CATALOGUE 12-548 OCCASIONAL

Models for time series

Statistics Canada Statistique Canada

STATISTICS CANADA
Current Economic Analysis Division
General Time Series Staff

MODELS FOR TIME SERIES

Published by authority of
The Minister of Industry Trade and Commerce

May 1974
3-1501-517

Price: $1.40

Information Canada
Ottawa

Foreword

This monograph on Models for Time Series was written by Dr. Estela Bee Dagum, Head of the Research and Development Unit, General Time Series Staff.

The main purpose of this study is the identification and analysis of the basic hypotheses of different kinds of models for time series.

Emphasis is given to univariate models for which well established methods of estimation already exist. By discussing the assumptions of these models, it is hoped to make users and researchers aware of their limitations.

The reason for this study lies in the fact that 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 (optimality defined here in the usual statistical sense). 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 solution.

The author is solely responsible for the views expressed in this monograph; however, she would like to acknowledge with thanks, Professor C. Dagum, Ottawa University, Mr. Phillip Smith, Current Economic Analysis Division, and Miss Mary Lennox, Chief, General Time Series Staff for their valuable comments.

P. N. TRIANDAFILLOU
Director,
Current Economic Analysis Division.

TABLE OF CONTENTS

1. INTRODUCTION AND SUMMARY

During the last decade, many statisticians have been devoted to the search for optimal methods of estimation for time series. 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. Thus, the work done to obtain accurate and robust estimates for historical and current data has been centered on methods of estimation. 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.

By a model, we shall not mean a theoretical system lacking empirical content. On the contrary, it is defined here (See Dagum and Dagum, 1972) as: "an empirical-theoretical system which satisfies the following requirements:

(1) it must represent a non-contradictory but possible world;

(2) it must represent a world of possible experience; and,

(3) it must be submitted to and has to pass tests of either corroboration or falsification".

Following the idea of Wold (1938), already incorporated in the current literature, a time series is seen here as a sample realization of a stochastic process, which from a non-mathematical point of view, is any process controlled by probabilistic laws. The observations made as the process continues indicates the way it evolves. For each point of time t, belonging to a period T, that is, $t \in T$, the observation made $x(t)$ is considered the outcome or observed value of a random variable $X(t)$ and a family of such random variables $\{X(t), t \in T\}$ defines a stochastic process or random function.

Given that an observed time series $\{x(t), t \in T\}$ is assumed to be a 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.

Therefore, in order to analyze 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 are stochastic processes and there are several ways to classify a stochastic process.

We shall consider a classification that allows us to discuss the assumptions on which each type of process is based as well as to distinguish those processes used for practical applications from the more theoretical ones.

Table I

Classification of a Stochastic Process

(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 interval.	(1) Markovian (2) Non-Markovian

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 $t_1, t_2, \ldots, t_n$, but not on the actual values of these quantities; that is, if the set of points $t_1, t_2 \ldots, t_n$ is shifted along the time axis t, the probability distributions remain the same.

A stochastic process is completely specified by its probability distributions but this procedure is, in general, not convenient because of its unwieldiness.

In practice, instead of considering the probability functions, only the first two order moments are specified. A process is then said to be stationary in the wide sense or second order stationary if the mean value and the variance are constants and the covariance function depends only on the difference between any two points in time.

Any process for which the conditions of stationarity in the wide sense are not fulfilled is called non-stationary. There is an important subclass of non-stationary processes that has been extensively developed for practical applications, namely, the homogeneous non-stationary processes, also called processes with stationary increments. These processes are non-stationary but, by taking differences of a finite order, the processes become stationary in the differences. In a way, they are a generalization of the theory of the stationary random processes.

The second attribute which can also be used as the basis of a classification of random functions is the form of the distribution functions that specify the process. The distribution law most frequently encountered is the normal law. Also, normal processes can use certain methods of calculations which cannot be applied in other cases. Hence, it is advantageous to divide the stochastic processes in two groups: normal and non-normal. It is important to observe that if the distribution functions are normal, only the first two order moments are necessary for a complete description of a normal random process.

Finally, for a classification of stochastic processes, one can take into consideration the dependence of the behaviour of a process on its value in the preceding time interval. A stochastic process $\{X(t), t \varepsilon t\}$ is said to be

Markovian if for any set of n points $t < t_2 < \ldots < t_n$, the conditional distribution of $X(t_n)$ for given values of $X(t_1)$, $X(t_2), \ldots, X(t_{n-1})$ depends only on $X(t_{n-1})$, the most recent known value. In other words, the probability of any particular future behaviour of the process when its present state is known is not altered by conventional knowledge concerning its past behaviour. Conversely, if the probability properties of the process at a given movement of time depend on its values in preceding time intervals the process is said to be non-Markovian.

These three main categories are compatible and any combination among them is possible. For example, a process may be normal stationary of a Markovian type, or stationary and non-normal, or normal and non-stationary, and so on.

However, for practical applications, not all the combinations are possible. The main constraints are the lack of a well-developed theory and of proper methods of estimation to make their applications feasible. Therefore, we shall discuss the assumptions of those types of processes for which empirical applications are possible at present; namely, (1) Normal Non-Stationary (in the mean) processes; (2) Normal Linear Stationary Processes and (3) Normal Homogeneous Linear Non-Stationary Processes. A stochastic process $X(t)$ is said to be <u>linear</u> if there exists a purely random process U_t (i.e., a process of independent random variables identically distributed) and a sequence of parameters $\alpha_0, \alpha_1, \alpha_2, \ldots$ such that $X(t) = \sum_{k=o}^{\infty} \alpha_k U_{t-k}$.

Processes (1), (2) and (3) have proved to be the easiest to deal with from a mathematical point of view. Also, they seem to describe quite accurately the generating mechanism of many physical problems.

The properties that make these types of processes very useful are that, by the assumption of normality, they are completely specified with only the first two moments of the distribution functions; by being assumed stationary or stationary in the differences, (homogeneous non-stationary) their means and variances are constants and their autocovariance functions depend only on the time lags. For processes (2) and (3), the assumption of linearity guarantees the fulfillment of the principle of ergodicity.

The importance of stationarity in a process resides in the fact that the ergodic theorem and the spectrum were first defined for stationary stochastic processes. All linear stationary processes are ergodic; i.e., one can obtain consistent estimators of the mean and covariance function of a given process with only one single realization. In other words, averages computed from one sample (one set of data) called time averages can ultimately be identified with corresponding ensemble averages, i.e., averages over the whole space of outcomes or realizations of the process at a given moment. The ergodic property is extremely important when dealing with economic time series, the data of which are the result of a single experiment.

The spectrum is the Fourier transform of the autocovariance function and, from a mathematical point of view, they are equivalent. Both provide the same type of probability information for determining the generating mechanism of an observed time series but the autocovariance function stresses the time domain whereas the spectrum conveys the same information in the frequency domain.

The spectral representation of a stochastic process is done with the use of Fourier series or Fourier integral, depending on whether the time parameter is discrete or continuous. The periodic functions in Fourier analysis are sines and cosines. They have the important properties that an approximation of a given number of terms gives the minimum mean square error between the function and its approximation, and also, they are orthogonal, so the coefficients may be determined independently of one another.

The spectral representation of a stationary process is then a "decomposition" of the process into separate pair-wise uncorrelated periodic oscillations and the total variance of the process is distributed over frequency. For discrete time, it is possible to determine the proportion of variance attributable to each component with a particular frequency λ and for time continuous processes, one refers to the contribution of a band of frequencies λ .

Spectral analysis has been used preponderantly for solving problems that require frequency response studies like the design of aircraft structures, or for experiments to optimize the performance of industrial processes. However, when dealing with time series, particularly with economic time series, the autocovariance function is more appropriate for model building. However we must point out, following Mandelbrot (1972), that the autocovariance function analysis is effective primarily for models that correspond to near normal or Gaussian stochastic processes with high frequencies or small lags; i. e., for models of short run near Gaussian effects.

In our study, we shall stress the analysis of the autocovariance (autocorrelation) function but we shall also give the corresponding spectrum (normalized spectrum) for the models considered. The latter is done more for the sake of completeness than usefulness.

Within the broad categories of Normal Non-Stationary (in the mean) processes, Normal Linear Stationary processes and Normal Homogeneous Linear Non-Stationary processes, several kinds of models have been built for description and for forecasting of time series. Although the assumption of normality is not always made explicit in the formulation of the models, it is indeed introduced when testing their hypotheses. The following classification is useful for our purposes despite its simplicity and a certain amount of overlapping in its elements.

Table II

Classification of Models for Time Series

(1) Concerning the set of hypotheses made on the signal and the noise.	(a) Error Models	
	(b) Linear Stationary Processes	(b. 1) Autoregressive Processes
		(b. 2) Moving Average Processes
		(b. 3) Autoregressive Moving Average Processes
	(c) Homogeneous Linear Non-Stationary Processes	Autoregressive Integrated Moving Average Processes
(2) Concerning the number of parameters of the model.	(a) Parametric Models	
	(b) Non-Parametric Models	

In an error model the generating mechanism of a time series is assumed to be composed of a systematic component or signal which is a deterministic function of time (in general, a polynomial of low degree) and a random component supposed to be purely random or white noise, i.e. identically distributed with constant mean, constant variance and zero autocorrelation. This set of assumptions regarding the random component of the model forces any effect of time to be made in the deterministic part and therefore, each observation is stochastically independent of the previous ones. The properties of the time series are then summarized in the expected value of the function and variations in the random element do not affect these properties. In a way, the order in which the observations of the time series appears is of no relevance and the analysis of such a time series reduces to a multi-dimensional statistical analysis. The process is non-stationary but if the signal is a polynomial of time, it is then a homogeneous non-stationary process. In this case, by taking a finite number of differences of the observations, the process is linear stationary in the differences, the order of which is given by the degree of the polynomial.

Historically, these models were elaborated first in the research domain of Astronomy. They were built to determine the position of a planet at a given moment of time. In such a case, errors in the observed time series are attributed to errors of observation due to atmospheric conditions or imperfections of the telescope. The errors will not affect later positions of the planets nor our observations of them.

These models are then acceptable when the errors can be attributed to errors of measurement. However, for many other time series, in particular economic time series, the errors are not only due to wrong observations but also to more serious irregularities, and once they appear, they are incorporated into the process and influence its future evolution. In other words, the errors are autocorrelated; the order of the observations is crucial and the current value of X_t will depend on the time which has elapsed since the process started.

For such cases, within the categories of normal linear stationary and of normal homogeneous linear non-stationary, several models were built. Among them, the Autoregressive-Moving Average, ARMA (p, q) is very useful for description and forecasting when the time series is assumed to follow a stationary normal or near normal process with few lags. Such a process is obtained by equating an autoregressive process of order p with a moving average process of order q. A process X_t is said to be an autoregressive process of order p if it can be expressed as a weighted sum or linear combination of the p previous values of the process plus a purely random component U_t.

On the other hand, a process X_t is said to be a moving average process of order q if it can be expressed as a weighted sum or linear combination of purely random variables U_t, U_{t-1}, ..., U_{t-q}.

An ARMA (p, q) process is then considered to be the output X_t obtained from a purely random input U_t summitted to a linear transformation with weights resulting from the quotient of two polynomials.

If the generating mechanism of a time series departs from the stationarity assumption only by observed differences in the level and/or the slope of different parts of the series, the process is then considered homogeneous non-stationary. Given the assumption of autocorrelated errors, the models used in these cases are called Autoregressive Integrated Moving Averages, ARIMA (p, d, q) models, where p stands for the order of the autoregressive process, d, for the order of the differences and q for the order of the moving average process. The assumptions of normality or near normality and of short run dependence are implicit in these models. The ARMA(p, q) and ARIMA(p, d, q) have been used for forecasting with quite good success when p, d, and q are no larger than 2.

It is traditionally assumed in the analysis of economic time series that the observed series results from the superposition of four processes that represent different types of evolution; namely, (a) trend, (b) cycle, (c) seasonal and (d) irregular.

The trend corresponds to a variation in some defined sense persisting over a long period of time, that is, a period which is long in relation to the cycle. The cycle is a movement of quasi-periodic appearance characterized by alternating periods of expansion and contraction. The seasonal movement corresponds to regular weekly, monthly or quarterly variations (repetitive intra-annual fluctuations) caused by the climate and other institutional practices.

The irregulars are unforeseeable movements connected with events of all kinds. In general, they have a stable random appearance. In much analytical work, the trend and cycle components are combined, since it is the standing of the series apart from seasonal variation upon which interest is centered.

The presence of a trend-cycle element introduces non-stationarity in the stochastic process. It is usually considered as a deterministic function of time (in general, a polynomial of a low degree) in the error models and as a function subjected to stochastic changes in the ARIMA models.

The seasonal movement plays a very important role in the analysis of economic time series. We have devoted Section 10 to the models that are usually built to describe the seasonality. We must point out, however, that the seasonal movement very rarely occurs on its own without being superimposed on a more or less regular trend. Besides, like all the other components of an economic time series, it is not an observable process. If the seasonal variation is assumed

to be stable and deterministic, it is represented by a strictly periodic function of time, usually a polynomial of sines and cosines of time with both constant amplitudes and phases. But if the seasonal movement is assumed stable and stochastic, then the amplitudes are described by a purely random process. When the seasonal pattern is evolving through time in a deterministic manner, the amplitudes are a function of time, while if it is assumed to change stochastically then the amplitudes follow a stationary stochastic process, not purely random.

In the error models, the seasonal variation is treated as deterministic while in the ARIMA models, it is treated as stochastic.

The irregular are assumed to follow a purely random process or sometimes, less restrictively, a non-autocorrelated process.

The error models, the ARMA (p, q), the ARIMA (p, d, q) and the seasonal models are parametric, i. e., a finite number of parameters are involved. Other ways to describe the probability structure of a time series are non-parametric models, so called because an infinite number of parameters are considered. This classification of parametric and non-parametric models for time series should not lead to confusion with the use of the words parametric and non-parametric in statistical analysis. In the latter, a method is said to be parametric if it is assumed that the random variable follows a given distribution, and is said to be non-parametric if the distribution is free.

The non-parametric models we discuss here are the autocovariance function and its Fourier transform, the power spectrum. The study of any one of these functions is very useful for exploratory purposes, in the early stage of a research when very little is known about a particular phenomenon. The

main purpose of the time series analysis at this stage is to look at the data in different ways to see what hypotheses may be suggested and an exploratory model would then help to build an appropriate parametric model.

Finally, we want the reader to bear in mind that the choice of the appropriate model for a time series is always function of both the a priori knowledge we may have of the nature of the phenomenon considered and the purpose of the analysis. But for immediate empirical applications, this choice is also constrained to the existence of numerical methods of solution readily mechanized.

2. A TIME SERIES AS A SAMPLE REALIZATION OF A STOCHASTIC PROCESS

From a non-mathematical point of view, a stochastic process is any process controlled by probabilistic laws. The observations made as the process continues indicate the way it evolves.

In most cases we are interested in temporal variation and the state of the process at any time t is described by the values of a certain number of observable quantities which are random variables.

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

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

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 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

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

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

The family of all these joint 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

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

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 subjected to the same permutation.

(b) The compatibility condition according to which

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

$$F_{t_1, t_2, \ldots, t_m}(x_1, x_2, \ldots, x_m)$$

for any $t_{m+1}, \ldots, t_n$ if $m<n$.

In other words, the compatibility condition requires that

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

It was proved by a celebrated theorem by Kolgomorov (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) (2.7) $X = \{X(\omega, t),\ \omega \in \Omega,\ t \in T\}$.

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 infinite set of functions which might have been generated by the stochastic process.

The feature of time series analysis which distinguishes it from other statistical analyses is the explicit recognition of the order 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 this study, 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 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.

3. STOCHASTIC PROCESSES MODELS: STATIONARY AND NON-STATIONARY PROCESSES

The most important 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 functions. The lower moments include the mean, the variance and the covariance function or its Fourier transform, the spectrum. An alternative approach to the above is that a stationary stochastic process can be also adequately described by a model with few parameters.

We shall first define what is meant by a stationary stochastic process.

From an intuitive point of view, a process is said to be stationary if it is in statistical equilibrium, in the sense that its properties do not change with time. In other words, it is a process without trend and all its properties can be summarized by computing certain functions from the data. (Of these functions, the autocorrelation function was the first to be studied.)

From a statistical point of view, a stochastic process $X = \{X(\omega, t), \omega \in \Omega, t \in T\}$ is defined as 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 axis 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), the (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 $t - t_j$ $(j = 2, 3, \ldots, n)$.

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,

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

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

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

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

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

setting $T = -t_1$, gives

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

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

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

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.

The simplest example of a strictly stationary process is the purely random process or white noise in which the random variables are assumed serially independent and identically distributed. This is the assumption underlying the random component used in error models which are non-stationary processes.

Any process for which conditions (i) and (ii) are not fulfilled defines a non-stationary stochastic process. Of the class of non-stationary processes, the so-called homogeneous non-stationary or processes with stationary increments were first studied by Kolgomorov (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 Kolgomorov (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,

$$(3.7)\quad D(\tau_1,\tau_2) = E[X(t+\tau_1) - X(t)]\,[X(t+\tau_2) - X(t)]$$

For the real case, instead of (3.7), we have a function of one variable,

$$(3.8)\quad D(\tau) = E\left[X(t+\tau) - X(t)\right]^2$$

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

$$(3.9)\quad D(\tau) = 2\sigma(0) - \sigma(\tau) - \sigma(-\tau) = 2\sigma(0) - 2\sigma(\tau)$$

The error models which we shall discuss in Section 4 are homogeneous non-stationary processes when the deterministic component (the signal) is assumed to be a polynomial of time. By taking successive differences, the process is reduced to a linear stationary process. Another important category of stochastic process is the homogeneous linear non-stationary process. The property of linearity here comes from the fact that the process is seen as the output from a linear filter, the input of which is white noise or purely random.

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. A simple example is given by a process called random walk in which, as the time increases, the random variables tend to oscillate about its 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. For example, let us assume that W_t is a purely random process and X_t is another process related to W_t as follows:

$$X_1 = W_1$$

$$X_2 = X_1 + W_2$$

$$\vdots$$

$$X_t = X_{t-1} + W_t$$

Then X_t can be expressed as a linear combination of the purely random process W_t with all the weights equal to 1. If the expected value of W_t is μ and the variance σ^2_W, it follows that:

$$(3.7)\quad E(X_t) = t\,\mu$$

and $$(3.8)\quad \text{var } X_t = t\,\sigma^2_W$$

The autocovariance of the X_t process is

$$(3.9)\quad \sigma_{XX}(t_1, t_2) = \min(t_1, t_2)\,\sigma^2_W$$

4. ERROR MODELS

The first studies on time series were done in the research domain of Astronomy. Models were built to determine the position of a planet at a given moment in time. The nature of the problems encountered by astronomers led to the building of very simple models which are today known as error models or simply, as regression models (Pannekoek, 1961).

Error models assume that successive observations of a time series are independent. The ordered set $\{x_t,\ t = \ldots -1,\ 0,\ 1 \ldots\}$ (1) can be expressed by a general model composed of a completely deterministic function of time f(t) which is called the signal or systematic part of the model and a random element U_t called white noise (or purely stochastic component) of the process which is assumed to be independent of f(t). The model can be written as follows:

$$(4.1)\quad X_t = f(t) + U_t \qquad t = \ldots -1, 0, 1, \ldots .$$

The assumption of independence among successive observations is introduced into the set of hypotheses that characterize the random component U_t which is supposed to be identically distributed with zero mean, constant variance at each point of time and zero autocorrelation. These specifications force any effect of time to be made in the signal f(t).

In some cases, f(t) is a known function of time or other observable quantities and of parameters. If f(t) is assumed to be linear in the parameters, then it reduces to a "regression function" and the estimation of its parameters is made with least squares methods.

(1) We shall use lower case letters when referring to observed quantities and upper case letters for the corresponding stochastic variable. We shall write t as a subindex when time is discrete, and as an argument when time is continuous.

When the signal is not linear in its parameters, as for example in growth functions, then the estimation of the parameters and the testing of the hypotheses require previous transformations or more complicated procedures.

In other cases, the analytical form of the systematic component is not known but can be fairly approximated by linear combinations of known functions of time, such as powers of time t or trigonometric functions of t.

In general, two types of functions of time may be distinguished for f(t). One is a polynomial of fairly low degree which fulfills the assumption that the systematic component moves slowly, smoothly, and progressively through time. The other, is a linear combination of sines and cosines of time with constant coefficients (a finite Fourier Series) which takes into account cyclical fluctuations, strictly periodic or not.

For economic time series, the signal f(t) very often follows a pattern of behaviour that requires the addition or multiplication of both types of function. The polynomial is then said to represent a long-term trend and the finite Fourier Series stands for cyclical fluctuations and seasonal variations.

In all cases mentioned above, the estimation of the parameters of the signal is done with the techniques of regression analysis (classical least squares theory) or analogies of them. Under the assumption that the random variables are normally distributed, all tests of hypotheses can be performed.

However, there are cases where the signal can not be approximated by a simple time function for all the time range although it behaves smoothly and does not fluctuate greatly in any small interval of time. The usual procedure is then to use non-parametric methods for smoothing.

The smoothing of a sequence x_t means going over to a new sequence x_t^* obtained from the original sequence by forming the moving average,

$$(4.2)\quad x_t^* = \sum_{-n}^{n} a_k x_{t+k} \qquad t=n+1, \ldots, T-n$$

For simplicity, we have written a finite sum but, if certain convergence conditions are met, we can extend the sum to ∞.

Also, the a_k's are usually normalized so $\sum_{-n}^{n} a_k = 1$. In practice, smoothing is widely used to filter sequences, in order to diminish the effects of measurement errors and other weakly correlated disturbances.

Assuming we have a finite number of observations in model (4.1), after smoothing we obtain,

$$(4.3)\quad x_t^* = \sum_{-n}^{n} a_k f(t+k) + \sum_{-n}^{n} a_k u_{t+k} \qquad t=n+1, \ldots, T-n$$

where $E(u_t)=0$, $E(u_t^2)=\sigma_u^2$ and $E(u_t u_s)=0$, $t \neq s$; Then, u_t^* has variance $\sigma_u^2 \sum_{-n}^{n} a_k^2$.

The mean square of u_t^* isgreatly decreased if the smoothing weights a_k are properly chosen whereas the signal f(t) is approximately the same. Hence, the weighted average of the observed values are used to estimate the trend and the rather irregular curve generated by the observed points is replaced by a smooth curve of the moving average. However, successive terms in the smoothed series are correlated. We have,

$$E(u_t^* u_{t+h}^*) = \begin{cases} \sigma_u^2 \sum_{-n}^{n} a_k a_{k-h} & ; \quad h = 0, 1, \ldots, 2n \\ 0 & ; \quad h = 2n+1, \ldots \end{cases}$$

One should observe that the assumption of smoothness is a local property whereas the assumption of a polynomial trend concerns the entire time range t = 1, 2, . . . , T. The assumption of smoothness uses only observations near a given point in time t to estimate the trend at that point, while the assumption of a polynomial trend implies that all observations are used to estimate the function that represents the trend over the entire time range considered.

The general basis for most smoothing procedures is to fit a polynomial(1) to 2n±1 successives observations and use this fitted polynomial to estimate the trend at the middle value. Since the estimates of the parameters of the polynomial are linear in the observed values x_{t+k}, the estimate of the trend has the form (4.2).

Suppose that the trend f(t+k) at time points t+k = t − n, ... , t+n, can be approximated by a polynomial on k of degree p, that is,

$$(4.4)\quad f_t(k)=\alpha_o+\alpha_1 k+\alpha_2 k^2+\ldots+\alpha_p k^p;\ k=-n,\ \ldots,\ n$$

then for k = 0, f(t) is approximately $f_t(0)=\alpha_o$.

We can estimate the α's in terms of the observed values $x_{t-n}, \ldots, x_{t+n}$ using the method of least squares. The normal equations for the estimates of $\hat{\alpha}_o, \hat{\alpha}_1, \ldots, \hat{\alpha}_p$ are

$$(4.5)\quad \hat{\alpha}_o \sum_{-n}^{n} k^j+\hat{\alpha}_1 \sum_{-n}^{n} k^{j+1}+\ldots+\hat{\alpha}_p \sum_{-n}^{n} k^{j+p}=\sum_{-n}^{n} k^j x_{t+k}$$

$$J = 0,\ 1,\ \ldots,\ p$$

Note that the sums Σk^j are functions of n only.

---

(1) A similar procedure rests on the use of a formula derived by fitting a harmonic function such as,

$$f(t)=\alpha_o+\sum_{J=1}^{p}[\beta_j \cos \omega_j k+\gamma_j \sin \omega_j k]$$

where p and the angles ω_j have to be chosen a priori and the fitting determines the coefficients α_o, β_j and γ_j

By symmetry, the sum of any odd power of k is zero. Since our estimate of f(t) will be $\hat{\alpha}_o$ we are interested only in the equations for j even, namely,

$$(4.6)\quad \hat{\alpha}_o \sum_{-n}^{n} k^j + \hat{\alpha}_2 \sum_{-n}^{n} k^{j+2} + \ldots + \hat{\alpha}_p \sum_{-n}^{n} k^{j+p} = \sum_{-n}^{n} k^j x_{t+p} \qquad j = 0, 2, 4, \ldots, p$$

The equations to be solved for $\hat{\alpha}_0$ are the same for p odd as for the next lower even value of p. We need to consider only p= 0, 2, 4, . . .

The system (4.6) can be split in two. One equation for j = 0 and a system of equations for j = 2, 4, . . . , p. Then (4.6) becomes

$$(4.7)\quad (2n+1)\hat{\alpha}_o + 2\sum_{1}^{n} k^2 \hat{\alpha}_2 + \ldots + 2\sum_{1}^{n} k^p \hat{\alpha}_p = \sum_{-n}^{n} x_{t+k}; \qquad j=0$$

$$2\sum_{1}^{n} k^j \hat{\alpha}_o + 2\sum_{1}^{n} k^{j+2} \hat{\alpha}_2 + \ldots + 2\sum_{1}^{n} k^{j+p} \hat{\alpha}_p = \sum_{1}^{n} k^j (x_{t-k} + x_{t+k})$$

$$j=2, 4, \ldots, p$$

The solution of (4.7) for $\hat{\alpha}_o$ is

$$(4.8)\quad \hat{\alpha}_o = \sum_{-n}^{n} a_k x_{t+k}$$

Where $a_{-k} = a_k$ and the a_k's are functions of n and p only and are polynomials in k. Then the estimation of the trend x_t^* given by (4.2) is $x_t^* = \hat{\alpha}_o$ as determined by (4.8).

Thus the process of fitting a polynomial by the moving average method consists of determining the weights $\underline{a}$ and calculating for each consecutive set of $2n+1$ terms in the series, a value $\hat{\alpha}_o$ given by (4.8).

It can be proved that, for a given p, the variance of the smoother sequence decreases as the number of observations or points are increased and that for a given number of points (n), the variance increases with increasing p. The selection of p and n that optimize the fitting is a statistical multiple decision problem (Anderson, T.W.,1971).

The advantage of smoothing to estimate the signal is its flexibility. However, since the method is not based on an explicit probability model it cannot be treated fully in terms of mathematical statistics and statistical inference is severely limited. For example, f(t) is not determined by a small number of parameters for which a confidence region can be given. One cannot perform either hypothesis testing about the trend or directly relate the estimated trend to a theory or model for the generation of the observed series.

We shall not pursue any further this topic which is concerned with the problem of estimation of the signal and not with the probabilistic model that generates the time series. We proceed to summarize the main properties of error models introduced in this section. Error models for time series assume that the observed series is made up of a well known function of time and an error component which is white noise. Therefore, the observations are serially independent and the effect of time is not incorporated in U_t but is assumed to affect only f(t). The analytical form given to f(t) is usually a linear combination of functions of time, such as powers of time t or sines and cosines functions of t. The estimation of the trend is done either by the least squares method or by a smoothing technique like the moving averages.

These kinds of models are acceptable when the errors in the observed time series can be attributed only to errors of measurement or of observation. In that case, any discrepancy is regarded as purely temporary, with no effect on the future motion of the process. This seems to be the case in many of the problems of Astronomy. For example, the errors in the observed position of the planets at a moment of time t could be attributed to errors of observations due to the atmospheric conditions or imperfections of the telescope. Such error will not affect future positions of the planets nor our observations of them. The properties of the time series are then summarized in the mean value of the signal, and variations in the random component do not affect these properties.

When the errors are autocorrelated, the error models are no longer representatives of the time series generating process. The observations are now stochastically dependent in time and the current value of t will depend on the time which has elapsed since the process started. The interpretation given to autocorrelated errors is that the errors are not simply due to incorrect independent observations, but to other systematic forces and, once they appear, they are incorporated in the process and influence its future evolution. This seems to be the case for most time series, in particular social and economic time series. The discovery of this aspect has been attributed to Yule (1921) (1927), Processes that take into consideration the assumption of autocorrelated errors are the Linear Stationary and Homogeneous Linear Non-Stationary Processes.

We deal with these types of models in the following sections.

5. LINEAR STATIONARY STOCHASTIC PROCESSES

A stochastic process X_t is said to be linear if there exists a purely random process U_t and a sequence of parameters $\alpha 0, \alpha 1, \alpha 2, \ldots$ such that,

$$(5.1)\quad X_t - m = U_t + \alpha_1 U_{t-1} + \alpha_2 U_{t-2} + \ldots = \sum_{k=0}^{\infty} \alpha_k U_{t-k};\ \alpha_o = 1$$

In continuous time, the relation (4.2.1) becomes,

$$(5.2)\quad X(t) - m = \int_o^{\infty} \alpha(v)\, U(t-v)\, dv.$$

The $E(U_t) = E[U(t)] = 0$ and $E(X_t) = E[X(t)] = m$

If the series (or the integral) is convergent, then 5.1 (or 5.2) defines a stationary linear stochastic process, where m is the expected value of the process.

If the series (or the integral) is not convergent, then (5.1) or (5.2) defines a non-stationary linear process and m has no precise meaning, except as a point of reference for the level of the process. The (5.1) is sometimes called an infinite moving average although the infinite sum of the coefficients is not constrained to be equal to 1.

For a finite number of terms, (5.1) defines a linear parametric stochastic process with a discrete time parameter.

In this section we shall deal only with discrete time parameter processes. (The analysis for continuous time follows a similar procedure.)

Using the lag operator L, where $L^o = 1$, $LU_t = U_{t-1}$ and $L^n U_t = U_{t-n}$, the (5.1) for $m = 0$ can be written in compact form as follows:

$$(5.3)\quad X_t = (1+\alpha_1 L+\alpha_2 L^2+\dots)U_t = \sum_{k=0}^{\infty} \alpha_k L^k U_t = \alpha(L)\, U_t$$

where $\sum_{k=0}^{\infty} \alpha_k L^k = \alpha(L)$ is the linear filter or operator that transforms the input U_t into the output X_t. It is also called the transfer function of the linear system or generating functions of the weights. In the latter, L can be considered as a dummy variable whose k-th power is the coefficient of α_k.

For (5.3) to be stationary in the wide sense, it is a necessary and sufficient condition that the transfer function ∂ (L) be convergent for $|L| \leq 1$ which implies that $\sum_{k=0}^{\infty} |\alpha| < M$, M finite.

5.1 Autoregressive Processes (AR)

Autoregressive processes were first introduced by U. Yule (1921) and are a subclass of linear processes. A process X_t is said to be an autoregressive process of order p, AR(p) if it can be expressed as a linear combination of p previous values of the process plus a purely random component U_t. It can be written in the following form

$$(5.1.1)\quad X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_p X_{t-p} = U_t$$

Using the lag operator L, (5.1.1) results in,

$$(5.1.2)\quad U_t = (1+\beta_1 L + \beta_2 L^2 + \dots + \beta_p L^p)\, X_t = \sum_{r=0}^{p} \beta_r L^r X_t ; \quad \beta_0 = 1$$

If the X_t process is stationary then (5.1.2) can be written as an infinite moving average;

$$(5.1.3)\quad X_t = \frac{1}{\sum_{r=0}^{p} \beta_r L^r}\, U_t = \beta^{-1}(L)\, U_t = \sum_{r=0}^{\infty} h_r L^r U_t$$

In effect, each X_{t-l}, l=1, .., p can be replaced by an autoregressive process and finally we end up with an infinite series of U's.

There are several equivalent conditions for the stationarity of (5.1.3). Written as an infinite moving average, we already said in the previous sections that the infinite series of weights must be convergent for $|L| \leq 1$, which implies $\sum_{r=o}^{\infty} |h_r|$ finite.

These conditions for convergence are equivalent to $\sum_{r=o}^{p} \beta_r L^r \neq o$

since if $\sum_{r=o}^{p} \beta_r L^r = o$, then the transfer function becomes infinite.

We shall see which are the conditions for $\sum_{r=o}^{p} \beta_r L^r \neq o$

$$(5.1.4) \quad \sum_{r=o}^{p} \beta_r L^r = (1 + \beta_1 L + \beta_2 L^2 + \ldots + \beta_p L^p) = \beta(L)$$

Observe that $\beta(L)$ is a polynomial in L of order p. Factoring out L^p and making $L = \frac{1}{G}$ we have

$$(5.1.5) \quad \beta\left(\frac{1}{G}\right) = \frac{1}{G^p}\left(G^p + \beta_1 G^{p-1} + \beta_2 G^{p-2} + \ldots + \beta_p\right) =$$

$$\frac{1}{G^p}\left(G-G_1\right)\left(G-G_2\right) \ldots \left(G-Gp\right)$$

Where G_1, G_2, ..., G_p are the p roots of the characteristic equation of $\beta(\frac{1}{G})$.

It is obvious that the (5.1.5) can also be written as

$$(5.1.6) \quad \beta(L) = (1-G_1L)(1-G_2L) \ldots (1-G_pL)$$

Therefore, if there is a $G_i = \frac{1}{L}$, $\beta(L) = 0$, the process (5.1.3) is non-stationary. For it to be stationary, the roots of $\beta(L)$, namely $L_i = G_i^{-1}$, $i = 1,2, \ldots, p$ must all be in absolute value greater than 1 or equivalently $|G_i| < 1$. The conditions $|G_i| < 1$ coincide with the requirements for the stability of the deterministic component of (5.1.1). Observe that if $U_t = 0$, then (5.1.1) reduces to a homogenous difference equations of order p, whose general solution is

$$(5.1.7)\quad X_t = A_1 G_1^t + A_2 G_2^t + \ldots + A_p G_p^t$$

Where G_i, $i = 1,2,..,p$ are the roots of its characteristic equation and for (5.1.7) to be stationary, $|G_i| < 1$, for $i = 1,2,..,p$.

For a continuous time parameter, the autoregressive process (5.1.1) is defined by a differential equation of order p and to be stationary, the roots of its characteristic equation must all have negative real parts.

5.2 Finite Moving Average Processes (MA)

The autoregressive model discussed above, expressed X_t an a finite weighted sum of p previous values of the process X_t, plus a purely random component U_t. Equivalently, if the process is stationary, it can be expressed as an infinite weighted sum of the U_t's, where $E(U_t) = 0$, $E(U_t)^2 = \sigma_U^2$ and $E(U_t\, U_{t+\tau}) = 0$ for all $\tau \neq 0$.

Now a process X_t is said to be a finite moving average of order q, MA(q) if it is a linear combination of purely random variables U_t, U_{t-1}, .., U_{t-q}. That is,

$$(5.2.1)\quad X_t = U_t + \alpha_1 U_{t-1} + \alpha_2 U_{t-2} + \ldots + \alpha_q U_{t-2} = \sum_{r=o}^{q} \alpha_r U_{t-r};\ \alpha_o = 1.$$

Using the lag operator L, (5.2.1) becomes,

$$(5.2.2)\quad X_t = \sum_{r=0}^{q} \alpha_r L^r U_t = \alpha(L)U_t$$

Where $\alpha(L) = 1 + \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_q L^q$, is the generating function of the weights.

The (5.2.1) is always stationary, that is $\alpha(L)$ is convergent for $|L| \leq 1$ but for (5.2.1) to be invertible i.e. to be expressed as an infinite autoregressive process,

$$(5.2.3)\quad U_t = \left(\sum_{r=o}^{q} \alpha_r L^r\right)^{-1} X_t = \alpha^{-1}(L)X_t = \sum_{r=o}^{\infty} \delta_r L^r X_t$$

the roots of $\alpha(L) = o$ must lie outside the unit circle.

5.3 Autoregressive-Moving Average Processes (ARMA)

For empirical applications, a combination of an autoregressive process, say of order p with a moving average, say or order q, has the advantage of involving very few parameters.

Thus, an ARMA (p,q) process can be written as,

$$(5.3.1)\quad X_t = U_t + \alpha_1 U_{t-1} + \ldots + \alpha_q U_{t-q} - \beta_1 X_{t-1} - \ldots - \beta_p X_{t-p}$$

or equivalently

$$5.3.2)\quad \left(\sum_{r=o}^{p} \beta_r L^r\right)X_t = \left(\sum_{r=o}^{q} \alpha_r L^r\right)U_t; \qquad \alpha_o = \beta_o = 1$$

and therefore,

$$(5.3.3)\quad X_t = \frac{\sum_{r=o}^{q} \alpha_r L^r}{\sum_{r=o}^{p} \beta_r L^r} U_t$$

The ARMA (p,q) process is considered as the output X_t obtained from an input U_t purely random or white noise, where the transfer function is the quotient of two polynomials. The number of parameters of the model (5.3.1) is p+q+2, including the mean of X_t and the variance of U_t. In most of the cases solved with this model p and q are no larger than 2 (Box and Jenkins, 1970).

For the ARMA (p,q) process (5.3.3) to be stationary the same conditions discussed in the previous sections are required namely, the roots of the characteristic equation $\sum_{r=o}^{p} \beta_r L^r = \beta(L) = 0$ must all be in absolute value greater than 1. The (5.3.3) is invertible if $\sum_{r=o}^{q} \alpha_r L^r = \alpha(L) = 0$ has all its roots outside the unit circle. Then,

$$(5.3.4)\quad U_t = \frac{\sum_{r=o}^{p} \beta_r L^r}{\sum_{r=o}^{q} \alpha_r L^r} X_t$$

6. NON-PARAMETRIC STATIONARY STOCHASTIC MODELS: THE AUTOCOVARIANCE FUNCTION AND THE POWER SPECTRUM

The models previously discussed are all parametric, that is they have a finite number of parameters. Another way to describe the generating process of a stationary time series is by means of non-parametric models - models with an infinite number of parameters. Among the non-parametric approaches, the analysis of the autocovariance and autocorrelation functions and their Fourier transforms, the power spectrum and the normalized spectrum are the most relevant.

From a mathematical point of view, these functions are Fourier pairs and consequently, they are equivalent. Both provide the same type of probability information, in the sense that both characterize all the second order moments of a stationary stochastic process. The use of either the autocovariance function analysis or the spectral analysis depends on the particular properties of the data that one needs to stress. The latter stresses the frequency domain, whereas the former conveys the same information in the time domain.

In the spectral representation, a stationary process is seen as a linear combination of random oscillatory components and the total variance is distributed over frequency. If the process is defined for discrete time parameter, it is possible to determine the proportion of variance attributable to each component with a particular frequency λ, but for time continuous processes, we refer to the contribution of a band of frequencies around a particular λ.

An important use of the autocorrelation function and the normalized spectrum is to permit the identification of linear filters that minimize the mean square error when the systematic component of a process if corrupted by a purely random component.

Both functions are also very useful as initial guides in constructing a probability model for the mechanism which has generated the time series. Thus, for example, an autocorrelation function that is positive for successive values of τ (time lag) and tends to zero as τ increases, will reflect both a smooth behaviour of the time series and the fact that the process is more a finite autoregressive type than purely random. The order of the autoregressive process can also be obtained from the partial autocorrelation function. The same information is given in the frequency domain by a normalized spectral density function (normalized spectrum) with predominancy of low frequencies. On the other hand, when adjacent values of $\rho(\tau)$ are negatively correlated, the process generating the time series will show a great fluctuation for short periods of time and the corresponding normalized spectral density function will have predominancy of high frequencies.

However, although important for model building (especially in engineering and physics) spectral analysis has shown to be more relevant in frequency response studies and in the area of design of experiments to optimize the performance of industrial processes.

In the analysis of economic time series the first non-parametric approaches were based on the autocovariance function. At the end of the fifties and during the decade of the sixties, the latter was almost abandoned

and spectral analysis became fashionable. Very recently, the autocovariance function began regaining acceptance, mostly because of the availability of new computer programs (Box and Jenkins, 1970).

The autocovariance function of a stationary process X_t is by definition,

$$(6.1)\quad \sigma_{XX}(\tau) = E(X_t\, X_{t+\tau}) \qquad \tau = .., -1,0,1,...$$

where τ is the time lag, assumed here to be an integer. If the time parameter of the process is continuous, then τ can assume any value between $\pm\infty$.

$$(6.2)\quad \rho_{XX}(\tau) = \frac{\sigma_{XX}(\tau)}{\sigma_{XX(o)}} = \frac{\sigma_{XX}(\tau)}{\sigma_X^2}\ ; \qquad \tau = .., -1,0,1,...$$

Observe that,

$$(6.3)\quad \sigma_{XX}(0)\ \rho_{XX}(\tau) = \sigma_{XX}(\tau)$$

and therefore, if we know the autocorrelation function and the variance of the process X_t we have all the information provided by the autocovariance.

The graph of (6.2) is also known as the correlogram. The basic properties of the autocorrelation function for a real process are: (We shall suppress the subindex X to abbreviate the notation):

(1) $\rho(0)=1$

(2) $\rho(-\tau) = \rho(\tau)$ The function is symmetric with respect to the origin because of the stationarity assumption and therefore it needs only to be calculated for positive lags.

(3) $|\rho(\tau)| \leq 1$. This is a consequence of the fact that the variance of a random variable or of linear combination of random variables is positive.(1)

(4) The autocorrelation matrix is positive semi-definite. That is, the determinant of the autocorrelation matrix and all its principal minors are positive or zero. Property 4 is a generalization of property 3 and shows that the autocorrelation function is always positive semi-definite. The converse is also true; that is, every positive semi-definite function of a real (or integral) argument is the autocorrelation function of a continuous (discrete) stochastic process. (This was proven by Khinchin and Kolmogorov, see Yaglom, 1962).

---

(1) In effect, assume $Y_t = \lambda_1 X_t + \lambda_2 X_{t-\tau}$ then the variance of Y_t is

(1) var. $Y_t = \lambda_1^2$ var. $X_t + \lambda_2^2$ var. $X_{t-\tau} + 2\lambda_1\lambda_2$ cov. $(X_t X_{t-\tau})$

The right member is non-negative for all λ_1, λ_2 real and the second member is a quadratic form in λ_1, λ_2. For it to be positive, its roots must be imaginary, which implies

(2) var. X_t var. $X_{t-\tau} \geq [\text{cov}(X_t X_{t-\tau})]^2$

or equivalently

$$(3)\quad \rho^2(X_t, X_{t-\tau}) = \frac{[\text{cov}\ (X_t X_{t-\tau})]^2}{\text{var}\ X_t\ \text{var}\ X_{t-\tau}} \leq 1$$

For a stationary process, (3) reduces to,

$$|\rho(\tau)| = \left|\frac{\sigma(\tau)}{\sigma(o)}\right| \leq 1$$

(5) If the process is continuous, then $\rho(\tau)$ is defined for τ taking values between $+\infty$ and $-\infty$ and it is a necessary and sufficient condition that the function be continuous at $\tau = 0$ since this implies continuity everywhere (Yaglom, 1962). If the process is assumed to be purely random, this continuity property poses problems.

For a discrete purely random process U_t, the autocorrelation function is $\rho(0) = 1$ and $\rho(\tau) = 0$, for all $\tau \neq 0$. If U_t is a time continuous process we would have a discontinuity at $\tau = 0$. To avoid this, the autocovariance function is redefined as

$$(6.4)\quad \sigma_{UU}(\tau) = \sigma(0)\ \delta(\tau) = \sigma_U^2\ \delta(\tau)$$

Where $\delta(\tau)$ is a Dirac delta or impulse function(1), interpreted as zero for $\tau \neq 0$ and infinite for $\tau = 0$, then the covariance between neighbouring points is zero but at the expense of making the variance of the process infinite (Jenkins and Watts, 1969).

Since any analytical function, periodic or not, can be approximated to any degree using any class of periodic functions, the spectral representation of a stochastic process can be done using Fourier series or Fourier integral, depending on the time parameter being discrete or continuous. In Fourier analysis, the periodic functions are sines and cosines. They have the important properties that an approximation of

(1) A delta function is defined as a sequence of function $\delta_n(t)$ such that $\int_{-\infty}^{\infty} \delta_n(t)\, dt = 1$, for every n and in the limit as n tends to ∞

$$\delta(t)\begin{cases} = 0 & t \neq 0 \\ = \infty & t = 0 \end{cases}$$

a given number of terms gives the minimum mean square error between the function and its approximation, and also that they are orthogonal, so the coefficients may be determined independently of one another.

The use of Fourier series to describe phenomena evolving through time was suggested in several studies by Lagrange (1772-78), Buys-Ballot (1847) and Stokes (1879), but the best known work was the periodogram method used by Schuster (1898) in the search of hidden periodicities in sunspot data. In economics, the periodogram was used by Moore (1914) and by Beveridge (1922). The use of the periodogram to describe time series failed because of the assumptions of fixed amplitudes, frequencies and phases in the Fourier components. The modern spectral analysis uses the Fourier series (or Fourier integral) assuming that the amplitudes and phases are random variables.

It is shown (Yaglom, 1962) that every stationary stochastic process $X(\omega,t)$ can be approximated by a linear combination or harmonic oscillations of form

$$(6.5)\quad X(\omega,t) = X_k(\omega)f(t)=X_k(\omega)Re^{i(\lambda_k t+\theta)}$$

where $X_k(\omega)$ is a time independent random variable and $f(t)$ is a numerical function of t. The numerical factor $Re^{i\theta}$ can be included in the random variable $X_k(\omega)$ and the product $X_k(\omega)Re^{i\theta}$ will be simply denoted here by X_k; then (6.5) becomes,

$$(6.6)\quad X_k(t) = X_k\, e^{i\lambda_k t}$$

where X_k is a complex random variable with mean value zero, and λ is a constant.

Then, each component of the form (6.6) describes a periodic oscillation of angular frequency λ_k(1), with random amplitude R and random phase θ.

If the process is defined for a discrete time parameter, we can represent it by

$$(6.7)\quad X_t = \sum_{k=1}^{\infty} X_k e^{i\lambda_k t}$$

and for continuous time parameter by,

$$(6.8)\quad X(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dZ(\lambda)$$

where $Z(\lambda)$ is a stochastic process indexed on λ.

The (6.7) is the spectral representation of a stationary process with a discrete spectrum and the set of numbers $\{\lambda_1, \lambda_2 \ldots\}$ is called the spectrum of the process.

The (6.8) is the spectral representation of a stationary process, where $Z(\lambda)$ is a continuous spectrum with mean value zero and uncorrelated increments. The possibility of such representations for arbitrary stationary processes was first shown by Kolgomorov (Yaglom, 1962).

The spectral representation of a stationary process is then a "decomposition" of the process into separate pairwise uncorrelated periodic oscillations. It is possible to separate spectral components corresponding to different parts of the spectrum by using suitable chosen linear operators or filters(2). In practice, the filters

(1) The angular frequency $\lambda = 2\Pi f = 2\Pi\frac{1}{T}$, is the number of cycles around the unit circle per unit of time. f is simply the frequency and it is the reciprocal of the period T or length of time required for one complete oscillation.

(2) In engineering, a filter is a device which passes harmonic oscillations in certain frequency range (the pass band) while suppressing oscillations with different frequencies.

used are of three types, the low-pass filters, passing all oscillations with frequencies less than a certain critical frequency λ_0, the high-pass filters, passing all oscillations greater than λ_0 and the band-pass filters passing only oscillation with frequencies λ that belong to a given interval (pass-band) $[\lambda_o, \lambda_1]$.

We shall see now, that the information contained in the autocovariance function is equivalent to the one given by its Fourier-Stieljes transform, the spectral distribution function.

Since any arbitrary stationary process can have a spectral representation, its corresponding autocovariance function can also be expressed in the spectral form.

Thus, for a process such as (6.7), which is assumed stationary and therefore $E(X_k\bar{X}_1)=0$, $k\neq 1$ (by $\bar{X}$ we denote the conjugate of X), the autocovariance function is,

$$(6.9)\quad \sigma(\tau)= \sum_{k=1}^{\infty} E|X_k|^2 e^{i\lambda_k\tau} = \sum_{k=1}^{\infty} b_k e^{i\lambda\tau}; \qquad b_k>o$$

The autocovariance function (6.9) exists if the series is convergent, that is, if

$$(6.10)\quad \sum_{k=1}^{\infty} E|X_k|^2 = \sum_{k=1}^{\infty} b_k < \infty$$

It was shown by Slutsky (1938) that the converse is true, every stationary stochastic process with autocovariance function of the form (6.9) can be represented in form of (6.7) with $E(X_k\bar{X}_1)=0$ for $k\neq 1$.

Setting $\tau=0$, the (6.9) becomes,

$$(6.11)\quad \sigma(0)= \sum_{k=1}^{\infty} E|X_k|^2 = \sum_{k=1}^{\infty} b_k$$

which shows that in the superposition of uncorrelated periodic oscillations, the total variance of the process is equal to the sum of the variances of the separate periodic components.

Observe that variances b_k of the separate periodic components are the mean values of the squares of the amplitude X_k of the harmonic components $X_k e^{i\lambda_k \tau}$ of the process X_t.

The formula (6.9) was generalized by Khinchin (1934) who proved that the autocovariance function of any stationary stochastic process can be represented in the form of an integral,

$$(6.12)\quad \sigma(\tau)= \int_{-\infty}^{\infty} e^{i\lambda\tau} dG(\lambda)$$

where $G(\lambda)$ is the spectral distribution function or the Fourier-Stieljes transform of the autocovariance function $\sigma(\tau)$. The spectral distribution function is a monotonically non-decreasing function, symmetric with respect to the origin and bounded $G(-\infty)=o$ and $G(\infty)=\sigma(0)$. When $G(\lambda)$ is normalized, that is, divided by the variance, then $F(\lambda)= \frac{G(\lambda)}{\sigma(0)}$ is called the normalized spectral distribution function which is the Fourier-Stieljes transform of the autocorrelation function $\rho(\tau)$. That is,

$$(6.13)\quad \rho(\tau)= \int_{-\infty}^{\infty} e^{i\lambda\tau} dF(\lambda)$$

The normalized distribution function $F(\lambda)$ is also non-decreasing, symmetric with respect to the origin and bounded $F(-\infty)=0$ and $F(\infty)=1$. It can be decomposed as,

$$(6.14)\quad F(\lambda)=F_1(\lambda)+F_2(\lambda)+F_3(\lambda)$$

where $F_1(\lambda)$, $F_2(\lambda)$ and $F_3(\lambda)$ are each non-decreasing, $F_1(\lambda)$ is a pure step function, $F_2(\lambda)$ is absolutely continuous, that is, $F_2(\lambda)=\int_{-\infty}^{\infty} F_2^1(u)du$ and $F_3(\lambda)$ is a singular function, continuous and with $F_3^1(\lambda)=0$ almost everywhere.

Thus $F(\lambda)$ can be seen as a distribution function and $\rho(\tau)$ as its characteristic function. Since $G(\lambda)$ and $F(\lambda)$ are odd functions, for every real process, the (6.12) and (6.13) are real integrals and can be written as

$$(6.15)\quad \sigma(\tau)= \int_{-\infty}^{\infty}\cos\lambda\tau dG(\lambda)=\int_{o}^{\infty} \cos\lambda\tau dG_1(\lambda)$$

where $G_1(\lambda)= 2G(\lambda)$ and

$$(6.16)\quad \rho(\tau)= \int_{-\infty}^{\infty}\cos\lambda\tau dF(\lambda)=\int_{o}^{\infty}\cos\lambda\tau dF_1(\lambda)$$

where $F_1(\lambda)=2F(\lambda)$.

When $F(\lambda)$ and $G(\lambda)$ have derivatives (which are the interesting cases)

$$(6.17.a)\quad dF(\lambda) = F(\lambda)d\lambda$$

$$(6.17.b)\quad dG(\lambda) = g(\lambda)d\lambda$$

then $f(\lambda)$ is called the normalized spectral density function or normalized spectrum, and $g(\lambda)$ the spectral density function or power spectrum.

$\sigma(\tau)$ and $\rho(\tau)$ are the inverse Fourier transforms of $g(\lambda)$ and $f(\lambda)$ respectively.

For τ discrete, the normalized spectral density function $f(\lambda)$ is the Fourier transform of a sequence of autocorrelations and we have

$$(6.20)\quad f(\lambda)= \frac{1}{2\Pi} \sum_{\tau=-\infty}^{\infty}\rho(\tau)e^{-i\lambda\tau}; \qquad -\Pi\leq\lambda\leq\Pi$$

and for τ a continuous time parameter,

$$(6.21)\quad f(\lambda)= \frac{1}{2\Pi} \int_{-\infty}^{\infty} \rho(\tau)e^{-i\lambda\tau}d\tau; \qquad -\infty \leq\lambda\leq\infty$$

Similarly the spectrum $g(\lambda)$ for a discrete process is

$$(6.22)\quad g(\lambda)= \frac{1}{2\Pi} \sum_{\tau=-\infty}^{\infty}\sigma(\tau)e^{-i\lambda\tau}; \qquad -\Pi\leq\lambda\leq\Pi$$

and for a continuous process,

$$(6.23)\quad g(\lambda)= \frac{1}{2\Pi} \int_{-\infty}^{\infty}\sigma(\tau)e^{-i\lambda\tau}d\tau \qquad -\infty\leq\lambda\leq\infty$$

Since $F(\lambda)= \int_{-\infty}^{\lambda}f(u)du$, integrating (6.20) and (6.21) we obtain the normalized spectral distribution function $F(\lambda)$.

For a real process X_t, the (6.20) reduces to

$$(6.24)\quad f(\lambda)= \frac{\rho(o)}{2\Pi} + \frac{1}{\Pi} \sum_{\tau=1}^{\infty}\rho(\tau)\cos\lambda\tau = \frac{1}{2\Pi} \sum_{\tau=-\infty}^{\infty}\rho(\tau)\cos\lambda\tau; \quad -\Pi\leq\lambda\leq\Pi$$

and the (6.21) takes the form,

$$(6.25)\quad f(\lambda)= \frac{1}{2\Pi} \int_{-\infty}^{\infty}\rho(\tau)\cos\lambda\tau d\tau \qquad -\infty\leq\lambda\leq\infty$$

Then, the corresponding normalized spectral distribution functions are:

$$(6.26)\quad F(\lambda)= \frac{\rho(o)}{2\Pi} + \frac{1}{\Pi} \sum_{\tau=-\infty}^{\infty} \frac{\rho(\tau)\sin\lambda\tau}{\tau} \qquad -\Pi\leq\lambda\leq\Pi$$

and

$$(6.27)\quad F(\lambda)= \frac{1}{2\Pi} \int_{-\infty}^{\lambda} \int_{-\infty}^{\infty}\rho(\tau)\cos\lambda\tau d\tau d\lambda; \qquad -\infty\leq\lambda\leq\infty$$

A similar procedure is followed to obtain $G(\lambda)$.

In the next section we deal with the autocorrelation function and the normalized spectral density corresponding to the linear processes in (5), (5.1), (5.2) and (5.3), and illustrate with some theoretical examples.

7. AUTOCOVARIANCE AND AUTOCORRELATION FUNCTIONS OF LINEAR STATIONARY PROCESSES AND THEIR FOURIER TRANSFORMS

We saw in Section (5) that a linear stationary process X_t can be interpreted as the output obtained from an input U_t (a purely random process) that has been passed through a transfer function $\alpha(L)$ which is a convergent infinite sum of weights α for $|L| \leq 1$. That is,

$$(7.1)\quad X_t = \sum_{k=o}^{\infty} \alpha_k U_{t-k} = \sum_{k=o}^{\infty} \alpha_k L^k U_t = \alpha(L) U_t ; \quad \sum_{k=o}^{\infty} |\alpha_k| < M; \qquad M \text{ finite}$$

Where L is the lag operator in the transfer function. When the transfer function is considered as a generating function of the α weights, L is then treated as a dummy variable where α is the coefficient of the k-th power of L. Then applying the formula for the autocovariance function and remembering that the process is stationary, we obtain,

$$(7.2)\quad \sigma_{XX}(\tau) = E(X_t X_{t+\tau}) = \sigma_U^2 \sum_{k=o}^{\infty} \alpha_k \alpha_{k+\tau}$$

The variance of X_t is then,

$$(7.3)\quad \sigma_{XX}(o) = \sigma_U^2 \sum_{k=o}^{\infty} \alpha_k^2$$

and the autocorrelation function is

$$(7.4)\quad \rho_{XX}(\tau) = \frac{\sigma_{XX}(\tau)}{\sigma_{XX}(o)} = \frac{\sum_{k=o}^{\infty} \alpha_k \alpha_{k+\tau}}{\sum_{k=o}^{\infty} \alpha_k^2}$$

The autocovariance function can be obtained in an easier way using the autocovariance generating function, which also can be used to obtain the spectrum of the process.

The autocovariance generating function is

$$(7.5)\quad \sigma_{XX}(L) = \sum_{\tau=-\infty}^{\infty} \sigma_{XX}(\tau) L^{\tau}, \qquad \tau = o, \pm 1, \pm 2, \ldots$$

Since for a stationary process $\sigma_{XX}(\tau)$ is an even function, then $\sigma_{XX}(k) = \sigma_{XX}(-k)$,

is the coefficient of L^k and L^{-k}.

For the linear process (7.1) the autocovariance generating function is shown to be

$$(7.6)\quad \sigma_{XX}(L) = \sigma_U^2 \alpha(L)\alpha(L^{-1})$$

For $L = e^{-i\lambda}$, the (7.5) becomes

$$(7.7)\quad \sigma_{XX}(L) = \sum_{\tau=-\infty}^{\infty} \sigma_{XX}(\tau)e^{-i\lambda\tau}; \qquad \tau = 0, \pm 1, \pm 2, \ldots$$

Comparing (7.7) with the spectrum $g_{XX}(\lambda)$ given in (6.22), namely $g_{XX}(\lambda)= \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} \sigma_{XX}(\tau)e^{-i\lambda\tau}$, we see that $2\pi g_{XX}(\lambda)=\sigma_{XX}(L)$. If we limit λ to be non-negative, then, $\pi g_{XX}(\lambda)=\sigma_{XX}(L)$.

Therefore, multiplying the autocovariance generating function by $\frac{1}{\pi}$ gives us the power spectrum of the process. Then, applying (7.6), the spectrum of the linear process in (5) can also be written in the form,

$$(7.8)\quad g_{XX}(\lambda)= \frac{\sigma_U^2}{\pi}|\alpha(e^{-i\lambda})|^2; \qquad -\pi\leq\lambda\leq\pi$$

$|\alpha(e^{-i\lambda})|^2$ is called the <u>filter gain</u> and it is the square of the filter transfer function. The (7.8) shows that the spectrum of the output X_t of the linear process (7.1) can be obtained from the constant uniform spectrum $\frac{\sigma_U^2}{\pi}$ of a white noise process U_t, multiplied by a factor (the filter gain) that depends only on the characteristics of the filter. The frequencies for which $|\alpha(e^{-i\lambda})|^2$ is large are magnified and those for which the filter gain is small are reduced. The variance of X_t is,

$$(7.9)\quad \sigma_U^2=\int_o^{\pi} g_{XX}(\lambda)d\lambda= \frac{\sigma_U^2}{\pi}\int_o^{\pi}|\alpha(e^{-i\lambda})|^2 d\lambda$$

Dividing (7.8) by (7.9) we obtain the normalized spectral density function $f_{XX}(\lambda)$ which is the Fourier transform of the autocorrelation function $\rho_{XX}(\tau)$. Then,

$$(7.10)\quad f(\lambda)=\frac{|\alpha(e^{-i\lambda})|^2}{\int_o^\pi |\alpha(e^{-i\lambda})|^2 d\lambda}; \qquad o\leq\lambda\leq\pi$$

It is easy to show that the autocovariance and autocorrelation functions of the autoregressive process AR(p) in (5.1.1) satisfy the same form of difference equation corresponding to the deterministic part of it. In effect, multiplying (5.1.1) by $X_{t-\tau}$ and applying (5.1.3) we obtain,

$$(7.11)\quad \sum_{r=o}^{p}\beta_r X_{t-r}X_{t-\tau}=\sum_{r=o}^{\infty}h_r U_t U_{t-\tau-r}$$

Since $E(X_{t-r}X_{t-\tau})=\sigma_{XX}(\tau-r)$; $E(U_t^2)=\sigma_U^2$; $E(U_t U_s)=0$ for all $t\neq s$; the expected values of the two sides of (7.11) satisfy for $\tau=o$ and for $\tau>o$, respectively,

$$(7.12)\quad \sum_{r=o}^{p}\beta_r\sigma_{XX}(-r)=\sigma_U^2$$

and

$$(7.13)\quad \sum_{r=o}^{p}\beta_r\sigma_{XX}(\tau-r)=o \qquad \tau = 1, 2, \ldots$$

These are often called the Yule-Walker equations. Thus, the sequence $\sigma_{XX}(\tau-1)$, $\sigma_{XX}(\tau-2)$, ..., $\sigma_{XX}(\tau-p)$, $\tau=1, 2, \ldots$. satisfies the homogeneous difference equation (7.13). Dividing (7.13) throughout by σ_X^2, we obtain the autocorrelation function which also satisfies a homogeneous difference equation analogous to the one of the process X_t itself. We can write the (7.13) using the lag operator L in the form of,

$$(7.14)\quad \sum_{r=o}^{p}\beta_r L^r\sigma_{XX}(\tau)=\beta(L)\sigma_{XX}(\tau)=o; \qquad \tau = 1, 2, \ldots..$$

and equivalently for the autocorrelation function,

$$(7.15)\quad \sum_{r=o}^{p}\beta_r L^r \rho_{XX}(\tau)=\beta(L)\rho_{XX}(\tau)=o; \qquad \begin{matrix}\tau = 1, 2, \ldots \\ \beta o = 1\end{matrix}$$

where L operates on τ.

The same conditions for stationarity that were required for the finite autoregressive processes of the form (5.1.1) apply here. If all the roots $|G_i|<1$ are distinct, we have two situations:

(1) A root G_i is real, in which case $A_i G_i^s$ (see 5.1.7) decreases geometrically to zero as s increases. If the root is positive, we will have a decreasing exponential function; and, if it is negative, we will have an exponential function alternating in sign and decreasing in absolute value.

(2) A pair of complex conjugate roots G_i, G_j, in which case they generate a term that is an oscillating trigonometric function, decreasing in absolute value, and whose period of oscillation depends on the argument of the complex roots.

The variance $\sigma_{XX}(o)$ of an autoregressive process X_t can be obtained from (7.12) and also can be expressed in terms of the autocorrelation function by dividing (7.12) by $\sigma_{XX}(0)$ and making $\sigma_{XX}(-r)=\sigma_{XX}(r)$. Then,

$$(7.16)\quad \sigma_X^2 = \frac{\sigma_U^2}{1+\beta_1\rho(1)+\beta_2\rho(2)+\ldots+\beta_p\rho(p)}$$

The spectrum $g_{XX}(\lambda)$ for the AR(p) process can be obtained using (7.8) where the filter gain is $|\beta(e^{-i\lambda})|^{-2}$.

Then,

$$(7.17)\quad g_{XX}(\lambda)=\frac{\sigma_U^2}{\Pi|\beta(e^{-i\lambda})|^2}=\frac{\sigma_U^2}{\Pi|1+\beta_1 e^{-i\lambda}+\ldots+\beta_p e^{-ip\lambda}|^2};\qquad o\leq\lambda\leq\Pi$$

The normalized spectral density function $f(\lambda)$ is obtained dividing $g_{XX}(\lambda)$ by σ_X^2.

Though the autocorrelation function of an AR(p) process is infinite in extent, by its own nature it can be described in terms of p non-zero functions of the autocorrelations. This information is provided by the partial autocorrelation function which helps to determine the order of an autoregressive process to fit to an observed time series. For an autoregressive process of order p, the partial autocorrelation function has a cutoff after the p lag.

Denoting by β_{kj} the jth coefficient in an AR(k) process so that β_{kk} is the last coefficient, then from (7.15) β_{kk} satisfies the set of equations.

$$(7.18)\quad \rho(j)=\beta_{k1}\rho(j-1)+\beta_{k2}\rho(j-2)+\ldots+\beta_{kk}\rho(j-k);\qquad j=1,\ 2,\ \ldots\ k$$

leading to the Yule-Walker equations, which may be written as;

$$(7.19)\quad \begin{bmatrix} 1 & \rho & \ldots & \rho(k-1) \\ \rho & 1 & \ldots & \rho(k-2) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \rho(k-1) & \rho(k-2) & & 1 \end{bmatrix} \begin{bmatrix} \beta_{k1} \\ \beta_{k2} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \cdot \\ \cdot \\ \cdot \\ \rho(k) \end{bmatrix}$$

solving (7.19) for k=1, 2, 3, ... we obtain the partial autocorrelations β_{11}, β_{22}, β_{33}, ... In general β_{kk} is the partial autocorrelation function of the lag k.

For the finite moving average process of order q (MA-q) in (5.2.1) the autocovariance function is

$$(7.20)\quad \sigma_{XX}(\tau)=E(X_t X_{t+\tau})=\sigma_U^2 \sum_{k=o}^{q-\tau} \alpha_k \alpha_{k+\tau}; \qquad \tau \leq q$$

and

$$\sigma_{XX}(\tau)=o; \qquad \tau>q$$

Then the variance of the MA(q) process is

$$(7.21)\quad \sigma_X^2=\sigma_U^2 \sum_{k=o}^{q} \alpha_k^2$$

and the autocorrelation function is

$$(7.22)\quad \rho_{XX}(\tau)= \frac{\sum_{k=o}^{q-\tau} \alpha_k \alpha_{k+\tau}}{\sum_{k=o}^{q} \alpha_k^2}; \qquad \tau \leq q$$

and

$$\rho_{XX}(\tau)=o; \qquad \tau>q$$

Consequently, the correlogram of a MA(q) process is zero for $\tau=q+1$ onwards.

According to (4.7.8) the spectrum of a MA(q) process is

$$(7.23)\quad g_{XX}(\lambda)= \frac{\sigma_U^2}{\Pi} |\alpha(e^{-i\lambda})|^2 = \frac{\sigma_U^2}{\Pi} \left| \sum_{k=o}^{q} \alpha_k e^{-i\lambda k} \right|^2; \quad \alpha_o=1; \quad o \leq \lambda \leq \Pi$$

and the normalized spectral density function $f_{XX}(\lambda)$ is $g_{XX}(\lambda)/\sigma_X^2$.

For the ARMA(pq) process X_t defined in (4.5.1), the $\sigma_{XX}(\tau)$, $\rho_{XX}(\tau)$, $g_{XX}(\lambda)$ and $f_{XX}(\lambda)$ can be obtained in a similar way. Thus,

$$(7.24)\quad \sigma_{XX}(\tau)=\sigma_{XU}(\tau)+\alpha_1\sigma_{XU}(\tau-1)+\ldots+\alpha_q\sigma_{XU}(\tau-q)-\beta_1\sigma_{XX}(\tau-1)-\ldots-\beta_p\sigma_{XX}(\tau-p)$$

where $\sigma_{XU}(\tau)$ is the cross covariance function between X_t and U_t and is defined by $\sigma_{XU}(\tau)=E(X_{t-\tau}U_t)$.

Since $X_{t-\tau}$ depends only on random components which have occurred up to time $t-\tau$ uncorrelated with U_t it follows that $\sigma_{XU}(\tau)=o$ for $\tau>o$ and $\sigma_{XU}(\tau)\neq o$, for $\tau \leq o$.

The (7.24) reduces to,

$$(7.25)\quad \sigma_{XX}(\tau)=-\beta_1\sigma_{XX}(\tau-1)-\ldots-\beta_p\sigma_{XX}(\tau-p); \qquad \tau\geq q+1$$

Hence, the autocorrelation function is:

$$(7.26)\rho_{XX}(\tau)=-\beta_1\rho_{XX}(\tau-1)-\ldots-\beta_p\rho_{XX}(\tau-p), \qquad \tau\geq q+1$$

or

$$\beta(L)\rho_{XX}(\tau)=o; \qquad \tau\geq q+1$$

where L operates on τ.

Then for an ARMA(p,q) process there will be q autocorrelations whose values depend on the choice of the q moving average parameters α as well as on the p autoregressive parameters β. Now for $\tau>q+1$, the p autocorrelations already obtained provide the initial values for the homogeneous difference equation $\beta(L)\rho_{XX}(\tau)=o$ which then entirely determines the autocorrelations of higher lags.

If $q<p$, the whole autocorrelation function will consist of a mixture of damped exponentials and/or damped sine functions. If $q\geq p$, the $q-p+1$ autocorrelations used as initial values will not follow this general pattern.

For $\tau=0$, the (7.24) gives the variance of the process

$$(7.27)\quad \sigma_X^2=\sigma_U^2+\alpha_1\sigma_{XU}(-1)+\ldots+\alpha_q\sigma_{XU}(-q)-\beta_1\sigma_{XX}(1)-\ldots-\beta_p\sigma_{XX}(p)$$

which has to be solved along with the p equations (7.24) for $\tau=1, 2, \ldots p$ to obtain $\sigma_X^2,\sigma_{XX}(1), \ldots, \sigma_{XX}(p)$.

The spectrum of the process is

$$(7.28)\quad g_{XX}(\lambda)=\frac{\sigma_U^2}{\Pi}\,\frac{\left|\sum_{r=o}^{q}\alpha_r e^{-i\lambda r}\right|^2}{\left|\sum_{k=o}^{p}\beta_r e^{-i\lambda r}\right|^2}; \qquad o\leq\lambda\leq\Pi$$

and the normalized spectral density function $f_{XX}(\lambda)$ is the quotient between $g_{XX}(\lambda)$ and σ_X^2.

We shall now illustrate with some theoretical examples adapted from A.A. Sveshnikov (1966) the connection between the autocorrelation function and the normalized spectral density function.

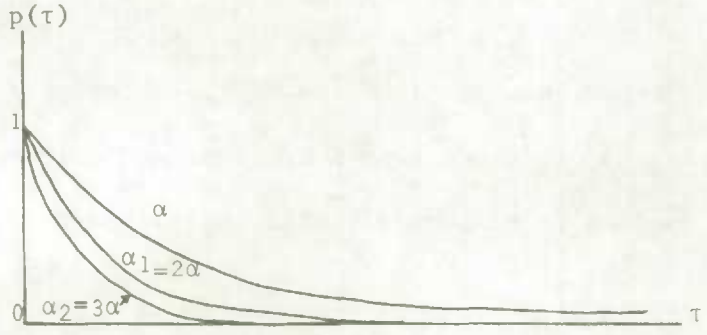
Example 1

Consider an autocorrelation function of the form

(1) $\rho(\tau)=e^{-\alpha|\tau|}$

where $o<\alpha<1$ serves as a measure of the rapidity of decrease of the $\rho(\tau)$ with the increase of the time lag τ. We see in Figure 1, that the greater the α, the more damped the autocorrelation function, thus implying a less smooth stochastic process.

Figure 1



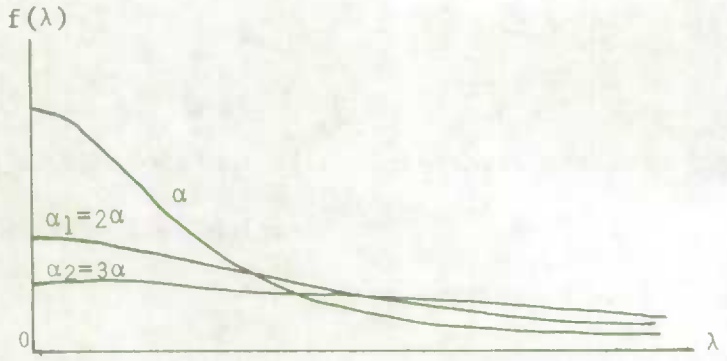
The corresponding normalized spectral density function $f(\lambda)$ is,

(2) $f(\lambda)=\frac{1}{2\Pi}\int_{\infty}^{\infty}\rho(\tau)e^{-i\lambda\tau}d\tau=\frac{1}{2\Pi}\int_{-\infty}^{\infty}e^{-i\lambda\tau-\alpha|\tau|}d\tau=\frac{1}{\Pi}\frac{\alpha}{\lambda^2+\alpha^2}$

and it gives the same information contained in (1) but in the frequency domain. The normalized spectral density function is shown in Figure 2. We can see that for small α the normalized spectrum has predominancy of low frequencies implying a smooth process, whereas as α increases, the curve is compressed toward the λ- axis, at the same time becoming flatter. This kind of behaviour of the function $f(\lambda)$ enables us to illustrate the

of a purely random process or white noise, the normalized spectral density of which is a constant equal to $\frac{1}{2\Pi}$ for $-\Pi \leq \lambda \leq \Pi$. Observe that the ordinate of $f(\lambda)$ at the origin is $\frac{1}{\Pi\alpha}$ and as α increases, the interception decreases. In fact, it is assumed that α can assume very high values and the $\rho(\tau)$ is transformed into a spike-shaped function, different from zero only in a very small neighborhood around $\tau=0$.

Figure 2



The total variance of the process is the area under $f(\lambda)$. For small α, a low frequency band accounts for most of the variance whereas for the large α , the variance is distributed almost uniformly in the frequency band capable of exerting an effect on the process under consideration. It is impossible for absolutely white noise to exist, since for the spectral density to be constant in the whole range of variations of λ, the autocovariance function for $\tau=0$, would have an infinite variance which cannot take place in any real process. In effect,

$$\sigma_{XX}(o)=\int_{-\infty}^{\infty} g(\lambda)d\lambda=c\int_{-\infty}^{\infty} d\lambda= \infty$$

Example 2

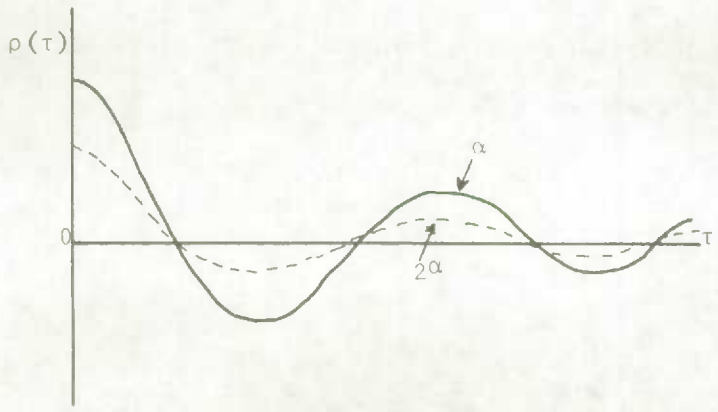
As a second example, consider an autocorrelation function $\rho(\tau)$ of the form

(3) $\rho(\tau)=e^{-\alpha|\tau|}\cos \omega\tau$

which differs from (1) by the presence of the factor $\cos \omega\tau$ that gives to $\rho(\tau)$ the form of a damped harmonic oscillation as shown in

Figure 3. If we were to observe the generating process, some periodicity would be apparent.

Figure 3

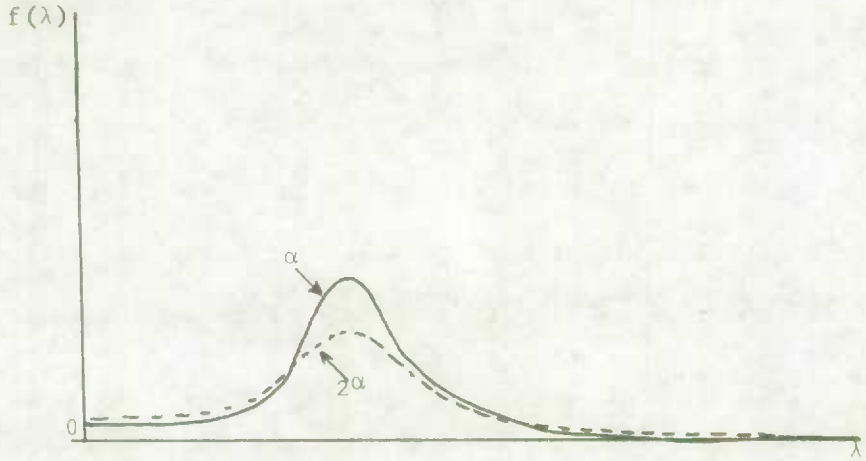


The corresponding normalized spectral density function can be obtained replacing cos ωτ by $\frac{1}{2}(e^{i\omega\tau}+e^{-i\omega\tau})$ and by replacing λ by (λ−ω) and (λ+ω) in the integrals of f(λ). Then,

$$(4)\quad f(\lambda)=\frac{1}{2\Pi}\left|\frac{\alpha}{(\lambda-\omega)^2+\alpha^2}+\frac{\alpha}{(\lambda+\omega)^2+\alpha^2}\right|=\frac{\alpha}{\Pi}\left|\frac{\lambda^2+\alpha^2+\omega^2}{(\lambda^2-\omega^2-\alpha^2)^2+4\alpha^2\lambda^2}\right|$$

The representation of f(λ) shown in Figure 4 presents peaks in the neighborhood of the angular frequency ω. For $\omega=\frac{\Pi}{3}$, the fundamental seasonal frequency in the time domain would correspond to a period of 6 months; for $\omega=\frac{\Pi}{6}$, to a period of 12 months. Then, the corresponding model for the generating process would have an oscillatory seasonal component whose amplitude and phase are changing slowly compared with the fundamental seasonal period. The larger the α (that is, the more rapid is the rate of change) the more obscure is the seasonal component and therefore, the less sharp is the peak in the normalized spectral density function.

Figure 4



8. 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 cant take the arithmetic mean of all the observed values X_j of X. Similarly, to determine the mean value and the autocovariance function $\sigma(t,s)$ of a stochastic process X(t), we must have a large number of realizations of the process X(t), that is, $X_1(t),\ldots,X_N(t)$ 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 stationary random processes.

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

$$(8.1)\quad \hat{\mu}_T = \frac{1}{T}\int_0^T X(t)dt; \qquad 1\leq t\leq T$$

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

$$(8.2)\quad \lim_{T\to\infty} E(\hat{\mu}_T-\mu)^2=0$$

since,

$$(8.3)\quad \lim_{T\to\infty} E(\hat{\mu}_T-\mu)^2=\lim_{T\to\infty}\frac{1}{T}\int_o^T\sigma(\tau)d\tau; \qquad \tau = s-t$$

then, the (8.2) will be verified if and only if

$$(8.4)\quad \lim_{T\to\infty}\frac{1}{T}\int_o^T\sigma(\tau)d\tau= o$$

This condition was first shown by Slutsky (1938) and any stationary process that fulfils the (8.4) 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

$$(8.5)\quad \hat{\sigma}_T(\tau)= \frac{1}{T}\int_o^T\{[X(t+\tau)-\mu]\ [X(t)-\mu]\}dt$$

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

$$(8.6)\quad \lim_{T\to\infty} E[\hat{\sigma}_T(\tau)-\sigma(\tau)]^2=o$$

For normal process, the (8.6) reduces to

$$(8.7)\quad \lim_{T\to\infty}\ \frac{1}{T}\int_o^T|\sigma(\tau)|^2d\tau= o$$

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

It is easy to show that <u>all linear stationary stochastic processes are ergodic</u>. However, not all the stationary processes are ergodic. Consider, for example, a simple harmonic process, where a and b are normally distributed random variables with zero means and common variance σ^2.

$$(8.8)\quad X(t)= a\cos\lambda t+b\sin\lambda t$$

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

$$(8.9)\quad \sigma(\tau)= \sigma^2\cos\lambda\tau$$

Using (8.4) we can show that this process is ergodic in the mean. We have

$$(8.10)\quad \lim_{T\to\infty} \frac{1}{T}\int_0^T \sigma(\tau)d\tau = \lim_{T\to\infty} \sigma^2 \frac{\sin\lambda T}{T} = 0$$

But the process is not ergodic for the autocovariance function. Applying (8.7),

$$(8.11)\quad \lim_{T\to\infty} \frac{1}{T}\int_0^T |\sigma(\tau)|^2 d\tau = \lim_{T\to\infty} \left(\frac{\sigma^4}{2} + \sigma^4 \frac{\sin 2\lambda T}{T}\right) = \frac{\sigma^4}{2}$$

9. HOMOGENEOUS LINEAR NON-STATIONARY PROCESSES

9.1 Autoregressive Integrated Moving Average - (ARIMA) models

The ARMA (p,q) model discussed in (5.3) can be generalized by including processes which are non-stationary in either their local level and/or slope. In other words, we will consider the case in which the linear operator that transforms an input U_t purely random into an output X_t, is homogeneous non-stationary. Assume X_t follows a process stationary in the dth-difference, and U_t follows a moving average process of order q, we can then write

$$(9.1.1) \qquad \phi(L)\, X_t = \alpha(L) U_t$$

where $\phi(L)$ is a non-stationary autoregressive operator and $\alpha(L)$ is an invertible moving average operator.

Since the process X_t is stationary in the dth-difference, d of the roots of $\phi(L) = 0$ are equal to 1 whereas the rest, say p, lie inside the unit circle. The $\phi(L)$ operator can then be written in the form

$$(9.1.2) \qquad \underset{p+d}{\phi}(L) = \underset{p}{\beta}(L)(1-L)^d = \beta_p(L)\Delta^d$$

where $1-L = \Delta$ is the difference operator, the subindices indicate the order of the operators, namely, p+d for the generalized autoregressive operator $\phi(L)$ and p for the stationary autoregressive operator $\beta(L)$.

Assuming the order of the moving average operator to be q, the (9.1.1) becomes,

$$(9.1.3) \qquad \underset{p+d}{\phi}(L) X_t = \beta_p(L)\Delta^d X_t = \alpha_q(L) U_t$$

and is called an Autoregressive Integrated Moving Average process - ARIMA (p,d,q). Note that for d = 0, the (9.1.3) defines an ARMA (p,q) process. Also by making $\Delta^d X_t = W_t$, (9.1.3) is stationary in W_t, or equivalently in the dth-difference of X_t.

The ARIMA (p,d,q) model can be written under the explicit form of a difference equation as follows

$$X_t = U_t + \alpha_1 U_{t-1} + \ldots + \alpha_q U_{t-q} - \phi_1 X_{t-1} - \ldots - \phi_{p+d} X_{t-p-d}; \quad (9.1.4)$$

$$\alpha_o = \phi_o = 1$$

The difference equation form is generally used for calculating the forecasts.

Since in (9.1.3), $\beta_p(L)$ and $\alpha_q(L)$ are respectively stationary and invertible operators, other explicit forms of an ARIMA model are possible, namely: (1) in function of the current and previous values of the random process U_t, and (2) in function of the previous values of X_t plus the current value of U_t. In the first case we have,

$$X_t = \phi^{-1}_{p+d}(L)\alpha_q(L)U_t = \psi(L)U_t \quad (9.1.5)$$

where

$$\psi(L) = 1+\psi_1 L + \psi_2 L^2 + \ldots \quad (9.1.6)$$

The ψ weights can be obtained by equating coefficients of L in the expansion of $\phi(L)\psi(L)U_t = \alpha_q(L)U_t$.

That is,

$$(1+\phi_1 L+\ldots+\phi_{p+d} L^{p+d})\ (1+\psi_1 L+\psi_2 L^2+\ldots) = 1+\alpha_1 L+\alpha_2 L^2+\ldots+\alpha_q L^q \quad (9.1.7)$$

For j greater than the larger of p+d-1 or q, the ψ weights satisfy the following difference equation,

$$(9.1.8) \qquad \phi_{p+d}(L)\psi_j=\theta_p(L)\Delta^d\psi_j=o$$

where L operates on the subscript j.

In the second case, X_t in function of the previous values of X_t plus the current value of U_t, we have

$$(9.1.9) \qquad \phi_{p+d}(L)\alpha_q^{-1}(L)X_t=\Pi(L)X_t=U_t$$

where

$$(9.1.10) \qquad \Pi(L)=1+\Pi_1L+\Pi_2L^2+\ldots$$

Since $\alpha_q(L)$ is an invertible operator, $\Pi(L)$ is convergent on or within the unit circle.

To obtain the Π weights, a procedure similar to the one described for the ψ weights, is followed. We equate coefficients of L in

$$(9.1.11) \qquad \phi_{p+d}(L)=\alpha_q(L)\Pi(L)$$

and for j greater than the larger of p+d-1 or q, the Π weights satisfy the difference equation

$$(9.1.12) \qquad \alpha_q(L)\Pi_j=o$$

where L operates on j.

Finally, the ARIMA (p,d,q) process in (9.1.1) can be modified by adding a deterministic function of time f(t), usually a polynomial of a degree equal to the order of the difference operator, which stands for a deterministic trend.

Also, it can happen that there is added noise to the process in the sense that we actually do not observe X_t but another random process $Z_t=X_t+U_t$, where U_t can be a purely random process or a

correlated random process. Then, if the observed process is Z_t,

$$(9.1.13) \qquad \beta_p(L)\Delta^d Z_t=\beta_p(L)\Delta^d(X_t+U_t^1) = \alpha_q(L)U_t+\beta_p(L)\Delta^d U_t^1$$

Now if we assume that U_t^1 is an ARMA (p_1,q_1) process, that is,

$$\beta_{p_1}(L)U_t^1 = \alpha_{q_1}(L)E_t$$

where E_t is a purely random process independent of U_t^1, then (9.1.13) becomes

$$(9.1.14) \qquad \beta_{p_1}(L)\beta_p(L)\Delta^d Z_t=\beta_{p_1}(L)\alpha_q(L)U_t+\beta_p(L)\alpha_{q_1}(L)\Delta^d E_t$$

Let $P = p_1+p$ and Q be equal to whichever of p_1+q or $p+q_1+d$ is larger, then the (9.1.14) can be written in the form

$$(9.1.15) \qquad \beta_p(L)\Delta^d Z_t = \alpha_Q(L)V_t$$

with V_t a purely random process and Z_t an ARIMA (P,d,Q) process since the sum of two independent moving average processes is another moving average process with order given by the component process of higher order.

9.2 Integrated Moving Average (IMA) Models

For $p = 0$, the ARIMA (p,d,q) process defined in (9.1.3) reduces to an integrated moving average process of order (0,d,q). That is, the stochastic process X_t is seen as generated by applying d times the infinite sum operator to a finite moving average process of order q. In symbols,

$$(9.2.1) \qquad \Delta^d X_t= \alpha_q(L)U_t$$

or equivalently

$$(9.2.2) \qquad X_t=\Delta^{-d}\alpha_q(L)U_t=S^d\alpha_q(L)U_t$$

where $S = (1-L)^{-1} = \Delta^{-1}$ is the infinite summation operator and the supraindex d indicates that the (9.2.1) process is summed d times.(1)

The (9.2.1) model can be written explicitly in the form of a difference equation as follows

$$(9.2.3) \qquad X_t = U_t + \alpha_1 U_{t-1} + \ldots + \alpha_q U_{t-q} + dX_{t-1} - \frac{1}{2}d(d-1)X_{t-2} + \ldots + (-1)^{d+1}X_{t-d}$$

It can also be written explicitly in function of the current value of U_t and its previous values, or in function of the previous values of X_t plus the current value of U_t.

(1). Note that

(1) $SU_t = \sum_{h=-\infty}^{t} U_h = (1+L+L^2+\ldots)U_t = (1-L)^{-1}U_t$

(2) $S^2U_t = S(SU_t) = \sum_{i=-\infty}^{t}\sum_{h=-\infty}^{i} U_h = SU_t + SU_{t-1} + \ldots$

and so on.

10. SEASONAL MODELS

In the analysis of economic time series, it is traditional to distinguish different types of evolution; namely, (a) the trend, (b) the cycle, (c) the seasonal variation, and (d) the irregular fluctuations.

The trend reflects long-run movements, lasting many years. The cycle is a quasi-periodic oscillation characterized by alternating periods of expansion and contraction. The seasonal represents the composite effect of climatic and institutional elements which repeat more or less regularly each year. These three types of fluctuation are assumed to follow systematic patterns. (They are the signals of the process.)

On the other hand, the irregulars are unforseeable movements related to events of all kinds. In general they have a stable random appearance.

In much analytical work, the trend and the cycle are combined since it is the standing of the series apart from seasonal variation upon which interest is centered. The information given by seasonally adjusted series plays a vital role in long-term decision-making for controlling the economic activity. It is therefore very important to identify the pattern followed by the seasonality and to estimate the seasonal factors with a high degree of accuracy. In this section, we will review briefly the most common models built for the seasonal variation.

The simplest and most often studied assumption is that the seasonal component has a stable pattern; that is, it is a strictly periodic function with periodicity of 12 months. Then, the seasonal component s_t can be represented for a monthly data as a series of twelve constants, one for each month, which sum zero. Let these constants be a_k, then we have;

$$(10.1)\quad s_t=\begin{cases} a_k; & \text{for } t=k \text{ or } t-k \text{ divisible by } 12, \\ o; & \text{otherwise.} \end{cases}$$

$$\sum_{k=1}^{12} a_k=o$$

The (10.1) can be represented also working with frequencies instead of periods as follows

$$(10.2)\quad s_t=\sum_{j=1}^{6}(\alpha_j\cos\lambda_j t+\beta_j\sin\lambda_j t); \qquad \lambda j=\frac{2\Pi j}{12}$$

Where the λ_j's are the seasonal frequencies for a monthly series and where components with a periodicity of less than 2 months are not being observed. However, all results can be easily modified for weekly or other series.

From (10.1) and (10.2) we obtain the following relationship

$$(10.3)\quad a_k = \sum_{j=1}^{6} (\alpha_j \cos\lambda_j k + \beta_j \sin\lambda_j k); \qquad k=1, \ldots, 12$$

Although s_t is assumed to be strictly periodic, it is not a deterministic function as long as α_j and β_j are considered purely random variables, with zero means and

$$E(\alpha_j \alpha_k) = E(\beta_j \beta_k) = \begin{cases} \sigma_j^2 & ; \quad j=k \\ o & ; \quad j \neq k \end{cases}$$

and $E(\alpha_j \beta_k) = o$ for all j and k.

Then, (10.3) will be a stable seasonal model that follows a covariance stationary process, but which will not be ergodic in the second order moments if the random variables are assumed to be normally distributed (see Section 8).

The spectral representation of such process s_t will have a spike of height σ_j^2 at each λ_j. We will have a concentration of spectral mass precisely at the seasonal frequencies. Although this will never occur in practice, it is a good approximation if the seasonal peaks in the spectrum are narrow.

Several attempts have been made to produce seasonal models of a more realistic kind. Because of institutional, technological and other exogenous changes in an economic structure, the seasonal patterns of economic time series tend to change through time.

In such cases, the best procedure will depend on the nature of the evolving pattern. Various possibilities have been discussed such as the following:

(1) There is a sudden change which divides the observed series in two fairly uniform parts. In such cases, one should obviously estimate a seasonal pattern for each part of the series. O. Lange (1963) and M. Abel (1962) give examples for such cases.

(2) There are changes in the amplitude while the phase still remains unchanged. In other words, the α_j and β_j coefficients of (10.3) depend on t. The simplest case is to assume that they are polynomials of a low degree. In the case of linearity, the s_t is,

$$(10.4) \qquad s_t = \sum_{j=1}^{6} \{\alpha_{1j} + \alpha_{2j} t) \cos\lambda_j t + (\beta_{1j} + \beta_{2j} t) \sin\lambda_j t\}$$

These types of models have been studies by Hannan (1963) Nettheim (1965) and H.M. Rosenblatt (1963).

Another way to approach changes in amplitudes is by defining the stochastic process that generates α_j and β_j. Hannan, Terrel and Tuckwell (1970) and Terrel and Tuckwell (1970) have worked empirically with the assumption that α_j and β_j follow a stationary autoregressive model of order 1. Grether and Nerlove (1970) have considered a similar approach.

However, there is no constraint, except from the point of view of the estimation procedures, to assume that α_j and β_j are described by other kind of processes.

(3) The seasonal pattern may change also in phase in which case, not only the amplitudes but the phase should be considered generated by a stochastic process. Then, the (10.2) written now under the cosine form becomes

$$(10.5) \quad s_t = \sum_{j=1}^{6} R_{jt}(\cos\lambda_j t + \theta)$$

Where $R_{jt} = \sqrt{\alpha_{jt}^2 + \beta_{jt}^2}$ is the random amplitude of the process and $\theta = \text{arcot.} \frac{\alpha_{jt}}{\beta_{jt}}$ is the phase and it is also random.

For forecasting time series where a changing seasonal pattern may exist, Box and Jenkins (1970) have suggested a kind of model known as multiplicative seasonal model. This model belongs to the category of homogeneous non-stationary linear processes and it is based on the assumption that, in periodic data, one expects relationships: (1) between successive monthly observations within a year, and (2) between observations for the same month in successive years. We are then faced with two time intervals, one for which the L lag is of one period and other for which the L lag period is given by the seasonality. Assuming that the seasonal pattern is of periodicity s, then $L^S X_t = X_{t-s}$. Looking at the observations of each month in successive years, and ARIMA (P,D,Q) model is proposed to describe them; that is,

$$(10.6) \quad B_P(L^S)\Delta_s^D X_t = A_Q(L^S)E_t$$

where the subindices P, D, and Q denote the order respectively of the stationary autoregressive operator $B(L^S)$; of the difference operator $\Delta_s = 1 - L^S$ and of the invertible moving average operator

$A(L^s)$. E_t stands for an autocorrelated random process. In other words, one expects that each month's observation is also related to previous months and therefore, that E_t is related to E_{t-1}, E_{t-2}, and so on. To take care of this relationship, a second ARIMA (p,d,q) model is introduced.

$$(10.7) \quad B_p(L)\Delta^d E_t = \alpha_q(L)U_t$$

where the U_t is now a purely random process and $\beta_p(L)$ is a stationary autoregressive operator of order p, $\Delta = 1-L$ is the difference operator of order d and $\alpha_q(L)$ is an invertible moving average of order q. Combining (10.6) and (10.7) one obtains

$$(10.8) \quad \beta_p(L)B_p(L^s)\Delta^d\Delta_s^D X_t = \alpha_q(L)A_Q(L^s)U_t$$

which is the general multiplicative seasonal model of order $(p,d,q)(P,D,Q)_s$.

The (10.8) is used for forecasting time series where a changing seasonal pattern may exist. K.R. Brewer (1969) has used the forecast functions or predictors corresponding to different identifications of (10.8) and divided them into trend and seasonal components for current seasonal adjustment. The idea is very interesting and well developed from a theoretical point of view but can not be used in practical applications on a production line basis. There are serious limitations stemming from both the validity of the Box and Jenkins models to describe different kinds of economic time series and the method of estimation for the decomposition of the forecast functions.

11. CONCLUSIONS

Models for time series are stochastic processes, that is, processes controlled by probability laws. From a theoretical point of view, a large class of models can be built to describe the behaviour of phenomena that evolve through time in a probabilistic way. For empirical applications, however, only a restricted subclass of models have been properly developed. These models belong to the categories of: (1) Normal Non-Stationary (in the mean) Processes; (2) Normal Linear Stationary Processes, and (3) Normal Homogeneous Linear Non-Stationary Processes. Very often, the assumption of normality is not explicitly introduced in the model but it is used in the testing of the hypotheses. When the normality is dropped, the stationarity is defined in the wide sense.

The building of time series models depends on both the behaviour of the phenomenon and the purpose of the analysis but for immediate empirical applications, it also depends on the existence of optimal methods of estimation.

If the purpose of the analysis is the description of the probability structure of a time series, then parametric (finite number of parameters, usually small) and non-parametric (an infinite number or a very large number of parameters) models are equally acceptable.

Among the parametric models, the error model which is a normal non-stationary (in the mean) process and the autoregressive integrated moving average model, which is a normal homogeneous linear non-stationary process, are the most frequently applied.

The non-parametric models most commonly used are the autocovariance function and the power spectrum. However, they are not useful for forecasting which can only be done with parametric models.

In an error model, the generating process of a time series is decomposed into a systematic component (signal) and a purely random process (white noise). The properties of the time series are then summarized in the expected value of the process, and variations in the random element do not affect these properties. In other words, any effect of time in the process is made on the deterministic part and therefore, each observation is stochastically independent of the previous ones.

This model is classical for the analysis of economic time series where the signal is assumed to have a trend (a systematic component that moves slowly smoothly and progressively through time), a cycle (a quasi-periodic oscillation) and a seasonal factor (an oscillatory component that tends to repeat more or less regularly each year).

The assumption of a non-autocorrelated error component, however, introduces serious limitations in the validity of this type of model to describe the behaviour of economic time series. The errors in the observed values of economic time series are not only due to wrong independent observations but also to other more serious irregularities, and once they appear, they are incorporated in the process and influence its future evolution. Yule (1921, 1927) was first attributed with pointing out this aspect; however, it has been almost completely neglected in the analysis of economic time series until very recently.

The decomposition of the signal into a trend-cycle component and a seasonal component is of extreme importance for the economic policy maker who, faced with the problem of controlling the level of economic activity, does not wish to mistake a seasonal movement for a long-term change.

A model that fulfills these purposes and, at the same time, introduces more realistic assumptions about the behaviour of the random component would be one in which the time series is decomposed on both a signal (trend-cycle and seasonality) and an error which is autocorrelated. Although from a theoretical point of view this is feasible, there is no method of estimation properly developed for time series following this type of model.

Computer programs for estimation of models with autoregressive residuals based on generalized least squares or maximum likelihood have been produced in the last few years. But they can only be used when the trend cycle follows a fairly low degree polynomial of time and when the seasonal factors are also polynomial functions of time for all the time range. Unfortunately, these requisites are fulfilled very rarely by economic time series.

On the other hand, the autoregressive integrated moving average model assumes that the generating mechanism of a time series has autocorrelated errors and that it is non-stationary either in its local level and/or slope. Therefore, by adequately differencing the process, one obtains a stationary process in the difference of a finite order. The stationary process is assumed to be the output obtained from an input that is white noise, which has been submitted to a linear transformation. This model is generalized by including a seasonal operator either in a multiplicative or additive way.

The identification of the parameters for this type of model is made by the analysis of the autocorrelation function. It has been shown however, that the autocorrelation function analysis is effective primarily for models that correspond to near normal or normal stochastic processes with high frequencies or small lags; that is, for models of short-run near-gaussian effects (Mandelbrot, 1972). Moreover, the final estimates for these models give the total for the signal and do not split this total into the amount that corresponds to the trend-cycle and the seasonal, so they are useful for different purposes than the ones of the error model. They are useful mainly for providing final estimates of the raw data and for forecasting. They can also be applied to estimate and forecast already deseasonalized series but then, great caution should be exercised in the identification and estimation since by correcting the raw data, distortions may be introduced that contradict the assumptions upon which the models rely.

The above considerations demonstrate that for empirical applications, particularly in economics, the types of time series models for which already exist optimal methods of estimation are very limited and they are far from being the best. It still requires a great amount of theoretical and empirical research for a thorough understanding of the problems raised by the analysis of economic time series.

BIBLIOGRAPHY

Books

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

(2) Bartlett, M.S. (1955) Stochastic Processes, Cambridge University Press

(3) Bensoussan, A. (1971) Filtrage Optimal des systèmes lineaires, Dunod, Paris

(4) Box, G. and G. Jenkins (1970) Time Series Analysis Forecasting and Control, Holden Day, California

(5) Bucy, R.S. and P.D. Joseph (1968) Filtering for Stochastic Processes with Applications to Guidance, Interscience, Wiley, N.Y.

(6) Cramer, H. and M.R. Leadbetter (1967) Stationary and Related Stochastic Processes, Wiley, N.Y.

(7) Doob, J.L. (1953) Stochastic Processes, Wiley, N.Y.

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

(9) Fishman, G. (1969) Spectral Methods in Econometrics, Harvard University Press

(10) Genedenko, B. (1966) The Theory of Probability, Chelsea Pub. Co. 3rd ed.

(11) Gikhman, I. and A.V. Skorokhod (1965) Introduction to the Theory of Random Processes, Nanka Moscow (in Russian)

(12) Gikhman, I. and A.V. Skorokhod (1968) Stochastic Differential Equations, Kiev (in Russian)

(13) Granger, C.W. and M. Hatanaka, (1964) Spectral Analysis of Economic Time Series, Princeton University Press

(14) Grenander, V. and M. Rosenblatt (1957) Statistical Analysis of Stationary Time Series, Wiley, N.Y.

(15) Hannan, E.J. (1960) Time Series Analysis, Methuen and Co., London

(16) Hannan, E.J. (1970) Multiple Time Series, Wiley, N.Y.

(17) Harris, B. (ed) (1967) Spectral Analysis of Time Series, Wiley, N.Y.

Books - Concluded

(18) Jazwinski, A. (1970) Stochastic Processes and Filtering Theory, Academic Press

(19) Jenkins, G. and D. Watts (1969) Spectral Analysis and Its Applications, Holden Day, California

(20) Kolmogorov, A. (1933) Foundations of the theory of probability (in Russian) translated in English by Chelsea Pub. Co. London - 3rd ed. (1956)

(21) Kushner, H.J. (1971) Introduction to Stochastic Control, Holt, Rinehart and Winston

(22) Lagrange, (1972-78) Oevres, Vol. 6

(23) Lange, O.C. (1963) Introduction to Econometrics, MacMillan

(24) Lions, J.L. (1969) Quelques Methodes de Résolution des Problèmes aux Limites Non-Lineaires, Dunod, Paris

(25) Moore, H.L. (1914) Economic Cycles: Their Law and Causes, MacMillan Co.

(26) Pannekoek, A. (1961) A History of Astronomy, Interscience, New York.

(27) Parzen, E. (1962) Stochastic Processes, Holden Day California

(28) Pugachev, U.S. (1966) Theory of Random Functions and its Applications to Control Problems, Pergamon Press

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

(30) Stuart, R.D. (1961) An Introduction to Fourier Analysis, Methuen, London

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

(32) Wainstein, L.A. and V.O. Zubakov (1961) Extraction of Signals from Noise, Prentice Hall

(33) Wiener, N. (1949) Extrapolation, Interpolation and Smoothing of Stationary Time Series, Wiley N.Y.

(34) Wold, H. (1938) A Study in the Analysis of Stationary Time Series. Almguist and Wiksell, Stochholm, 2nd ed. (1954)

(35) Yaglom, A.M. (1962) Theory of Stationary Random Functions, Prentice Hall

Articles

(1) Abel, M.E. (1962) "Harmonic Analysis of Seasonal Variation with an Application to Hog Production" J.A.S.A., 57, p. 655.

(2) Beveridge, W.H. (1922) "Wheat Prices and Rainfall in Western Europe" J. Roy. Stat. Soc. 85, p. 412.

(3) Brewer, K.R.W. (1969) "Box-Jenkins Models For Prediction and Seasonal Adjustment", Commonwealth, Bureau of Census and Statistics, Canberra.

(4) Buys-Ballot, G. (1847) "Les changements périodique de temperature" Utrecht.

(5) Dagum, C. and E.B. Dagum (1972) "The Meaning of Theory and Model in Social Sciences", Working Paper Series No. 72-9, The University of Iowa, 39 pages.

(6) Grether, D.M. and M. Nerlove (1970) "Some properties of Optimal Seasonal Adjustment", Econometrica, V. 38 - N. 5 p. 682.

(7) Hannan, E.J. (1960) "The Estimation of Seasonal Variation" the Australian Jour. of Stat. II p. 1.

(8) Hannan, E.J. (1963) "The Estimation of Seasonal Variation in Economic Time Series" J.A.S.A. LVIII - p. 31.

(9) Hannan, E.J. (1964) "The Estimation of a changing Seasonal Pattern" J.A.S.A. LIX p.p. 1063-77.

(10) Hannan, E.J., R.D. Terrell and N.E. Tuckwell (1970) "The Seasonal Adjustment of Economic Time Series" Intern. Ec. Review p. 24.

(11) Kalman, R.E. (1960) "A New Approach to Linear Filtering and Prediction Problems" Jour. Basic. Engineering Series D82-35 p. 35.

(12) Kalman, R.E. and R.S. Bucy (1961) "New Results in Linear Filtering and Prediction Theory" Jour. Basic. Engineering, Series D83-V. 35 p. 95.

(13) Khinchin, A.Y. (1934) "Korrelations Theorie der Stationarem Stochastischen Prozesse" Math. Ann. 109, p. 604.

(14) Khinchin, A.Y. (1938) "The Theory of Damped Spontaneous Effects" Izv. Akad. Nauk. SSR, Ser. 3 p. 313.

(15) Kolgomorov, A.N. (1939) "Sur l'interpolation et l'extrapolation des Suites stationaires" Compt. tend. acad. sci. Paris, v. 208 p. 2043.

(16) Kolgomorov, A.N. (1941) "The Local Structure of Turbelence in Incompressible Viscous Fluid for Very Large Reynolds numbers, Dokl, Akad, Nauk. SSR, Vol. 30, p. 301 - Reprinted in: Turbulence: Classic Papers on Statistical Theory, edited by S.K. Friedlander and L. Topping, Interscience Pub. N.Y. (1961).

Articles - Concluded

(17) Mandelbrot, B. (1972) "Statistical Methodology for Non-periodic Cycles from the Covariance to R/S Analysis, Annals of Ec. and Soc. Measurement, NBER, V. 1 No. 3, p. 259.

(18) Nerlove, M. (1964) "Spectral Analysis of Seasonal Adjustment Procedures" Econometrica, 32, No. 3, p. 241.

(19) Nerlove, M. (1965) "A Comparison of a Modified Hannan and the BLS Seasonal Adjustment Filters" J.A.S.A. 60, No. 310, p. 442.

(20) Nettheim, N.F. (1965) "Fourier Methods for Evolving Seasonal Patterns" J.A.S.A. Co. N. 310, p. 492.

(21) Parzen, R. (1967) "The Role of Spectral Analysis in Time Series Analysis" Rev. of the I.S.I., 35, No. 2.

(22) Rosenblatt, H.M. (1963) "Spectral Analysis and Parametric Methods for Seasonal Adjustment of Economic Time Series" Proc. Bus. Ec. Stat. Section of A.S.A., p. 94.

(23) Schuster, A. (1898) "On the Investigation of Hidden Periodicities with Application to a Supposed Twenty-six Day Period of Meteorological Phenomena", Terrestial Magnetism, Vol. 3, p. 13.

(24) Stokes, G.C. (1879) "Note on Searching for Periodicities" Proc. Roy. Soc. V. 29, p. 122.

(25) Slutsky, E.E. (1938) "Sur les fonctions aléatoires presque périodiques et sur la décomposition de fonctions aleatoires stationnaires en composantes" Actualités Scient. et Ind. 738, Paris.

(26) Terrel, R.D. and N.E. Tuckwell (1971) "Application of an Evolving Seasonal Pattern to an Australian Economic Time Series" Australian Economic Papers, Vol. 9, p. 186.

(27) Yule, V. (1921) "On the Time Correlation Problem" J. Roy. Stat. Soc. Vol. 84, p. 497.

(28) Yule, V. (1927) "On a Method of Investigating Periodicities in Disturbed Series" Trans. Roy. Soc. (A) p. 226.

(29) Zadeh, L. and J. Ragazzini (1950) "An Extension of Wiener's Theory of Prediction" J. of Applied Physics, Vol. 21, p. 645.

(30) Yaglom, A.M. (1955) "The Correlation Theory of Processes whose n.th difference constitute a Stationary Process" Matem. Sb. 37(79) p. 141.