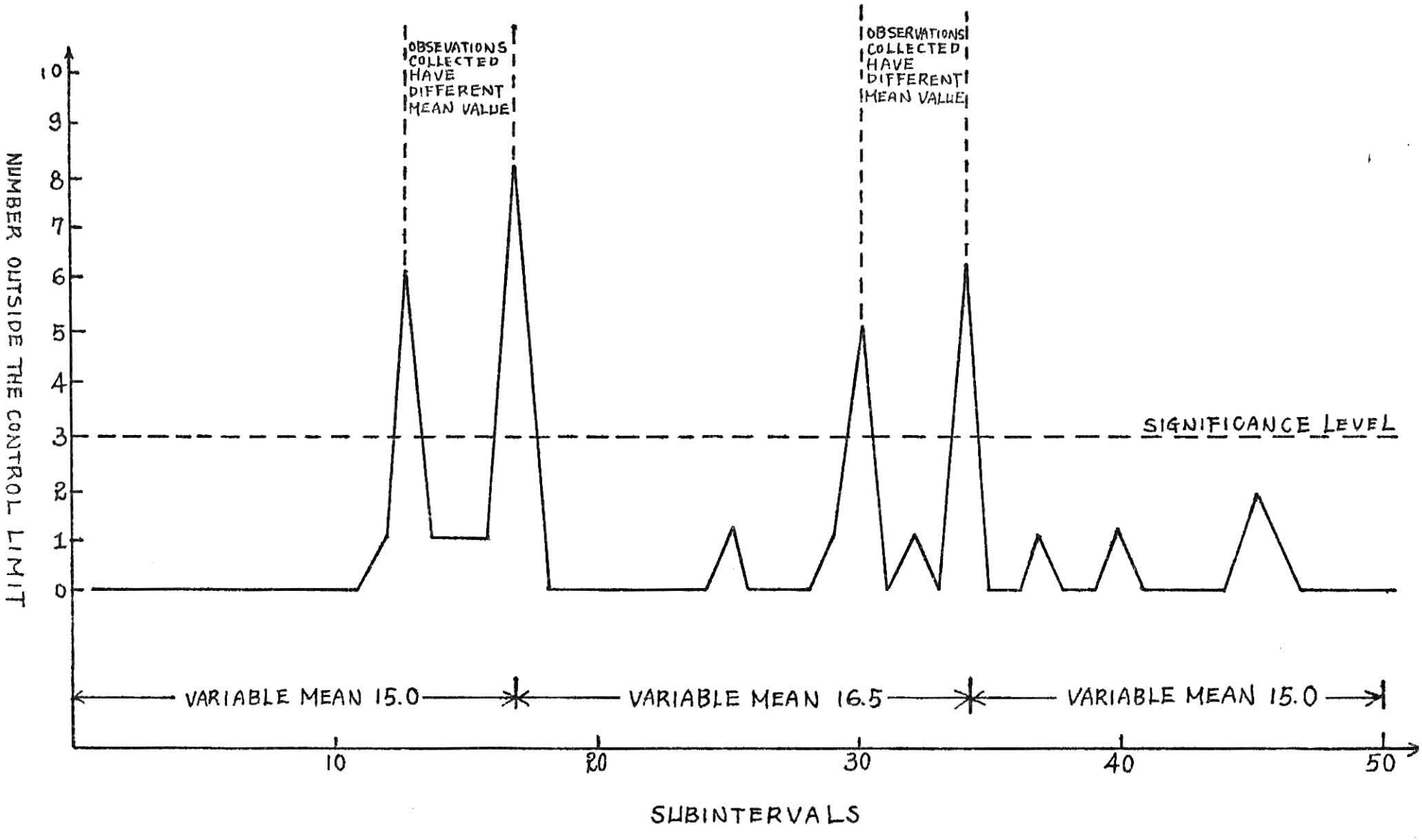

Figure 2.3a THE Δ BEHAVIOR OF A STEP PROCESS FOR THE PARAMETERS N = 50, m = 3

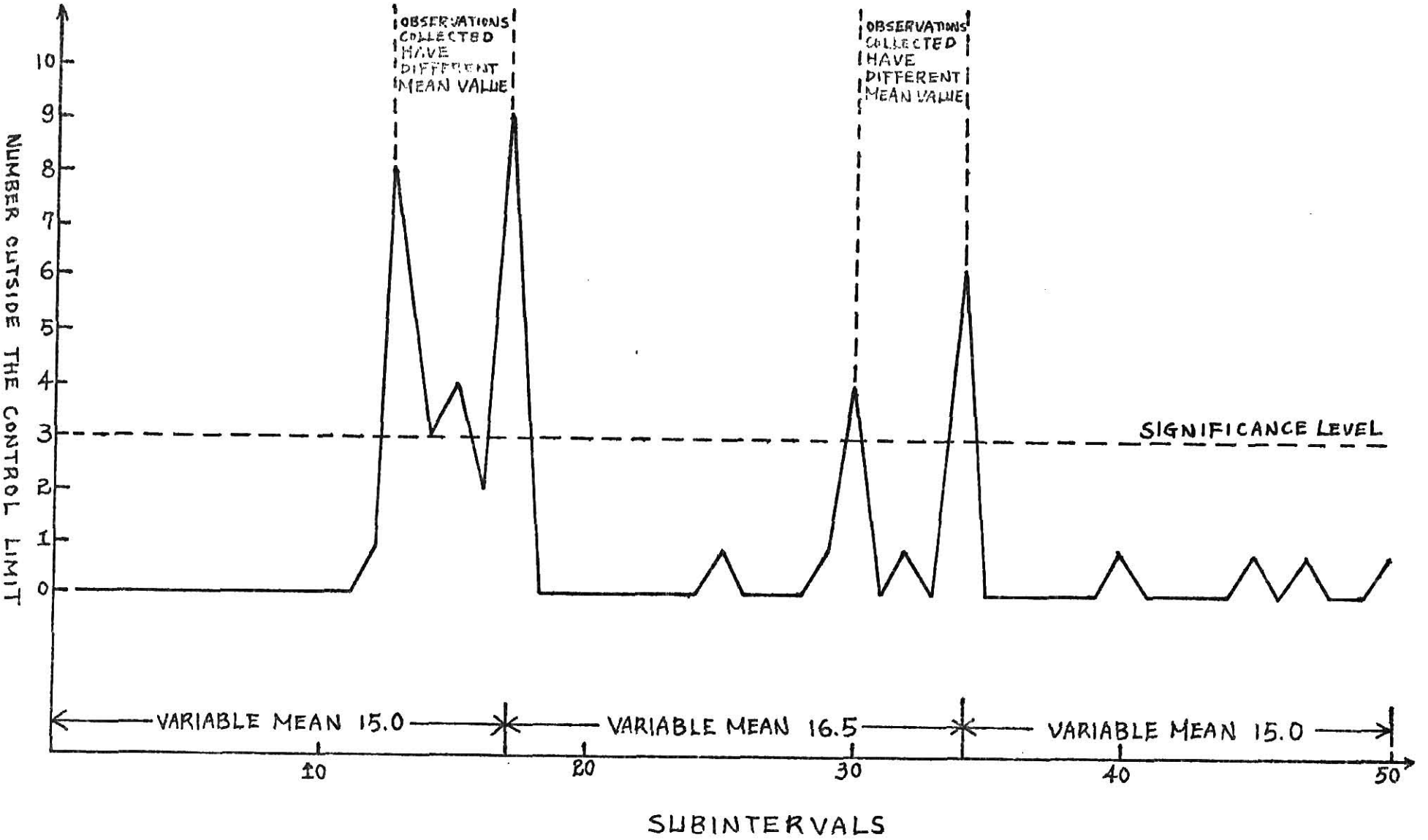

Figure 2.3b THE Δ BEHAVIOR OF A STEP PROCESS FOR THE PARAMETERS N = 50, m = 5

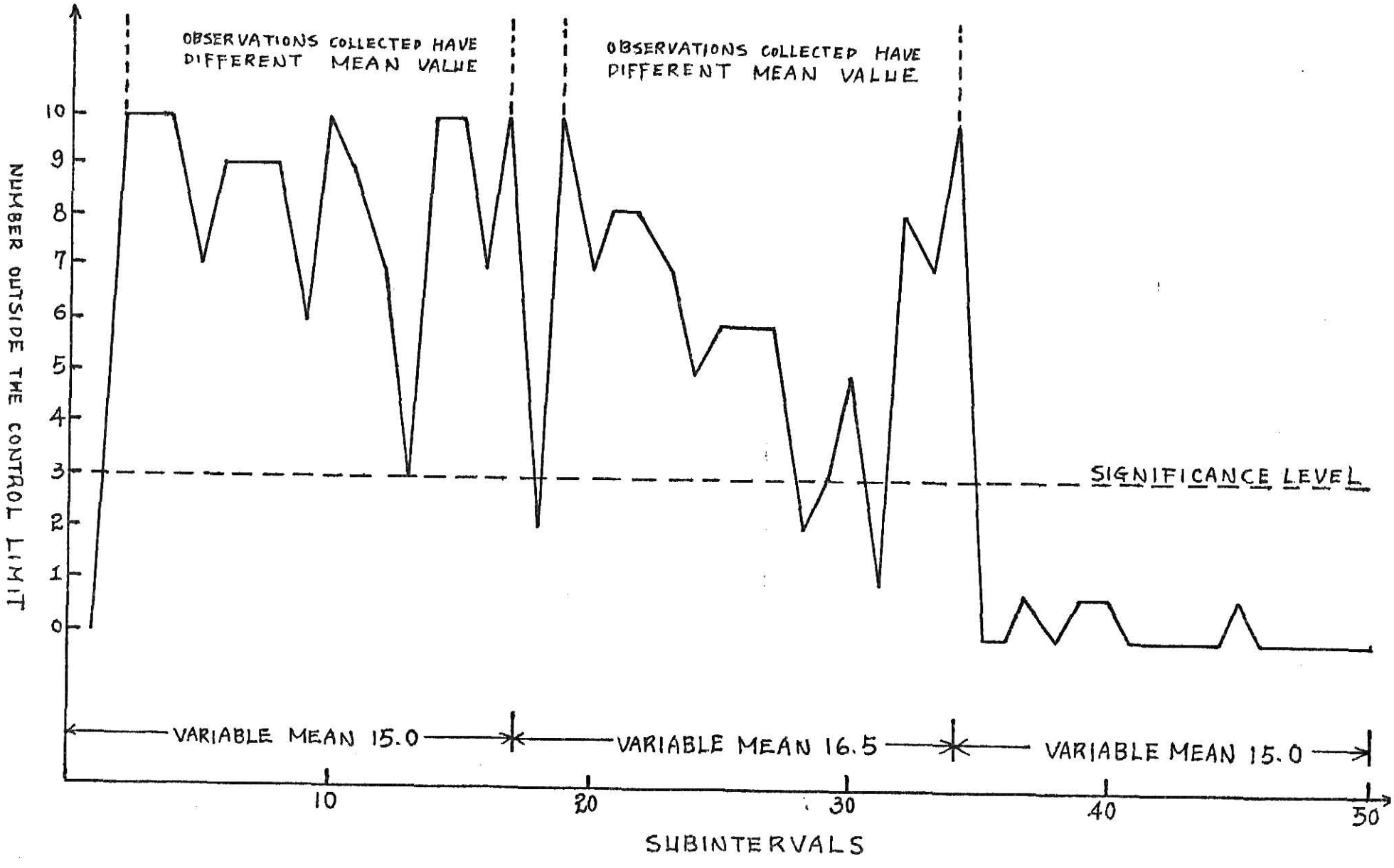

Figure 2.3c THE Δ BEHAVIOR OF A STEP PROCESS FOR THE PARAMETERS N = 160, m = 10

CHAPTER 3

COMPUTATIONAL EXPERIMENTATION AND CONCLUSION

3.1. The Experiment

A generated time series is introduced to represent a simulation for simplicity of comparison and economy in computing. This simulation is called a pseudo-simulation. The generated time series has a decaying exponential trend and is formulated as

$$X_t = \text{Constant} + \text{Trend Function} + \text{Error Term}$$

$$= A + Be^{-C.t} + Z_t\sigma,$$

where A, B, C are constants (arbitrary chosen) which could be used to control this generated process. We could let the process reach stationarity quickly, or not at all, by assigning those constants (A, B, C) with bigger or smaller values. The random normal variable is generated by use of a modified random number generator. This generator yields uniformly distributed random numbers from the interval (0,1). In conversion to a normally distributed random number, the formula from Burr (2) is used:

$$Z = \left\{[(1-R)^{-0.16239} - 1]^{0.20517} - (R^{-0.16239} - 1)^{0.20517}\right\} / 0.324,$$

where R is a uniformly distributed random number on the interval (0,1), and Z is a random variate from the standard normal distribution. Thus,

for a fixed value σ, $Z_t\sigma$ represents an error term (white noise) and is also normally distributed with $N(0, \sigma^2)$. Figure 3.1 shows the shape of the generated process. This process is very easy to perform and costs little to simulate. If this pseudo-simulation works well for the X value, the real simulation could be tried.

The experiment will be run by feeding different values of the parameters N and m. Each run collected 500 Δ observations for analyzing the transient problem. Similarly as mentioned in section 2.4, a binary sequence can be obtained by comparing the Δ series to the control limit, $C_{50}(0.95)$. Here, two decision rules have been used to investigate the end of the transient stage;

(1) We divide the binary sequence (length of 500) into 50 sub-intervals, and count the number of Δ points outside the control limit, $K(n, \alpha)$. Thus, the significance level, $SL(n,\alpha,\beta)$ can be computed for various confidence limits of $K(n,\alpha)$, $1-\beta$. For instance, let $n = 10$, $\alpha = 0.95$, $\beta = 0.05$, then from the Binomial Table, the significance level is 3. Therefore, for each sub-interval it is easy to determine whether it is significant or not by comparing $K(10,0.95)$ with $SL(10, 0.95, 0.05)$. Theoretically, the sub-intervals should be significant (i.e., $K(n,\alpha) \geq SL(N,\alpha,\beta)$) at the beginning, because the observations were generated from a trend at that time. If two successive sub-intervals are both non-significant, then the generated process has reached steady state, i.e., the transient stage is passed. Figure 3.2 shows that the $K(n,\alpha)$ are plotted against

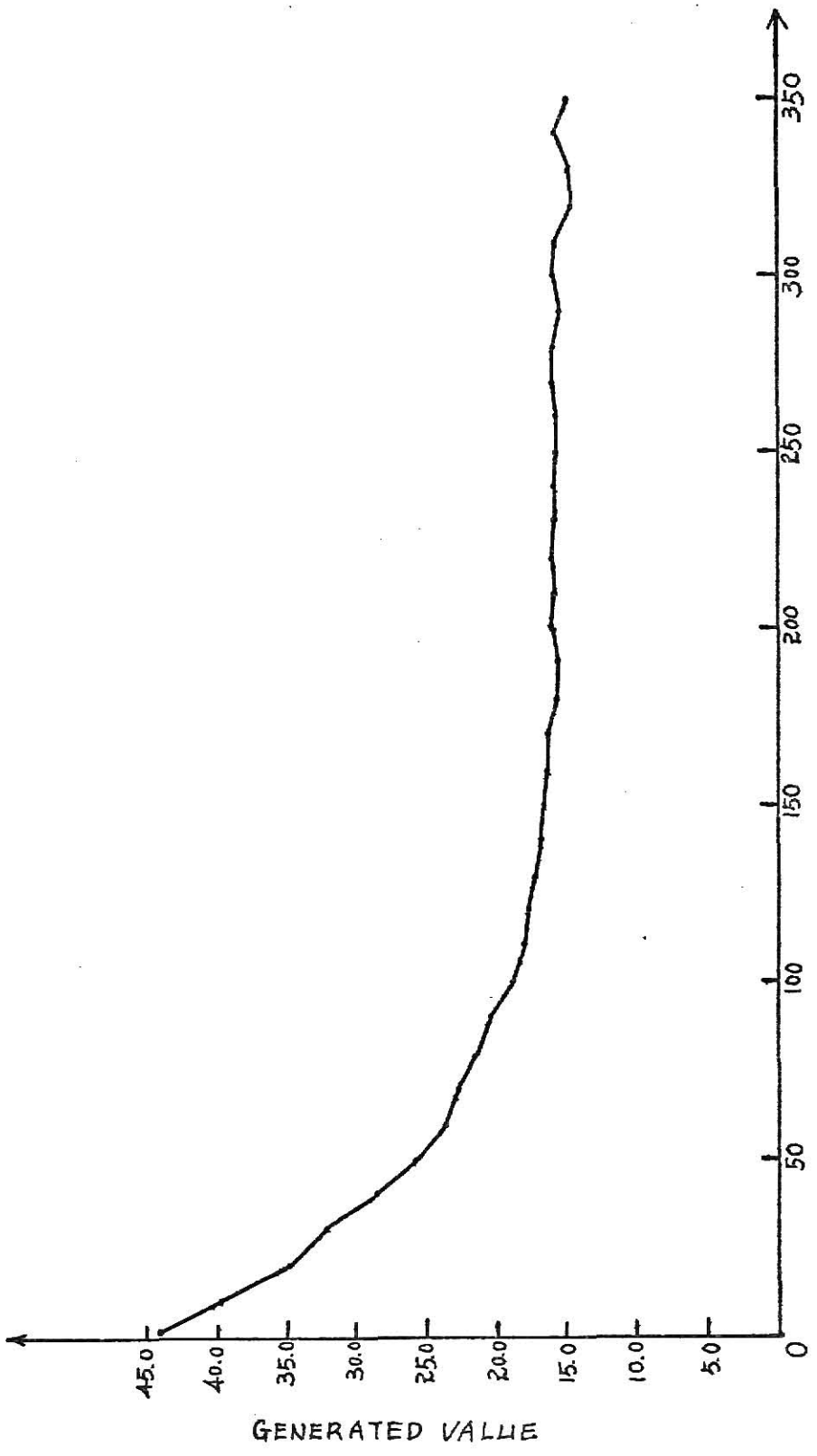

Figure 3.1 THE FIRST 350 GENERATED SAMPLE POINTS FOR THE PARAMETERS N = 70, m = 5

sub-intervals. From Figure 3.2, sub-interval 19 and 20 are first two successive non-significant sub-intervals. Thus, the transient stage is terminated at sub-interval 18, i.e., the length of the transient stage is 180.

(2) We take 30 elements of the binary sequence each time and obtain $K(n,\alpha)$. This is similar to a recursive form. First, we observe the binary sequence from 1 to 30, then 2 to 31, and so on. The significance level is 5 for $\alpha = 0.95$, $\beta = 0.05$. If there is a set of 30 elements whose $K(n, \alpha)$ is less than the significance level, then the generated process has reached steady state, i.e., no more transient stage left in the process. The transient stage is said to terminate at the last element of the previous set. Hopefully, this decision rule will consider a shorter length of the transient stage because it is decided element by element of the binary sequence instead of by a sub-interval.

3.2. Test Low Noise Level ($\sigma/A = 0.01$)

In this test, we set $A = 15.0$, $\sigma = 0.15$, thus the noise level (σ/A) is small. Basically, this noise level will not disturb the behavior of the Δ series. The length of the transient stage will be decided by both of two decision rules which were explained in previous section.

Table 3.1a to 3.1d shows the length of the transient stage for various combinations of N and m. This length was decided by first decision rule for the significance level, $SL(10, 0.95, 0.05) = 3$.

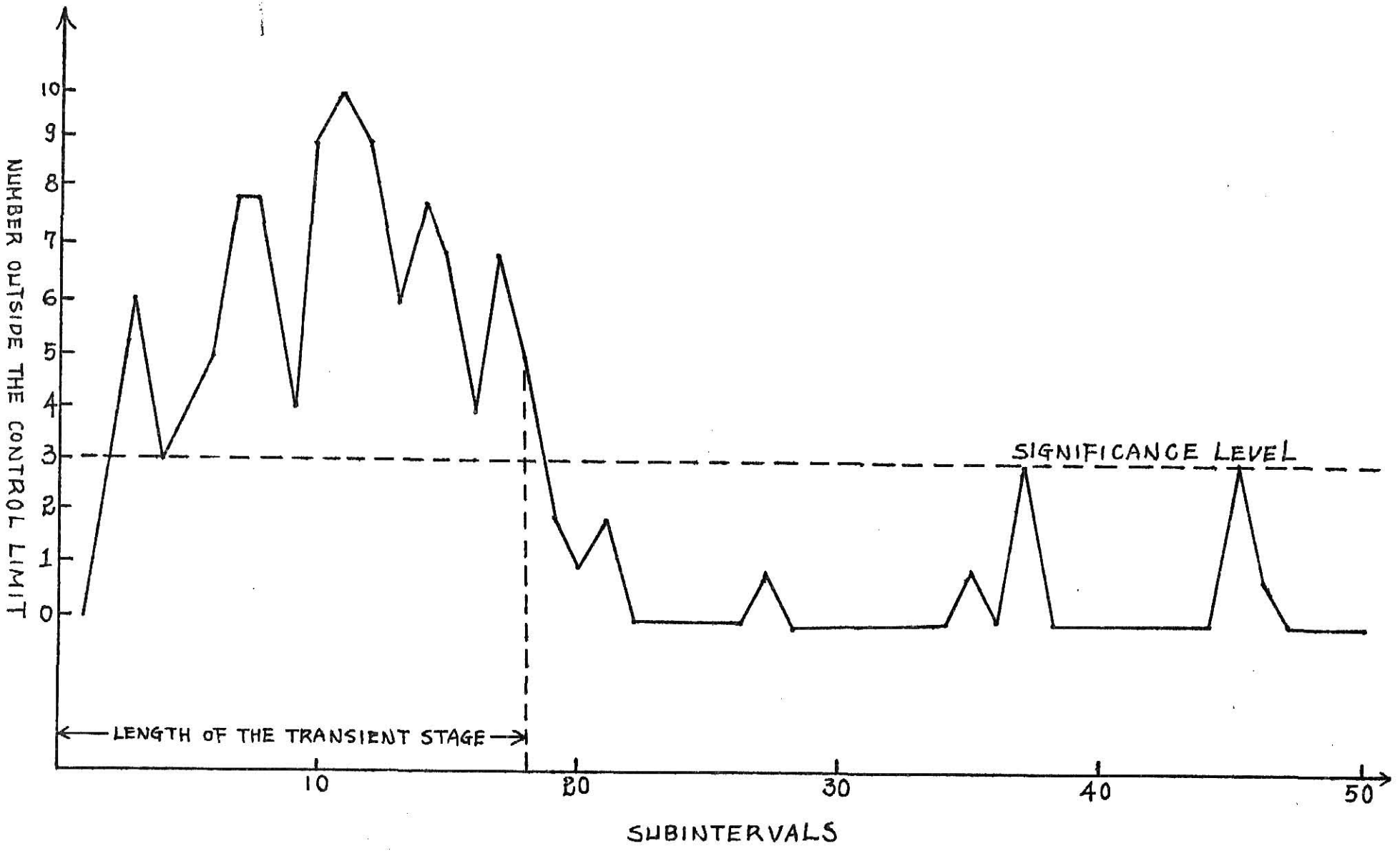

Figure 3.2 THE Δ BEHAVIOR OF A GENERATED PROCESS FOR THE PARAMETERS N = 100, m = 6

Table 3.1 LENGTH OF THE TRANSIENT STAGE USING FIRST DECISION RULE

SL(10, 0.95, 0.05) = 3		
MAXIMUM NUMBER OF LAG	NUMBER OF OBSERVATIONS	LENGTH OF THE TRANSIENT STAGE
3	40	110
3	50	110
3	60	110

Table 3.1a

SL(10, 0.95, 0.05) = 3		
MAXIMUM NUMBER OF LAG	NUMBER OF OBSERVATIONS	LENGTH OF THE TRANSIENT STAGE
5	30	140
5	50	140
5	70	170
5	80	180
5	90	180
5	100	180

Table 3.1b

SL(10, 0.95, 0.05) = 3		
MAXIMUM NUMBER OF LAG	NUMBER OF OBSERVATIONS	LENGTH OF THE TRANSIENT STAGE
6	10	110
6	20	130
6	30	140
6	40	130
6	50	150
6	60	150
6	70	170
6	80	190
6	90	190
6	100	180
6	120	190

Table 3.1c

SL(10, 0.95, 0.05) = 3		
MAXIMUM NUMBER OF OF LAG	NUMBER OF OBSERVATIONS	LENTH OF THE TRANSIENT STAGE
10	40	140
10	60	180
10	80	190
10	100	210
10	140	190
10	150	230
10	160	220
10	200	190

Table 3.1d

Again, Table 3.2a to 3.2d showed the length of the transient stage using second decision rule. The significance level, SL(30, 0.95, 0.05) = 5.

Table 3.2 LENGTH OF THE TRANSIENT STAGE USING SECOND DECISION RULE

SL(30, 0.95, 0.05) = 5		
MAXIMUM NUMBER OF LAG	NUMBER OF OBSERVATIONS	LENGTH OF THE TRANSIENT STAGE
3	40	112
3	50	107
3	60	112

Table 3.2a

SL(30, 0.95, 0.05) = 5		
MAXIMUM NUMBER OF LAG	NUMBER OF OBSERVATIONS	LENGTH OF THE TRANSIENT STAGE
5	30	124
5	50	139
5	70	167
5	80	174
5	90	174
5	100	181

Table 3.2b

SL(30, 0.95, 0.05) = 5		
MAXIMUM NUMBER OF LAG	NUMBER OF OBSERVATIONS	LENGTH OF THE TRANSIENT STAGE
6	10	103
6	20	128
6	30	132
6	40	124
6	50	142
6	60	147
6	70	169
6	80	181
6	90	182
6	100	175
6	120	181

Table 3.2c

SL(30, 0.95, 0.05) = 5		
MAXIMUM NUMBER OF LAG	NUMBER OF OBSERVATIONS	LENGTH OF THE TRANSIENT STAGE
10	40	139
10	60	174
10	80	295
10	100	202
10	140	182
10	150	225
10	160	214
10	200	181

Table 3.2d

From Figure 3.1, we know that the process starts at the value A+B, then goes down to A gradually and stays there for the remaining time periods. Thus, when the process reaches stationarity, there is no more trend in the process, i.e., the transient stage is terminated. The range for the trend function is B. If we consider

$$R.E. = \frac{Be^{-C \cdot t}}{B} = e^{-C \cdot t},$$

where R.E. is relative error, and

$$\lim_{t \to \infty} e^{-C \cdot t} \longrightarrow 0.$$

Therefore, the estimated length of the transient stage can be transformed into the form of the relative error. Figure 3.3 and 3.4 are plotted with the relative error (transformed length of the transient stage) against N/m ratio. Each Figure contains four curves, m = 3, 5, 6, 10. These curves are decreasing as the N/m ratio is increasing. The advantage of using this figure is to give the researcher an easier choice of the parameters N and m.

Suppose one decides that the maximum lag is to be m = 10, and relative error to be 0.02 as an acceptable level for considering the process a stationary time series. Thus, from Figure 3.3, 85 should be chosen as the number of the observations (N). In one way or antoher, if two out of three parameters (N, m, and R.E.) are known, the third can be obtained from the curve. However, Figure 3.3 and 3.4 only provide a rough idea, so it is not advisable to regard the value precisely.

Unfortunately, the curves in Figure 3.3 and 3.4 showed that there were some points which do not behave as they should be, i.e., the curve

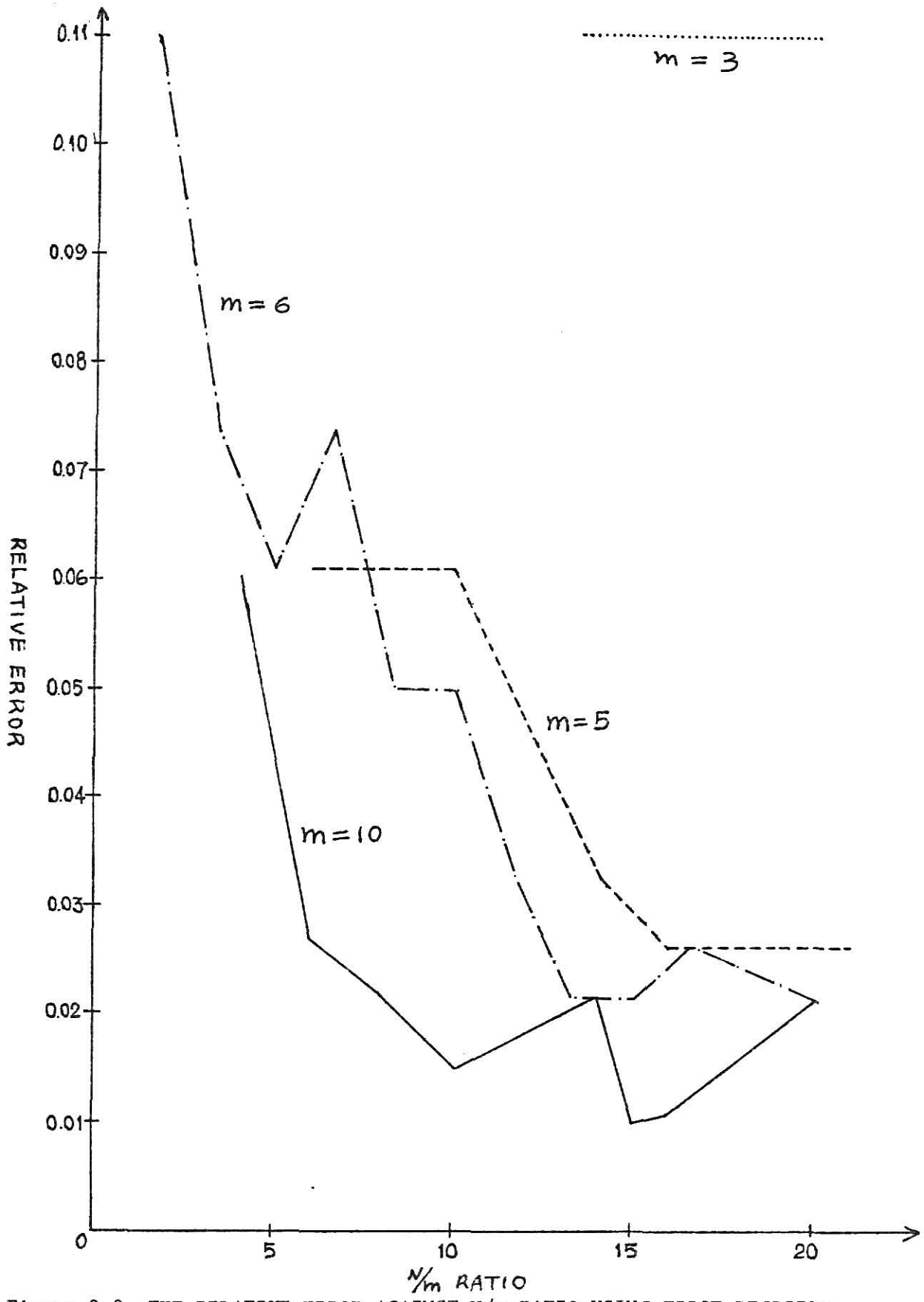

Figure 3.3 THE RELATIVE ERROR AGAINST N/m RATIO USING FIRST DECISION RULE

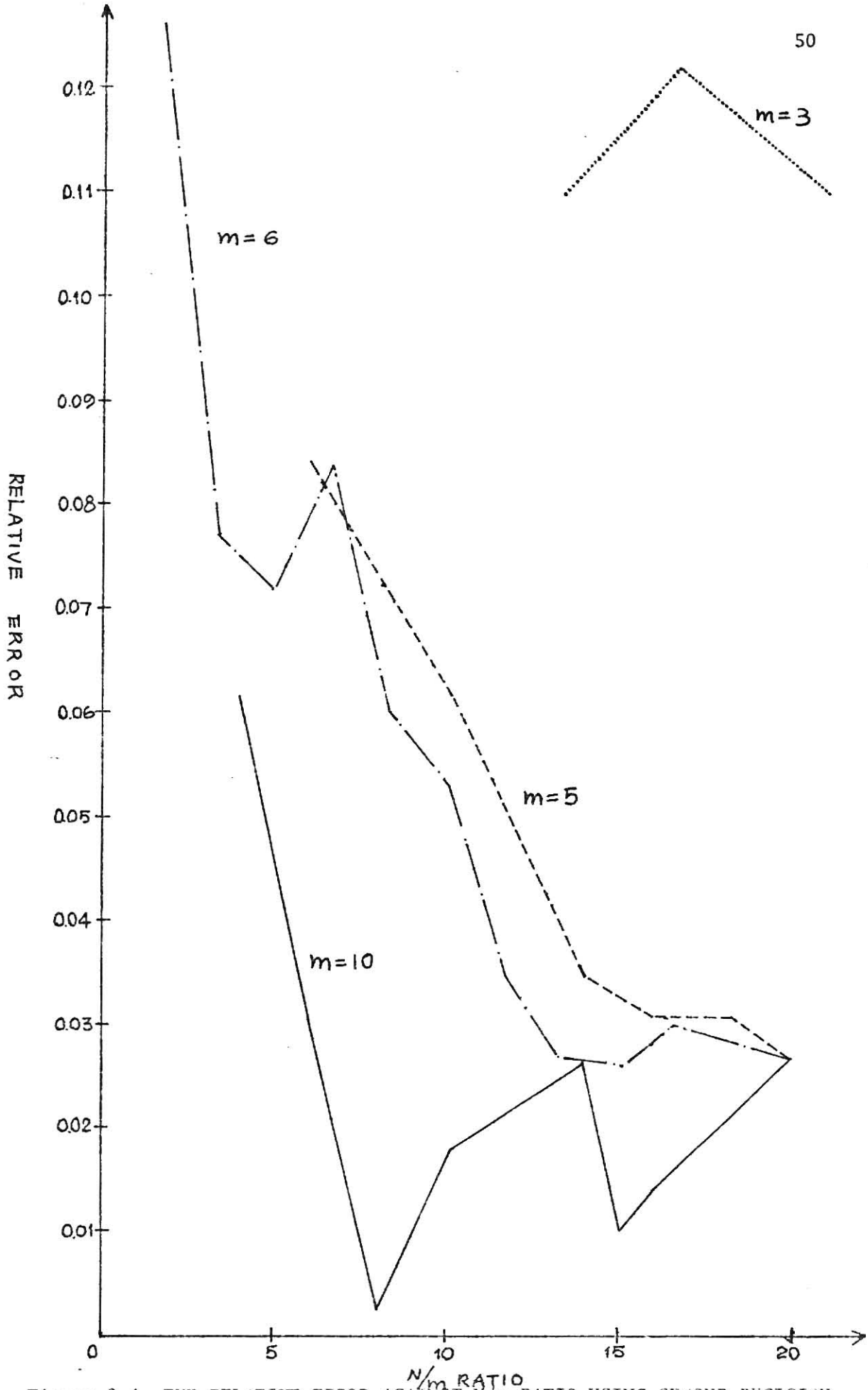

Figure 3.4 THE RELATIVE ERROR AGAINST N/m RATIO USING SECOND DECISION RULE

does not decrease as the N/m ratio increases. Thus, to clarify the curve, we made additional runs with same N and m combination but with a different process value to test those irregular points. For instance, when m = 10, N = 80, the estimated length of the transient stage is 295 (see Table 3.2d). This seemed to be an unreasonable length for the transient stage estimated as noted by comparing it to the length of the transient stage estimated by first decision rule for the same parameters. Four different process values have been run to explore this point. The results showed that the length of the transient stage for m = 10, N = 80 is between 190 and 225. Obviously, the 295 figure is a biased length of the transient stage. Thus sometimes, the low noise level will interfere with the estimation of the length of the transient stage

3.3. Variation of Noise Level

Now, we want to see what will happen to the behavior of the series if we raise the noise level to 0.05 and 0.10 (i.e., A = 15, σ = 0.75 and A = 15.0, σ = 1.5). Results are shown in Table 3.3 and 3.4.

Both of the Tables 3.3, 3.4 showed that the noise term does disturb the estimated length of the transient stage when the noise level is high, and the length of the transient stage is not proportional to the noise level.

3.4. Error of the Second Type

A phenomenon has occured in that the Δ series collected at the beginning of the time periods usually fall within the control limit. Theoretically, most of the Δ values should be outside the control

SL(30, 0.95, 0.05) = 5		
MAXIMUM NUMBER OF LAG	NUMBER OF OBSERVATIONS	LENGTH OF THE TRANSIENT STAGE
6	40	31
6	50	42
6	60	34
6	70	65
6	80	41
6	90	70
6	100	63

Table 3.3 NOISE LEVEL = 0.1

SL(30, 0.95, 0.05) = 5		
MAXIMUM NUMBER OF LAG	NUMBER OF OBSERVATIONS	LENGTH OF THE TRANSIENT STAGE
6	40	43
6	50	53
6	60	82
6	70	81
6	80	84
6	90	79
6	100	86

Table 3.4 NOISE LEVEL = 0.05

limit while the process still remains in the transient stage. However, the phenomenon is regarded as the second type of error which means that after testing the samples, we accept the samples as being good although it is in fact a bad. The outputs showed that this error of the second type will be reduced when the m value is increasing, i.e., the larger the m value, the earlier the Δ points appeared outside the control limit.

3.5. Conclusion

(1) This procedure does indicate the change of stage for the time series. There is no clear-cut point estimate to determine the end of the transient stage, thus, the approach presented above provides a rough estimate of the transient length.

(2) The length of the transient stage will be longer if more observations (N) are taken.

(3) The maximum number of lag (m) taken can not be too small, otherwise, this approach will not be able to detect the change of the time series clearly because of the noise term interference. It is advisable to use the value $m \geq 5$ in this approach.

(4) The noise level (σ/A) affects the behavior of the Δ series. The high noise level reduces the length of the transient stage, but the noise level is not proportional to the estimated length of the transient stage.

(5) The Δ series gives a false test at the beginning of the transient stage. This tendency is reduced by choosing a sufficiently large m, $m \geq 10$.

(6) For a better estimation of the length of the transient stage, a ratio of N/m around 15.0 is suggtested. Too small or too big (the N/m ratio) will cause noise trouble in the Δ series.

A reasonable technique which is to be used to inspect the end of the transient stage should be dynamic and should not cost too much in terms of computing time. The study as presented above, has achieved both requirements. The Δ series is programmable and is capable of indicating the end of the transient stage dynamically. The time for calculating 500 Δ points is 0.018 hours on the IBM 360/50.

APPENDIX A

THE COMPUTER PROGRAM LIST OF DERIVING Δ VALUES AND A SAMPLE OUTPUT

ILLEGIBLE DOCUMENT

THE FOLLOWING DOCUMENT(S) IS OF POOR LEGIBILITY IN THE ORIGINAL

THIS IS THE BEST COPY AVAILABLE

```
      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
      C                                                                      C
      C     X.........GENERATED RANDOM NUMBER                                C
      C     N.........NUMBER OF TERMS                                        C
      C     M.........MAXIMUM NUMBER OF LAGS (P)                             C
      C     NP........TOTAL SPACE STORED FOR RANDOM NUMBER                   C
      C     MP........NUMBER OF COMPONENTS                                   C
      C     NN........TOTAL NUMBER GENERATED                                 C
      C     IY........INITIAL VALUE FOR GENERATING RANDOM NUMBER             C
      C     PI........3.14159                                                C
      C     TM........FLOATING POINT OF M                                    C
      C     TN .......FLOATING POINT OF N                                    C
      C     IO........THE INDEXING OF OLDEST RANDOM NUMBER                   C
      C     IN........THE INDEXING OF NEWEST RANDOM NUMBER                   C
      C     II........THE NUMBER OF NEWEST GENERATING RANDOM NUMBER          C
      C     NC .......THE NUMBER OF CALLING THIRD SUBROUTINE                 C
      C     NS........INDEXING OF SPECTRUM                                   C
      C     TD .......TOTAL NUMBER OF CAPITAL DELTA                          C
      C     SSUM......THE SUMMATION OF X(I),I=1,N-P                          C
      C     DSUM .....THE SUMMATION OF X(I),I=P,N                            C
      C     SUPRD.....SUM OF CROSS PRODUCTS, X(I)*X(I-P)                     C
      C     COSN......THE COSINE VALUE AT EACH FREQUENCE, COS(PI*P*H/M),WHERE P=0 OR MC
      C                 2COS(PI*P*H/M),WHERE P=1,M-1                         C
      C     CORV......CORVARIANCE                                            C
      C     ESTF...... STIMATED SPECTRUM FUNCTION VALUE                      C
      C     SPEC......LN SPECTRUM VALUE                                      C
      C     SIGMA.....ABSOLUTE LN SPECTRUM VALUE AFTER MOVING AVERAGE        C
      C     DEVI .....THE MAXIMUM SPECTRUM AT EACH LAG                       C
      C                                                                      C
      CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
0001        DIMENSION ACC(1000),TAPC(1000),SIP(11),SIMA(11),DEVIA(11),AMEAN(11
           1)
0002        COMMON N,M,NP,MP,NN,IY,PI,TM,TN,
           1IO,IN,II,NS,NC,DELTA,AVERA,
           1SSUM(11),DSUM(11),SUPRD(11),COSN(150),CORV(11),ESTF(11),SPEC(11,3)
           1,SIGMA(11),X(200),YY(200),ZZ(200),AA,BB,CC
0003        READ(5,7)AA ,BB,CC
0004      7 FORMAT(3F10.5)
0005        DO   LBJ=1,2
0006        READ (5,1) N,MP,NP,NN,IY,TN,TM,PI
0007      1 FORMAT(6I5,3F10.5)
      C
      C     ECHO CHECK
      C
0008        PRINT 33,N,M,NP,MP,NN,IY,TM,TN,PI
0009     33 FORMAT(3X,'N=',I5,5X,'M=',I5,5X,'NP=',I5,5X,'MP=',I5,5X,'NN=',I5,5
           1X/5X,'IY=', I5  ,5X,'TM=',F10.5,5X,'TN=',F10.5,5X,'PI=',F10.5)
      C
      C
      C     GENERATE THE FIRST N VALUES AND GET THE SUM , SUM OF PRODUCT
      C
      C
0010        CALL FIRST
0011        II=N
0012        NA=0
0013        DO 510 I=1,MP
0014        SIMA(I)=0.
0015    510 SIP(I)=0.
```

```
      CALL [illegible]
      DO 2 [illegible]
      [illegible]
C
C
C     [illegible] THE NEW VALUE
C
C
      [illegible]
      [illegible]
      [illegible]
C
C
C     [illegible] RANDOM NUMBER AND APPLY TO TIME SERIES
C
C
      YY(IN)=[illegible](IY)
      ZZ(I,J)=(((1.-YY(IN))**(-0.16239)-1.)**0.20517-(YY(IN)**(-0.16239)
     1-1.)**0.20517)/5.3[illegible]
      X(I,J)=AA+[illegible]2.7172)**(-[illegible])*1.5
      IF([illegible])3,4,3
    3 [illegible]+1
      GO TO 5
    4 [illegible]
    5 CALL [illegible]
      CALL [illegible]
      IF([illegible]
      [illegible]
      [illegible](I)+SIMA(I)
  520 [illegible](I)**2+SIG(I)
      [illegible]
      [illegible]
    2 CONTINUE
      [illegible]
      [illegible]
      [illegible](I)=[illegible](I)/TCC
  530 DEVIA(I)=(([illegible](I)-SIMA(I)**2)**0.5/TCC
      PRINT [illegible],I=1,MP)
  5[illegible] FORMAT([illegible],'THE MEAN VALUE AT EACH FREQUENCY IS'/11F10.5//)
      PRINT 540,(DEVIA(I),I=1,MP)
  540 FORMAT([illegible],'THE STANDARD DEVIATION AT EACH FREQUENCY IS'//11F10.5/
     1/)
      [illegible]
      PRINT [illegible]
  560 FORMAT([illegible],'THE SERIES OF THE CAPITAL DELTA VALUE',//)
      WRITE 40,([illegible](K),K=1,10)
   40 FORMAT([illegible])
    [illegible] CONTINUE
      STOP
      END
```

```
0001          SUBROUTINE FIRST
        C
        C     CALCULATE INITIAL VALUE
        C
0002          COMMON N,M,NP,MP,ND,IY,PI,TM,TN,
             1I ,IF,II,NA,NB,DELTA,AVERA,
             1SSUM(11),DSUM(11),SUPRD(11),COSN(250),CORV(11),ESTF(11),SPEC(11,3)
             1,SIGMA(11),X(200),YY(200),ZZ(200),AA,BB,CC
0003          DO 101 I=1,N
0004          AI=I
0005          YY(I)=RANDU(IY)
0006          ZZ( I)=((((1.-YY( I))**(-0.16239)-1.)**0.20517-(YY( I)**(-0.16239)
             1-1.)**0.20517)/0.3240
0007      101 X(I )=AA*(P*(2.71828)**(-CC*0. 01*AI)+ZZ(I )*1.5
0008          DO 102 J=1,MP
0009          Y=0.
0010          Z=0.
0011          W=0.
0012          JM=J-1
0013          TJM=JM
0014          ND=NP-J
0015          NDD=ND
0016          DO 103 I=1,ND
0017      103 Y=Y+X(I)
0018          SSUM(J)=Y
0019          DO 104 K=J,N
0020      104 Z=Z+X(K)
0021          DSUM(J)=Z
0022          DO 105 I=1,ND
0023      105 W=W+X(I)*X(I+JM)
0024          SUPRD(J)=W
        C
        C
        C     CALCULATE THE COSINE VALUE AT EACH FREQUENCE
        C
        C
0025          DO 102 K=1,MP
0026          KP=K-1
0027          TKP=KP
0028          JK=M*KP+J
0029          COSN(JK)=COS(PI*TKP*TJM/TM)
0030          IF((K. EQ.1).OR.(K.EQ.MP)) GO TO 102
0031          COSN(JK)=2.*COSN(JK)
0032      102 CONTINUE
0033          RETURN
0034          END
```

```
0001          SUBROUTINE SECOND
         C
         C    ALLOCATE THE X VALUE AND UPDATING
         C
0002          COMMON N,M,NP,MP,NN,IY,PI,TM,TN,
             1IC,ID,II,IA,M,DELTA,AVERA,
             1SSUM(11),CSUM(11),SUPRD(11),COSN(150),CORV(11),ESTF(11),SPEC(11,3)
             1,SIGMA(11),X(200),YY(200),ZZ(200),AA,BB,CC
0003          DO 201 J=1,MP
0004          JM=J-1
0005          IA=II-JM
0006          IF(IA.LE.0)IA=NP+IA
0007          IB=I+JM
0008          IF(IB-MP)204,204,205
0009      205 IB=IB-MP
0010      204 SSUM(J)=SSUM(J)+X(IA)-X(IC)
0011          CSUM(J)=CSUM(J)+X(ID)-X(IB)
0012      201 SUPRD(J)=SUPRD(J)+(X(IN)*X(IA)-X(IC)*X(IB))
0013          RETURN
0014          END
```

```
0001            SUBROUTINE THIRD
          C
          C     CALCULATE THE VALUE OF CAPTITAL DELTA
          C
0002            COMMON N,M,NP,MP,MM,IY,PI,TM,TN,
               1IC,IN,II,NA,NB,DELTA,AVERA,
               1SSUM(11),DSUM(11),SUPRD(11),COSN(150),CORV(11),ESTF(11),SPEC(11,3)
               1,SIGMA(11),X(200),YY(200),ZZ(200),AA,BB,CC
0003            NA=NA+1
0004            NB=MOD(NA,3)
0005            IF(NB.EQ.0)NB=3
          C
          C
          C     CALCULATE THE CORVARIANCE VALUE
          C
          C
0006            DO 301 J=1,MP
0007            JM=J-1
0008            TJM=JM
0009      301   CORV(J)=(1./(N-TJM))*SUPRD(J)-(1./(N-TJM)**2)*SSUM(J)*DSUM(J)
0010            DO 302 J=1,MP
0011            JM=J-1
0012            A=0.
          C
          C
          C     CALCULATE THE ESTIMATE SPECTRUM
          C
          C
0013            DO 303 K=1,MP
0014      303   A=A+CORV(K)*COSN(K+JM*MP)
0015      302   ESTF(J)=A
          C
          C
          C     SMOOTHING THE ESTIMATE SPECTRUM
          C
          C
0016            SPEC(1,NB)=.5*ESTF(1)+.5*ESTF(2)
0017            DO 304 K=2,M
0018      304   SPEC(K,NB)=.25*ESTF(K-1)+.5*ESTF(K)+.25*ESTF(K+1)
0019            SPEC(MP,NB)=.5*ESTF(M)+.5*ESTF(MP)
0020            DO 310 J=1,MP
0021            IF(SPEC(J,NB).LE.0.)SPEC(J,NB)=0.00001
0022      310   SPEC(J,NB)=ALOG(SPEC(J,NB))
0023            IF(NA.LT.3)GO TO 308
0024            DO 311 J=1,MP
0025      311   SIGMA(J)=ABS(SPEC(J,1)/3.+SPEC(J,2)/3.+SPEC(J,3)/3.-SPEC(J,NB))
0026            SIGMA(1)=.5*SIGMA(1)
0027            SIGMA(MP)=.5*SIGMA(MP)
0028            DELTA=SIGMA(1)
0029            DO 309 I=2,MP
0030      309   DELTA=AMAX1(SIGMA(I),DELTA)
0031      308   RETURN
0032            END
```

```
FORTRAN IV G LEVEL 18                    RAND                    DATE = 72060          00/43/24          PAGE 0001

0001                  FUNCTION RAN(IX)
                      C
                      C     GENERATING THE RANDOM NUMBER
                      C
0002                  COMMON N,M,AP,MP,NA,IY,PI,TM,TA,
                     1 I,J,II,IIA,D,DELTA,AVERA,
                     2 S(11),COV(11),SUMS(11),COSM(150),CORV(11),ESTF(11),SPEC(11,3)
                     3, ISMA(11),X(200),YY(300),Z(200),AA,BP,CC
0003                  IX=IX*3125
0004                  IF(IX)5,6,6
0005                5 IX=IX+2147483647+1
0006                6 YFL=IX
0007                  RAN=YFL*.4656613E-9
0008                  RETURN
0009                  END
```

N= 50 M= 7 NP= 51 OP= 7 NN= 551
IY=54765 IP= 7.00000 TH= 50.00000 PI= 3.14159

THE MEAN VALUE AT EACH FREQUENCY IS

0.02570 0.04078 0.03803 0.03938 0.03992 0.04122 0.02808

THE STANDARD DEVIATION AT EACH FREQUENCY IS

0.02469 0.04140 0.04456 0.04243 0.04362 0.04471 0.02895

THE SERIES OF THE CAPITAL DELTA VALUE

0.14117	0.05565	0.05602	0.12543	0.20705	0.17077	0.05037	0.17622	0.17746	0.07751
0.16070	[illegible]	0.14424	0.06359	0.04554	0.02404	0.03237	0.19234	0.15192	0.12064
0.05111	0.17581	0.15508	0.07575	0.17881	0.17189	0.35574	0.11097	0.14400	0.12711
0.17719	[illegible]	0.11312	0.01971	0.12298	0.10618	0.08695	0.07943	0.06596	[illegible]
[illegible]	[illegible]	[illegible]	0.12015	0.06081	0.11941	0.09137	0.11949	0.14141	[illegible]
0.06151	[illegible]	0.26497	0.14268	0.04542	0.05583	0.03987	0.02660	0.08032	[illegible]
0.05218	0.10374	0.05180	0.07737	0.13424	0.06173	0.04514	0.15657	0.10444	[illegible]
0.11330	0.05769	0.10532	0.12946	0.09503	0.01659	0.04747	0.04528	0.06987	[illegible]
0.01867	[illegible]	0.01679	0.05130	0.02469	0.09440	0.02247	0.04091	0.06502	[illegible]
[illegible]	[illegible]	0.05774	0.05509	0.08972	0.11096	0.37685	0.16616	0.09194	0.21476
[illegible]	[illegible]	[illegible]	0.04192	0.02250	0.01540	0.10404	0.04740	0.03644	[illegible]
0.07514	[illegible]	[illegible]	0.06016	0.07546	0.06955	0.14750	0.10192	0.17503	0.11705
[illegible]	[illegible]	0.07631	[illegible]	0.03550	[illegible]	0.14126	0.14231	0.06192	[illegible]
[illegible]	[illegible]	[illegible]	0.07515	0.06174	0.07769	0.03329	[illegible]	0.04850	[illegible]
[illegible]	[illegible]	[illegible]	0.07112	0.05840	[illegible]	0.05709	0.02491	0.00856	[illegible]
[illegible]	[illegible]	0.07045	[illegible]	0.05160	[illegible]	0.02732	0.02095	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.05776	0.06749	[illegible]	[illegible]	0.01948	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.01509	0.03571	0.13163	0.14907	0.07200	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.05446	0.04756	[illegible]	0.06937	0.08851	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.03225	0.05218	0.04684	0.04264	0.03951	[illegible]
[illegible]	[illegible]	[illegible]	0.04573	[illegible]	0.01669	0.02651	[illegible]	0.01222	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.14004	[illegible]	0.07505	0.03087	[illegible]	[illegible]
0.19471	[illegible]	[illegible]	[illegible]	0.04980	0.03863	0.05175	0.02116	0.02648	0.02660
[illegible]	[illegible]	[illegible]	[illegible]	0.04872	[illegible]	0.07119	[illegible]	0.04157	[illegible]
[illegible]	[illegible]	[illegible]	0.12523	[illegible]	0.07957	0.07737	0.06625	0.15427	[illegible]
[illegible]	[illegible]	[illegible]	0.04562	0.01944	0.02677	0.05950	0.14649	0.10160	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.12121	[illegible]	0.12710	0.02927	0.03418	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.04777	0.04400	0.09128	0.06385	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.03512	0.03507	0.05010	[illegible]
[illegible]	[illegible]	0.05205	0.03594	0.07201	0.01465	0.08327	0.04164	0.13927	0.07750
[illegible]	[illegible]	[illegible]	[illegible]	0.00767	[illegible]	[illegible]	0.13762	0.06643	[illegible]
[illegible]	[illegible]	0.13311	[illegible]	[illegible]	[illegible]	0.05595	0.06566	0.03197	0.09109
[illegible]	[illegible]	[illegible]	0.01566	[illegible]	0.05679	[illegible]	0.03342	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.06582	0.03653	0.01367	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.02155	0.04701	[illegible]	0.01395	0.06275
[illegible]	[illegible]	0.01201	0.00677	0.02514	[illegible]	0.02742	0.06077	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.16668	[illegible]	0.05932	0.07045	0.09605	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.09698	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	0.03089	[illegible]	0.01154	0.05579	[illegible]	0.07383	[illegible]
[illegible]	[illegible]	[illegible]	0.12625	[illegible]	[illegible]	0.08211	0.02772	0.09058	[illegible]
[illegible]	[illegible]	[illegible]	0.01579	0.09540	[illegible]	0.02729	0.03341	[illegible]	0.04809
[illegible]	[illegible]	[illegible]	0.09181	0.06433	[illegible]	0.05193	0.03711	0.04426	0.02504
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.03019	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.01904	0.01550	0.04089	0.01631	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.01923	0.07694	0.03205	0.05288	0.02197	[illegible]
0.02497	[illegible]	[illegible]	[illegible]	0.13482	0.06377	0.03508	0.01401	[illegible]	[illegible]
[illegible]	[illegible]	0.14662	0.11612	0.02306	0.05213	0.05817	0.05146	0.09596	0.09246
[illegible]	[illegible]	[illegible]	[illegible]	0.05264	0.11729	0.05401	[illegible]	0.11505	[illegible]
[illegible]	[illegible]	[illegible]	0.14106	0.08133	0.03330	0.10116	0.06443	[illegible]	0.09251

APPENDIX B

THE COMPUTER PROGRAM LIST OF CONTROL LIMIT AND THE RESULTS

```
FORTRAN IV G LEVEL  19                  MAIN              DATE = 72121          16/12/38              PAGE 0001

0001              DIMENSION SD(2000),PROB(200),CHI(200),X(100)
0002              READ(5,70) M1,M2,M3,N1,N2,N3,MP1,MP2
0003           70 FORMAT(8I5)
0004              I=1
0005              DO 10 J=M1,M2
0006              TJ=J
0007              DO 10 K=N1,N2
0008              TK=K
0009              TN=5.*TJ*TK
0010              R1=(1.-1./TN)**2
0011              R2=(1.-2./TN)**2
0012              SD(I)=(10.75*TJ/TN)*(2./3.-(2./9.)*R1-(4./9.)*R2))**0.5
0013           10 I=I+1
0014              I=1
0015              DO 20 II=N1,N2
0016              READ(5,30) P
0017           30 FORMAT(F10.5)
0018              DO 20 J=MP1,MP2
0019              TJ=J
0020              PROB(I)=(1.-P)*ALOG(P)/TJ
0021              IF(PROB(I).GE.0.05.AND.PROB(I).LT.0.95) CHI(I)=0.004+(1.-PROB(I)-0
                 1.05)/(1.9*3.63
0022              IF(PROB(I).GE.0.02.AND.PROB(I).LT.0.05) CHI(I)=3.841+(1.-PROB(I)-0
                 1.05)/(1.05-0.05)*1.71
0023              IF(PROB(I).GE.0.01.AND.PROB(I).LT.0.02) CHI(I)=5.412+(1.-PROB(I)-0
                 1.99)/(1.01-0.99)*1.273
0024              IF(PROB(I).GE.0.001.AND.PROB(I).LT.0.01) CHI(I)=6.635+(1.-PROB(I)-
                 1.999)/(0.001-0.199
0025              IF(PROB(I).GE.0.000865.AND.PROB(I).LT.0.001) CHI(I)=10.827+(1.-PRO
                 B(I)-0.999)/(0.001-0.000135)*0.273
0026           20 I=I+1
0027              I=1
0028              DO 40 LL=1,M3
0029              LM=LL-1
0030              PRINT 100,I
0031          100 FORMAT(///,30X,'THE CONTROL LIMIT AT LAG ',I2,2X,'FOR CONFIDENCE LE
                 1VEL 0.99,0.95,0.90',/)
0032              I=I+1
0033              DO 40 L=N1,N2
0034              LN=L-1
0035              DO 50 J=M1,M2
0036           50 X(J)=CHI(LN*M3+LL)**0.5*SD(LM*N2+J)
0037           40 PRINT 60,(X(J),J=M1,M2)
0038           60 FORMAT(5X,19F6.3)
0039              STOP
0040              END
```

THE CONTROL LIMIT AT LAG 3 FOR CONFIDENCE LEVEL 0.99,0.95,0.90

0.579 0.430 0.348 0.279 0.234 0.201 0.176 0.157 0.141 0.128 0.118 0.109 0.101 0.094 0.088 0.083 0.079 0.075 0.071
0.535 0.362 0.274 0.220 0.184 0.158 0.139 0.123 0.111 0.101 0.093 0.086 0.080 0.074 0.070 0.066 0.062 0.059 0.056
0.481 0.326 0.246 0.198 0.165 0.142 0.125 0.111 0.100 0.091 0.083 0.077 0.072 0.067 0.063 0.059 0.056 0.053 0.051

THE CONTROL LIMIT AT LAG 4 FOR CONFIDENCE LEVEL 0.99,0.95,0.90

0.793 0.537 0.407 0.326 0.273 0.234 0.205 0.183 0.165 0.150 0.137 0.127 0.118 0.110 0.103 0.097 0.092 0.087 0.083
0.653 0.429 0.324 0.261 0.218 0.187 0.164 0.146 0.131 0.120 0.110 0.101 0.094 0.088 0.082 0.078 0.073 0.070 0.066
0.573 0.386 0.292 0.235 0.196 0.169 0.148 0.131 0.118 0.108 0.099 0.091 0.085 0.079 0.074 0.070 0.066 0.063 0.059

THE CONTROL LIMIT AT LAG 5 FOR CONFIDENCE LEVEL 0.99,0.95,0.90

0.893 0.605 0.457 0.368 0.307 0.264 0.231 0.206 0.186 0.169 0.155 0.143 0.133 0.124 0.116 0.110 0.103 0.098 0.093
0.745 0.505 0.381 0.307 0.256 0.220 0.193 0.172 0.155 0.141 0.129 0.119 0.111 0.103 0.097 0.091 0.086 0.082 0.078
0.658 0.446 0.337 0.271 0.226 0.195 0.171 0.152 0.137 0.124 0.114 0.105 0.098 0.091 0.086 0.081 0.076 0.072 0.069

THE CONTROL LIMIT AT LAG 6 FOR CONFIDENCE LEVEL 0.99,0.95,0.90

0.943 0.636 0.504 0.405 0.338 0.291 0.255 0.227 0.204 0.186 0.170 0.157 0.146 0.137 0.128 0.121 0.114 0.108 0.103
0.847 0.574 0.434 0.349 0.291 0.250 0.219 0.195 0.176 0.160 0.147 0.136 0.126 0.118 0.110 0.104 0.098 0.093 0.088
0.740 0.501 0.379 0.305 0.255 0.219 0.192 0.171 0.154 0.140 0.128 0.119 0.110 0.103 0.096 0.091 0.086 0.081 0.077

THE CONTROL LIMIT AT LAG 7 FOR CONFIDENCE LEVEL 0.99,0.95,0.90

1.066 0.722 0.546 0.439 0.367 0.315 0.276 0.246 0.222 0.202 0.185 0.171 0.159 0.148 0.139 0.131 0.124 0.117 0.111
0.939 0.636 0.481 0.387 0.323 0.278 0.243 0.217 0.195 0.178 0.163 0.150 0.140 0.130 0.122 0.115 0.109 0.103 0.098
0.814 0.552 0.417 0.335 0.280 0.241 0.211 0.188 0.169 0.154 0.141 0.130 0.121 0.113 0.106 0.100 0.094 0.089 0.085

THE CONTROL LIMIT AT LAG 8 FOR CONFIDENCE LEVEL 0.99,0.95,0.90

1.143 0.775 0.586 0.471 0.393 0.338 0.296 0.264 0.238 0.216 0.198 0.183 0.170 0.159 0.149 0.140 0.132 0.126 0.119
1.024 0.694 0.524 0.421 0.352 0.303 0.265 0.236 0.213 0.194 0.178 0.164 0.152 0.142 0.133 0.126 0.119 0.112 0.107
0.883 0.598 0.452 0.364 0.304 0.261 0.229 0.204 0.183 0.167 0.153 0.141 0.131 0.123 0.115 0.108 0.102 0.097 0.092

THE CONTROL LIMIT AT LAG 9 FOR CONFIDENCE LEVEL 0.99,0.95,0.90

1.216 0.824 0.623 0.500 0.418 0.359 0.315 0.280 0.253 0.230 0.211 0.195 0.181 0.169 0.158 0.149 0.141 0.133 0.127
1.102 0.747 0.565 0.454 0.379 0.326 0.286 0.254 0.229 0.208 0.191 0.177 0.164 0.153 0.144 0.135 0.128 0.121 0.115
0.947 0.642 0.485 0.390 0.326 0.280 0.245 0.218 0.197 0.179 0.164 0.152 0.141 0.132 0.123 0.116 0.110 0.104 0.099

THE CONTROL LIMIT AT LAG 10 FOR CONFIDENCE LEVEL 0.99,0.95,0.90

1.252 0.875 0.662 0.532 0.445 0.382 0.335 0.298 0.268 0.244 0.224 0.207 0.192 0.179 0.168 0.158 0.150 0.142 0.135
1.176 0.797 0.602 0.484 0.405 0.348 0.305 0.271 0.244 0.222 0.204 0.188 0.175 0.163 0.153 0.144 0.136 0.129 0.123
1.018 0.690 0.521 0.419 0.350 0.301 0.264 0.235 0.211 0.192 0.177 0.163 0.151 0.141 0.133 0.125 0.118 0.112 0.106

APPENDIX C

THE CONTROL LIMIT CHART AT 0.99, 0.95, 0.90 CONFIDENCE LEVEL

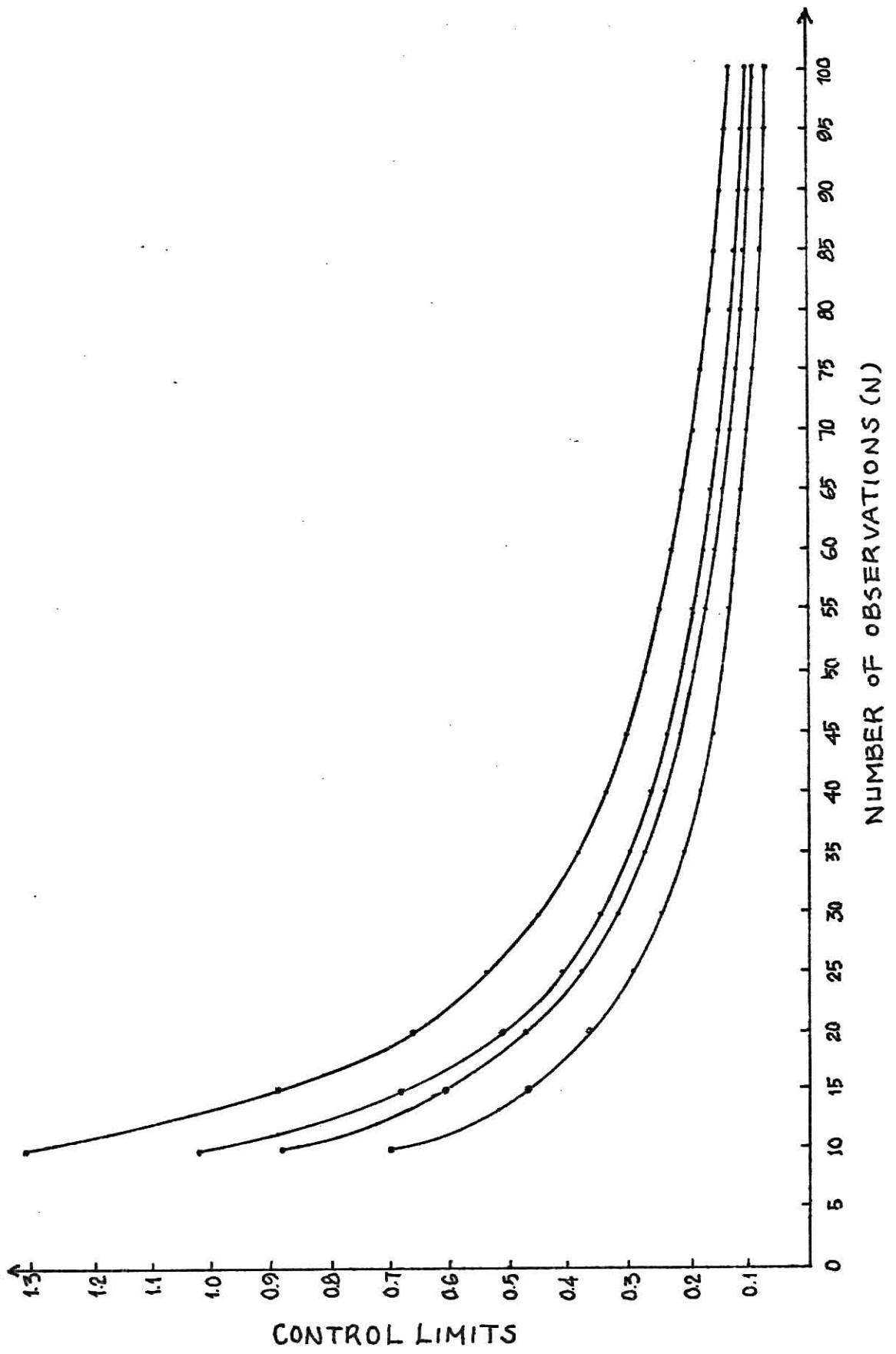

Figure A THE CONTROL LIMIT FOR 0.99 CONFIDENCE LEVEL

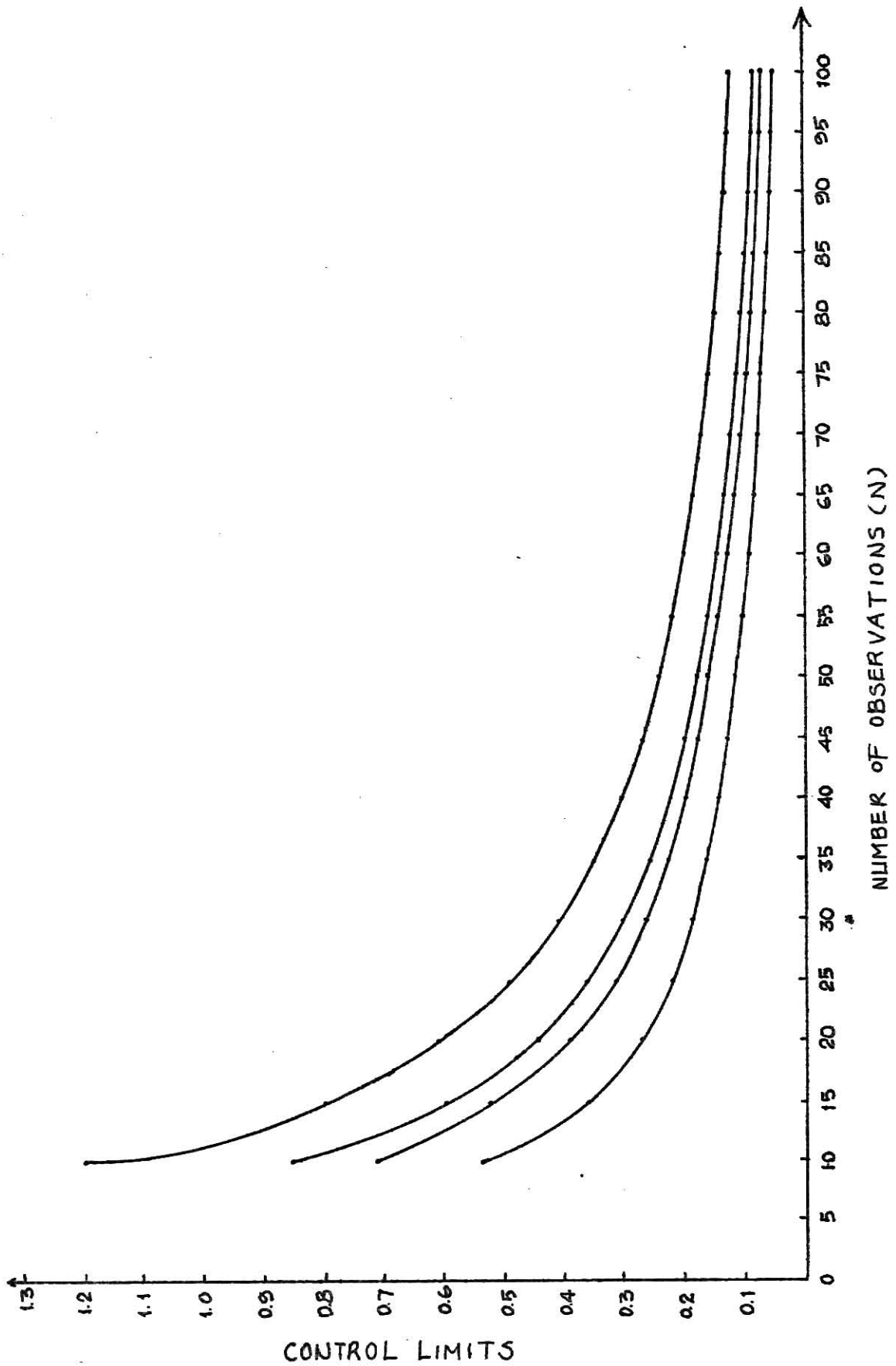

Figure B THE CONTROL LIMIT FOR 0.95 CONFIDENCE LEVEL

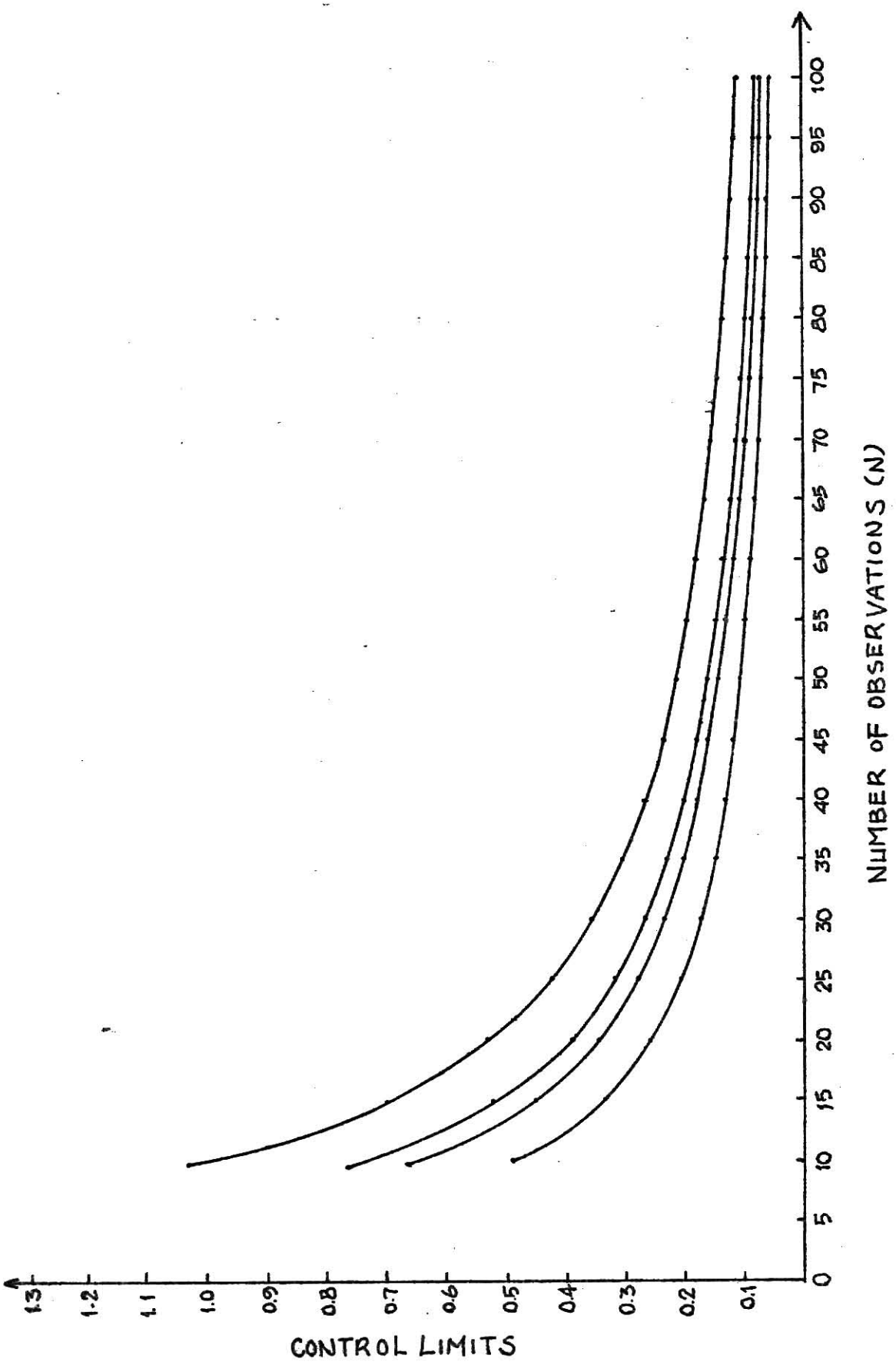

Figure C THE CONTROL LIMIT FOR 0.90 CONFIDENCE LEVEL

BIBLIOGRAPHY

1. Box, G.E. and Jenkins, G.M., Time Series Analysis, Holden-Day, Inc., San Francisco, California, 1970.

2. Brown, Robert Goodill, Smoothing Forecasting and Prediction of Discrete Time Series, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1962.

3. Bueno, Ramon J., Some Practical Solutions to Two Statistical Problems in Simulation, Master of Science Thesis in Engineering, University of California, Los Angeles, 1966.

4. Burr, I.W., A Useful Approximation to the Normal Distribution Function with Application to Simulation, Technometrics, Vol. 9, No. 4, Nov. 1967, Page 647-651.

5. Conway, R.W., Some Tactical Problems in Digital Simulation, Management Science, Vol. 10, No. 1, 1963.

6. Emshoff, James R., and Sisson, Roger L., Design and Use of Computer Simulation Models, The Macmillan Company, London, 1970.

7. Fisher, Ronald A., Statistical Methods for Research Workers, Oliver and Boyd, 14th edition, 1970.

8. Fishman, G.S., Problems in the Statistical Analysis of Simulation Experiments: The Comparison of Means and the Length of Sample Records, The Rand Corporation, RM 4880-PR, Feb. 1966.

9. Fishman, G.S., and Kiviat, P.J., The Analysis of Simulation-Generated Time Series, Management Science, Vol. 13, No. 7, Mar. 1967.

10. Granger, C.W.J., Spectral Analysis of Economic Time Series, Princeton University Press, Princeton, New Jersey, 1964.

11. Hannan, E.J., Time Series Analysis, London, 1960.

12. Harris, F.R., A Study of the Transient Stage Problems in Simulations, Master of Science Thesis in Industrial Engineering, Kansas State University, Manhattan, Kansas, 1971.

13. Jenkins, G.M., General Considerations in the Analysis of Spectra, Technometrics, Vol. 3, No. 2, pp. 133-166.

14. Jenkins, G.M., Cross-Spectral Analysis and the Estimation of Linear Open Loop Transfer Function, Proceedings Symposium of Time Series Analysis, John Wiley and Sons, 1963.

15. Jenkins, G.M. and Watts, D.G., Spectral Analysis and Its Applications, Holden-Day, Inc., 1968.

16. Kirtane, S.V., An Experimental Investigation of the Transient Phase in Digital Simulation, Master of Science Thesis in Industrial Engineering, Kansas State University, Manhattan, Kansas, 1971.

17. National Bureau of Standards, Tables of the Binomial Probability Distribution, United States Government Printing Office, Washington, 1949.

18. Parzen, Emanuel, Mathematical Considerations in the Estimation of Spectra, Technometrics, Vol. 3, No. 2, May, 1961.

19. Rao, Ambar G. and Shapiro, Arthur, Adaptive Smoothing Using Evolutionary Spectra, Management Science, Vol. 17, No. 3, Nov. 1970.

20. Reese, Peter A., Steady State Parameter Estimation in Computer Simulated Systems, Arizona State University Bulletin, Jan. 1969.

21. Rosenblatt, Murray, Time Series Analysis, The SIAM Series in Applied Mathematics, Brown University, Jun. 11-14, 1962.

22. Trigg, D.W. and Leach, A.G., Exponential Smoothing with an Adaptive Response Rate, Operational Research Quarterly, Vol. 18, No. 1, 1967, pp. 53-59.

25. Watson, G.S., Extreme Value in Samples from m-Dependent Stationary Stochastic Processes, Annual Mathematics Statistics, Vol. 25, No. 4, 1954, pp. 798-800.

AN INVESTIGATION OF THE TRANSIENT STAGE IN SIMULATION DATA BY SPECTRAL ANALYSIS

by

PETER SHIH-YU CHANG

B.A., CHUNG-YUAN COLLEGE OF SCIENCE AND ENGINEERING, 1969

AN ABSTRACT OF A MASTER'S THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1972

ABSTRACT

In simulation a problem is encountered in identifying the transient stage. The information gathered during the transient stage is of little use if the researcher's main goal is to estimate a steady state level of performance. When analyzing statistics gathered during the transient stage, the mean is biased and the variance is inflated. Therefore, a sensitive method of indicating the end of the transient stage is desirable. By identifying the end of the transient stage one may collect true statistics of the simulated process.

The output of a simulation can be interpreted as a time series. This time series is nonstationary when the process is in the transient stage and is stationary when the process reaches steady state. The basic idea of this study is to determine this stationarity using spectral analysis. An exponential decay time series is generated to represent the output of a simulation. The spectra of the generated observations was computed and compared to the spectra of a stationary time series. If two successive spectra are different, the generated time series is nonstationary and the process is still in the transient stage. On the other hand, if two spectra are the same, the generated time series is a stationary time series and the process is said to be in steady state. This method is applied to a generated time series. The results show this technique to be a possible means of determining the end of the transient stage.