WORKING PAPER TSRA 85-027E

TIME SERIES RESEARCH AND ANALYSIS DIVISION

CAHIER DE TRAVAIL RASC 85-027F

DIVISION DE LA RECHERCHE
ET DE L'ANALYSE DES CHRONIQUES

BASIC CONCEPTS OF TIME SERIES ANALYSIS

by

Estela Bee Dagum

I. BASIC CONCEPTS OF TIME SERIES ANALYSIS

1.0 Introduction

During the last decade, many statisticians have been devoted to the search for optimal methods of estimation for time series analysis and forecasting. The reason for this search is not hard to perceive. The need for accurate statistical data is crucial for decision-making. Policymakers, faced with the responsibility of controlling the economic activity, will hardly base their decisions on poor estimates or on estimates subject to significant revisions whenever new information is available.

Several methods of estimation, and corresponding computer programs, have been produced. For the most part, they are based on classical statistical techniques such as least squares or moving averages, although some other types of linear filters have also been considered. However, when using these methods of estimation, a fundamental and simple principle is forgotten, the optimality of any method of estimation strictly depends on the fulfillment of the assumptions upon which the methods rely.

In other words, behind each method of estimation, there is a model for which it is optimal. But since there is no unique model that should be applied to all time series in all situations, no method of estimation should be used uncritically as the one giving unique optimal solutions.

The more the behaviour of the generating process of a time series departs from the assumptions of the model, the less its validity to describe the phenomenon and consequently, no matter how optimal its corresponding method of estimation is, the final results will be seriously distorted.

This kind of negligence should be attributed not only to users but also to researchers engaged in wasteful discussions on the superiority of one method with respect to another (superiority usually illustrated with a given time series for which the method proved to be optimal). This criticism is extended to the theoretical statisticians, for most of the new methods of estimation are based on very simple models, the assumptions of which are sometimes so general that it is very hard to find the real phenomenon that properly fits them. Yet a large class of models exists which, from a mathematical point of view, are feasible to deal with and better fitting the behaviour of many phenomena evolving through time, but for which methods of estimation have not been developed. We specifically refer to those models based on the assumptions of non-stationarity, and non-normality. We shall not discuss here these types of models but shall concentrate our analysis on univariate models for which well-established methods of estimation exist. By discussing their assumptions, we hope to make users and researchers aware of their limitations. Our approach is concerned strictly with the theoretical empirical foundation of the models and not with their estimation procedures or other aspects of statistical inference.

We hope in this manner to provide an insight into the most realistic and/or mathematically tractable assumptions to be made concerning the behaviour of a phenomenon that evolves through time in a probabilistic way.

2. Time Series as a Sample Realization of a Stochastic Process

Following the definition of Wold (1938) already accepted in the current literature, a time series is a sample realization of a stochastic process, which from a non-mathematical viewpoint, is any process controlled by probabilistic laws. The most important characteristic for time series is that the data are dependent and the nature of this dependence is of interest in itself. The observations made as the process continues indicates the way it evolves.

Assume the process is described by only one random variable $X(t)$, then for each t, the numerical value of $X(t)$ will not be uniquely determined as in the case of a deterministic system but will depend on the random influences that have been acting upon the process up to the time t. Then, for every fixed t, $X(t)$, is a random variable defined on a probabilistic space (Ω, F, P). When t varies over the time range under consideration, we obtain a family of random variables $X(t)$ depending on the parameter t and defined on the same probablistic space. This family $\{X(t), t \in T\}$ of random variables $X(t)$ is a random function or stochastic process. If T is a finite set, then we have a finite number of random variables in the process which can be described by using multidimensional distribution functions. In effect, if $X = \{X_1, X_2, \ldots, X_n\}$, then this is an n-dimensional random-vector and it is specified by its multidimensional distribution function.

$$F(x_1, x_2, \ldots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\} .$$

If T is the set of the integers, the process is infinite and is called a stochastic process with a discrete time parameter (discrete here refers to the nature of the set T) or a stochastic sequence. If T is the set of the real numbers, the process is also infinite but uncountable or non-

denumerable and is called a continuous stochastic process or simply a stochastic process. The term stochastic process, has usually been applied to indicate processes with an infinite (countable or uncountable) number of random variables. The complete specification of a stochastic process requires something more than a mere extension of the finite case.

For every fixed t, say $t=t_1$, we have a random variable $X(t_1)$ which is completely specified by its distribution function (d.f) denoted by

$$(2.1) \qquad F_{t_1}(x_1)=P\{X(t_1)\leq x_1\}$$

For every element t in the set T, we have

$$(2.2) \quad F_t(x)=P\{X(t)\leq x\}$$

For each pair of points t_1, t_2 in the set T, we have the following bivariate distribution function,

$$(2.3) \quad F_{t_1,t_2}(x_1,x_2)=P\{X(t_1)\leq x_1,\ X(t_2)\leq x_2\}$$

of the two-dimensional random variable $X = (X(t_1), X(t_2))$.

In general, for any arbitrary finite set of t values, we have

$$(2.4) \quad F_{t_1,t_2,\ldots,t_n}(x_1, x_2,\ldots,x_n)=P\{X(t_1)\leq x_1, X(t_2)\leq x_2,\ldots,X(t_n)\ x_n\}$$

corresponding to an n-dimensional random variable $X=(X(t_1), X(t_2),\ldots,X(t_n))$.

The family of all these joint probability distributions for n=1, 2,... and all possible values t_j constitutes the family of finite dimensional distributions associated with the stochastic process $X=\{X(t),\ t\in T\}$. For (2.4) to specify a stochastic process, it has to satisfy the following two conditions:

(a) The symmetry condition, according to which

$$(2.5)\quad F_{t_{j_1},t_{j_2},\cdots t_{j_n}}(x_{j_1},x_{j_2},\ldots,x_{j_n}) = F_{t_1,t_2,\ldots t_n}(x_1,x_2,\ldots x_n)$$

where $j_1,j_2,\ldots j_n$ is any permutation of the indices 1, 2,...,n.
In words, the symmetry condition requires that the n-dimensional distributions F given in (2.4) should be symmetric in all pairs (x_j,t_j) so that F remains invariant when the t_j and x_j are subject to the same permutation.

(b) The compatibility condition according to which

$$(2.6)\quad F_{t_1,t_2,\cdots t_m,t_{m+1},\ldots,t_n}(x_1,x_2,\ldots,x_m,\infty,\ldots,\infty) = F_{t_1,t_2,\ldots,t_m}(x_1,x_2,\ldots,x_m)$$

for any $t_{m+1},\ldots,t_n$ if $m<n$.
In other words, the compatibility condition requires that

$$\lim_{x_m \to +\infty} F_{t_1,t_2,\ldots,t_m} f(x_1,x_2,\ldots x_m) = F_{t_1,t_2,\cdots t_{m-1}}(x_1,x_2,\ldots,x_{m-1})$$

It was proved by a celebrated theorem by Kolmogorov (1933), that (2.4) with (2.5) and (2.6) uniquely defines the probability distribution of the sample space of the stochastic process.

The converse is also true; that is, any family of finite distribution functions (2.4) satisfying conditions (2.5) and (2.6) can be regarded as defining some stochastic process.

We have defined a stochastic process or random function on T as a family of random variables $\{X(t),\ t\in T\}$. It is important to point out that, since the X(t)'s are random variables, this means that they are real valued functions of the outcome ω of the sample space Ω. For this

reason, sometimes a stochastic process is written explicitly as a collection of functions with two arguments, namely, ω (the outcome or point) which is an element of Ω and t (time) which is an element of T. In symbols

(2.7) $X = \{X(\omega,t), \omega\varepsilon\Omega, t\varepsilon T\}$

This was not done above because in probability theory, the dependence on ω of a random variable X is traditionally suppressed.

Looking at (2.7), two interpretations of a stochastic process can be given depending on which of the two argument variables the emphasis is placed. For a given ω, (2.7) reduces to a family of functions of time, which are indexed on ω. Thus to each outcome ω of a given experiment, there is a corresponding well-defined real function of the variable t. This function is called a realization or sample function of the stochastic process. This approach is of great interest to us since a realization of a stochastic process is precisely an observed time series.

On the other hand, if t is given, (2.7) reduces to a collection of random variables indexed on t. In this case, to specify the stochastic process, we have to give the probability of occurrence of the various realizations, which leads to the definition of a probability measure P on the function space of realizations.

If t and ω are both given, then (2.7) reduces to a number.

The process that generates the observation of a time series is thus seen as a random process where one of the infinite many ω, members of Ω, could have been the observed outcome at an instant of time t, say t_1. Since this is done for all t elements of T, one realization (observed time series) of the process is one function of a doubly infinity set of functions which might have been generated by the stochastic process.

Figure 1 shows a monthly time series consisting of all agricultural

male employees 20 years and older from January 1967 up to October 1980.

(Place Figure 1 about here)

Figure 2 shows the number of employees from **January 1975 to December 1977** together with other time serie which might have been generated from the population of time series defined by the underlying stochastic process. Therefore, each observation x_t at a given time t, say t=January 1975 can be seen as a realization of a random variable X_t with a probability density function $P(x_t)$. Similarly, the observations at any two instants of time, say t_1=January 1975 and t_2=June 1975 may be seen as realizations of two random variables X_{t_1} and X_{t_2} with joint probability density functions $P(x_{t_1}, x_{t_2})$ and in general, the observations of an equispared time series can be described by an n-dimensional random vector $X = \{X_1, X_2,\ldots,X_n\}$ and its multi-dimensional cumulative distribution functions $F(x_1,x_1,\ldots,x_n$.

(Place Figure 2 about here)

The feature of time series analysis which distinguishes it from other statistical analyses is the explicit recognition of the <u>order</u> in which the observations are made. In several areas of study, successive observations of a time series are dependent, particularly in social and economic time series. There are, however, cases where the observations are statistically independent, especially in some physics and astronomy problems. The hypothesis of dependence or independence among successive observations of a time series determines the kind of model to use for describing the generating process of the series.

When a stochastic process is specified according to the family of finite distribution functions (2.4) that fulfill the symmetry condition

(2.5) and compatibility condition (2.6), the model that generates a time series is said to be non-parametric in the sense that an infinite number of parameters are involved.

There exist however other methods of specifying a stochastic process. It is often convenient to define a random function by an analytical model, containing a finite number of parameters which are random variables. When this is the case, the model is said to be parametric. In other chapters, we analyse the parametric and non-parametric models which are more often encountered in time series analysis. One parametric model that has proved to be useful for description and forecasting in empirical cases is the autoregressive integrated moving average (ARIMA) process. On the other hand, the autocovariance and autocorrelation functions and their Fourier transforms the spectral density (spectrum) and the normalized spectral density (normalized spectrum) are non-parametric approaches to describe a random function.

There are several classical books that discuss extensively the theory of stochastic process and to which the interested reader may refer; see, for example, Bartlett (1978), Bucy and Joseph (1968), Doob (1953), Grenander and Rosenblatt (1957), Jazwinski (1970), Parzen (1952), Skorokhod (1965), and Sveshnikov (1966).

3. Classification of Stochastic Processes

Given that an observed time series $\{x(t),\ t\in T\}$ is assumed to be the sample realization of a stochastic process $\{X(t),\ t\in T\}$ one attempts to infer from the observed time series the generating mechanism or probability structure of the process. It follows that in order to analyse a time series, one must first assume a model for it which must be completely specified except for the values of its parameters which one proceeds to estimate on the basis of observed samples. Models for time series are stochastic processes and there are several ways to classify stochastic processes. Next, we introduce in table 1 a classification that enables us to discuss the underlying assumptions of each type of process as well as to distinguish those processes used for practical applications from the more theoretical ones.

Table I

Classification of Stochastic Processes

(1) According to the independence or not of the properties of the stochastic process on the time origin.	(a) Stationary (b) Non-Stationary
(2) According to the distribution functions that characterize the process.	(a) Normal (b) Non-Normal
(3) According to the independence or not of the behaviour of the process on its values in the preceding time intervals.	(1) Markovian (2) Non-Markovian

3.1. Stationary Stochastic Processes

The most important and common assumptions made about a time series are that the corresponding stochastic process is stationary and that a

stationary stochastic process can be described by the lower moments of its distribution function, namely, the mean, the variance and the autocovariance function or its Fourier transform, the spectrum.

Intuitively, a stationary process is one whose distribution remains the same as time progresses because the random mechanism producing the process does not change with time. In other words, all the probability distributions depend only on the mutual positions of the instants of time t_1, $t_1,\ldots,t_n$, but not on the actual values of these quantities.

The theory of stationary stochastic processes is extensively discussed in Cramer and Leadbetter (1967), Wiener (1949), Wold (1938) and Yaglom (1962) and only a brief summary will be given here.

A stochastic process $X = \{X(\omega,t), \omega\in\Omega \quad , t\in T\}$ is said to be stationary or strictly stationary if all the finite dimensional distribution functions (2.4) remain the same when the set of points $t_1,t_2,\ldots,t_n$ is shifted along the time series t. That is, if

$$(3.1)\quad F_{t_1,t_2,\ldots,t_n}(x_1,x_2,\ldots,x_n) = F_{t_1+T,t_2+T,\ldots,t_n+T}(x_1,x_2,\ldots,x_n)$$

for any n, $t_1,t_2,\ldots,t_n$ and T.

In particular, this implies that for a stationary stochastic process, all the one-dimensional distributions of $F_t(x)$ (2.2) must be identical. In other words, they are independent of the values of time t. In the case of the two-dimensional distribution functions of (2.3), equation (3.1) implies that they can only depend on the time difference $t_2 - t_1$, but not on the values of t_1 and t_2 and, in general, according to (3.1) the finite n-dimensional distribution functions depend only on the differences $t_1 - t_j$ $(j=2,3,\ldots,n)$.

Figure 3 below illustrates the definition of a stationary stochastic process with values of t as abscissa and values of z_t as ordinates. Resulting from one realization the stochastic process will be a sequence of values which can be represented on the graph (dotted line). Given a

sequence of numbers $x_1, x_2,\ldots,x_n$ and a sequence of times $t_1,t_2,\ldots,t_n$ we can associate with them a curve (unbroken line). The distribution function $F_{t_1,t_2,\ldots,t_n}(x_1,x_2,\ldots,x_n)$ gives the probability that the realization of the stochastic process yields values which do not exceed the real numbers x. It follows from (3.1) that this probability does not vary when X is displaced laterally. In other words, the distribution is the same in the interval $[t_1+T, t_n+T]$ and in the interval $[t_1,t_n]$.

(Place Figure 3 about here)

In practice, instead of defining a process as strictly stationary, it is very useful to consider it as stationary in the wide sense or second order stationary. In this case, only the properties of the first two moments are specified. Thus, a stochastic process is defined as stationary in the wide sense if:

(i) the mean value or first order moment $\mu(t)$ is a constant. In symbols,

(3.2) $\mu(t)=E[X(t)]=\int_{-\infty}^{\infty} x dF_t(x) = m$

(ii) the autocovariance function $\sigma_{XX}(t_1,t_2)$ defined as,

(3.3) $\sigma_{XX}(t_1,t_2) = E[X(t_1) - \mu(t_1)][X(t_2) - \mu(t_2)]$

is finite and depends on a single variable which is the difference between any two points in time t_2,t_1; that is,

(3.4) $\sigma_{XX}(t_1,t_2)=\sigma_{XX}(t_1+T,t_2+T)$

setting $T = -t_1$, gives

(3.5) $\sigma_{XX}(t_1,t_2) = \sigma_{XX}(0,t_2-t_1)=\sigma_{XX}(t_2-t_1)=\sigma_{XX}(\tau)$

where $\tau= t_2-t_1$ is the time lag between the two random variables. For $t_2-t_1 = 0$ (3.5) gives the variance $\sigma_{XX}(0)$ of the process and dividing (3.5) by the

variance we obtain, the autocorrelation function $\rho(\tau)$

$$(3.6)\quad \rho(\tau) = \rho_{XX}(t_1 t_2) = \frac{\sigma_{XX}(\tau)}{\sigma_{XX}(0)}$$

The mean value is an important characteristic of a stochastic process but it only gives the coarsest properties of the process: it is only a measure of location. A better description is provided by the autocovariance function. For $t_1=t_2$, $\sigma_{XX}(\tau)$ reduces to the variance, that is, a measure of dispersion in the mean square sense and for all $t_2 \neq t_1$, $\sigma_{XX}(\tau)$ is a measure of the linear association of the random variables through time.

For the comparison of the autocovariance functions of two different stochastic processes, it is convenient to use the autocorrelation function which eliminates the influence of the unit of measure of the random variables involved.

Observe that a second order stationary process coincides with a strictly stationary process when the process is assumed Gaussian or normal, with finite second order moments.

We now give some examples of stationary stochastic processes with a discrete time parameter.

Example 1 - A process of independent random variables or white noise

Let U_t's be independently and identically distributed random variables with

$$(3.7)\quad E(U_t) = m \ , \ \sigma_{uu}(0) = \sigma^2$$

and

$$(3.8)\quad \sigma_{uu}(\tau) = 0 \qquad \forall \tau \neq 0$$

This is a strictly stationary process but if the requirement of identical distribution is not fulfilled, the process is then stationary in the wide sense. This process is known as a purely random process by

statisticians and white noise by engineers.

Example 2. A Moving Average Process Let the independent random variables of example 1 all be identically distributed with zero mean and unit variance. We now define a new stochastic process that is a linear combination of these variables, that is,

(3.9) $Z_t = \alpha_0 U_t + \alpha_1 U_{t-1} + \alpha_q U_{t-q}$; $t=\ldots-1,0,1\ldots$

where the α_j's are given constants, then every Z_t will be identically distributed with zero mean and $\sigma^2 = \alpha_0^2 + \alpha_1^2 + \ldots + \alpha_q^2$. For the covariance of $Z_t Z_{t+s}$ we obtain

$$\sigma_{ZZ}(s) \quad (3.10) \qquad \begin{cases} \alpha_0\alpha_s + \alpha_1\alpha_{s-1} + \ldots + \alpha_{q-s}\,\alpha_q & \text{for } s \leq q \\ 0 & \text{for } s > q \end{cases}$$

Thus, the covariance depends only on the time lag s so that two pairs of random variables (Z_t, Z_{t-s}) and (Z_{t+h}, Z_{t+s+h}) will have the same joint probability distribution for any integer h. Thus, the moving average process (3.9) is strictly stationary.

Figure 4 shows a moving average process of order q=2 generated from $\alpha_0 = 1$, $\alpha_1 = .80$; $\alpha_2 = -.60$ and $\sigma^2_u = 1.0$

(Figure 4 about here)

Example 3 An Autoregressive Process. This is a process satisfying the following stochastic difference equation,

(3.11) $Z_t + \beta Z_{t-1} + \ldots + \beta_p Z_{t-p} = U_t$; $t=\ldots-1,0,1\ldots$

distributed with zero mean. If the roots of the associated polynomial equation

(3.12) $\beta_0 x^p + \beta_1 x^{p-1} + \ldots + \beta_p = 0$

are less than one in absolute value, (3.11) is said to be an invertible process, such that,

$$(3.13) \qquad Z_t = \sum_{j=0}^{\infty} \alpha_j U_{t-j} \qquad t = \ldots, -1, 0, 1, \ldots$$

where the right-hand side converges in the mean. Thus, the autoregressive process (3.11) is strictly stationary. If the U's are uncorrelated and have common mean and variance (but are not necessarily independently and identically distributed) the process is stationary in the wide sense.

Figure 5 shows an autoregressive process of order p=2 and β_0=1, β_1 =.60 β_2 =-.50 and σ_u = 1.0.

(Place Figure 5 about here)

Example 4 A process that is stationary in the wide sense but not in the strict sense is

$$(3.14) \quad Z_t = \cos \omega t \qquad t = 1, 2, \ldots$$

Where ω is uniformly distributed in the interval $(0, 2\pi)$. Then the mean of the process is zero with $\sigma^2 = 1/2$ and $\sigma(\tau) = 0$ for all $\tau \neq 0$. These random variables are uncorrelated although dependent functionally and statistically.

3.2 Non-stationary Stochastic Processes

A non-stationary stochastic process is one that does not fulfill conditions (i) and (ii) as expressed by the equations (3.2) and (3.3) for a stationary process. Of the class of non-stationary processes, the so-called homogeneous non-stationary or processes with stationary increments were first studied by Kolmogorov (1941) and Yaglom (1955).

Processes of this kind are non-stationary but, by adequately differencing the process, we obtain a stationary process in the difference

of a finite order. In the empirical applications, very often the distinction between one part of the observed series and another part of the same series is only in the local level and/or slope of the curve; therefore, the order of the difference is low.

The homogeneous non-stationary processes generalize the theory of stationary random functions. It is obvious that every stationary process is also a process with stationary differences.

The derivative (difference) of a random function (sequence) of stationary increments is a stationary process and conversely, the indefinite integral (infinite sum) of a stationary process is a process of stationary increments.

We give now the following definition for a process with stationary increments (Yaglom 1955): The random process $X(t)$ is called a process with stationary increments if the mathematical expectation of the increment of $X(t)$ during any time interval is proportional to the length of the interval, so that $E[X(s) - X(t)] = a(s-t)$; $\underline{a}$, constant; and the structure function $D(t; u,v)$ of the process $X(t)$ depends only on the differences $u-t$ and $v-t$; i.e., $D(t; u,v) = D(u-t, v-t) = D(\tau_1, \tau_2)$.

The structure function is more appropriate than the autocovariance function for the description of this type of process and was first used by Kolmogorov (1941). Thus, a process with stationary increments is characterized by a constant $\underline{a}$ (which in practice can be taken to be zero) and by the structure function, which is a function of two variables.

$$(3.15)\quad D(\tau_1, \tau_2) = E[X(t+\tau_1) - x(t)]\,[X(t+\tau_2) - x(t)]$$

For the real case, instead of (3.15) we have a function of one variable,

$$(3.16)\quad D(\tau) = E[X(t+\tau) - X(t)]^2$$

For a real stationary process, a is equal to zero and D( ) can be expressed in terms of the autocovariance function $\sigma(\tau)$ as follows:

$$(3.17)\quad D(\tau) = 2\sigma(0) - \sigma(\tau) - \sigma(-\tau) = 2\sigma(\tau)$$

An important category of homogenous non-stationary processes is the homogenous linear non-stationary where the process is seen as the output from a linear filter, the input of which is white noise. In this type of process, previous values of the random variable X(t) will contribute to the determination of its value at time t. The non-stationarity may be present in the mean and/or in the variance.

We now give an example of non-stationary stochastic processes.

Example 5. Random Walk Process. A simple example of non-stationary processes is given by the random walk process in which, as the time increases, the random variables tend to oscillate about the mean value (a line) with an ever increasing amplitude. This type of process has been used for time series of economic levels and in particular, to fit stock market price data. Let us assume that U_t is a purely random process and X_t is another process related to X_t as follows:

$$X_1 = U_1$$

$$X_2 = U_1+U_2 = X_1+U_2$$

$$\vdots$$

$$(3.7)\qquad X_t = U_1+U_2+\ldots+U_t = X_{t-1}+U_t$$

Then X_t can be expressed as a linear combination of the purely random

process U_t with all the weights equal to 1. If the expected value of U_t is μ and the variance σ^2_U, it follows that:

(3.8) $E(X_t) = t\mu$

and (3.9) $\text{var } X_t = t\,\sigma^2_U$

The autocovariance of the X_t process is

(3.10) $\sigma_{XX}(t_1,t_2) = \min(t_1,t_2)\,\sigma^2_U$

Equation (3.7) is also called a process with <u>independent increments</u>, since

(3.11) $U_t = X_t - X_{t-1}$

is a purely random process or white noise and hence equation (3.11) has stochastically independent increments $U_t - U_{t-1}$, $U_{t-1} - U_{t-2}$, ..., $U_2 - U_1$.

3.3. <u>Normal and Non-Normal Stochastic Processes</u>

Another attribute that can be used for the classification of stochastic processes is the form of the distribution functions that specify the process. The distribution function most frequently encountered is the multivariate Gaussian or Normal distribution. Since the multivariate Normal distribution is fully characterized by its moments of the first and second order, the existence of a constant mean and an autocovariance matrix where each component is function only of the time lag, would suffice for this type of process to be stationary in the wide sense; and if the random variables are real-valued, then the process will be wide stationary and strictly stationary at the same time. Most of the processes to be studied here belong to this class. If the distribution functions that characterize a stochastic process are not normal, the process is then said to be a Non-Normal process. One of the most commonly

applied Non-Normal process is the Poisson process. The Poisson process arises in situations in which one is interested in the total number of occurrences of a specified type of event up to time $t \geq 0$ as, for example, the telephone calls originating in a given locality, the occurrence of accidents at a certain intersection or in a mine, the arrival of customers, for services and the breakdowns of a machine. The justification for viewing these cases as Poisson processes is based on the concept of the law of rare events. We have a situation of many Bernoulli trials with small probability of success where the expected number of successes is constant. Under these conditions the actual number of events occurring follows a Poisson distribution.

A process $X(t)$, $t \geq 0$ is said to be a Poisson process with rate $\lambda > 0$ if it satisfies the following conditions:

(i) $X(0) = 0$

(ii) $X(t)$ is a process of independent increments

(iii) the number of events in any interval of length t is Poisson distributed with rate λt, that is, for all s, $t \geq 0$,

(3.18) $P\{X(t+s) - X(s) = x\} = e^{-\lambda t} (\lambda t)^x / x!$; $x = 0,1,2,...$

It follows from condition (iii) that a Poisson process has stationary increments and also that $E\{X(t)\} = \lambda t$ which explain why λ is called the rate of the process. Since it is not clear how we would determine that conditions (iii) is satisfied, an equivalent definition is given where (3.18) is changed to:

(3.19) $P\{X(t+h) - X(t) = 1\} = \lambda h - o(h)$

(3.20) $P\{X(t+h) - X(t) \geq 2\} = o(h)$

where a function $f(x)$ is said to be of order $o(h)$ if $\lim_{h \to 0} f(h)/h = 0$.

Equation (3.19) gives the probability of at least one event happening in a time period h and equation (3.20) states that the probability of two or more events occurring in time h is o(h). Thus, condition (iii) excludes the possibility of simultaneous occurrence of two or more events.

Next we give an example of a Poisson process and for a more detailed discussion of this kind of stochastic process the reader may refer to Karlin and Taylor (1981), Rolski (1981) and Ross (1980)

Example 6. Total Claims on a Life Insurance Company

Let W_1, W_2,... denote the occurrence times of the death of the policyholders of a given life insurance company. Considering these times as the arrival times of insurance claims, the number of deaths can be described by a Poisson process $\{X(t)\}$.

3.4. Markovian and Non-Markovian Processes

Another important classification of a stochastic process is that of Markovian and Non-Markovian processes made according to the independence or not of the behaviour of the process on its values in preceding time intervals.

Consider a stochastic process $\{X_n,\ n=0,1,2,\ldots\}$ that takes on a finite or countably number of possible values. Let us assume this set of possible values is that of the non-negative integers $\{0,1,2,\ldots\}$. A discrete stochastic process is called a Markov Chain if, for any sequence $\{x_0, x_1, \ldots, x_{n+1}\}$ of states

$$P\{X_{n+1} = x_{n+1} | X_n = x_n,\ X_{n-1} = x_{n-1}, \ldots X_0 = x_0\} =$$

$$P\{X_{n+1} = x_{n+1} | X = x_n\} = P(n,\ x_n,\ x_{n+1}) \tag{3.21}$$

Equation (3.21) can be interpreted as stating that for a Markov Chain the conditional distribution of any future state x_{n+1} given the past states x_{n-1}, ..., x_0 and the current state x_n is independent of the past states and depends only on the present state. The values of $P(n,\ x_n,\ x_{n+1})$

represents the probability that the process, when in state x_n, will make a transition into state x_{n+1}.

The probabilities P(n, x, y) are called one-step transition probabilities at time n. Let us fix an $n \in T = \{0, 1, 2, \ldots\}$, then the transition probabilities P(n,x,y), $x,y \in S$ where S denotes the state space, can be treated as the elements of a matrix

$$(3.22) \quad M(n) = \begin{bmatrix} P_{oo} & P_{o1} & P_{o2} \ldots \\ P_{10} & P_{11} & P_{12} \ldots \\ \vdots & & \\ P_{io} & P_{i1} & P_{i2} \ldots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

The matrix M(n) is called the one-step transition matrix at time n. Since the probabilities depend on n, the Markov Chain is said to be non-homogenous or temporarily non-homogenous. If the one-step transition probabilities P(n,x,y) are independent of n; that is M(n) = M = [P(x,y)] for all $n \in T$ then the Markov Chain is said to be homogenous or with stationary transition probabilities. Therefore, if a Markov Chain is homogenous, the process is completely determined by its one-step transition matrix M = [P(x,y)] ($x,y \in S$) and an initial distribution ($P_0(x)$, $x \in S$). The concept of one-step transition probabilities can be extended to n-step transition probabilities defined by,

$$(3.23) \quad P^n(x,y) = P\{X_{m+n} = y \mid X_m = x\} \; ; \quad n \geq 2$$

Equation(3.23)gives the probability of transition from state x to state y in n steps. Accordingly, an n-step transition matrix is defined by $M(n) = [P^n(x,y), \; x,y \in S]$.

The Poisson processes discussed in the previous section are also Markov processes with discrete states in continuous time. The main characteristic of this prototype of Markov processes is that in a small time interval there is either no change or a radical change of state. Therefore, in a finite interval there is either no change or a finite or countably number of discontinuous changes. Realizations of such processes are step functions. If change of states occur all the time the process is called a diffusion process or a Markov process with continuous state space.

Next, we give a few examples of commonly encountered Markov processes and for further discussion on these processes the reader may refer to Cox and Miller (1965), Basawa and Rao (1980), Joffe (1978), Dynkin (1965) and Ross (1980).

Example 7. A process of Independent Random Variables.
If $X = \{X_n, n \geq 0\}$ is a sequence of independent random variables, X is trivially a Markov process.

Example 8. A First-Order Autoregressive Process. This kind of process is defined by

$$(3.24)\quad Z_t = \beta Z_{t-1} + U_t; \quad t \geq 0$$

and is a Markov process because the distribution function of Z_t is completely characterized by a knowledge of Z_{t-1} and is independent of the random variables preceding Z_{t-1}. If U_t is Normal with mean zero and variance σ_u^2 the process is also Gaussian and the conditional probability $P\{Z_t = z_t \mid Z_{t-1} = z_{t-1}\}$ is Normal with mean βZ_{t-1} and variance σ_u^2. If $\beta = 1$ and Z_0 is finite, then (3.24 defines a random walk (see example 5). The random walk is a non-stationary process whereas the first-order autoregressive process is stationary if $|\beta| < 1$.

Example 9. Wiener or Brownian motion Process. A stochastic process $\{X(t), t \in R\}$ is called a Wiener or Brownian motion process if:

(i) $X(0) = 0$

(ii) the increment $X(t) - X(s)$ over the interval $[s,t]$ is normally distributed with mean zero and variance $\sigma^2(t-s)$. In other words, $X(t)$, $t \geq 0$ is a process with stationary increments; and

(iii) $X(t)$ is a process with independent increments.

The Brownian motion is a Gaussian-non-stationary-Markov process. It is the limiting case of random walk processes, where the state space is the continuum of real numbers and in which changes of states are occurring all the time.

A process is said to be non-Markovian if the probability properties of the process at a given point in time depend on its values in preceding time intervals. Among the non-Markovian processes the Homogeneous Linear Non-stationary processes are very important for describing socio-economic phenomena. These processes will be extensively discussed in another section.

The three main categories of stochastic processes here introduced are compatible and several combinations are possible. Thus, a process may be normal-stationary-Markovian, or non-normal and non-stationary or normal-non-stationary-non-Markovian and so on.

4. Ergodicity

In the applications of probability theory, one ordinarily deals with events that repeat themselves many times. Hence, as the mean value of a random variable X characterizing an observed event, we can take the arithmetic mean of all the observed values X_j of X. Similarly, to determine the mean value and the autocovariance functions $\sigma(t,s)$ of a stochastic process X(t), we must have a large number of realizations of the process X(t), that is, $X_1(t)...,X_N(t)$ and then we calculate the mean for every value t, and the autocovariance function for every pair of values t and s. However, in practice there are many time series and particularly, economic time series, the data of which is the result of a single experiment. For this kind of series, it is necessary to specify the conditions under which a single realization of the process allows the calculation of consistent estimators for all the characteristics of the distribution of the process. The possibility of doing the latter is because the Ergodic theorem (or law of large numbers) is applicable to a class of homogeneous linear non-stationary processes that can be made linear stationary after taking finite differences of a relatively low order.

According to the ergodic theorem, the mathematical expectation of X(t) and of X(t)X(s) obtained by taking the average of the corresponding quantities over the whole space of outcomes Ω (called the ensemble average or sometimes spatial average) can be replaced by the time averages of the same quantities.

Given stationary stochastic process X(t), the time average for the mean of X(t) is defined by,

$$(4.1) \qquad \hat{\mu}_T = \frac{1}{T}\int_0^T X(t)dt; \qquad 1 \leq t \leq T$$

For $T\to\infty$, the time average $\hat{\mu}_T$ converges to the ensemble average μ in mean square (also in probability) if and only if

$$(4.2)\quad \lim_{T\to\infty} E(\hat{\mu}_T-\mu)^2 = 0$$

since,

$$(4.3)\quad \lim_{T\to\infty} E(\hat{\mu}_T-\mu)^2 = \lim_{T\to\infty} \frac{1}{T}\int_0^T \sigma(\tau)d\tau; \qquad \tau = s-t$$

then, the (4.2) will be verified if and only if

$$(4.4)\quad \lim_{T\to\infty} \frac{1}{T}\int_0^T \sigma(\tau)d\tau = 0$$

This condition was first shown by Slutsky (1938) and any stationary process that fulfills the (4.4) is said to be ergodic in the mean. For ergodicity in the second order moments, we need the time average of the autocovariance function, namely

$$(4.5)\quad \hat{\sigma}_T(\tau) = \frac{1}{T}\int_0^T \{[X(t+\tau)-\mu]\ [X(t)-\mu]dt$$

to converge in mean square to the ensemble average of the autocovariance function $\sigma(\tau)$. That is,

$$(4.6)\quad \lim_{T\to\infty} E[\hat{\sigma}_T(\tau)-\sigma(\tau)]^2 = 0$$

For normal process, the (4.6) reduces to

$$(4.7)\quad \lim_{T\to\infty} \frac{1}{T}\int_0^T |\sigma(\tau)|^2 d\tau = 0$$

The existence of time averages and their convergence properties were proved in the famous ergodic theorem of Birkhoff and Khinchin (see Genedenko (1966)).

It is easy to show that all linear stationary stochastic processes are ergodic. Similarly, the homogeneous linear non-stationary processes which are linear stationary after taking differences of a finite order fulfill the principle of ergodicity. However, not all the stationary processes are ergodic. Consider, for example, the following simple stationary process, where a and b are normally distributed random variables with zero means and common variance σ^2.

$$X(t) = a\cos\lambda t + b\sin\lambda t \tag{4.8}$$

The autocovariance function $\sigma(\tau)$ is

$$\sigma(\tau) = \sigma^2\cos\lambda\tau \tag{4.9}$$

Using (4.4) we can show that this process is ergodic in the mean. We have

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T\sigma(\tau)d\tau = \lim_{T\to\infty}\sigma 2\ \frac{\sin\lambda T}{T} = 0 \tag{4.10}$$

But the process is not ergodic for the autocovariance function. Applying (4.7),

$$\lim_{T\to\infty}\ \frac{1}{T}\int_0^T|\sigma(\tau)|^2 d\tau = \lim_{T\to\infty}\left(\frac{\sigma^4}{2} + \frac{\sigma^4\sin 2\lambda T}{T}\right) = \frac{\sigma^4}{2} \tag{4.11}$$

REFERENCES

Basawa, I.V. and B.L.S.P. Rao (1980): Statistical Inference for Stochastic Processes, Academic Press, New York

Bartlett, M.S. (1978): An Introduction to Stochastic Processes with Special References to Methods and Applications, Cambridge University Press, England, 3rd edition.

Bucy, R.S. and P.D. Joseph (1968): Filtering for Stochastic Processes with Applications to Guidance, Interscience, John Wiley, New York.

Cox, D.R. and H.D. Miller (1965): The Theory of Stochastic Processes, Chapman and Hall Ltd., England.

Cramer, H. and M.R. Leadbetter (1967): Stationary and Related Stochastic Processes, John Wiley, New York.

Doob, J.L. (1953): Stochastic Processes, John Wiley, New York.

Dynkin, E.B. (1965): Theory of Markov Processes, Pergamon Press, England.

Genedenko, B. (1966): The Theory of Probability, Chelsea Publishing Co., London, 3rd edition.

Grenander, V. and M. Rosenblatt (1975): Statistical Analysis of Stationary Time Series, John Wiley, New York.

Jazwinski, A. (1970): Stochastic Processes and Filtering Theory, Academic Press, New York.

Joffe, A. (1978): Branching Processes, Marcel Dekker, New York.

Karlin, S. and H.M. Taylor (1981): A Second Course in Stochastic Processes, Academic Press, New York.

Kolmogorov, A. (1933): Foundations of the Theory of Probability (in Russian) translated in English by Chelsea Publishing Co., London, 3rd edition (1956).

Kolmogorov, A. (1941): The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds Numbers, Daklady Akad.Nank. SSR, Vol. 30 - (reprinted in: Turbulence: Classic Papers on Statistical Theory, edited by S.K. Friedlander and L. Topping, Interscience Publishing, New York, 1961).

Parzen, M. (1962): Stochastic Processes, Holden Day, California.

Rolski, T. (1981): Stationary Random Processes Associated with Point Processes, Springer-Verlag, New York.

Ross, S. (1980): Stochastic Processes, John Wiley, New York.

Skorokhod, A.V. (1965): Studies in the Theory of Random Processes, Addison Wesley.

Slutsky, E.E. (1938): "Sur les fonctions aléatoires presque périodiques et sur la décomposition de fonctions aléatoires stationnaires en composantes" Actualités Scient. et Ind., Paris.

Sveshnikov, A.A. (1966): Applied Methods of the Theory of Random Functions, Pergamon Press, England.

Wiener, N. (1949): Extrapolation, Interpolation and Smoothing of Stationary Time Series, John Wiley, New York.

Wold, H. (1938): A Study in the Analysis of Stationary Time Series, Almquist and Wiksell, Stockholm (2nd ed. 1954), Sweden.

Yaglom, A.M. (1962): Theory of Stationary Random Functions, Prenctice Hall, New Jersey.

Yaglom, A.M. (1955): "The Correlation Theory of Processes whose nth Difference constitute a Stationary Process" Matem.Sb 37(79) p.141.

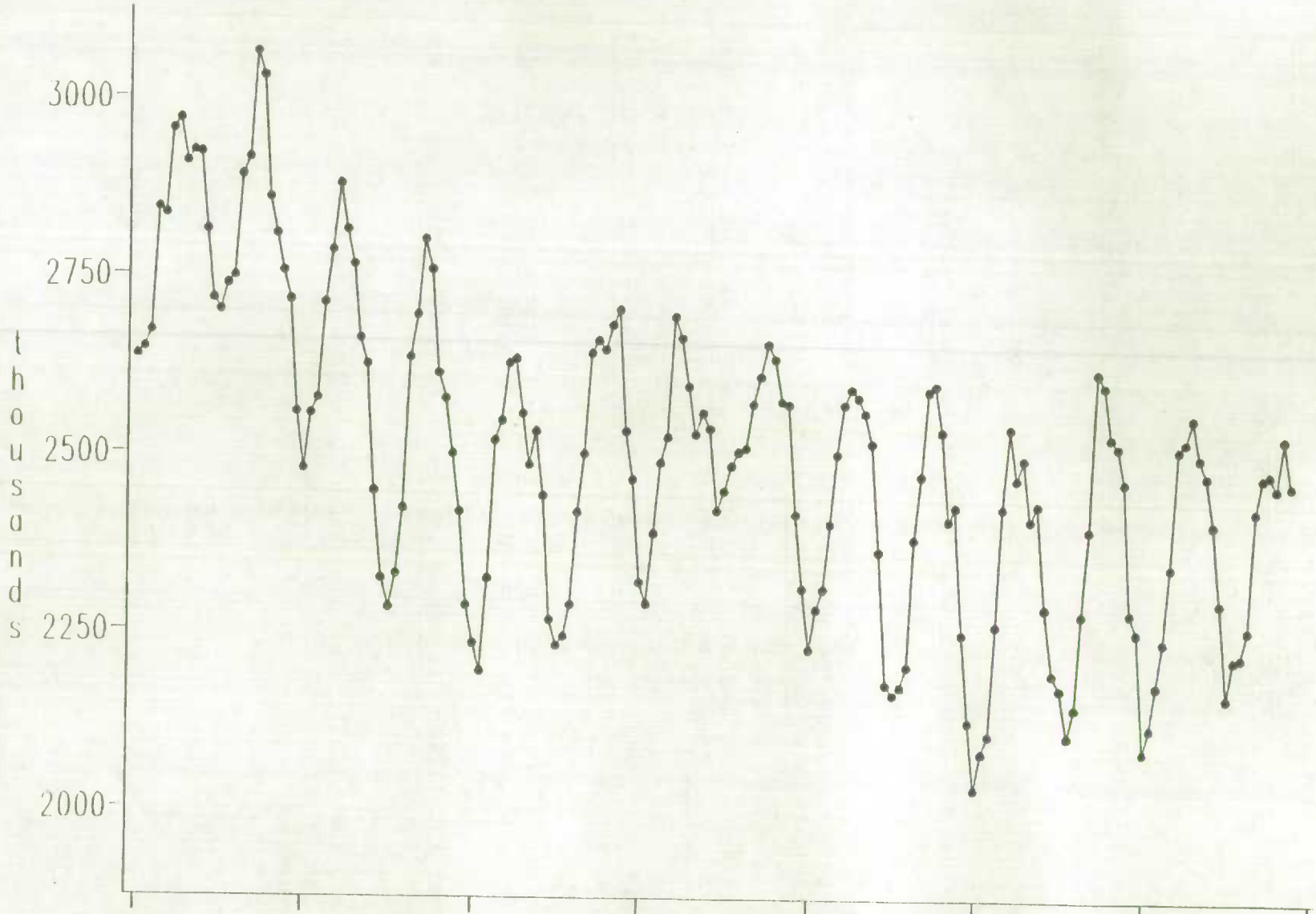

Figure 1 – Agricultural Employment Men 20+
Source: U.S. Bureau of Labour Statistics

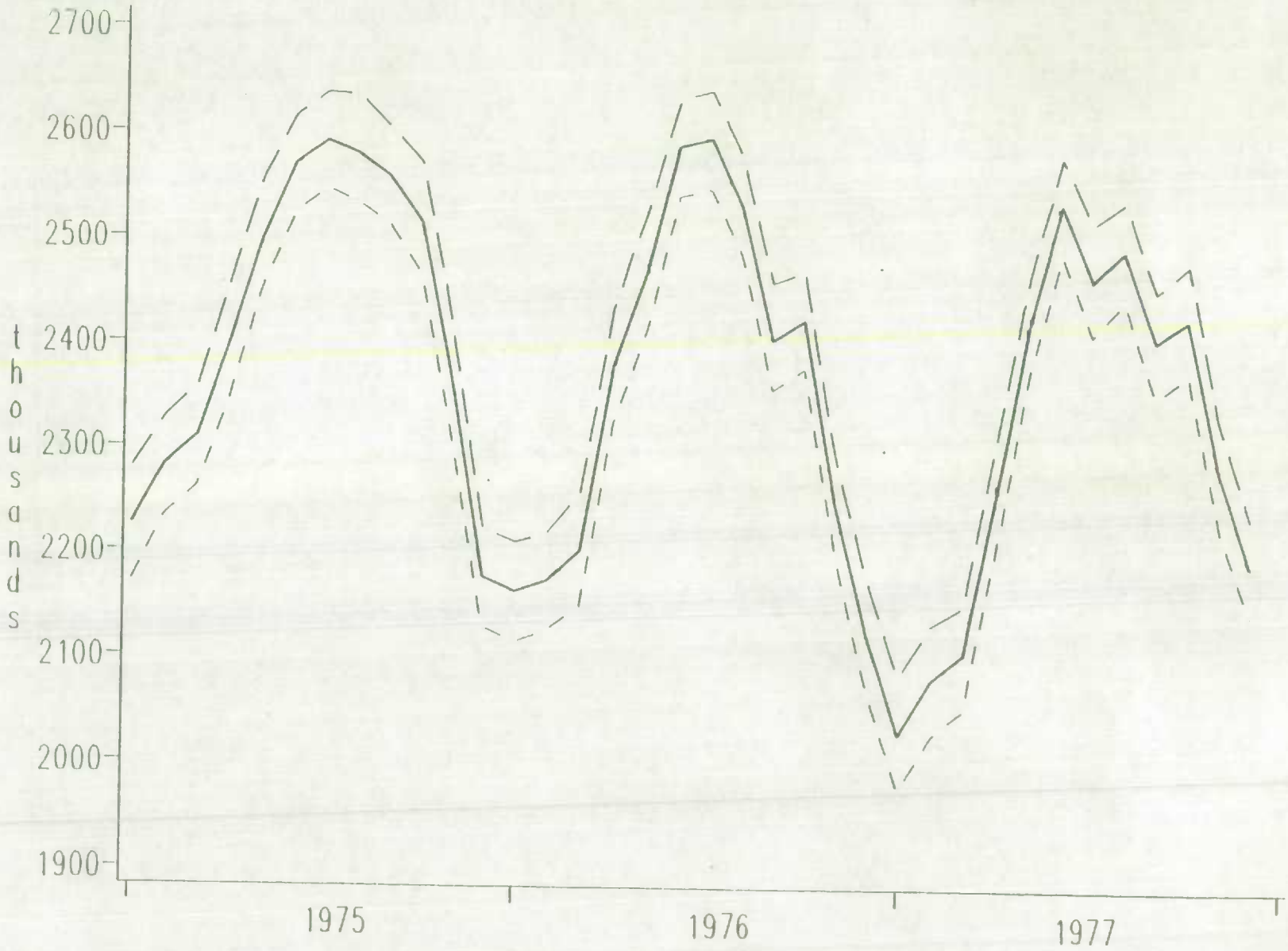

Figure 2 – An observed time series (thick line) with other time series representing realizations of the same stochastic process

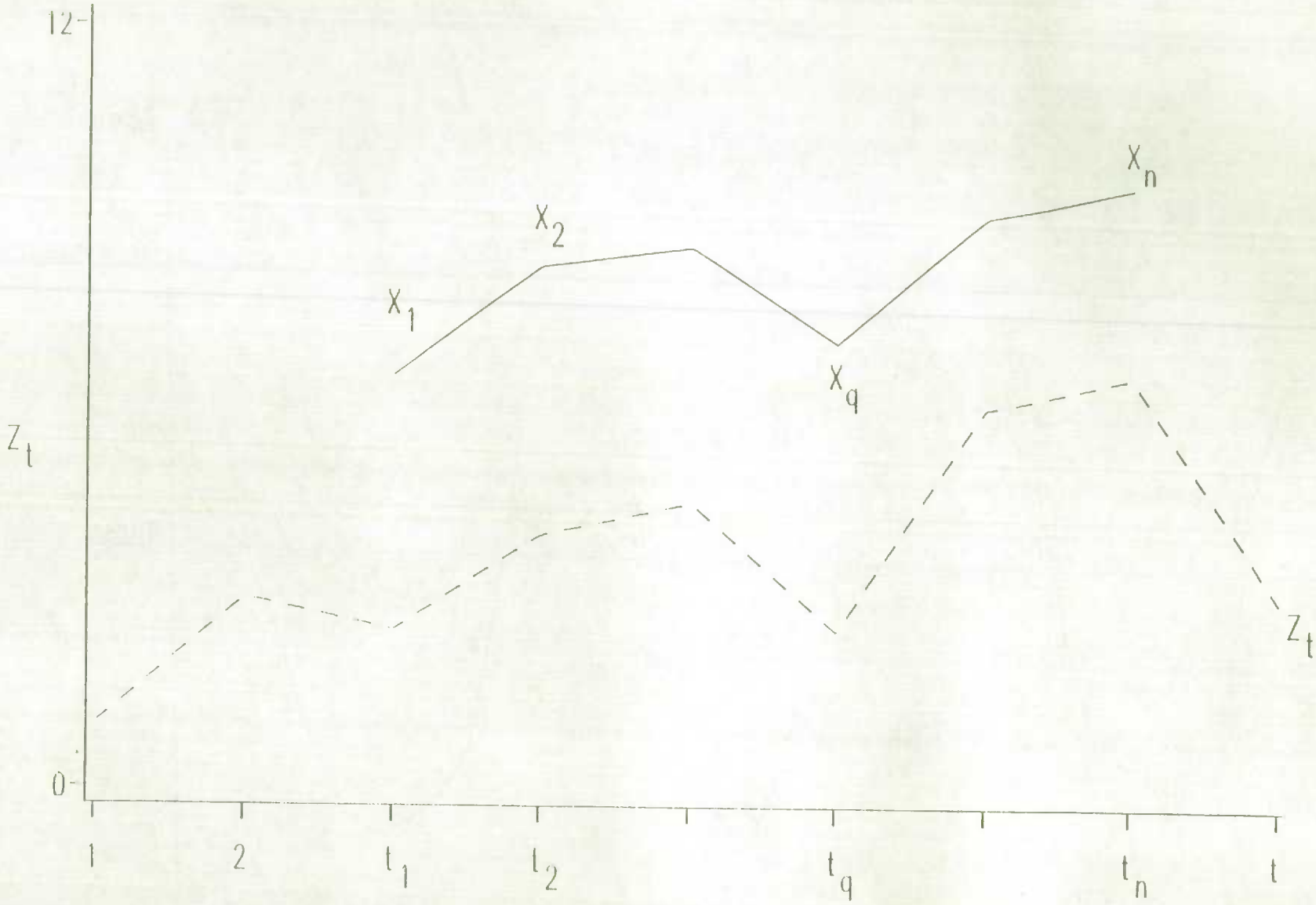

Figure 3 - Graphical Illustration of the Definition of a Stationary Stochastic Process

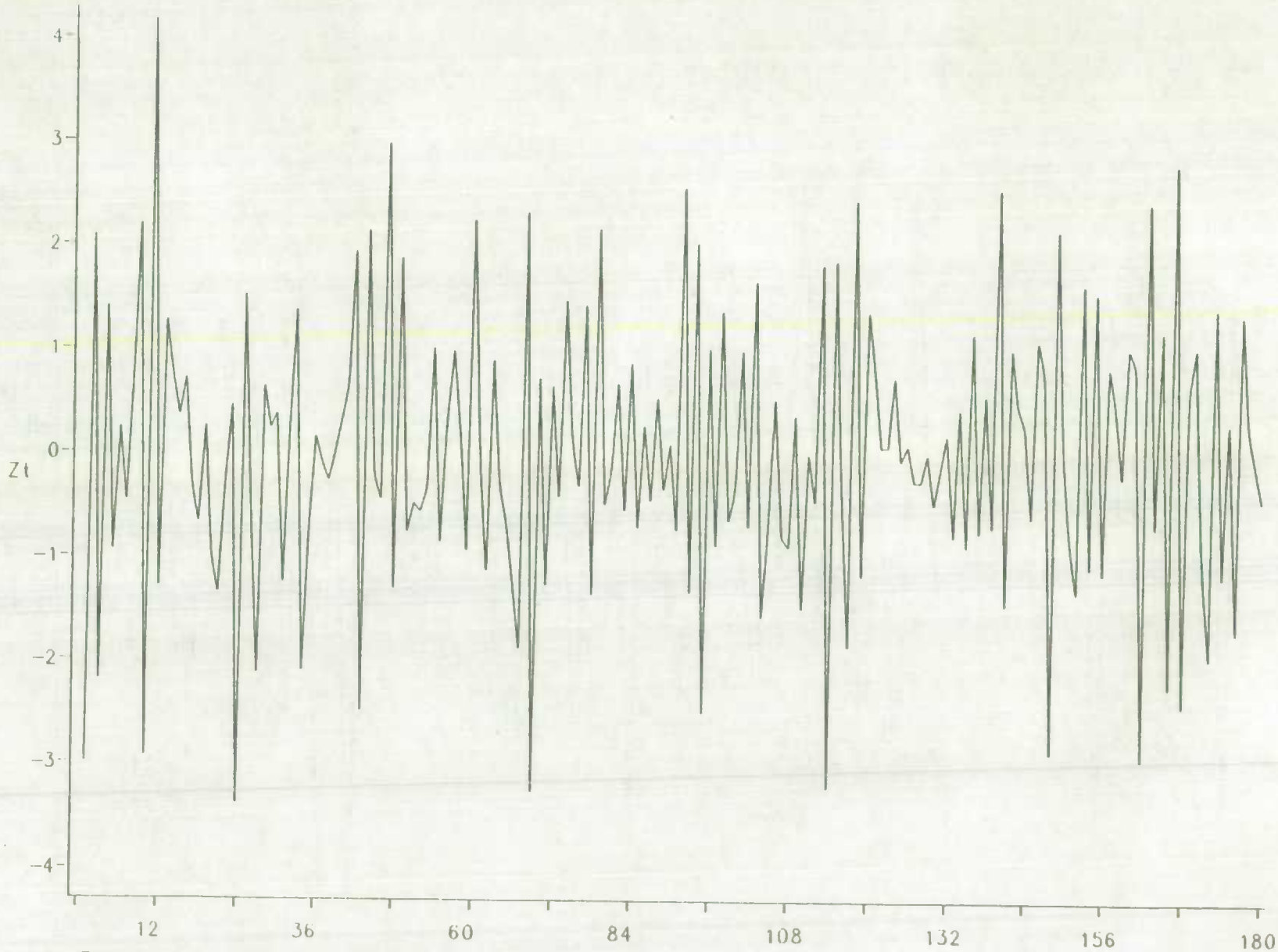

Figure 4 – A Realization of a Moving Average Process of Order 2
Generated From $\alpha_1 = .80$ $\alpha_2 = -.60$ and $\sigma_u^2 = 1.0$

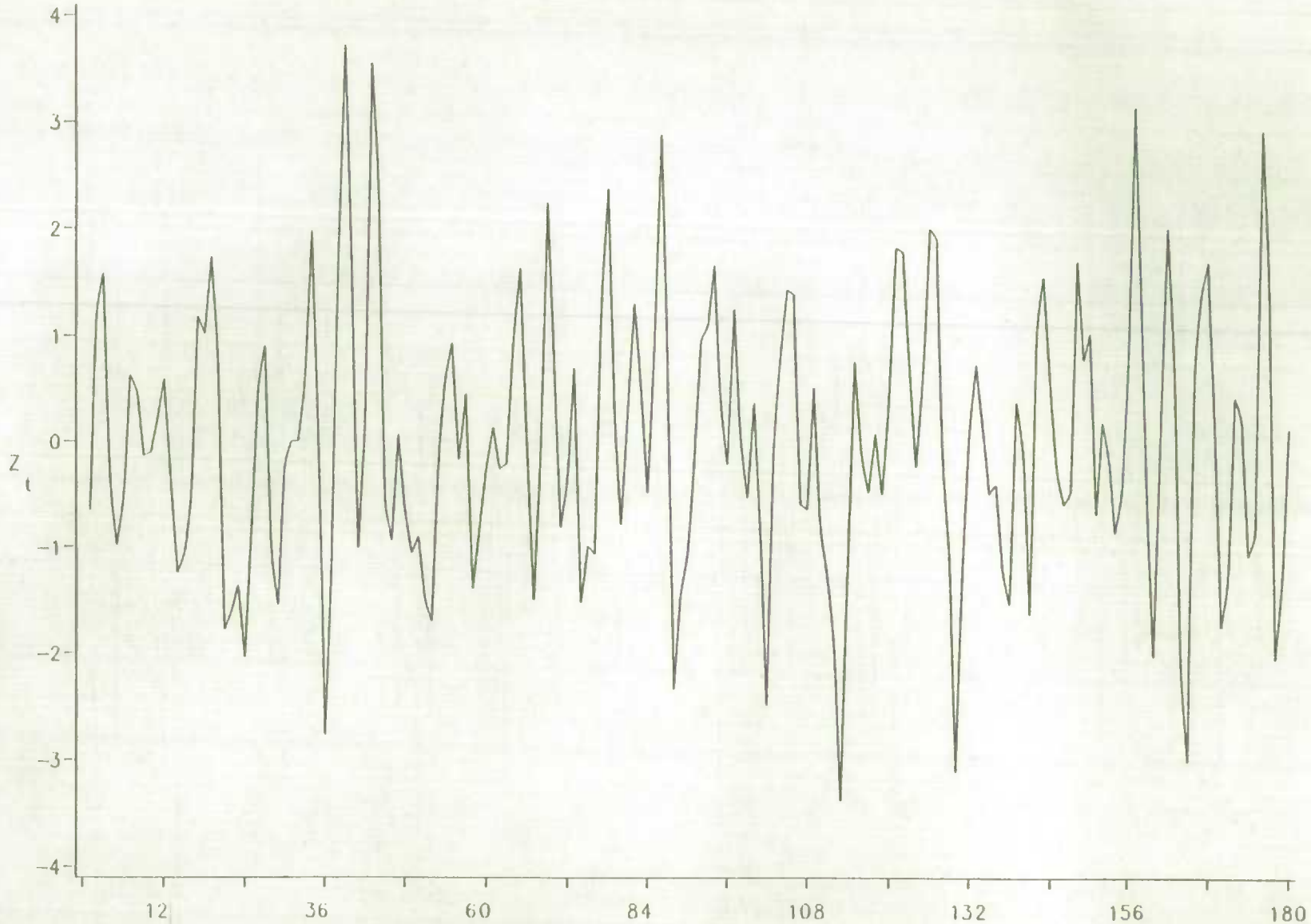

Figure 5 – A Realization of an Autoregressive Process of Order 2
Generated From $\phi_1 = .60$ $\phi_2 = -.50$ and $\sigma_a^2 = 1.0$