Int. Stat. Rev., Vol. 44, No. 1, 1976, pp. 29–70/*Longman Group Ltd/Printed in Great Britain*

A Survey of Time Series

Spyros Makridakis

The European Institute of Business Administration, Fontainebleau 77305, France

"*All our measurements and observations are nothing more than approximations to the truth*".

(Gauss, *Theoria Motus*)

Summary

This paper surveys time-series methods developed and used in the fields of statistics, engineering, operation research and economics. It will describe each of the major methods now in existence and synthesize them into an integrated whole by pointing out similarities and differences existing among them. The applicability of the different methods to various real-life requirements will be discussed and further research needs will be raised.

Introduction

The last 40 years have witnessed an unprecedented growth in the area of time series. Statisticians, engineers, operation researchers and economists have departed from different starting points with a variety of needs and objectives in mind to develop a new and highly important area. Their work has resulted in a theoretically sound body of knowledge, with considerable practical value and widespread usage in many fields of endeavour.

It will be the purpose of this paper to survey and synthetize all major categories of time-series methods developed in the areas of statistics, engineering, operations research and economics. A main objective of the work that follows is to allow the time-series specialists of different fields to know of relevant work already available in others. It is hoped that this can facilitate communication and allow existing developments to be shared across disciplines; needless to say that such a task is very difficult at present because of the fragmentation of knowledge along field lines and the existence of a different terminology in each. Other objectives of the paper are to serve both as a guide to newcomers and as a reference to all important contributions to those already familiar with time series. Part I of this paper will provide the reader with a short background and a brief history of time series. Part II will examine autocorrelation and spectral techniques as the two major approaches to time-series analysis. It will then move to Part III which will describe the different time-series methods classified into the five categories of Autoregressive and Moving Averages (ARMA) processes, Filtering methods, Smoothing models, Time-series Decomposition methods and Miscellaneous techniques. Finally, a discussion of the main areas of time-series applications will follow and further areas of research will be suggested.

Time-series considerations originated in 1807, when the French mathematician Fourier claimed that any series can be approximated as the sum of sine and cosine terms. This idea was used by Schuster (1906) who applied Fourier's expansion to estimate the length of hidden periodicities and who widely utilized periodogram analysis in his work. The modern era in time series started in 1927 with Yule (1927) and achieved its major advances in 1938 when Wold (1954) developed a comprehensive theory of Autoregressive/Moving Averages (ARMA) schemes, around 1940 when Wiener and Kolmogoroff (1941) solved the estimation problem for continuous and discrete filters correspondingly, and in the early '60s when Kalman (1960) and Kalman and Bucy (1961) extended Wiener and Kolmogoroff's estimation procedures to non-stationary series involving systems in the time domain. On the Operation-Research side,

the late '50s and early '60s saw the development of exponential smoothing models (1961) which, simple as they were, became utilized by business firms and the military. In the area of decomposition, the utilization of digital computers opened a new era (1957) by allowing the cumbersome computations, done beforehand on a desk calculator, to be easily performed by the computer. Finally, spectral analysis appeared as a viable method of time-series analysis (1948; 1950). As with decomposition methods, the appearance of the digital computer made both spectral and autocorrelation methods practicable in the analysis of time series.

The progress made in the development of time-series methods has been due to the following factors:

(*a*) deep dissatisfaction with periodogram analysis as a practical tool;
(*b*) peripheral advances in mathematics, probability theory and statistical techniques;
(*c*) the need for reliable methods (mainly in engineering, economics and business) to estimate and predict the pattern, or behaviour, of different types of phenomena or systems.

The aim of the time-series approach is to separate the noise (randomness) from the pattern (behaviour) of a real process (system). This is achieved through two phases: (i) The first phase is called "time-series analysis" whose objective is to discover the characteristics of the real process or phenomenon generating the series. This is done by the utilization of either autocorrelation or spectral methods. The second phase includes several categories of methods which can be classified as autoregressive/moving average, transfer function, filtering, exponential smoothing, and decomposition. Each of the methods attempts to separate the noise from the observed value of time series in its own way; however, the ultimate objective is different: Autoregressive/moving-average schemes, exponential smoothing models and decomposition methods aim at forecasting; transfer-function models are usually used for control purposes, while filters are primarily used to estimate current states of real processes.

Before we proceed any further, a very brief description of the two phases and the time-series methods within each will be given to serve as an overall overview of what is to follow:

I. *Time-Series Analysis*, as the name implies, involves the analysis of data so that their characteristics (level of stationarity, length of seasonality, frequency, amplitude, phase) can be discovered. The analysis can be done in the time domain through the utilization of the *autocorrelation* function or in the frequency domain through the use of the *spectral function*.

II. *Time-Series Methods* include the following categories of techniques:

1. (*a*) *Autoregressive* (*AR*) *and Moving-Average* (*MA*) *Schemes* which assume that a given value of a time series is a weighted linear sum of past values and/or residual deviations, e_t. Thus, (1) is an autoregressive scheme (when v is equal to zero), (2) is a moving-average process, while (3) is of mixed autoregressive/moving-average character.

$$X_{t+v} = \sum_{i=1}^{p} W_i X_{t-i} + e_{t+v} = W_1 X_{t-1} + W_2 X_{t-2} + \ldots + W_p X_{t-p} + e_{t+v} \qquad (1)$$

where W_i represents the autoregressive parameters (also called "weighting", or "impulse response function").

$$X_t = e_t - \sum_{j=1}^{q} \theta_j e_{t-j} = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \qquad (2)$$

where e_t and e_{t-j} are the residual errors at period t and $t-j$ respectively, and j is the jth moving-average parameter.

$$X_t = \sum_{i=1}^{p} W_i X_{t-i} + e_t - \sum_{j=1}^{q} \theta_j e_{t-j}$$

$$= W_1 X_{t-1} + W_2 X_{t-2} + \ldots + W_p X_{t-p} + e_t - \theta_{t-1} - \theta_2 e_{t-2} \ldots - \theta_q e_{t-q} \quad (3)$$

where W_i, θ_j, e_t and e_{t-j} are defined as before.

1. (*b*) *Transfer-function Models* aim at linking the output of some process, Y_t, by several inputs X_t, X_{t-1}, X_{t-2}, ..., in such a way that the relationship between the inputs and outputs at previous time intervals can be estimated and subsequently used to control the process. A transfer functions is of the form:

$$Y_t + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_5 Y_{t-5} = \alpha_0 X_{t-d} - \alpha_1 X_{t-d-1} - \ldots - \alpha_r X_{t-d-r} \quad (4)$$

where α_i and β_i are the parameters of the process of the input and output correspondingly.

2. *Filtering Models* are special cases of (1) when $v = -1$. They are of particular interest to engineers in that they can be designed through electrical circuit, consisting of inductances and capacitancies, in such a way that no inputs beyond a certain range will be accepted.
3. *Exponential Smoothing* methods are special cases of autoregressive schemes when $v < -1$. Its weights or parameters, W_i, [see (1)] decrease according to some exponential fashion, thus the name "exponential" is utilized.
4. *Time-Series Decomposition* methods work by either "breaking up" the series into trend, seasonal and irregular components, or by passing varying frequency filters through the data to separate low, medium or high frequencies.

The above-mentioned methods present a rather wide variety of techniques which can be used in different areas of applications. Their cost accuracy, and complexity vary considerably in a way that can satisfy various needs existing in real-life applications. They will be described in Parts II and III of the paper.

I. A Short Background into Time Series

A time series, $X(t)$, is a set of observations ordered sequentially in time. Khintchine (1932, 1934) showed that it can be viewed as a random process of the variables X_1, X_2, X_3, ..., X_n, sampled at equidistant time intervals t_1, t_2, t_3, ..., t_n. Furthermore, Kolmogoroff (1933) showed that it is equivalent to a probability distribution in n number of dimensions. Thus, the theory of probability and statistics can be applied when dealing with time series. Each time point corresponds to one dimension in the distribution, and consists of two parts: X'_t, generated by the real process represented through the time series, and u_t, a white noise term. Thus,

$$X_t = X'_t + u_t \quad (5)$$

where X_t is usually expressed as deviation from the overall mean of the time series. The distribution of u_t is often assumed to be normal with:

$$E(u_t) = 0 \quad (6)$$

and

$$E(u_t u_{t+i}) = \begin{cases} \sigma_u^2 & \text{if } i = 0 \\ 0 & \text{if } i \neq 0 \end{cases} \quad (7)$$

1. *Fourier Series*

In 1807, Joseph Fourier showed that any time series, $X(t)$, can be expanded into a sum of sine and cosine terms which can approximate the series $X(t)$, as close as one wishes, provided

that the number of sine/cosine terms is large enough. In tribute to his highly significant discovery, the result of the expansion is called a "Fourier series".[1] Thus,

$$X(t) = \sum_{i=1}^{\infty} a_i \cos \frac{2\pi t}{T} i + \sum_{j=1}^{\infty} b_i \sin \frac{2\pi t}{T} j. \tag{8}$$

The estimation of coefficients a_i and b_i, called the "Fourier coefficients", can be calculated from the following definite integrals.

$$a_i = \frac{2}{T} \int_0^T X(t) \cos \frac{2\pi t}{T} \, dt \tag{9}$$

$$b_i = \frac{2}{T} \int_0^T X(t) \sin \frac{2\pi t}{T} \, dt. \tag{10}$$

From (8), (9), and (10), it seems that identifying the behaviour of a time series, or predicting its future values is indeed a trivial affair. But in practice, there are many problems such as the number of sine and cosine terms to be included, how to specify the value of T, or how to estimate the integrals of (9) and (10). The main difficulty, however, is that (8) cannot accommodate for the existence of noise [Bartlett (1935)]. In other words, it assumes that the time series is deterministic, or the noise, u_t, in (5) is zero. This is a great disadvantage because there are few noiseless series in the social sciences or engineering. This drawback, therefore, makes the utilization of Fourier series for prediction purposes nil.

2. *Search for Hidden Periodicities*

Since early in the nineteenth century [Herschel (1801)], interest has centred on the existence of common periodicities in certain natural phenomena and their length. A popular subject of the investigation has been the variation in sunspot activity and its influence on terrestrial phenomena. The method of finding the length of a hidden periodicity consisted of dividing the data into a number of subgroups [arranged horizontally one below the other called the "Buys-Ballot table" (1847)] whose length was somewhat longer than the suspected length of the periodicity, and then calculating the mean of the first element, of all subgroups, second, third, ..., last; subsequently the deviations of each of the calculated means from the overall mean were found. As the next step, the length of periodicity was assumed to be first shorter and then longer around the suspected length. The same process of finding the differences of the shorter and longer means from the overall mean was utilized. Finally, these deviations were plotted and the highest one was assumed to correspond to the real length of periodicity [Steward and Dodgson (1879)]. Such a job was very laborious and time-consuming. Consequently, it was suggested [Stokes (1879), p. 120] that the data should be analyzed in terms of a Fourier series whose coefficients, (9) and (10), could be calculated and then their square sum plotted. Since a_i and b_i of (9) and (10) will correspond to different periods of time, the length corresponding to the highest square sum: $c_i^2 = a_i^2 + b_i^2$ (called "intensity of the ith harmonic"; it is highest at the c_i corresponding to the true length of periodicity – see *Spectra Analysis*) of a_i and b_i will be the true length of periodicity. Although such a suggestion did not prove to be less time consuming, it opened a new and important approach to discovering hidden periodicities in time series: the periodogram.

The periodogram, as a way of discovering the length of hidden periodicities, carried considerable esteem and was used widely, but it also troubled its users. The drawbacks of periodograms became evident little by little: Schuster (1906), for example, found out that once he divided his data into two halves, the periodogram of each was quite dissimilar. He,

---

[1] See Carslaw (1930), Lanczos (1966) and Zymund (1959).

of course, attempted to explain the fact but with little success since the real reasons were not logical but mathematical.

Another classical utilization of periodogram analysis was made by Beveridge (1922). The result again showed wide discrepancies between the first and second halves of the data but in addition, it became clear how erratic a periodogram can be and, at the same time, sensitive to small changes in the length of period: a small decimal increase can result in a high peak, while another small increase can bring a trough. The obvious consequence was, of course, that one had to apply an extremely large number of decimal periods to make sure that no important peak was missed. These problems existed in addition to questions raised about the ability of periodogram analysis to make sure that a given peak was statistically significant.

3. *The autoregressive Moving-Average Approach*

Yule, in 1926, claimed that a Fourier series expansion cannot do an adequate job in terms of predicting a series or estimating the length of its periodicity because the length and amplitude (see section II.3 for definition) of real life series are not constant, an assumption made in the periodogram analysis. Yule (1927), in another article, argued that it would be more realistic to assume a type of disturbed harmonic series where "both length and amplitude can vary by reason of chance". He further showed that such a series can be better described as a function of its past values. In other words, he introduced the concept of autoregression, even though he restricted himself to an order of four or fewer terms [$p \leqq 4$ in (1)]. Yule's approach was extended by Walker (1931) who defined the general autoregressive scheme of equation (1). Later, Slutsky (1937) presented the moving-average scheme described by (2), and finally, Wold (1954) showed that any discrete, stationary time series X_t can be broken down into:

$$X_t = \sum_{i=1}^{p} W_i X_{t-i} - \sum_{j=1}^{q} \theta_j e_{t-j} - \sum_{r=1}^{h} \sum_{k=1}^{s} \gamma_{rh} \cos\left(\frac{2\pi tk}{T} + \lambda_{rk}\right). \tag{11}$$

For practical purpose, the last term of (11) is ignored, thus (3) is actually an approximation of (11) where its last term is random noise equal to e_t.

The works of Walker, Slutsky and Wold, in addition to the theoretical developments in mathematics, statistics and computational sciences, open the road to the modern approach to time series. This approach is based on the assumption that a time series can be adequately described in terms of previous values of the series itself and/or previous error terms.

Finally, Kendall (1945) showed beyond any doubt that the periodogram can give very misleading results. This was proved theoretically 5 years later by Bartlett (1950), who showed that sampling effects can distort the periodogram of a time series.

However, it was not the end of periodogram analysis. Drawbacks were overcome [see *Spectra Analysis* (II.3)] and the concept, slightly modified, is again used to analyze a time series in terms of its frequency characteristics. The improvement has been mainly in the direction of smoothing the periodogram (or spectrum as it is known at present) in such a way that it corresponds more accurately to its theoretical values [Bartlett (1948), (1950); Blackman (1958)].

II. Time-Series Analysis

If we are interested in the over-time behaviour of some phenomenon, we should observe and record its states at equidistant time intervals. This will result in a time series, as if was defined in Part I. It may be that our instruments are exact, yielding no measurement error, that there are no sampling errors, and that the output has always exactly the same pattern, i.e. there are no outside disturbances affecting the generating process. Such time series, called "deterministic", are very easy to deal with. The majority of time series, however, are not

deterministic. In real life situations, measurement errors, sampling effects and outside disturbances do always exist and are intermixed with the output of the generating process. The result is a corrupted time series called "stochastic" (if we assume that the errors, white noise, or randomness involved have zero mean, a constant variance and are not correlated with each other). That is, the assumptions (6) and (7) are sufficient to make (5) a stochastic time series. It is only the contaminated, with noise series, we can observe and study. In fact, the major purpose of analyzing and modelling time series is to be able to separate the noise from the output, thus, obtaining an uncorrupted series which "truly" represents the outcome of the generating process. Communication engineers call this outcome "the message", while statisticians prefer the term "pattern".

Once the separation of noise and message or pattern has been achieved, we can attempt to satisfy one of several objectives. We may, for example, be interested in part values of the generating process (intrapolation or smoothing) as in ballistics where we must know the past states of a target; we may want to know its present value (estimation or filtering) as in a receiver where the separation of the noise from the real message is sought; we may desire future values of the process (extrapolation or prediction), a typical example of which could be sales forecasting, or we may want to control the generating process through manipulation of "controllable" inputs. However, the achievement of the above four objectives will be accomplished as a byproduct of our ability to separate the noise from the observed values of the time series.

In order to be able to "uncontaminate" a time series from noise, we must (*a*) fit a model to the series and (*b*) make sure that the model fitted can describe the series correctly. The process of model fitting is like a double-edged knife: it can result in a model which does not account for all the existing noise, or equally bad, it can end up with a model which includes noise as part of its pattern. A Fourier series is a clear illustration of the inability of the model builder to draw the thin line between under and overestimation of the noise factor. He does not, in fact, have the ability to decide how many terms to include in order to specify the "correct" model.

The process of model building for the different time-series models available will be discussed in the next section; needless to say, the choice of models is rather large. Time-series analysis aims at providing guidance so that the model builder will not wander indefinitely, trying to choose the right model, or specifying the number of parameters needed. Furthermore, once the model is chosen and its parameters estimated, time-series analysis of the residuals can be employed to determine whether or not the model chosen represents adequately the time series and consequently the generating process. This is an indispensable task which makes the utilization of time-series model practical.

The above-mentioned aims of time-series analysis (clues for model specification, and tests for model adequacy) are fulfilled by two different but complementary methods: Autocorrelation Analysis (AA) and Spectral Analysis (SA). The former is mainly used by statisticians and operation researchers; it analyses the data in terms of their time characteristics (existence of seasonality, length of seasonality, stationarity, possible types of generating processes) The latter is mainly utilized by engineers to discover the frequency response characteristics of the generating process (amplitude, frequency, phase). In addition, AA enables us to determine whether or not the series is a stationary one, while SA allows us to estimate the gain function of a filter. Both are extremely useful for reasons which can be seen later on in this section.

Stationarity

A time series is stationary if it fluctuates around a constant mean. On the other hand, a non-stationary series includes a longer-term secular trend. The advantage of dealing with stationary

series is that their statistical properties are independent of time. The implications are two-fold: on the theoretical side, it is much easier to develop a theory of stationary, stochastic time series – in fact, it is the only theory in existence for such processes; from a practical point of view, it requires less work to estimate the parameters of a stationary model and, most important, once estimated, they are the same over a considerable time interval.

On the other hand, a large number of actual time series are not stationary; however, this does not present an insolvable problem, since there are several methods which, in fact, allow us to transform a non-stationary series into one which is indeed stationary.

A non-stationary time series includes a trend element which, in practice, can be represented by a function of time, a polynomial of the form:

$$T_t = a + b_1 t + b_2 t^2 + b_3 t^3 + \ldots. \tag{12}$$

It is obvious that if we can estimate (12) and then subtract it from (or divide it into) the time series X_t, the result will be a de-trended series; the theory of stationary stochastic process will therefore be applicable in such a series.

Estimating T_t in (12) can present a problem since we will have to decide on the degree of polynomial to be fitted. Furthermore, we must decide as to how many terms (observations) we would like to use in fitting the polynomial. It could, of course, be decided to use all terms but such a procedure is troublesome to update, since every time a new observation becomes available, the parameters of the polynomial in (12) must be re-estimated. Furthermore, the far distant observations will carry the same weight as the most recent ones. In other words, the fitting will not be responsive to current changes in trend. These disadvantages of fitting a trend to the entire range of observations have necessitated the utilization of polynomial trend fitting to a constant number of points. This implies the necessity of incorporating a new observation while dropping the least recent one, thus keeping the number of terms used constant. This approach is called the "method of moving average" by statisticians and "low-pass filter" by engineers.

(a) *Moving Averages*

A polynomial of second or third degree is usually fitted to a constant number of terms, its value being subtracted from the middle value of the moving average (the number of terms is usually chosen to be odd). This process starts in the beginning of the series and "rolls" until the last term has been included in the average. The result is a de-trended series with $n-m+1$ terms, where n is the number of observations, and m the number of terms included in the moving average. We can also have double moving averages (i.e. a moving average of a moving average) and triple moving averages (a moving average of a double moving average). For example, a five-term linear moving average takes the form

$$M_{t-2} = \frac{X_{t-4} + X_{t-3} + X_{t-2} + X_{t-1} + X_t}{5} \tag{13}$$

where M_{t-2} denotes the moving average and t varies from 1 to n, where n is the total number (past, present and future) of time-series terms.

The choice of the appropriate length of the moving averages cannot be determined by any set of rules; it is rather a subjective selection depending upon the specific time series or method. The better known moving averages are the Spencer's (1904) 15- and 21-point formulae [see Kendall (1946), pp. 371-394] about derivation and example of Spencer's and other types of moving-average formulae); and recently Henderson's 5-, 9-, 13- and 23-point weighted

moving averages [Shiskin *et al.* (1967)]. The Spencer's 15-point weighted[1] moving average [see (14)], for example, can remove up to cubic trend and requires a total of 15 terms which remain constant as new values appear and old ones are thrown away:

$$\begin{aligned} M'_{t-7} = & -0{\cdot}009X_{t-14}-0{\cdot}019X_{t-13}-0{\cdot}016X_{t-12}+0{\cdot}009X_{t-11}+0{\cdot}066X_{t-10}+0{\cdot}144X_{t-9} \\ & +0{\cdot}209X_{t-8}+0{\cdot}231X_{t-7}+0{\cdot}209X_{t-6}+0{\cdot}144X_{t-5}+0{\cdot}66X_{t-4}+0{\cdot}009X_{t-3} \\ & -0{\cdot}016X_{t-2}-0{\cdot}019X_{t-1}-0{\cdot}009X_t. \end{aligned} \tag{14}$$

(b) *Differencing*

First and second-order differencings are sufficient, in practice, to remove any kind of trend existing in a time series. A first-order difference will be:

$$X_t = Y_t - Y_{t-1} \tag{15}$$

where Y_t is a non-stationary time series, while a second-order difference takes the form:

$$X_t = Y_t - 2Y_{t-1} + Y_{t-2}. \tag{16}$$

Equations (15) and (16) act as high-pass filters on the trended data.

In addition to non-stationarity in the mean of time series, we can also have non-stationarity in the variance. The latter can become stationary more easily by transforming the data into a logarithmic scale (we may sometimes have to transform the time series into a fraction of a power).

Different methods deal differently with non-stationarity, ARMA schemes, for example, detect stationarity through the examination of the autocorrelation function. The Census II method of decomposition assumes that a trend exists and then applies Spencer's 15-point formula (the X-11 version of the Census II uses a Henderson's 5-, 9-, 13- or 23-point weighted moving averages depending upon the extent of the random component) to remove it, while exponential smoothing methods can account directly for trend by introducing an appropriate number of polynomial parameters in the model. Whatever the procedure, the outcome should be obvious: time-series theory cannot deal directly with trend because it introduces spurious correlations, so it must be removed, or accounted for separately, before any method can be applied. We shall, therefore, in the rest of this paper, assume that the series is or has been converted to a stationary one.

1. Autocorrelation Analysis

Given the stationary time series:

$$X_t = X'_t + u_t \tag{17}$$

if we assume that u_t is normally distributed, then X_t can be described by its lower order moments (mean, variance, autocovariance and autocorrelation), whose estimates are:

$$E(X_t) = \frac{\sum_{t=1}^{n} X_t}{n} = \bar{X} \tag{18}$$

$$E((X_t - \bar{X})^2) = \frac{\sum_{t=1}^{n} (X_t - \bar{X})^2}{n-1} = \sigma^2 \tag{19}$$

$$E((X_t - \bar{X})(X_{t+k} - \bar{X})) = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{n} = \gamma_k \tag{20}$$

[1] The weights of the 15-term moving average result from a 5-term [like the one of (13)] of a 5-term, of a 4-term, of a 4-term quadruple moving average.

$$\frac{E\left((X_t-\bar{X})(X_{t+k}-\bar{X})\right)}{E((X_t-\bar{X})^2)} = \frac{\sum_{t=1}^{n-k}(X_t-\bar{X})(X_{t+k}-\bar{X})}{\sum_{t=1}^{n}(X_t-\bar{X})^2} = \frac{\gamma_k}{\sigma^2} = \rho_k. \tag{21}$$

If we accept the convention to substitute $X_t = X_t - \bar{X}$, (20) and (21) become:

$$\gamma_k = \frac{\sum_{t=1}^{n-k} X_t X_{t+k}}{n} \tag{22}$$

$$\rho_k = \frac{\sum_{t=1}^{n-k} X_t X_{t+k}}{\sum_{t=1}^{n} X_t^2}. \tag{23}$$

Autocorrelations are measures of the relationship between successive values of a variable ordered in time, and as such, they are the backbone of time-series analysis. They vary from -1 to $+1$, and because of stationarity, are even, i.e. $\rho_{-k} = \rho_k$. The utilization of the concept of correlation goes back to Poynting (1884) who attempted to ascertain the relationship between cotton and silk imports into England, and several other factors over time. However, it was Hooker, (1901) who, in order to measure the relationship between the marriage rate and the index of trade in Britain employed (21). Since then, autocorrelations have been used widely to determine the following:

(a) *Existence of Stationarity*

The autocorrelation coefficients of a stationary time series damp out to zero quickly. If this is not the case, the first difference (or some form of moving average applied to remove the trend) should be taken and the autocorrelations of the de-trended series found, If they damp out to zero quickly, it means that the differenced series is stationary; otherwise the second difference should be found. Any series met in practice need no more than two differences to become stationary.

(b) *The Length of Seasonality*

This can be detected by looking at a graphical plot of the autocorrelation coefficients. Unless the random component of (17) is extremely strong, it will be quite easy to recognize constant periodic patterns, if they exist, of course. If the length of the seasonal pattern is equal to T, then T, $2T$, $3T$, ... should be the higher valued autocorrelation coefficients in relation to the rest.

(c) *Various Tests of Significance*

Ordinary tests of significance are not valid in time series because successive observations are not independent of one another [Bartlett (1935); Yule (1921)]. However, special tests which take dependence into account have been developed. The distribution of a single correlation coefficient was derived by Anderson (1942) who showed that it is normally distributed with zero mean and $\frac{1}{n}$ variance, while Dixon (1944) developed likelihood-ratio criteria for testing different hypotheses. This was followed by Bartlett (1946) who approximated the standard error to be:

$$\sigma(\rho_k) = \sqrt{\frac{1}{n}\left(1+2\sum_{k=1}^{q}\rho_k^2\right)} \tag{24}$$

where q is the highest order non-random autocorrelation coefficient.

In the case of the series being completely random, (24) reduces to

$$\sigma(\rho_r) = \frac{1}{\sqrt{n}} \tag{25}$$

which is the standard error derived by Anderson.

Quenouille (1949) extended the results to include joint distributions of ρ_1, ρ_2, ..., ρ_k and as well as partial correlation coefficients (see definition below), whose standard errors above order q are independently distributed and can be approximated by:

$$\sigma(\theta_r) = \frac{1}{\sqrt{n}}. \tag{26}$$

Finally, Box and Pierce (1970) developed approximate tests (χ^2) capable of determining whether or not several autocorrelation coefficients, say the first F, are significantly different from zero. This χ^2-test has $F-p-q$ degrees of freedom (where p is the order of the Auto-Regressive (AR) models, and q, that of the Moving-Average (MA) models). It can be approximated by:

$$\chi^2 = n\sum\rho_i^2. \tag{27}$$

(27) can easily be used to determine whether the autocorrelations of the residuals are white noise. This will be true if the χ^2 calculated by (27) is smaller than the corresponding value for the degrees of freedom, from the table of χ^2 values, at some significant level.

(d) *Initialization of the Parameters of the AR and MA Models*

If both sides of (1) are multiplied by X_{t+k} and then expected values are taken [see Brown (1963) or Buys Ballot (1847)], we have:

$$\gamma_k = W_1\gamma_{k-1} + W_2\gamma_{k-2} + \ldots + W_p\gamma_{k-p}. \tag{28}$$

If (28) is divided by σ^2 [see Surman (1965) or Carslaw (1930)], the result will be:

$$\rho_k = W_1\rho_{k-1} + W_2\rho_{k-2} + \ldots + W_p\rho_{k-p}. \tag{29}$$

Substituting $k = 1, 2, 3, \ldots, p$, we have:

$$\left.\begin{aligned}
\rho_1 &= W_1 + W_2\rho_1 + W_3\rho_2 + \ldots + W_p\rho_{p-1} \\
\rho_2 &= W_1\rho_1 + W_2 + W_3\rho_1 + \ldots + W_p\rho_{p-2} \\
\rho_3 &= W_1\rho_2 + W_2\rho_1 + W_3 + \ldots + W_p\rho_{p-3} \\
&\vdots \\
\rho_p &= W_1\rho_{p-1} + W_2\rho_{p-2} + W_3\rho_{p-3} + \ldots + W_p
\end{aligned}\right\} \tag{30}$$

(30) is a system of p simultaneous equations known as the Yule-Walker equations [Yule (1926); Walker (1931)] with p unknowns (W_i) whose values can be found by solving (30). These values are used as initial estimates for the AR (p) model to be fitted in the data. For an AR(1) model, for example,

$$W_1 = \rho_1. \tag{31}$$

For an AR (2):

$$W_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \tag{32}$$

$$W_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}. \tag{33}$$

For an AR(3):

$$W_1 = \frac{\begin{vmatrix} \rho_1 & \rho_1 & \rho_2 \\ \rho_2 & 1 & \rho_1 \\ \rho_3 & \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \tag{34}$$

$$W_2 = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & \rho_2 & \rho_1 \\ \rho_2 & \rho_3 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \tag{35}$$

$$W_3 = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \tag{36}$$

etc., where the bracket $\|$ denotes determinants.

Similarly, initial estimates for an MA (q) process can be determined. Multiplying both sides of (2) by X_{t+k} and taking expected values, we have:

$$\gamma_k = (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \ldots + \theta_{q-k}\theta_q)\sigma_e^2. \tag{37}$$

Thus, the autocorrelation coefficients are:

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \ldots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2} \tag{38}$$

for $K = 1, 2, \ldots, q$.

Equation (38) is a system of q simultaneous equations with q unknown. However, unlike (30) it is not linear. Thus, solving for θ_k is more difficult and must be done recursively. However, there are tables [Box and Jenkins (1970), pp. 517–519] which provide initial estimates for low-order MA processes, given the autocorrelation coefficients.

Partial autocorrelations are similar to partial correlations (measuring the degree of association between two variables, keeping all other variables constant) and are defined as the last coefficient, W_i, of (29) for different values of k. Thus, the partial coefficient of time lag 1 is given by (31), of time lag 2 by (33), time lag 3 by (36), etc. The partial autocorrelations are usually denoted as W_{11}, W_{22}, W_{33}, W_{44}, etc.

Determining the Order of the ARMA Model

By examining the partial and autocorrelation function we can identify a tentative model (AR, MA, or mixed ARMA) and its order (1, 2, 3, ..., p; 1, 2, 3, ..., q). In general, "whereas the autocorrelation function of an autoregressive process of order p tails off, its partial-autocorrelation function has a cut-off after lag p. Conversely, the autocorrelation function of a moving-average process of order q has a cut-off after lag q, while its partial autocorrelation tails off. If both the autocorrelations and partial autocorrelations tail off, a mixed process is suggested. Furthermore, the autocorrelation function for a mixed process, containing a pth-order autoregressive component and a qth-order moving-average component, is a

mixture of exponential and damped sine waves after the first $q-p$ lags. Conversely, the partial-autocorrelation function for a mixed process is dominated by a mixture of exponentials and damped sine waves after the first $p-q$ lags" [Box and Jenkins (1970), p. 175]. More details and illustrations of how to identify an ARMA process and its degree can be found in Box and Jenkins (1970, pp. 53–84).

2. Spectral Analysis

The Fourier series described by (8) consists of a large number of harmonics of frequency i/T, where i corresponds to harmonic terms and T is the length of periodicity ($1/T$ is called "the fundamental harmonic").

Each of the ith harmonic is orthogonal, with amplitude (value of ordinate at its peak or through taken from the mean value): $c_i = \sqrt{a_i^2 + b_i^2}$. Assuming a finite number of terms, n, the equivalent of (8) expressed in terms of its amplitude, c, is:

$$X(t) = \sum_{i=1}^{n/2-0\cdot5} c_i \cos\left(\frac{2\pi t}{T} i + \alpha_i\right) \tag{39}$$

where n is odd, and α_i is the phase angle of the ith harmonic defined as:

$$\tan \alpha_i = -\frac{b_i}{a_i}. \tag{40}$$

The equivalents of (9) and (10) for a_i and b_i are:

$$a_i = \frac{2}{n} \sum_{t=1}^{n} X_t \cos \frac{2\pi t}{T} i \tag{41}$$

$$b_i + \frac{2}{n} \sum_{t=1}^{n} X_t \sin \frac{2\pi t}{T} i \tag{42}$$

and

$$c_i = \sqrt{a_i^2 + b_i^2}. \tag{43}$$

The maximum number of harmonic terms is $\frac{n}{2} - 0\cdot5$ [*see* (39)].

The periodogram is a histogram-type graph with a function of c_i as its ordinate (different linear functions of c_i have been used by various writers), and the discrete frequencies i/T as its abscissa. Obviously, assuming no sampling or other types of errors, the highest value of c_i, or any function of it, will correspond to, say, frequency $\frac{r}{T}$, and period $\frac{T}{r}$. The term T/r is, therefore, the length of hidden periodicity. Subsequent lower peaks of the ordinate will correspond to secondary periodic occurrences. Such is the information one can obtain from the periodogram.

A common way to define the ordinate of periodogram is through its intensity, $I(i/T)$ of the fundamental frequency $\frac{1}{T}$. That is:

$$I(i/T) = \frac{n}{2} c_i^2. \tag{44}$$

An important property of (44) is that each term of the periodogram adds to the overall variance by $\frac{1}{2}I\left(\frac{i}{T}\right)$, i.e.:

$$\sigma = \tfrac{1}{2}I\left(\frac{1}{T}\right) + \tfrac{1}{2}I\left(\frac{2}{T}\right) + \ldots \tfrac{1}{2}I\left(\frac{\frac{n}{2}-0\cdot5}{T}\right).$$

(44) assumes that the frequencies $\frac{i}{T}$ are integer harmonics of the fundamental frequency $\frac{1}{T}$. Obviously, the frequency will vary as $\frac{1}{T}, \frac{2}{T}, \frac{3}{T}, \ldots, \frac{\frac{n}{2}-0{\cdot}5}{T}$. If we lift this assumption and allow the frequency to vary continuously in the range of zero to 0·5 (where 0·5 is the highest frequency, called "Nyquist", corresponding to a period of 2) we obtain the spectrum of frequency f. Thus:

$$I(f) = \frac{2}{n}(a_f^2 + b_f^2), \quad 0 \leqq f \leqq \tfrac{1}{2}. \tag{45}$$

Expression (45) relates to the autocovariance function through equation (46), which expresses the correspondence between the autocorrelation and spectral functions [Kolmogoroff (1941); Wiener (1930); Wise (1955); Wold (1954)].

$$I(f) = 2\left(\gamma_0 + 2\sum_{k=1}^{n-1} \gamma_k \cos 2\pi f k\right) \quad 0 \leqq f \leqq \tfrac{1}{2}. \tag{46}$$

That is, the spectrum is the Fourier cosine transform of the autocovariance function. Furthermore, (46) is the starting point on which spectral analysis is based and is also much easier to calculate than (41), (42) or (43), since the autocovariance or autocorrelation coefficients have been computed anyway.

If we divide both sides of (46) by the variance of the process σ^2 we will get the following standardized measure:

$$\frac{I(f)}{\sigma^2} = S(f) = 2\left(1 + 2\sum_{k=1}^{n-1} \rho_k \cos 2\pi f k\right) \tag{47}$$

where $S(f)$ is called the "spectral density function". (From now on, f will always vary continuously from 0 to 0·5, unless otherwise specified.)

The spectral density function (47) is the Fourier transform cosine of the autocorrelation function and has two advantages over (44): (*a*) it is continuous over the entire range of frequencies, and (*b*) it can be applied as an expected value, which means that the amplitude, frequency and phase need not be fixed as in Fourier analysis. That is, (47) can be applied to subparts of data; thus obtaining several values for $S(f)$, the average $\bar{S}(f)$ of each frequency f can be plotted, resulting in a smoothed spectral density function [Bartlett (1948) (1950); Daniell (1946)] which has none of the disadvantages of the classical periodogram analysis. Smoothing the spectral density function is done through kernels or windows by which one can vary the "bandwith" of the estimates until all the information contained in the spectrum has been revealed.

The smoothing of the spectral density function (47) is done by the application of:

$$\bar{S}(f) = 2\left(1 + 2\sum_{k=1}^{n-1} \zeta_k \rho_k \cos 2\pi f k\right) \tag{48}$$

where ζ_k is called a "lag window" [Blackman and Tukey (1958)] and takes different values depending upon the number of groups, H, into which one would like to divide the data and the type of spectral window one would like to utilize. Obviously, H, the number of data subsets, can vary from 1 to $\frac{n}{2} + 0{\cdot}5$; in the last case, every subgroup includes two observations. Jenkins and Watts [Jenkins and Watts (1968), pp. 239–243] have shown that the variance of $\bar{S}(f)$ reduces as H increases. However, for large H the bandwidth is small, which means smoothing which hides the details of the spectrum. In addition, it introduces bias. Thus, one has to

compromise between smaller variance (large H) and too much smoothing or larger bias (small H). The compromise is not always easy to make, however, and depends upon several factors; nevertheless, there are some "empirical" rules one can follow [Jenkins and Watts (1968), pp. 258–298].

The number of subgroups, H, into which the data must be divided, should be decided upon by examining the autocorrelation function. H should be large enough to embrace all non-zero autocorrelations. A set of three different H values, H_1, H_2 and H_3, should be chosen to cover a fairly wide range of subgroups. This can be done by fixing the ratio H_3/H_1 to around 4. Another factor to be determined is the spacing of the frequency f. It should be equal to $0{\cdot}5/F$, where 0·5 is the highest frequency and $F = 2H$ or $F = 3H$. The smoothed spectral density function can then be calculated by:

$$\bar{S}\left(\frac{h}{F}\right) = \sum_{h=0}^{F} 2\left(1+2\sum_{k=1}^{H-1} \zeta_k \rho_k \cos 2\pi \frac{hk}{F}\right). \tag{49}$$

(49) should be used to calculate the smoothed spectrum for H_1, H_2 and H_3 which can be plotted on the same graph with the bandwith corresponding to each H_i.

The final choice of H will be made for the subgroup H_i which provides enough details and cannot be considerably improved (in terms of more details) by decreasing its bandwith.

A final consideration is the choice of ζ_k whose value is specified by what kind of spectral window one would like to use. For example, in Bartlett's spectral window, the lag window ζ_k is:

$$\zeta_k = 1 - \frac{h}{H}. \tag{50}$$

Thus, if $H = 3$, the value of ζ_k in (49) will be:

$$\bar{S}\left(\frac{h}{F}\right) = \sum_{h=0}^{F} 2\left(1+2\left(\frac{2}{3}\right)\rho_1 \cos 2\pi \frac{h}{f} + 2\left(\frac{1}{3}\right)\rho_2 \cos 2\pi \frac{2h}{F}\right). \tag{51}$$

A survey of the different types of windows, as well as review of spectral methods, in general, can be found in Jenkins (1961), pp. 145–148), (1956, pp. 11–16) and Jenkins and Priestley (1957, pp. 5–7).

It was said that the spectrum is the Fourier transform of the autocorrelation function [see Harrison (1965)]. The opposite is also true; therefore, in mathematical terms, knowledge of either the autocorrelation of spectrum functions is equivalent. It depends upon the user and his need to decide whether to utilize one or the other. Statisticians are usually inclined to calculate the autocorrelation coefficients, arguing that they are much simpler to compute, easier to interpret and provide as much information as the spectrum. Furthermore, they say that one needs to know the autocorrelation function before one can calculate the spectrum; so calculating the spectrum and going through the laborious and inexact process of constructing windows is a waste of time.

Engineers, on the other hand, insist that the spectrum is a more natural quantity to compute because it expresses a time series in terms of its frequency response which must be known for design purposes. They say that statisticians play down the value of spectral analysis mainly because they cannot think in frequency terms. Whatever the truth, the fact is that spectral analysis has found little use in the social sciences in general, mainly because it is troublesome to calculate and interpret. A few exceptions are those of Granger (1964), Burman (1965), Nerlove (1965), Rosenblatt (1963), [also see Grether and Nerlove (1970) for discussion], all of which are in the area of decomposition. However, it should be emphasized that the autocorrelation and spectral functions are related to each other, and that knowledge of one implies a knowledge of the other.

In engineering, spectral-analysis applications centre on the fact that the spectra of the input and output of a linear system are connected by (52), where $G(f)$ is the gain function between input(s) and output(s).

$$\gamma_{1k}(f) = \gamma_{2k}(f)\, G^2(f). \tag{52}$$

If both the input and output spectra are known, then the gain function can be easily estimated as:

$$G^2(f) = \frac{\gamma_{1k}(f)}{\gamma_{2k}(f)}. \tag{53}$$

Similarly, if the spectral of the input is known, together with the gain function, the spectral of the output can be computed and subsequently used to design the impulse response function (the coefficients of (1), W_i) of linear filters. As such, spectral analysis plays a critical role in univariate and multivariate linear-system designs.

Multivariate Time-Series Analysis

So far, we have been concerned with the autocorrelation and spectral functions of single series; however, the analysis can be extended to cover m time series:

$$X_{(j)t}, \quad j = 1, 2, \ldots, m. \tag{54}$$

The covariance between the $X_{(i)t}$ and $X_{(j)t}$ series will be denoted by $\gamma_{(ij)k}$ and will be equal to:

$$\gamma_{(ij)k} = \frac{\sum_{t=1}^{n-k} X_{(j)t} X_{(i)t+k}}{n} \quad k = 0, 1, 2, \ldots, n-1 \tag{55}$$

$$\gamma_{(ji)k} = \frac{\sum_{t=1}^{n+k} X_{(i)t} X_{(j)t-k}}{n} \quad k = 0, -1, -2, \ldots, -n+1 \tag{56}$$

where $X_{(j)t}$ is expressed as a deviation from its mean.

The autocorrelation between series $X_{(i)t}$ and $X_{(j)t}$ will be called "cross autocorrelation", or just "cross correlation", and will be equal to:

$$\rho_{(ij)k} = \frac{\sum_{t=1}^{n-k} X_{(j)t} X_{(i)t+k}}{\sqrt{\sum_{t=1}^{n} X_{(j)t}^2} \sqrt{\sum_{t=1}^{n} X_{(i)t}^2}} \quad k = 0, 1, 2, \ldots, n-1 \tag{57}$$

$$\rho_{(ij)k} = \frac{\sum_{t=1}^{n+k} X_{(i)t} X_{(j)t-k}}{\sqrt{\sum_{t=1}^{n} X_{(j)t}^2} \sqrt{\sum_{t=1}^{n} X_{(i)t}^2}} \quad k = 0, -1, -2, \ldots, -n+1. \tag{58}$$

The cross correlation has all the properties of the autocorrelation function and in addition:

$$\rho_{(ij)k} = \rho_{(ji)-k}. \tag{59}$$

Thus, the cross-correlation function needs not be symmetrical. Like the autocorrelation function, it can be used to analyse a univariate set of data, whereas the cross-correlation function can achieve the same results for multivariate sets. Its use is mainly directed to transfer-function models where the dynamic relationship between inputs and outputs of different variables is analyzed. Finally, the cross-spectral function is similar to the univariate spectral in the sense that it is the Fourier transform of the cross-correlation function. However,

the formulae involved are more complicated; the interested reader should be referred to Jenkins and Watts (1968, pp. 363–412).

III. Time-Series Method

Time-Series Methods can be divided into five categories (see *Introduction*) of Autoregressive and Moving Average (ARMA), Filtering, Smoothing, Decomposition and a category including some models not otherwise classified. Wold's decomposition theorem (1954) as expressed in equation (11) is the basis of all categories, with the exception of decomposition methods and some individual models which are based on other considerations.

The major task of time-series methods is to separate the noise from the observed time series in such a way that the pattern or message will become known. Thus, the characteristics of the generating process will be revealed and could, consequently, be used for smoothing, filtering, predictions or control. Each of the categories separates the noise and pattern through its own procedure; even then, there are considerable similarities. ARMA models are the most general of all the categories since they assume that the last term of (11) is not important in determining the pattern of the series which is therefore approximated by:

$$X_t = \sum_{i=1}^{p} W_i X_{t-i} - \sum_{j=1}^{q} \theta_j e_{t-j} + e_t. \tag{60}$$

Whatever is left between X_t and the first two terms of (11) will be considered the residual e_t. The aim of ARMA models is to choose the appropriate order of p and q in such a way that the residual error e_t is random. When e_t is tested and shown to be random, the separation is assumed to be complete and the pattern is expressed as an ARMA (p, q) model.

Filtering models are autoregressive in character only, that is:

$$X_t = \sum_{i=0}^{p} W_i X_{t-i} + e_t. \tag{61}$$

The separation of the noise is done in two steps: first some initial estimates, which isolate as much of the noise as possible, are made and subsequently recursive algorithms which continuously update the parameters W_i are applied to make e_t as random as possible.

Smoothing models, too, are mainly of moving average type. They smooth (average) residual error values of many periods according to some exponential weighting scheme. The outcome of the averaging (assuming non-systematic errors) will result in errors close to zero, which will therefore reveal the true pattern. Smoothing models are special cases of ARMA ones; however, there is a major difference: whether (60) can be utilized to fit any time series, there is not a single smoothing model, but many. Thus, the appropriate model has to be chosen and utilized for each specific series.

The principle of decomposition models is different from those already discussed. It is based on the assumption that economic or business time series consist of three main subpatterns or components. These are Trend (T), Seasonality (S) and Cyclicality (C), in addition to Randomness (R). Thus, if we assume a multiplicative relationship:

$$X_t = T \times S \times C \times R \tag{62}$$

the objective of decomposition is to isolate each of the three components of (62) until the remaining residual (R) becomes white noise. The process of "breaking down" X_t into T, S and C is done differently by the various methods which will be discussed later. However, the basic aim is always similar in terms of isolating each one of the three components as accurately as possible.

Transfer-function models, finally, are similar to ARMA ones, except that they apply to multiple, rather than single, time series.

When the noise of a time series has been eliminated, the uncontaminated series can be used to approximate past, present and future values of the series or to control the generating process. These are the end results of utilizing time-series methods and will be discussed in section IV after a survey of the five categories of methods has been completed in this section.

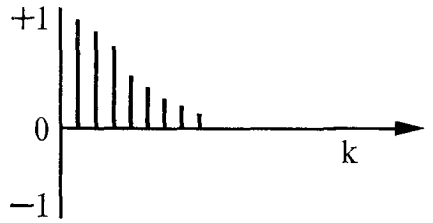
1. *Autoregressive and Moving Averages (ARMA) Schemes*

Yule (1927), Walker (1931) and Slutzky (1937) introduced the concept of autoregressive/moving-averages schemes. It was the work of Wold, however, which proved the theoretical validity of the method and devised general representation for times series. Wold's contribution in this field was of the utmost importance and in many respects he should be considered the founder of ARMA models. Since then several major contributions have been made, which have enabled the practical utilization of ARMA schemes. These contributions have been achieved in two directions: first, in devising an efficient estimation procedure for the AR, MA, and mixed ARMA processes, and secondly, in extending the results to include seasonal time series.

On the estimation side, to which Wold did not give much consideration, Kolmogoroff (1941) suggested general solutions to the smoothing and prediction problem. It was, however, Mann and Wald (1943) who used the method of maximum likelihood estimation for the solution of autoregressive processes. They extended the method of maximum likelihood estimation used in regular regression to include the parameters of autoregressive processes. The difficulty existing in the latter category is, of course, that the variables are not independent of one another. Furthermore, Mann and Wald showed that on the limits the maximum likelihood estimates are normally distributed. Finally, they discussed their belief that the estimates have the same optimal properties as when independence among variables does exist; they did not, however, provide proof of the last statement. Whittle (1953) extended the concept of ARMA models to cover multiple time series, while Durbin (1959; 1960), devised efficient methods of estimating the AR and MA parameters. Later, Walker (1962) extended the result to mixed ARMA schemes.

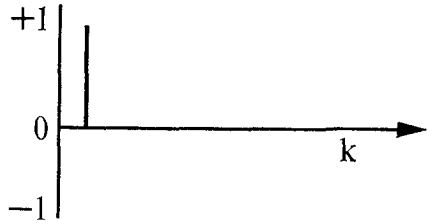
More recently, Jenkins and Watts (1968) and Box and Jenkins (1970) proposed models capable of dealing with seasonal series, devised efficient computational formulae for digital computers and provided procedures to deal with any kind of series, whether stationary or not. Box and Jenkins have greatly contributed to making ARMA models more widely accepted by those other than specialists in control engineering or statistics. This reason, together with the purely theoretical contributions, has resulted in their name being used synonymous with autoregressive and moving-average models. However, Box and Jenkins have been neither the originators nor the most important contributors in this field. This has also been true with Kalman whose name has been used synonymous with filters, although there have been others who had made significant contributions in the field [see Anderson (1971), p. 98; or Sorenson (1973), p. 66].

The Box-and-Jenkins procedure in utilizing autoregressive/moving-averages schemes is applied in two stages. The first stage includes a general class of model called "integrated autoregressive moving-average schemes". The models included can apply to seasonal or non-seasonal data as well as stationary or non-stationary series. The most general model can be seen in equation (63), where any parameter can be zero, thus allowing a more specific model. The second step is to identify an adequate model to be fitted into the series. This is done through an examination of the autocorrelation and partial-correlation functions (see Figure 1), which completes Stage 1. Once a tentative model has been identified, its parameters are estimated, and its adequacy is tested against the real time series. The residuals, that is, the difference between the actual and the predicted by the tentative-model values, are analyzed through autocorrelation to determine if they are randomly distributed. If they are, this will

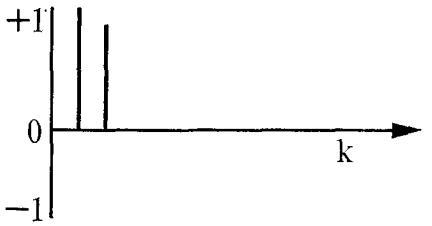


Autocorrelation function of an AR(1) model (exponentially-damped autocorrelation function, cut-off after k time lags).

Autocorrelation function of an AR(2) model (exponentially-damped autocorrelation function, cut-off after k time lags).

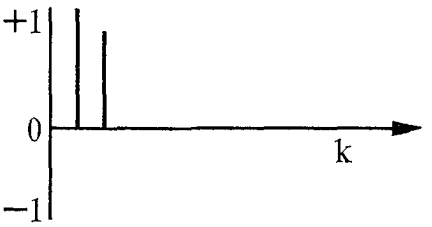


Partial-autocorrelation function of an AR(1) model (one partial autocorrelation significantly different from zero).

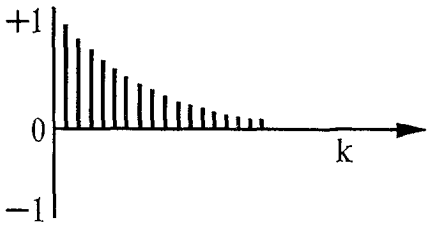


Partial-autocorrelation function of an AR-(2) model (two partial autocorrelations significantly different from zero).

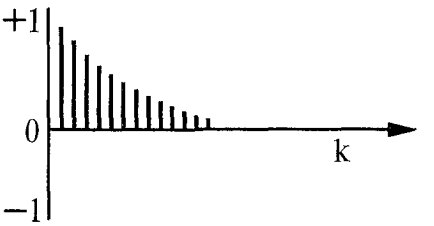
Autocorrelation function of an MA(1) model (one autocorrelation significantly different from zero).



Autocorrelation function of an MA(2) model (two autocorrelations significantly different from zero).



Partial-autocorrelation function of an MA (1) model (exponentially damped autocorrelations, cut-off after k time lags).



Partial-autocorrelation function of an MA-(2) model (exponentially-damped autocorrelations, cut-off after k time lags).

Figure 1. *Examples of model identification through the Autocorrelation and Partial Autocorrelation Functions.*

imply that the tentative model is correct, and therefore, should be used for forecasting or control purposes. Otherwise, an alternative model should be identified and subsequently estimated, because the one initially selected is not an adequate representation of the time series. Through this process, we will eventually find a model, whose residual will be random noise and which could be used in forecasting or control. The Box-and-Jenkins approach,

through this step-by-step procedure (see Figure 2), provides sufficient clues for identification of an initial model and then at later stages, gives us sufficient information to know if the appropriate model has indeed been chosen.

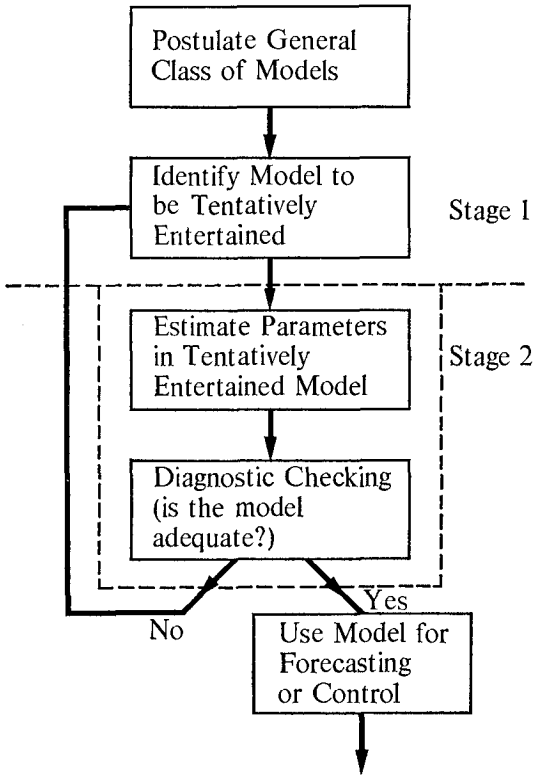


Figure 2. *The Box–Jenkins methodology to ARMA models.*

The Box-and-Jenkins approach can deal with seasonal time series by specifying a seasonal parameter in either the autoregressive ϕ_s, or the moving-average aide, θ_s:

$$(1+\phi_1 B+\phi_2 B^2+\ldots+\phi_p B^p)(1+\phi_s B^s)\, X_t = (1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q)(1-\theta_s B^s)e_t \quad (63)$$

where

$$BX_t = X_{t-1}$$
$$B^2 X_t = X_{t-2}$$
$$B^3 X_t = X_{t-3}$$
$$\cdot\ \cdot\ \cdot$$
$$\cdot\ \cdot\ \cdot$$
$$\cdot\ \cdot\ \cdot$$
$$B^m X_t = X_{t-m}.$$

(63) can include any number of short (i.e. $(1-B)\, X_t$, $(1-2B+B^2)\, X_t$, etc.) or long

(i.e. $(1-B^s)\, X_t$, $(1-2B^s+B^{s+1})\, X_t$, etc.)

differences so that X_t will be stationary. That is, (63) can deal with short-term non-stationarity (that is, month-to-month) or long-term non stationarity (that is, year-to-year) by taking the appropriate number of short or long differences. In final analysis, the Box-and-Jenkins approach can deal with the most general of the time-series cases. It is the generality and theoretical

vigorousness of the method which has resulted lately in widespread interest in the method and its applicability.

Transfer-function models are generalized presentations of (63), when there are more than one variable included in the equation [see Barham and Humphries (1970)]. Transfer-function models involve the utilization of the cross-autocorrelation and cross-partial-autocorrelation functions and considerably more complicated procedures for model identification and estimation than those already described for univariate time series. The major application of transfer-function models is in the area of control and the forecasting with leading indicators [Box and Jenkins (1968)].

2. *Filters*

The pioneer work of Wiener (1949) and the independent but similar developments of the Russian mathematician Kolmogoroff (1941) have provided the starting point and basis of the modern filtering theory. Both Wiener's and Kolmogoroff's works centred around ways of estimating the white noise, u_t, in (64), so that the original message, or pattern, X'_t, could be recaptured. (Wiener treated the continuous case while Kolmogoroff dealt with discrete time series.) They developed an estimation procedure in the frequency domain which allowed the design of optimal filters operating on a continuous basis. Their work was an extension of Fisher's (1912) estimation procedures but instead of being based on the maximum likelihood function, it was based on the frequency response characteristics of a system.

$$X_t = X'_t + u_t. \tag{64}$$

Recapturing X'_t can be achieved by solving the integral Wiener-Hopf equation (65), for the weighting function (or impulse-response function) $W(v)$. However, such a solution is difficult to obtain; thus the applicability of the Wiener-Kolmogoroff filters is limited. Originally, the solution was obtained by using spectral factorization.

$$\gamma_{xx'}(k) = \int_{-\infty}^{+\infty} W(v)\gamma_{xx}(k-v)dv \tag{65}$$

where γ denotes, as before, the autocovariance.

Levinson (1947) provided some discrete approximations in order to obtain $W(v)$ in (65). Furthermore, he devised recurrence equations to obtain updated estimates of $W(v)$ when new data became available. However, the matrix inversion still required is computationally impractical when there are many observations involved. Zadeh and Ragazzini (1950) simplified the problem of solution at least for the finite memory case, independently by Bode and Shannon (1927) and they provided a simplified approach for the solution of the Hopf-Wiener equation by introducing the idea of "shaping" filters. This was followed by Booton (1952) who dealt with non-stationary time series and by several others who discovered and theoretically proved relationships between the weighting function and the variance of the white noise, u_t. It was, however, the work of Kalman (1960) and Kalman and Bucy (1961) who broke new ground in filtering theory and opened the road to the widespread utilization of filters in a variety of engineering applications.

Kalman and Bucy cirumvented the problem of solving the Wiener-Hopf integral equation by transforming it into its equivalent differential equation. Then, instead of an analytic solution, obtained laboriously through the calculus of variation, they passed the burden of solution to the computer. Furthermore, they introduced efficient recurrence equations which make the updating of the equations particularly easy by a digital computer. Unlikely Wiener and Kolmogoroff who worked in the frequency domain, they involved their system description in the time domain and introduced state-space notation which offers considerable conceptual and computational advantages.

The Kalman filtering system is described by equations (66) and (67) and can be solved by equations (68), (69), (70) and (71).

$$X'_{k+1} = \phi_{k+1,k} X'_k + Z_k \tag{66}$$

$$X_k = H_k X'_k + u_k \tag{67}$$

where X'_k is the true state of the system (uncorrupted of noise);

Z_k is the white noise affecting the linear system at time t_k (process noise);

$\phi_{k+1,k}$ is the transition matrix relating X'_k to X_{k+1};

X_k is the measurement (actual) data obtained at time t_k;

H_k is the observation matrix at t_k;

and u_k is the white noise caused by measurement errors at time t_k (measurement noise).

An estimate of $\hat{X}'_{k/k}$ is computed from the data X_0, X_1, X_2, ..., X_k so as to minimize the mean-square error. This is accomplished by the utilization of X_k and all previous estimates, $\hat{X}'_{k-1/k-1}$, $\hat{X}'_{k-2/k-2}$, ..., of X_k.

That is,

$$\hat{X}'_{k/k} = \phi_{k,k-1} \hat{X}'_{k-1/k-1} + K\,(X_k - H_k \phi_{k,k-1} \hat{X}'_{k-1/k-1}) \tag{68}$$

where K is calculated by (69) and is chosen in such a way as to minimize the mean-square error in the predicted estimate.

$$K_k = P_{k/k-1} H_k^t\,(H_k P_{k/k-1} H_k^t + R_k)^{-1} \tag{69}$$

where

$$P_{k/k-1} = \phi_{k/k-1} P_{k-1} P_{k-1/k-1} \phi_{k,k-1} + Q_{k-1} \tag{70}$$

(Q: covariance of Z_k)

and

$$P_{k/k} = P_{k/k-1} - K_k H_k P_{k/k-1}. \tag{71}$$

Equations (68), (69), (70) and (71) are known as the Kalman filter, although it was reported by Sorenson (1968) and Anderson (1971) that others had produced similar results before Kalman. In particular, they refer to a paper by Swerling (1958) which, in all but a quite insignificant respect, was identical to the later work of Kalman and Bucy (1961) on discrete time filtering. However, as with Box and Jenkins, Kalman and Bucy received most of the credit mainly by making their respective approach known to wider audiences.

The Kalman-Bucy filters exhibit a smaller variance than the Wiener-Kolmogoroff but they are not always easier to implement than the latter [Singer and Frost (1969)]. Apart from this, the Kalman-Bucy approach is different from that of Wiener-Kolmogoroff in four major respects:

(*a*) The Wiener-Kolmogoroff theory requires stationarity whereas the Kalman-Bucy does not.

(*b*) The Kalman-Bucy filters demand that the process-noise and the measurement-noise covariances be known. The Wiener-Kolmogoroff does not.

(*c*) The Kalman-Bucy method requires detailed knowledge of the transition matrix $\phi_{k+1,k}$ while the Wiener-Kolmogoroff requires knowledge of the cross autocovariances γ_{xx} (k) and γ_{xx} (k) (the autocovariance).

(*d*) The Kalman-Bucy method requires initial estimates of all system parameters.

(*a*) is an advantage for the Kalman-Bucy approach, (*b*) is for the Wiener-Kolmogoroff, (*c*) depends upon the information available while (*d*) can be a negative factor for the Kalman-Bucy approach.

Requirements (*b*) and (*d*) above have been partially resolved by obtaining on-line estimates of both the noise and system parameters. Mehra (1971) distinguishes four different approaches to get round the problem (linear-least square, stochastic approximations, maximum-likelihood correlation techniques) and provides ways of efficient estimation [Mehra (1970)]. A survey of such techniques can be found in Sage and Husa (1969).

Finally, the utilization of filters outside engineering of space applications has been limited mainly because of (*b*), (*c*) and (*d*) above. The few exceptions [Harrison and Stevens (1971); Schneeweiss (1971)], however, have been extremely promising and should initiate a wider research effort of how to profitably apply filters to such areas as business or economics where there is a great need and unlimited potentials of computationally easy procedures.

3. *Smoothing models*

Smoothing models are moving average in character although there are special cases involving parameters which weight exponentially past values. Their main advantages are simplicity and low utilization cost. Sometimes, they are referred to as "adaptive", implying their ability to adjust their smoothed value depending upon the magnitude and nature of the error [see equation (75) or (76)]. Finally, smoothing models are of infinite memory type because of their ability to store all past information in a single, or a few, summary measures. Equation (73) is, in this respect, an expansion of (72) and shows how the measure $\hat{X}_{t-1}$ contains all information from the second term backwards. On the negative side, there are many smoothing models available which implies that (*a*) the right smoothing model must be selected, and (*b*) once a change in data has taken place it must be sensed, so that another model can be employed.

Exponential smoothing was first introduced by Holt (1957) and was elaborated and widely applied by Brown (1959) and Holt *et al.* (1960). Other important contributors in the area have been Winter (1960), who introduced the seasonal-exponential-smoothing model. Brown and Meyer (1963) who presented a theoretical backing of the smoothing methods, Theil and Wage (1964) who express smoothing models stochastically and therefore show how the weights can be selected in such a way that the mean-square-prediction error will be minimized. Finally, Trigg (1964) introduced tracking signals to indicate when the characteristics of the process have changed significantly and Harrison (1965) elaborated Harmonic-Smoothing models. Next we shall examine the major exponential-smoothing models now in existence.

Single Exponential Smoothing

This smoothing can be expressed through equation (72) which can by itself become the forecasting equation too.

$$\hat{X}_t = \alpha X_{t-1} + (1-\alpha)\hat{X}_{t-1} \tag{72}$$

$$F_{t+1} = \hat{X}_t$$

where F_{t+1} denotes forecasting for period $t+1$. If (72) is expanded by substituting the value of $\hat{X}_{t-1}$, we have:

$$\hat{X}_t = \alpha X_t + \alpha(1-\alpha) X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \alpha(1-\alpha)^3 X_{t-3} + \ldots \tag{73}$$

or

$$\hat{X}_t = (1-\alpha) \sum_{i=0}^{\infty} a^i X_{t-i}. \tag{74}$$

α can be chosen to minimize the sum of squared errors between F_t and X_t.

To show the idea of adaptiveness we can write (72) by rearranging terms as:

$$\hat{X}_t = \hat{X}_{t-1} + \alpha (X_{t-1} - \hat{X}_{t-1}) \tag{75}$$

or

$$\hat{X}_t = \hat{X}_{t-1} + \alpha e_{t-1}. \tag{76}$$

Since $X_{t-1} - \hat{X}_{t-1}$ is the residual error, (76) shows how the smoothed (or forecast) value $\hat{X}_t$ will adapt depending upon the direction of error.

Non-Stationary Exponential Smoothing

Equations (72) or (73) can be used for stationary data only, a fact which can be seen by taking expected values of both sides of (72), in which case $E(\hat{X}_t) = \overline{X}$. To deal with non-stationary data which can be linear, quadratic, cubic, etc., special models have been developed to account for each non-stationarity correspondingly.

Holt's Two-Parameter Linear Model

Holt (1957) proposed the following model:

$$\hat{X}_t = \alpha X_{t-1} + (1-\alpha)(\hat{X}_{t-1} + \hat{b}_{t-1}) \tag{77}$$

where

$$\hat{b}_t = \beta (\hat{X}_t - \hat{X}_{t-1}) + (1-\beta)\, \hat{b}_{t-1} \tag{78}$$

and

$$F_{t+\tau} = \hat{X}_t + \hat{b}_t (\tau) \tag{79}$$

where τ is the number of periods ahead (lead time) we would like to forecast.

Brown's One-Parameter Linear Model

Brown's (1959) model is similar to Holt's, except that there is a single parameter for the basic pattern as well as for the trend smoothing. Harrison (1965, p. 111) concludes what many had suggested: that Brown's method is at least as good as Holt's and it should therefore be preferred for non-seasonal forecasting. Brown's model can be expressed through the following three equations:

$$\hat{X}_t = \alpha X_{t-1} + (1-\alpha)\hat{X}_{t-1} \tag{80}$$

$$\hat{\hat{X}}_t = \alpha \hat{X}_t + (1-\alpha)\hat{\hat{X}}_{t-1} \tag{81}$$

and where $\hat{\hat{X}}_t$ is a double exponential smoothing and

$$F_{t+\tau} = 2\hat{X}_t - X_t + \frac{\alpha}{1-\alpha}(\hat{X}_t - X_t)\tau.$$

Brown's One-Parameter Quadratic Smoothing

Quadratic smoothing should be utilized when there is a higher order trend than linear. The equations involved are the following:

$$\hat{S}_t = \alpha X_t + (1-\alpha) S_{t-1} \tag{82.1}$$

$$\hat{\hat{S}}_t = \alpha \hat{S}_t + (1-\alpha)\hat{\hat{S}}_{t-1} \tag{82.2}$$

$$\mathring{S} = \alpha \hat{\hat{S}}_t + (1-\alpha)\mathring{S}_{t-1} \tag{82.3}$$

$$F_{t+r} = F_1 + rF_2 + \tfrac{1}{2}r^2 F_3 \tag{82.4}$$

where

$$F_1 = 3\hat{S}_t - 3\hat{\hat{S}}_t + \mathring{S}_t \tag{82.5}$$

$$F_2 = \frac{\alpha}{2(1-\alpha)^2}((6-5\alpha)\,\hat{S}_t - (10-8\alpha)\,\hat{\hat{S}}_t + (4-3\alpha)\mathring{S}_t) \tag{82.6}$$

$$F_3 = \frac{\alpha^2}{(1-\alpha)^2}(\hat{S}_t - 2\hat{\hat{S}}_t + \mathring{S}_t) \tag{82.7}$$

and

$\hat{S}_t$, $\hat{\hat{S}}_t$, $\mathring{S}_t$ are simple, double and triple smoothing correspondingly.

Similarly, cubic smoothing will involve quadruple smoothed statistics and include four smoothing equations. Quadratic smoothing, however, can deal with almost all practical cases.

Higher orders of non-stationarity will require similar formulae with higher algebraic complexity. Their development, a generalized approach, formulae, and examples, can be found in Brown's (1963) book.

Winter's Linear and Seasonal Model

Winter's (1960) model is an extension of Holt's; it includes an additional parameter as well as an extra equation for seasonality. Thus, we have:

$$\hat{X}_t = \alpha \frac{X_t}{\hat{S}_{t-L}} + (1-\alpha)(\hat{X}_{t-1} + \hat{b}_{t-1}). \tag{83}$$

where L equals the length of seasonality.

$$\hat{b}_t = \beta\,(\hat{X}_t - \hat{X}_{t-1}) + (1-\beta)\hat{b}_{t-1} \tag{84}$$

and

$$\hat{S}_t = \gamma \frac{X_t}{\hat{X}_t} + (1-\gamma)\,\hat{S}_{t-L} \tag{85}$$

In order to forecast we apply (86);

$$F_{t+\tau} = (\hat{X}_t + \tau\hat{b}_t)\,\hat{S}_{t-L+\tau}. \tag{86}$$

In addition to Winter's three-parameter models, there are single-parameter models capable of dealing with both seasonality and trend: however, they are less known, although according to Groff (1973), they are at least as accurate as Winter's multiparameter model.

Harrison's Harmonic Smoothing Model

Harrison's (1965) model combines characteristic of decomposition methods, harmonic analysis and smoothing. It first removes a linear trend by an L-month moving average (if the simple moving average is converted into a weighted one, higher trends than straight lines can be dealt with. This has already been done in some programmes [see Harrison (1965), p. 134]. It then divides the moving average into the original data to obtain crude seasonal coefficients, S_j, the average of which is forced to L and then used for the following calculations:

Smoothed Seasonal Coefficients

This is done through (87):

$$\hat{S}_j = 1 + \sum_{k=1}^{L} (a_k \cos kx_j + b_k \sin kx_j) \tag{87}$$

where

$$a_k = \frac{2}{L} \sum_{j=1}^{L} S_j \cos kx_j \tag{88}$$

$$b_k = \frac{2}{L} \sum_{j=1}^{L} S_j \sin kx_j \tag{89}$$

and

$$x_j = 2(j-1)\pi/12 - \pi$$

while the summation of (87) includes only those harmonics, a_k and b_k, which are statistically significant.

The updating of the seasonal factors $\hat{S}_j$ can be achieved on a continuous basis as new data becomes available. Given a new data point, X_t, for example, we calculate

$$P_t = \frac{X_t}{M_t}$$

where M_t is the general level of the series at time t (trend-cycle).

We can then calculate updates for (88), (89) and (87) as follows:

$$a_k(t) = a_k(t-1) + \frac{2\delta}{n}(P_t - \hat{S}_j(t-1)\cos kx_j \tag{90}$$

$$b_k(t) = b_k(t-1) + \frac{2\delta}{n}(P_t - \hat{S}_j(t-1)\sin kx_j \tag{91}$$

and

$$\hat{S}'_i(t) = \hat{S}'_i(t-1) + \frac{2\delta}{L}(P_j - \hat{S}_j(t-1) \sum_{k=1}^{L} \cos(x_i - x_j). \tag{92}$$

(The notation for equation (90) to (92) has been slightly modified to accommodate the subscript i – denoting a seasonal period – and t denoting time.)

In addition to the multiplicative model just described, Harrison developed formulae for the additive case.

Tracking Signals

Tracking signals aim at automatically monitoring the forecasting results in order to ensure that they remain in control, i.e. they fluctuate around zero. Brown (1963) introduced a tracking signal which could be used in association with the single exponential smoothing of (72). Its form is:

$$TS_t = \frac{E_t}{M_t} \tag{93}$$

where TS_t is the tracking signal at period t.

$$E_t = \sum_{i=1}^{t} e_i$$

and is updated by:

$$E_t = \sum_{i=1}^{t-1} e_i + e_t \tag{94}$$

and M_t is the mean absolute deviation which is updated by:

$$M_t = \gamma \mid e_t \mid + (1-\gamma) M_{t-1}. \tag{95}$$

Trigg (1964) improved (93) by defining

$$E_t = \gamma e_t + (1-\gamma) E_{t-1}. \tag{96}$$

(96) fluctuates between -1 and $+1$, and has a standard deviation of

$$\sigma = \frac{1.2\alpha}{\sqrt{2\alpha - \alpha^2}}. \tag{97}$$

(97) allows us to establish confidence intervals and to know when the forecasting errors are out of control. Furthermore, it is stable and standardized between -1 and $+1$.

Trigg and Leach (1967) added another feature in the usage of equation (93) which greatly improved the response rate of exponential smoothing. They proposed that the smoothing constant α should be equal to the modulus of the tracking signal. Thus,

$$\alpha_t = | TS_t |. \tag{98}$$

(98) increases the value of α when the error becomes large, while decreasing it in cases where they are small. However, this is exactly the process in which one needs to have a more accurate and responsible smoothing system. Thus, (93), (95), (96) and (98), can be used in conjunction with (72) and achieve a truly adaptive forecasting method which is easy to compute and update and requires little data in order to utilize. However, it may be that the resulting behaviour of the forecast has some undesirable characteristics, as it was pointed out in [McClain and Thomas (1973)].

Trigg and Leach attempted to utilize the idea of (98) in higher order smoothing models but with less success – the model was often unstable. Furthermore, there have been other efforts [Rao and Shapiro (1970); Farley and Hinich (1970)] which, however, have been more complex to utilize and less successful than those of Trigg (1964) and Trigg and Leach (1967) in their attempts to track significant changes in the data.

4. *Time-Series Decomposition*

Time-series decomposition is one of the oldest methods in the area of time series. It was originated around the beginning of this century and initiated from two different directions: in the first place, it was recognized that in order to study the serial correlation within or between variable(s); it was necessary to eliminate the spurious correlation present because of the existence of trend. Poynting as early as 1884, attempted to eliminate the trend as well as some seasonal fluctuations by averaging prices over several years. Its purpose was to eliminate both gradual and more rapid movements as much as possible from the data. Hooker (1901) followed his example but he was more precise in the way of eliminating trend. He was followed by Anderson (1914) and Spencer (1904) who generalized the procedure of trend elimination to include higher order polynomials. On the other hand, there were the economists who were worried about the effect of depressions and who wanted ways of predicting them. By doing so, they felt that the elements of economic activity should be separated so that changes in the business cycle could be isolated and hopefully known beforehand. In France, a committee was appointed, which in 1911, presented a report [Government Report (1911)] dealing with the reasons and causes of the economic crisis of 1907. It introduced the idea of leading and coincidental indicators [Government Report (1911), p. 38] and attempted to separate the trend from the cycle so that the movement of the latter could be followed.

The method was elaborated upon in the USA, and the idea of constructing barometers of business activity was developed. Furthermore, an attempt to separate the seasonal fluctuation from the rest of the components was made [see Copeland (1915), pp. 554–556]. The process of decomposition was refined by Macauley (1930) who, in the 1920s, introduced the ratio-to-moving-averages method, variations of which are the most widely used today. It consists of three basic steps: first, the calculation of the seasonal factors, which are the ratios of the actual to a 12-month moving average, then, through some form of averaging the seasonal factors, we obtain 12 seasonal indices; secondly, the trend is calculated for each point and

thirdly, the trend is divided into the moving averaged data with the result of obtaining the cyclical factors.

That is, assuming a multiplicative relationship (an additive model will behave in a similar way), whose seasonal pattern is of 12-period duration, we have:

$$X_t = S_t T_t C_t R_t \tag{99}$$

where S_t is the seasonal component,

T_t is the trend component,

C_t is the cyclical component,

and R_t is the random component,

the first step consists of:

$$M_t = T_t C_t \tag{100}$$

where M_t is a 12-month moving average of X_t which of course eliminates the seasonality and a great deal of randomness by the process of averaging itself:

$$\hat{S}_t = \frac{X_t}{M_t} = \frac{S_t T_t C_t R_t}{T_t C_t} = S_t R_t \tag{101}$$

where $\hat{S}_t$ are the seasonal factors.

Through averaging (101) we eliminate the randomness R_t; so the remaining can be considered the seasonal indices, i.e.

$$\hat{S}_j = \hat{\hat{S}}_t \tag{102}$$

Step two is:

$$\hat{T}_t = a + bt \tag{103}$$

(or any higher order trend fitting), where $\hat{T}_t$ is the trend line at each point t, fitted to the original data X_t, and a and b are the least-squares regression coefficients.

Finally, in step three, we divide (103) into (100) to obtain $\hat{C}_t$, an estimate of the cyclical factors, i.e.

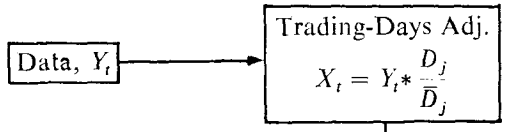
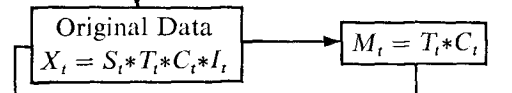
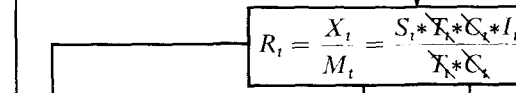
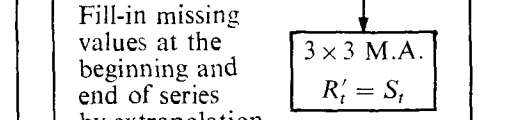
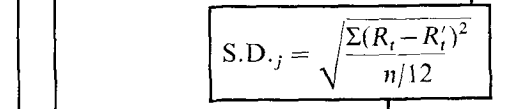
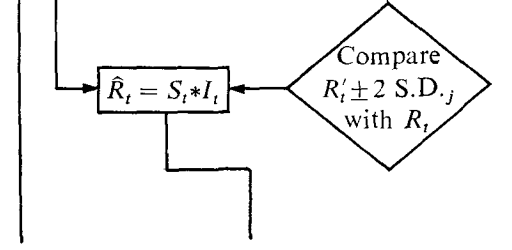
$$\hat{C}_t = \frac{T_t C_t}{T_t} = C_t. \tag{104}$$

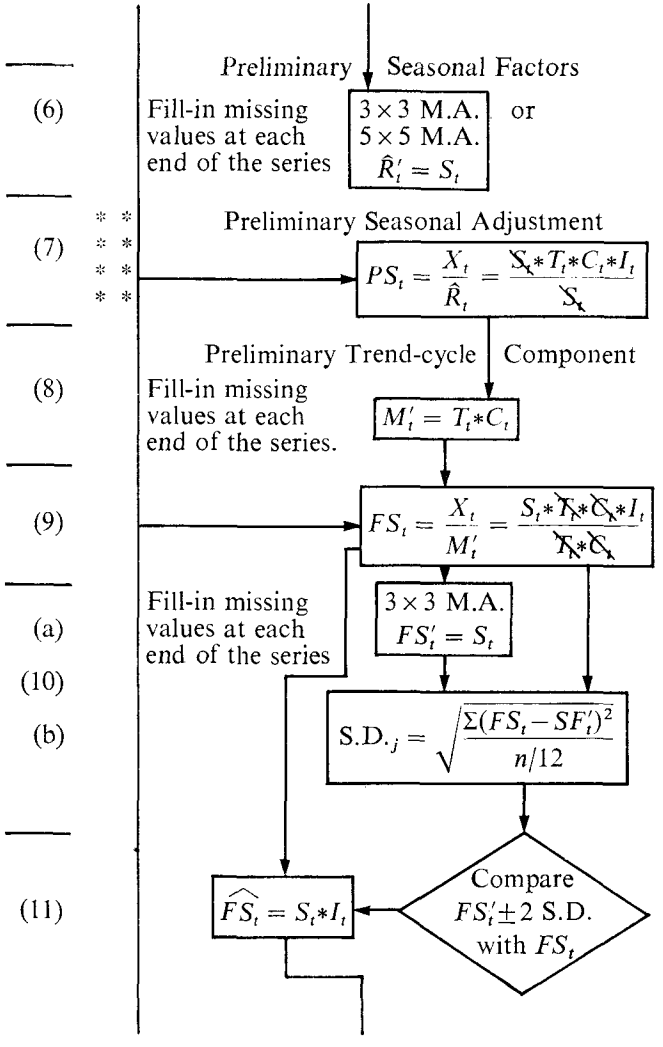
Equations (102), (103) and (104) present, therefore, the decomposition of original time series X_t into seasonal, trend and cyclical elements, correspondingly.

The ratio-to-moving-averages is the basis of the Census II method which has been developed and used all over the world, mainly by governmental agencies. In 1954 the manual procedure was replaced by a computer program (Census I) which was modified and enlarged in 1955, and since then known as Census II. Since 1955 there have been several variants of Census II starting from X-1 to the currently used X-11. They are different from Macaulay's original method in that (*a*) they use improved moving-averages schemes in order to compute the seasonal factor and the trend-cycle curves, (*b*) they replace the values lost in the beginning and end of the series because of the moving averages, (*c*) they eliminate extreme values, (*d*) they take into account trading-day variations, (*e*) most important, they isolate each component one at a time, by first calculating crude estimates which are further refined, and finally (*f*) they allow the user several options [see Shiskin (1957) (1961); Shiskin *et al.* (1967); McLaughlin (1962). (See also Figure 3.)

The Census II method first appeared in 1955 and was "essentially an electronic-computer adaptation of manual methods that have been developed and tested over a period of 30 to 40 years" [Shiskin (1961), p. 1]. Between 1955 and 1960 the method was extensively tested and in 1960 the X-3 version was made available to the public. During 1960 and 1961 several variations of the X-3 version of Census II were tested, both by the bureau of the Census of

Figure 3. *Major tasks in Census II.*

Steps		Description
(1)		Multipiy the data of each month by the Trading-Days (T-D) adjustment coefficient, $D_j/\bar{D}_j$ where D_j is the T-D for each of the months, and $\bar{D}_j$ is the average, over all years, for this particular month.
(2)		Calculate a 12-month, centred, Moving Average (12-MMA), its purpose is to eliminate the seasonality and some of randomness. Thus the Trend-cycle will remain.
(3)		Divide the M.A. into the original data, (1)/(2). The result is: Ratios of seasonal – irregular components.
(a)		Calculate a 3 × 3 M.A. (a 3-term M.A. of a 3-term Moving Average – double M-A) to each month separately. Its effect will be to eliminate the randomness in (3).
(4) (b)		Calculate the standard deviation (S.D.) for each month separately (12 in all) by summing up the square difference between R_t (M.A. with randomness ± (3)) and R'_t (M.A. without randomness (4.a)).
(5)		Replace extreme values of R_t, (3). Compare R'_t plus, or minus, two standard deviations with R_t, if R_t is greater than $R'_t + 2$S.D. or smaller than $R'_t - 2$S.D., its value is replaced as the average of the preceding and following value. Thus, the effects of extraordinary events (strikes, floods, wars, etc.) are eliminated.



(6) Calculate a 3×3 or a 5×5 M.A. of each month separately, in (5); the result is the elimination of randomness.

(7) Divide the data by the seasonal factors of (6). The outcome is a series which is preliminarily adjusted for seasonality.

(8) Apply the Spencer's 15-month weighted M.A. to (7). The result is a smooth series (with almost all of randomness eliminated) which highlights the Trend-cycle component.

(9) Divide M'_t, (8), into the original data; the effect is the same as in Step (3). However, (8) did not include seasonality to start with.

(a) as in Step (4a)

(10)

(b) as in Step (4b)

(11) as in Step (5)

Figure 3 (*continued*).

Steps		Description
(12)	Final Seasonal Factors Fill-in missing values at each end of the series 3×3 (or 5×5) M.A. $\hat{F}S'_t = S_t$	as in Step (6)
(13)	Final Seasonally Adj. Series $FA_t = \frac{X_t}{\hat{F}S'_t} = \frac{S_t * T_t * C_t * I_t}{S_t}$	Series adjusted for seasonality (final result).
(14)	Fill-in missing values at each end of the series $FA'_t = T_t * C_t$	Apply Spencer's 15-point weighted M.A. on (13); the outcome is to eliminate the randomness, thus achieving a refined estimate of Trend-cycle (final result).
(15)	$RC_t = \frac{FA_t}{FA'_t} = \frac{S_t * C_t * I_t}{T_t * C_t}$	RC_t is an estimate of the random component. It is found by dividing (13) by (14).
(16)	– Perform Different Test to make sure that the estimation of the Seasonal, Trend-cycle, and random components are correct. – Compute Months for Cyclical Dominance (MCD). – Calculate Summary Statistics.	Use original data and those of steps (13), (14) and (15) to obtain results.
(17)	– Compute a Moving Average whose length is equal to the months for cyclical dominance. Update this average to obtain estimates of trend-cycle. – Use the seasonal factors projected in (12) as seasonal indices.	Use the seasonally adjusted data of step (13). Update this moving average to easily obtain current estimates of Trend-cycles (i.e. it is used instead of FA'_t, in (14)).

the United States government and many other organizations. This resulted in minor revisions and modifications which were labelled X-4 to X-8.

By the end of 1961 the development of Census II reached a stable state and two versions became available to a wide number of users: the X-9 and X-10. X-9 is the standard program while X-10 can be used for erratic data. Since 1965 another version was introduced, the X-11, which is widely used today by government agencies and private organizations. X-11 differs from X-9 and X-10 in (*a*) providing different options to the user, (*b*) allowing the choice between an additive or multiplicative model, (*c*) producing full output or summary measures, (*d*) permitting choice of the σ limits for identifying and replacing extreme values, (*e*) specifying the type of moving average to be used and (*f*) enabling sequential additive and multiplicative adjustments. However, the differences between X-9 and X-11 are not great in essence but rather in the options allowed the user and the greater volume of output and complexity of use.

Figure 3 describes briefly the major tasks involved in Census II. Though the description does not refer to any specific version, it is closer to X-9 than the currently used X-11. As the reader can see from Figure 3, there is a large number of empirical rules which cannot be justified, except by saying that the method has been tested in tens of thousands of cases and has produced satisfactory results, from the point of view of accuracy.

The Census II method has been attacked on several fronts, Apart from having a large number of empirical rules applied, the method has been criticized on the following:

(*a*) It can overadjust the time series (i.e. it can attribute part of the randomness to some of the component(s) of the time series [Nerlove (1965)].
(*b*) It can distort the different components through the moving averages utilized [Burman (1965)].
(*c*) It eliminated only major seasonal peaks, leaving the remaining ones unchanged [Nerlove (1964)].
(*d*) Seasonally adjusted series distort the relationship between variables [Wallis (1974)].
(*e*) On technical grounds, Hannan (1963) and Durbin (1957) showed that the repetition of moving averages in order to remove the trend is unnecessary.

Several attempts have been made to correct the weaknesses of Census II. However, up to now, there has been no method which has been put into practice or widely accepted, either as a substitute or as a modification to Census II. The methods proposed [Granger (1954); Nerlove (1965); Burman (1965)] are based on applying some form of spectral analysis on the data. It is argued that filtering the data will evolve removing frequencies at any desired level without altering the basic form of the data. This is impossible with moving averages which introduce several biases in the data [see Kendall (1946), ch. 29].

Finally, McLaughlin and Boyle (1968) introduced a decomposition method which borrows heavily from Census II which, however, is well suited for business applications, thus complementing Census II which is mainly used for macro data. The major attempt of McLaughlin and Boyle's approach is to discover the direction of change in the cyclical component and to forecast it as much as possible. It is a new method in the sense that it can be used as an umbrella under which several time-series or regression methods can be continually tested for their accuracies and contributions.

5. *Miscellaneous Methods*

(a) *Trend Analysis*

The name "trend analysis" is used to denote simple regression with time as the independent variable. That is,

$$X_t = a + bt + u_t \tag{105}$$

where a and b are the regression coefficients and t is time. The parameters a and b are calculated through classical linear least squares. A variation of (105) will be made by introducing dummy variables which can absorb seasonality or

$$X_t = a + bt + b_2 D_1 + b_3 D_2 + \ldots + b_{L-1} D_{L-1} + u_t \tag{106}$$

where D_i is ith variable introduced for the seasonality, and L is the length of seasonality.

(b) *S-curve Analysis*

The term "*S*-curve", or "life-cycle analysis", is used to denote a trend fitting through (107).

$$X_t = e^{a+\frac{b}{t}}(u_t). \tag{107}$$

Equation (107) can become equivalent of (105) if we transform X_t and t, by:

$$\tilde{X}_t = \log eX_t$$

$$\tilde{t} = \frac{1}{t}$$

(107) then becomes:

$$\tilde{X}_t = a + b\tilde{t} + u_t \tag{108}$$

which can be solved as (105) by least squares. Both trend and *S*-curves are quite useful for medium and longer-term prediction; they are quite simple in both the building of the model and the estimation. For further references, the reader should refer to any standard statistical, econometric or forecasting books [for example, see Johnston (1972); Wheelwright and Madridakis (1973)].

(c) *Quality-Control Charts (QCC)*

Quality-control charts (QCC) were used extensively in the past as an aid to quality control. They consist basically of two control limits, which are $+3\sigma$ above the mean. If new values are within these limits, the process is considered under control; otherwise, the reasons for the significant deviation are sought. The literature in this field is extensive, starting in 1931 with the work of Shewhart (1931) and continuing until the early '60s, when the shift was made from QCC to statistical adaptive optimization and control, which is a more powerful control procedure. However, the interested reader should refer to Page (1957, 1961), Barnard (1959), Roberts (1959) and Golsmith and Whitfield (1970) for additional information about QCC whose usefulness, as a practical tool, is still highly significant.

An interesting application of control charts was made by Harrison and Davies (1964) who used them to detect (*a*) when the pattern of a time series has changes and (*b*) when a given forecasting method is no longer adequate. This is similar to Trigg's (1964) and Trigg's and Leach's (1967) tracking signals, except that they can be applied to other than single exponential smoothing time-series methods.

(d) *Combining the Results of Several Time-Series Methods*

Bates and Granger (1969) report that if separate sets of forecasts are combined to form a composite forecast the mean square error of the latter is lower than either of the mean square errors of the separate predictions. The weighting factor, K, used to weight the forecasts, is calculated by:

$$K = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{109}$$

where σ_1^2 is the variance of the first forecast,

σ_2^2 is the variance of the second forecast and ρ is the correlation between the two forecasted values.

(e) *Choosing and Comparing Various Forecasting Methods*

A computerized set of programs containing over 15 of the major forecasting methods is in existence [see Makridakis *et al.* (1974)]. It performs standard time-series-data analysis, and allows the user to select a method(s) which is best suited to his accuracy and costs preferences, to run this method(s), or any other available, to compare the results, to combine the results of several methods, and analyse his data in order to discover and estimate the trend-cycle component.

IV. Applicability of Time-Series Methods

Time-series applications are still in a period of consolidation after a tremendous volume of theoretical material has become available in the last 40 years. The potential applicability of time-series methods has not been fully utilized or even explored as yet. Much more has to be done. It is a kind of paradox that exponential smoothing models, the least sophisticated of time-series methods, are those which have been utilized the most. On the other hand, there are autoregressive/moving-averages schemes which are the least applied although their theoretical rigorousness is the most complete of all methods in the area of time series. In between, there are the filtering and the decomposition techniques which have been used by engineers and economists respectively, with considerable success and satisfaction.

Exponential smoothing models were introduced around 1960 by operation researchers, and within a few years they were utilized on a grand scale by business enterprises and the military. Their main contribution has been in the area of short-term forecasting of inventories and the sale of individual items or products. Their utilization cost is extremely low, while their accuracy is at least as good as that of autoregressive/moving averages processes according to one researcher [Groff (1973)]. Their disadvantage is that there are several models available, from which the user must select the one best suited to his data and personal preferences.

Filtering models became practical in the early 1960s with Kalman's (1960) contribution. His approach and the availability of the computer were the decisive factors for a wider-scale utilization. Filters have since then successfully been utilized in navigational systems, the space programme and in specialized areas such as guidance and control devices. It is the author's opinion that filtering techniques are the most promising of time-series methods and that their usage will increase in the future and eventually pass to disciplines other than engineering.

Decomposition methods have been used mainly by economists for the purpose of adjusting time series for seasonality and predicting the business cycle. They are the oldest of time-series methods (excluding those in time-series analysis) and the most empirically oriented of all. However, from the point of view of accuracy they seem to be producing adequate results; even though this last statement has not been proved by extensive research findings. Decomposition methods are very intuitive and intended for the non-specialist practitioner. As with smoothing techniques, they require little previous knowledge for being used or for interpreting the results.

Finally, autoregressive/moving-average methods, or the Box-Jenkins approach, as it is more widely known, have been used extensively by statisticians in the fields of forecasting and control. As it has already been mentioned, although their theoretical rigorousness and completeness are surpassed they have lagged in actual application. In my opinion, it is the complexity of the method which hinders utilization. There are not many expert statisticians in places where the method should be utilized to use it properly. Some characteristic problems of using the method can be seen in Chatfield and Prothero (1973) where the authors supposedly used the wrong transformation of the data, thus getting inaccurate results. In a discussion

following the presentation of the paper, given at a meeting of the Royal Statistical Society, a practitioner [Chatfield and Prothero (1973), p. 329] expressed his concern as follows:

"The authors (Chatfield and Prothero) have used the Box-Jenkins procedures in a way which is, at least, not obviously and blatantly erroneous, and the result have been worse than those obtained even by guestimating . . . I do not want to judge whether or not the criticisms made of the authors are valid, but I am sure that the authors represent ability far above that available for forecasting in most organizations and if they have not succeeded in applying the Box-Jenkins procedure correctly, then this is, in a sense, a criticism of the Box-Jenkins procedures."

On the other hand, there is a lot of room for improving autoregressive/moving-average methods in terms of their accuracy and usefulness, not only to a very small segment of expert statistically but also to the wide audience of users.

The time-series approach to the modelling of real systems is one of two alternative approaches. In next section, a brief examination of the two approaches will be given.

Time-Series Versus Explanatory Models

The time-series methods examined in section III of this paper can all be used for predictive purposes, while those of Part II aim at facilitating the utilization of Part III methods. In addition, some of the methods can be used for filtering or purely smoothing purposes. Time-series methods look at forecasting filtering or smoothing as a "black box"; they are not interested in why something is happening. The only other alternative to such a purely mechanistic approach is to attempt to explain the behaviour of a certain factor through variations in a number of independent variables. This can be done through single or a system of multiple regression equations which, by their nature, are explanatory.

Arguments as to whether to use a mechanistic or explanatory model centre on philosophical as well as accuracy-oriented grounds. The proponents of time-series models say that the future, at least in the short term, is a continuation of the past. They are only interested in discovering its past behaviour which could be consequently extended to predict the future. They are not concerned with the reasons for what is happening; they only want to know that it will take place. Explanatory models are the exact opposite. They search for casual relationship(s) which underline the reasons for what is going on. Their proponents argue that the basic structure of events changes with time, thus you cannot forecast the future without taking into consideration the factors which will influence its course.

Such an approach was very popular in the '50s and '60s and came mainly from economists; it resulted in huge econometric models, often consisting of hundreds or even thousands of variables. It is difficult to doubt the power of these models to explain in an ex-post manner what has happened. However, there are serious doubts as to their ability to forecast better than time-series models. The same point was made by Evans who talks about the utility of leading indicators: "The series serves as a valuable, historical record and sheds light on the causes of past depressions". But as a practical method of forecasting, the leading indicators cannot be used very effectively or accurately.

Explanatory models, by their nature, require a number of independent variables whose magnitude, for some future time, must be estimated before any predictions about the future can be made. The critics of the econometric approach saying that this is not forecasting but guess-work. What is actually done is to shift the burden of forecasting from that of directly predicting some factor of interest to another one which attempts to estimate several independent variables before it can forecast.

If we are to compare time-series and regression models on purely predictive grounds, the former have been doing better than the latter group [see Cooper (1972); Naylor (1972); Granger and Newbold (1972); Nelson (1972)]. In addition, time-series models are cheaper to develop, easier to utilize and update, and in general, less complex to use. Furthermore, they provide

the user with a wider choice of models from which to select the most appropriate for his cost and accuracy preferences. It is the author's opinion that since the mid-'60s, for the above reasons, there has been a gradual shift from regression to time-series models. Although such a statement cannot be documented with research findings, it can be said with a higher degree of confidence that time-series models provide a cheaper and much simpler alternative to forecasting than any of the regression techniques. This does not imply that multiple regression or econometric models are useless, but that they are better suited to answering "if then" type of questions. Their utilization is best suited when several alternatives must be explored and their effect estimated and used for planning purposes. As such, explanatory models are invaluable and perform a task which mechanistic approaches, such as time series, cannot handle.

In practical terms, time-series models are better suited for short and longer-term predictions. This leaves the range of medium-term forecasting open, which must be filled in mainly with explanatory models. In the short term, events are correlated, quite often highly, which makes time-series forecasting easier to perform. Furthermore, drastic changes are less likely. This results in more accurate forecasting. In the longer run, the two methods mentioned in section III.5 are also suitable although their accuracies are not very high, nor can they be well assessed. However, they are the only viable alternatives for forecasting future trends and saturation points. Regression models which often require the estimation of hundreds of independent variables cannot be used for such purposes because of the impossibility of accurately estimating the future values of the independent variables involved.

Selecting an Appropriate Time-Series Method

The author does not pretend to have made an exhaustive survey of all the time-series methods in existence. But even if we assume that a selection from the alternative methods described in Part III has to be made, the potential user must choose from a great many methods. The selection process is not easy because (*a*) there is little evidence to guide us as to the accuracies, costs and other factors which influence the choice of methods and (*b*) because several time-series methods are neither well known nor well documented. The last point can be easily demonstrated by the autoregressive/moving average models which did exist for a long period of time, but only became known and widely used since Box and Jenkins published their book. There are several other methods in the same position as with autoregressive-moving-average schemes before 1970.

There are a number of factors which influence the selection of a specific time-series method and will have to be examined separately before a specific choice is made; these factors relate to the number of data points, costs, accuracies and complexity required to utilize a specific method. Each of the factors will be examined in more detail next.

(a) *Data Requirements*

Different time-series method require varying numbers of data points. These can range from a few, required for simple exponential smoothing, to more than 72, which is the minimum for methods like the Census II or the seasonal autoregressive/moving-average models. It may seem trivial that data could be a restricting factor in terms of selecting a desired model, but in practice it is often the case. Data covering a long-time span often do not exist; at other times they are not in a compatible form, or are hard to gather. In addition, the cost of collecting and storing can be a prohibitive factor when a large number of time series is involved. These considerations make the number of available data points a crucial factor which, in turn, determines the method to be used.

We can distinguish two types of data requirements: the first is needed to fit initially a model

and to estimate its parameters, and the second in order to update the model or its parameters. In the first category, we can include all methods except the smoothing and the filtering techniques whose parameters, once calculated, can be easily updated as new data points become available. The data storage and computations required for updating such a method are minimal, an advantage of great practical value which makes the desirability of smoothing

Method	Minimum data requirements		
	Initial parameters – Estimation	Subsequent estimations	
		DATA	Parameters or summary measures
Autocorrelation analysis	6 to 7 times the suspected length of seasonality		not applicable
Spectral analysis	6 to 7 times the suspected length of seasonality		not applicable
Simple exponential smoothing	10 to 20	1	2
Linear exponential smoothing	10 to 20	1	3
Higher forms of exponential smoothing	15 to 25	1	3+order of trend
Harmonic smoothing	2L to 3L	1	3L
Autoregressive/moving-average schemes	30 plus 3 to 4 times the highest order parameter	same as initial	as with estimation
Decomposition	72	same as initial	as with estimation
Trend analysis	15–20	same as initial	as with estimation
S-curves	15–20	same as initial	as with estimation
Winter	$2L$ to $3L$	1	$L+5$
Adaptive filtering	$4L$ to $5L$	1	$L+1$
Foran	18	18	between 50–100
Kalman-Bucy	30–50	1	4 for each time series
Transfer function	40 plus 3 to 4 times same as the highest order parameter initial		as with estimation

Figure 4. *Minimum data requirements.*

and filtering methods high. Figure 4 shows the approximate (according to the author) minimum number of data points required to utilize the different time-series methods discussed in Parts II and III. Unfortunately, there are no objective rules to determine minimal data requirements.

(*b*) *Costs*

The costs associated with developing and running a given time-series method vary considerably. At one extreme, one can distinguish single smoothing techniques whose cost is insignificant,

and at the other we can place transfer-function models which are extremely costly both to develop and to run. As with everything else associated with time-series applications, there is very little information about the exact costs associated with developing and running time-series methods. Such information is vital to the user who wishes to have an approximate cost figure before committing himself to utilizing times series for either forecasting or control purposes. Chambers (1971) has suggested some cost figures; similarly Makridakis and Wheelwright (1974) have derived some overall cost estimates. However, both cost estimates are based on person experience rather than extensive research into actual development and running costs.

(c) *Accuracies*

Accuracies vary according to the type of methods being utilized, but matters are even more complicated here than with the number of data points or costs. While the latter is independent of external factors, accuracies are influenced by how many data points are used and by the ability of the user to select the most appropriate from the different categories of time-series methods as well as the specific model within each category. There are few cases of comparative research with the accuracy of the different time-series methods. Reid (1971), Granger and Newbold (1972) and Groff (1973) have tested a limited number of methods for their accuracies. However, the results are somewhat conflicting. Reid and Granger and Newbold report that the Box-and-Jenkins method was superior to the others when there were more than 50 data points and when one-step-ahead forecasts were made. On the other hand, Groff reports the Box-and-Jenkins approach to autoregressive/moving-average models to be less accurate than exponential and harmonic-smoothing methods. Furthermore, he does not discover that the accuracy of the Box-and-Jenkins method is influenced by the number of data points. Obviously, there are several factors, not yet known to us, which influence accuracy, thus causing discrepancies as those just mentioned. Reid, and Granger and Newbold, for example, used industry-wise and macro-data respectively, while Groff tested micro-series. It will be difficult to know the exact factors affecting accuracy unless more research is done in this area.

(d) *Complexity*

Complexity is a factor which determines if a potential user is willing to employ a certain time-series method. If the complexity of the method requires an expert statistician to develop it and interpret the result once it has been utilized, it will be a prohibiting factor for a great number of users. Thus, in addition to the data cost and accuracies of each method, its complexity should be taken into account when determining its applicability.

Some secondary factors involved are the time horizon of the forecasting or planning application (short, medium and long) and the type of the time series involved (stationary, non-stationary, seasonal, non-seasonal). It is not, however, worth examining these in more detail since their influence on the applicability of a given time-series method is rather obvious.

Conclusion

In all, time-series methods have gone a long way since they were first introduced. There is now a rather wide variety of methods which can be utilized to fit the majority of present-day estimation, forecasting or control needs involving time series. They now operate under varying monetary budgets, and can be employed in such a way that a high degree of accuracy can be achieved. The problem is not technical, but that there is a lack of communication between the experts in time-series method and the potential users. Jenkins and Watts (1968) and Box and Jenkins (1970) rendered a great service in making these methods more widely known and for which they deserve a lot of credit. The same is true of Kalman (1960) who

made the introduction of filtering techniques known to a wider audience which was persuaded as to their importance.

There are still practical problems, however. For once there is no adequate documentation, both in terms of verbal descriptions of how to use each method and how to interpret the results and in terms of computer programs, which will enable a non-technically-oriented user to utilize time-series methods. In both cases, there is little done for the non-specialist. The description of the methods is too mathematical and the different programs in existence are not standardized. Finally, the potential user is not convinced as to the benefits of utilizing time-series filtering, forecasting or control methods. There is a lack of information to allow the assessment of costs, accuracies and, of course, in final analysis, benefits from using time-series techniques. In all, the criticism is similar to that made on management science by Grayson (1973). These points can be again viewed in relation to the problems referred to in Chatfield and Prothero (1973), mainly that using time-series methods is an art requiring expertise knowledge which only few specialists know how to apply and which scare off practitioners in that case.

In the author's opinion, the above-mentioned drawbacks must be corrected; there are several research priorities urgently needed. First, research findings should establish the level of accuracy, preferably expressed in absolute mean percentage errors, of each individual time-series method. Secondly, a detailed figure of the actual cost of developing and running each time-series technique must be found. Such information would be invaluable for any potential user. The job, though not easy, because of the large number of factors influencing accuracy and cost, is not as difficult as similar tasks performed by physicists or other natural scientists.

Another area of research should be directly towards expanding filters to include Moving-Average schemes. Such a job will undoubtedly increase their efficiency with little extra computational burden. Furthermore, the applicability of filters outside of engineering situations seems highly promising. Harrison and Stevens (1971) took the first step in this direction, although much more should be done to make the utilization of the Kalman filters more possible for forecasting and control purposes. Filters combine the characteristics of both the smoothing methods (in terms of cost) and of autoregressive schemes in terms of theoretical rigorousness, without being as difficult or complex to use. Thus, their utilization in non-engineering applications could be extremely useful.

Another area requiring attention is that of ways of predicting the business cycle in a more satisfactory way. This is an area of extreme importance in business forecasting in which no time-series method has been successful. There are, of course, some isolated attempts such as those by the designers of Census II, and McLaughlin and Boyle (1968). However, the results have been mediocre. There must also be creative programming work which will result in the creation of forecasting packages easy to use, understand and which present their results in such a way that a non-specialist can utilize them for his planning, forecasting, control or estimating needs.

Finally, the most important job lying ahead is that of making time-series methods known to non-specialists. Such a task will create a large number of users and therefore increase the interest and activity in the area of time series. The author believes that the usefulness of time-series methods is poorly understood by only a tiny segment of potential users. In his opinion, it is the job of educating potential users about the merits of the time-series approach to chronological data which is the most urgently needed.

Acknowledgements

The author would like to thank Professors H. Boyd, C. Brumat, J. Cropper, G. Dyas, C. Faucheux, J. Lyons, R. Hogarth, J. Miller and D. Weinstein for their extremely helpful comments.

References

Anderson, B. D. (March 1971). A qualitative introduction to Wiener and Kalman-Bucy filters. *Proceedings, The Institution of Radio and Electronics Engineers* (Australia), pp. 93–103.

Anderson, O.; Nochmals über (1914). The elimination of spurious correlation due to position in time or space. *Biometrica*, X, 269.

Anderson, R. L. (1942). Distribution of the serial correlation coefficient. *Animals of Mathematical Statistics*, **13**, 1–13.

Barham, P. M.; Humphries, D. E. (1970). Derivation of the Kalman filtering equations from elementary statistical principles, in (ed.) Leondes, C. T., *Theory and Applications of Kalman Filtering*, Advisory Group for Aerospace Research and Development, Paris.

Bartlett, M. S. (1935). Some aspects of the time-correlation problem in regard to tests of significance. *Journal of the Royal Statistical Society*, **98**, 353–543.

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society*, **B, 8**, 27.

Bartlett, M. S. (1948). Smoothing periodograms from time series with continuous spectra. *Nature*, **161**, 686–687.

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrica*, **37**, 4–11.

Barnard, G. A. (1959). Control charts and stochastic processes. *Journal of the Royal Statistical Society*, B, **21**, 239–271.

Bates, J. M.; Granger, C. W. J. (1969). Combination of forecast. *Operation Research Quarterly*, **20** (4), 451–468.

Battin, R. H.; Levine, G. M. (1972). Application of Kalman filtering techniques to the Apollo programs. *Paper E*-2401, MIT Instrumentation Laboratory, Cambridge, Mass.

Beveridge, W. H. (May 1922). Wheat prices and rainfall in Western Europe. *Journal of the Royal Statistical Society*, **85**, 412–459 (followed by a discussion, pp. 460–478).

Blackman, R. B.; Tukey, J. W. (1958). *The Measurement of Power Spectra from the Point of View of Communications Engineering*. Dover, New York.

Bode, H. W.; Shannon, C E. (1950). A simplified derivation of linear least square smoothing and prediction theory. *Proceedings, I.R.E.*, **38**, 417–425.

Booton, R. C. (1952) An optimization theory for time-varying linear systems with non-stationary Statistical inputs. *Proceedings, I.R.E.*, **40**, 977–981.

Box, G. E. P.; Jenkins, G. M. (1970). *Time-Series Analysis: Forecasting and Control*. Holden Day, San Francisco.

Box, G. E. P.; Jenkins, G. M. (1968). Some recent advances in forecasting and control. *Applied Statistics*, **17**, 91.

Box, G. E. P.; Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving-average time-series models. *Journal of the American Statistical Association*, **65**, 1509–1526.

Brown, R. G. (1959). *Statistical Forecasting for Inventory Control*. McGraw-Hill, New York.

Brown, R. G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series*. Prentice Hall, New Jersey.

Brown, R. G. ; Meyer, R. F. (1961). The fundamental theorem of exponential smoothing. *Operations Research*, **9**, 673–685.

Burman, J. P. (1965). Moving seasonal adjustments of economic time series. *Journal of the Royal Statistical Society*, A, **128**, 534–558.

Buys Ballot, C. H. D. (1847). *Les Changements Périodiques de Température*. Utrecht.

Carslaw, H. S. (1930). *Fourier Series and Integrals*. Dover, New York.

Chambers, J. C. *et al.* (July–August 1971). How to choose the right forecasting technique. *Harvard Business Review*, pp. 45–74.

Chatfield, C.; Prothero, D. L. (1973). Box-Jenkins seasonal forecasting: problems in a case-study. *Journal of the Royal Statistical Society*, A, **136** (3), 295–336.

Cooper, R. L. (1972). The predictive performance of quarterly econometric models of the United States, in Hickman, B. G. (ed.), *Econometric Models of Cyclical Behavior*. National Bureau of Economic Research, New York.

Copeland, M. T. (May 1915). Statistical indices of business conditions. *Quarterly Journal of Economics*, **XXIX**, 522–562.

Daniell, P. J. (1946). Discussion on "Symposium on autocorrelation in time series". Supp. *Journal of the Royal Statistical Society*, **B, 8**, 88.

Davis, H. T. (1963). *The Analysis of Economic Time Series*. The Principia Press of Trinity University, San Antonio, Texas, first edition, 1941.

Dixon, W. J. (1944). Further contributions to the problem of serial correlations. *Annals of Mathematical Statistics*, **15**, 119–144.

Durbin, J. (1957). Trend elimination for the purpose of estimating seasonal and periodic components, in Rosenblatt, M. (ed.), *Time-Series Analysis*. Wiley, New York.

Durbin, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrica*, **46**, 306–316.

Durbin, J. (1960). Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society*, **22**, 139–153.

Durbin, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica*, **38**, 410–421.

Evans, M. K. (1969). *Macro-Economic Analysis*. Harper and Row, New York.
Farley, J. U.; Hinich, M. J. (November 1970). Detecting "small" mean shifts in time series. *Management Science*, **17** (3), 189–199.
Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics*, **41**, 155.
Gelb, A. (ed.) (1974). *Applied Optimal Estimation*. MIT Press, Cambridge, Mass.
Goldsmith, P. L.; Whitfield, H. (1961). Average run lengths in cumulative chart quality control schemes. *Technometrics*, **3**, 11–20.
Granger, C. W. J. (1974). *Spectral Analysis of Economic Time Series*. Princeton University Press, Princeton.
Granger, C. W. J.; Newbold, P. (1972). Economic forecasting the atheist's viewpoint. *Nottingham University Forecasting Project*, Note 11.
Grayson, C. J. (July-August 1973). Management science and business practices. *Harvard Business Review*, 41–48.
Grether, D. M.; Nerlove, M. (September 1970). Some properties of "Optimal" seasonal adjustments. *Econometrica*, **38** (5), 682–703.
Groff, G. K. (September 1973). Empirical comparison of models for short-range forecasting. *Management Science*, **20** (1), 22–31.
Hannan, E. J. (1963). The estimation of seasonal variation in economic time series. *Journal of the American Statistical Association*, **58**, 31–44.
Harrison, P. J.; Davies, O. (1964). The use of cumulative sum (CUSUM) techniques for the control of routine forecasts of product demand. *Operation Research Quarterly*, **12**, 325–333.
Harrison, P. J. (1965). Short-term forecasting. *Applied Statistics*, **14**, 102–139.
Harrison, P. J.; Stevens, C. F. (1971). A Bayesian approach to short-term forecasting. *Operational Research Quarterly*, **22**, (4), 341–362.
Herschel, W. (1801). Observations tending to investigate the nature of the sun. *Phil. Transactions of Royal Society of London*, pp. 265–313.
Holt, C. C. (1957). Forecasting seasonal and trends by exponentially weighted moving averages. *Carnegie Institute of Technology*, Pittsburgh, Pennsylvania.
Holt, C. C. *et al.* (1960). *Planning Production Inventories and Work Force*. Prentice-Hall, New York, New Jersey.
Hooker, R. H. (1901). The suspension of the Berlin produce exchange and its effect upon corn prices. *Journal of the Royal Statistical Society*, **64**, 574.
Hooker, R. H. (1901). Correlation of the marriage-rate with trade. *Journal of the Royal Statistical Society*, **64**, 485–492.
Jenkins, G. M.; Priestley, M. B. (1957). The spectral analysis of time series. *Journal of the Royal Statistical Society*, B, **19** (1), 1–12.
Jenkins, G. M. (1961). General considerations in the analysis of spectra. *Technometrics*, **3** (2), 133–166.
Jenkins, G. M. (1965). A survey of spectral analysis. *Applied Statistics*, **14** (1), 2–32.
Jenkins, G. M.; Watts, D. G. (1968). *Spectral Analysis and Its Applications*. Holden Day, San Francisco.
Johnston, J. (1972). *Econometrica*. McGraw-Hill, New York.
Kalman, R. E. 1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, D, **82**, 35–44.
Kalman, R. E.; Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, D, **83**, 95–107.
Kendall, M. G. (1945). On the analysis of oscillatory time series. *Journal of the Royal Statistical Society*, **VIII**, 93–120.
Kendall, M. G. (1946). *The Advanced Theory of Statistics*, Vol. II. C. Griffin & Company, London.
Kendall, M. G. (1946). *Contributions to the Study of Oscillatory Time Series*. Cambridge University Press, Cambridge, England.
Kendall, M. G.; Doig, A. G. (1962, 1965, 1968). *Bibliography of Statistical Literature* (3 vol.). Oliver and Boyd, Edinburgh and London.
Kendall, M. G. (1973). *Time Series*. Hafner Press, New York.
Khintchine, A. (1932). Sulle Succesioni Stationarie di Eventi. *Giorn. Ist. Ital. Attuari*, **3**, 267.
Khintchine, A. (1933). Uber Stationäre Reihen Zufälliger Variables. *Rec. Mathem. Moscou*, **40**, 124.
Khintchine, A. (1934). Korrelationstheorie der Stationaren Stochastischen Prozesse. *Math. Ann.*, **109**, 604.
Kolmogoroff, A. (1931). Uber die Analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematical Annals*, **104**, 415.
Kolmogoroff, A. (1933). *Giorm Ist Ipal Attuari*, Vol. 4, no. 83.
Kolmogoroff, A. (1941). Interpolation and extrapolation of stationary random sequences. *Bulletin of Academy of Science, USSR, Math. Ser.*, **5**, 3–14.
Lanczos, C. (1966). *Discourse on Fourier Series*. Oliver and Boyd, London.
Leser, C. E. V. (1968). A survey of econometrics. *Journal of the Royal Statistical Society*, A, **131**, 530–566.
Levinson, N. (January 1947). The Wiener RMS (root mean square) error criterion in filter design and prediction. *Journal of Mathematical Physics*, XXV, no. 4, 261–278.
Macauley, F. R. (1930). The smoothing of time series. *National Bureau of Economic Research*, pp. 121–136.
McLaughlin, R. L. (1962). *Time-Series Forecasting*. Marketing-Research Technique, Series no. 6, of American Marketing Association.
McLaughlin, R. L.; Boyle, J. J. (1968). *Short-Term Forecasting*. American Marketing Association.

McClain, J. O.; Thomas, L. J. (March–April 1973). Response-variance tradeoffs in adaptive forecasting. *Operations Research*, **21** (2), 554–568.

Makridakis, S. *et al.* (1974). An interactive forecasting system. *American Statistician*, **28**, 153–159.

Makridakis, S.; Wheelwright, S. (1974). *Quantitative and Technological Methods of Forecasting and Planning Applications* (mimeographed). INSEAD, Fontainebleau, France. John Wiley & Sons, New York, in press.

Mann, H. B.; Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometric*, **11** (383), 173–220.

Mathieu, A. (1972). Techniques de Contrôle de Prévision à Court Terme. *Revue Française d'Automatique, Informatique et Recherche Opérationnelle*, **B–1**, 29–47.

Mehra, R. K. (April 1970). On the identification of variances and adaptive Kalman filtering. *IEEE, Transactions on Automatic Control*, Vol. 10, AC-15, no. 2, pp. 175–184.

Mehra, R. K. (February 1971). On line identification of linear dynamic systems with applications to Kalman filtering. *IEEE, Transactions on Automatic Control*, Vol. 10, AC-16, no. 1, pp. 12–21.

Naylor, T. H.; Seaks, T. G. (1972). Box-Jenkins methods: an alternative to econometric models. *International Statistical Review*, **40** (2), 123–137.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the US economy. *American Economic Review*, **LXII**, 902–917.

Nerlove, M. (July 1964). Spectral analysis of seasonal adjustment procedures. *Econometrica*, **32**, 241–286.

Nerlove, M. (June 1965). A comparison of a modified "Hannan" and the BLS seasonal adjustment filters. *Journal of the American Statistical Association*, **60**, 442–491.

Page, E. S. (1957). On problems in which a change in parameter occurs at an unknown point. *Biometrica*, **4**, 249.

Page, E. S. (1961). Cumulative sum charts. *Technometrics*, **3**, 1–10.

Persons, W. M. (December 1916). Construction of a business barometer based upon annual data. *The American Economic Review*, **VI** (4), 739–769.

Poynting, J. H. (1884). A comparison of the fluctuations in the price of wheat and in the cotton and silk imports into Great Britain. *Journal of the Royal Statistical Society*, **47**, 345–364.

Quenouille, M. H. (1949). The joint distribution of serial correlation coefficients. *Annals of Mathematical Statistics*, **20**, 561–571.

Rao, A. G.; Shapiro, A. (November 1970). Adaptive smoothing using evolutionary spectra. *Management Science*, **10** (3), 208–218.

Reid, D. J. (1971). Forecasting in action: a comparison of forecasting techniques in economic time series. *Joint Conference of O.R. Society's Group on Long-Range Planning and Forecasting.*

Roberts, S. W. (1959). Control-chart tests based on geometric moving averages. *Technometrics*, **1**, 239.

Rosenblatt, H. M. (1963). Spectral analysis and parametric methods of seasonal adjustment of economic time series. *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, pp. 94–133.

Sage, A. P.; Husa, G. W. (November 1969). Algorithms for sequential adaptive estimation of prior statistics. *Proceedings, 8th IEEE Symposium on Adaptive Processes.* Penn. State Univ.

Schneeweiss, C. A. (March 1971). Smoothing production by inventory – an application of the Wiener filtering theory. *Management Science*, **17** (7), 472–499.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26-day period of metereological phenomena. *Terrestrial Magnetism*, **3**, 11–41.

Schuster, A. (1906). On the periodicities of sunspots. *Philosophical Transactions*, A, **206**, 69–81.

Shiskin, J. (1957). Electronic computers and business indicators. *National Bureau of Economic Research*, Occasional Paper, no. 57.

Shiskin, J. (1961). Tests and revisions of Bureau of the census methods of seasonal adjustments. *Bureau of the Census*, Technical Paper, no. 5.

Shiskin, J. *et al.* (1967). The X-11 variant of the Census II method seasonal adjustment program. *Bureau of the Census*, Technical Paper, no. 15.

Shewhart, W. A. (1931). *The Economic Control of the Quality of Manufactured Product.* Macmillan, New York.

Singer, R. A.; Frost, P. A. (August 1969). On the relative performance of the Kalman and Wiener filters. *IEEE, Transactions on Automatic Control*, Vol. AC-14, pp. 390–394.

Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica*, **5**, 105–146.

Sorenson, H. W. (1968). Kalman filtering techniques, in (ed.) Leondes, C. T., *Advances in Control Systems.* Academic Press, New York.

Sorenson, H. W. (July 1973). Least-squares estimation: from Gauss to Kalman. *IEEE Spectrum*, **7**, 63–68.

Spencer, J. (1904). On the graduation of the rates of sickness and mortality. *Journal of the Institute of Actuaries*, **38**, 334.

Steward, B.; Dodgson, W. (1879). Preliminary report to the committee on solar physics on a method of detecting the unknown inequalities of a series of observations. *Proceedings of the Royal Society*, **29**, 106–122.

Stokes, G. C. (1879). Note on the report of the committee on solar physics, *Proceedings of the Royal Society*, **29**, 122–123.

Swerling, P. (January 1958). A proposed stage-wise differential correction procedure for satellite tracking and prediction. *Report P 1292, Rand. Corp., Santa Monica*, California.

Theil, H.; Wage, S. (January 1974). Some observations on adaptive filtering. *Management Science*, **10** (2), 198–224.
Trigg, D. W. (1964). Monitoring a forecasting system. *Operational Research Quarterly*, **15**, 271–274.
Trigg, D. W.; Leach, D. H. (1967). Exponential smoothing with an adaptive response rate. *Operational Research Quarterly*, **18**, 53–59.
Tukey, J. W. (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics*, **3** (2), 191–219.
Walker, A. M. (1931). On the periodicity in series of related terms. *Proceedings of the Royal Society of London*, A, **131**, 518–532.
Walker, A. M. (1962). Large-sample estimation of parameters for autoregressive processes with moving average residuals, Parts I, II. *Biometrika*, **49**, 117–131.
Wallis, K. F. (March 1974). Seasonal adjustment and relation between variables. *Journal of the American Statistical Association*, **69** (345), 18–31.
Weiss, I. M. (August 1970). A survey of discrete Kalman-Bucy filtering with unknown noise covariances. *AIAA, Guidance, Control and Flight Mechanics Conference*, Paper no. 70-955, Santa Barbara, California.
Wheelwright, S.; Madridakis, S. (1973). *Forecasting Methods for Managements*. Wiley and Sons, New York.
Whittle, P. (1953). The analysis of multiple stationary time series. *Journal of the Royal Statistical Society*, B, **15**, 125–139.
Widrow, B. (1966). Adaptive filtering I: Fundamentals. *Center for Systems Research*, Technical Report no. 6764–6, Stanford University.
Wiener, N. (1930). Generalized harmonic analysis. *Acta-Math.*, **55**, 117.
Wiener, N. (1964). *Time Series*. MIT Press, Cambridge, Mass. (first edition 1949).
Winters, P. R. (April 1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, pp. 324–342.
Wise, J. (1955). The autocorrelation function and the spectral density function. *Biometrica*, **42**, 151–159.
Wold, H. (1954). *A Study in the Analysis of Stationary Time Series*. Almqrist & Wiksell, Stockholm (first edition 1938).
Wold, H. O. (1965). *Bibliography on Time Series and Stochastic Processes: An International Team Project*. Oliver and Boyd, London.
Yule, G. U. (1921). On the time-correlation problem with special reference to the variate difference correlation method. *Journal of the Royal Statistical Society*, **84**, 497–526.
Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time series? A study in sampling and the nature of time series. *Journal of the Royal Statistical Society*, **89**, 1–64.
Yule, G. U. (1927). On the method of investigating periodicities in disturbed series, with special reference to Wölfer's sunspot numbers. *Philosophical Transactions*, A, **226**, 267–298.
Zadeh, L. A.; Ragazzini, J. R. (1950). An extension of Wiener's theory of prediction. *Journal of Applied Physics*, **21**, 645–655.
Zymund, A. (1959). *Trigonometric Series*, Vols. I and II. Cambridge University Press, Cambridge, England.
Government Report (1911), Rapport sur les Indices des Crises Economiques et sur les Mesures Résultant de ces Crises.
Bureau of Census (1969). A spectral study of "overadjustment" for seasonality, Technical Paper, no. 21.
Bureau of Census (1970). Spectral analysis and parametric methods for seasonal adjustment of Economic time series, Technical Paper, no. 23.

Survey Articles and Books

1. Autoregressive-moving-average schemes [Box and Jenkins (1968); Jenkins and Watts (1968)].
2. Filters [Anderson (1971); Sorenson (1973); Weiss (1970)].
3. Smoothing models [Brown (1963)].
4. Decomposition models [Shiskin *et al* (1967)].
5. Miscellaneous
   (*a*) Tracking signals and quality control [Mehra (1970)].
   (*b*) Econometrics [Leser (1968); Johnston (1972)].
6. Spectral analysis [Jenkins and Priestley (1957); Jenkins (1965); Jenkins and Watts (1968)].
7. Bibliographies [Kendall and Doig (1962); Wold (1965)].
8. Time series (overall) [Kendall (1973); Makridakis and Wheelwright (1974)].

Résumé

Les méthodes de séries chronologiques ont été utilisées de façon extensive ces vingt dernières années à des fins d'analyse (par exemple pour trouver les caractéristiques d'une série de données), de prédiction ou de contrôle. Ce domaine est parfaitement interdisciplinaire puisqu'on y a développé des méthodes utilisées en statistiques (processus ARMA), le génie (filtres), la recherche opérationnelle (modèles de lissage exponentiel), l'économie (méthodes de décomposition). Le but de cet article est de décrire chacune des principales méthodes existant actuellement et de les synthétiser en un tout intégré faisant ressortir leurs analogies et leurs différences. Enfin, le problème de l'applicabilité des différentes méthodes aux situations concrètes est abordé et l'accent est mis sur les directions à envisager pour de futures recherches.