Bureau of Business Research
Graduate School of Business Administration
University of Michigan

December, 1970

BUREAU OF BUSINESS RESEARCH WORKING PAPER
NO. 16

TIME SERIES FORECASTING PROCEDURES

FOR AN ECONOMIC SIMULATION MODEL

by

Kenneth O. Cogger
Research Fellow
University of Michigan

BACKGROUND OF THIS PAPER

I. Introduction

In order to introduce the topic of time series analysis with emphasis on forecasting, it is necessary to clarify and define several concepts. A time series will, for our purposes, be narrowly defined as a family of real-valued random variables $\{X(t), t \varepsilon T\}$ where the index set T is the set of all integers. That is, $T = \{\ldots, -1, 0, 1, \ldots\}$. It will be assumed that every finite collection $\{X(t_1), \ldots, X(t_n)\}$ is a set of random variables with a specified joint distribution function $F_{t_1, \ldots, t_n}(X(t_1), \ldots, X(t_n))$. This distribution function must satisfy two consistency requirements. First, if $\alpha_1, \ldots, \alpha_n$ is any permutation of 1, 2, ..., n, then

$$F_{t_1, \ldots, t_n}(X(t_1), \ldots, X(t_n)) = F_{t_{\alpha_1}, \ldots, t_{\alpha_n}}(X(t_{\alpha_1}), \ldots, X(t_{\alpha_n})).$$

Second, for any integer $m < n$,

$$F_{t_1, \ldots, t_m}(X(t_1), \ldots, X(t_m)) = \lim_{X(t_j) \to \infty} F_{t_1, \ldots, t_n}(X(t_1), \ldots, X(t_n)).$$
$$j = m + 1, \ldots, n$$

Under these definitions and restrictions, our discussion is necessarily limited to real-valued time series defined over a discrete time domain. The class of all such time series is sufficiently large, however, to include any time series appearing in the economic sector of the real world. In particular, this class includes all time series that would appear in a simulation model of the automobile industry.

The definition of a time series given above leads to the equivalence of the terms "time series" and "stochastic process". The theory of stochastic processes has developed rapidly over the past thirty years. During this period, time series analysis has utilized two distinct, but overlapping approaches. The first of these approaches is restricted to what may be termed the frequency domain. The object of time series analysis in the frequency domain is to obtain

estimates of what is called the spectrum in order to determine whether a given

time series exhibits certain fundamental rhythms or harmonics. This method of

"hidden periodicities" has given way in recent years to utilizing the spectrum

as a broad characterization of any time series regardless of the existence of

harmonics. Time series analysis in the frequency domain, usually called spectral

analysis, has been utilized most effectively in physics, engineering, and other

non-economic areas. In fact the early theoretical developments in time series

analysis by Kolmogorov[1] and Wiener[2] were based mainly on spectral analysis.

The second approach to time series analysis is restricted to the more

familiar (at least to economists) time domain. There are several important

reasons for the fact that most applied economic time series analysis have taken

place in the time domain. The first of these reasons has already been suggested.

The time domain is simply a more familiar setting for economists interested in

time series analysis. A second reason is that spectral analysis does not provide

ready answers to questions related to forecasting. While spectral analysis may

have merits for characterizing a given time series in a general way, the spectrum

has no direct connection with time series forecasting. Since forecasting is

often an ultimate goal, as in the automobile study, the decision is usually made

to conduct economic time series analyses in the time domain, where statistical

results are immediately applicable to forecasting. A third reason for preferring

analyses in the time domain over spectral analysis is particularly pertinent to

the automobile study. Within a simulation model for an industry, managerial de-

cision rules must be formulated in an unambiguous manner. These rules can be

implemented only after the input of various estimates and forecasts based upon

time series analysis. From the standpoint of operating the simulation model,

---

[1]Kolmogorov, A. N., "Aufangsgrvude der Theorie der Markoffschen ketten mit unendlich vieleu möglichen Zuständen," Mat. Sbornik (N.S.), 1 (1936), 607-10.

[2]Wiener, N. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. New York: Wiley, 1949.

both the decision rules and their inputs must be capable of unambiguous description in terms of an algorithm. It is precisely this requirement of a simulation model that leads to the rejection of spectral analysis as a procedure for providing these inputs. Spectral analysis yields primarily graphical results which require a great deal of visual (and therefore subjective) interpretation. This lack of objectivity leads to considerable ambiguity in practice regarding the interpretation of sample spectrums. To be sure, graphical results abound within the field of time series analysis in the time domain also, but these results are not usually interpreted visually. A fourth reason for preferring analyses in the time domain over spectral analysis is that the latter is strictly valid only for stationary time series. The concept of stationarity will be fully described in the next section but it suffices to state now that many economic time series exhibit non-stationarity, effectively limiting the usefulness of spectral analysis within an economic simulation model.

II. Time Series Representations or Models

Time series may be characterized according to several different properties. Many of these properties are summarized in detail in Doob[3], a valuable reference source for the theory of stochastic processes. One of the most frequently discussed properties of time series is that of stationarity. A strictly stationary time series is defined to be any time series possessing the property that every finite collection of random variables $\{X(t_1),...,X(t_n)\}$ has the same joint distribution function as $\{X(t_1+h),...,X(t_n+h)\}$ for any integer h. A consequence of this definition is that if the above collection of random variables has a multivariate normal distribution function, then the time series is strictly stationary if and only if the expected value of $X(t)$ is independent of t and the covariance of $X(t_j)$ and $X(t_k)$ is a function only of the difference $t_j - t_k$. This

---

[3]Doob, J. L., Stochastic Processes. New York: Wiley, 1953.

characterization of multivariate normal distributions by their second-order, or covariance, properties leads to a less restrictive distinction between types of time series. A weakly stationary or second-order stationary time series (not necessarily normally distributed) is defined to be any time series such that the expected value of X(t) is independent of t and the covariance of X(t) and X(t+h) is a finite-valued function only of h. For future reference it is convenient to introduce the notation

$$\gamma_x(t_j, t_k) = C(X(t_j), X(t_k))$$

for the covariance of the random variables $X(t_j)$ and $X(t_k)$. If the time series of interest is weakly stationary, we shall write

$$\gamma_x(h) = C(X(t), X(t+h))$$

as the covariance of the random variables X(t) and X(t+h). It is important to note that many economic time series are non-stationary since their expected values are not constant. Other types of non-stationarity are possible of course, including the possibility that $\gamma_x(t_j, t_k)$ is not a function of $t_j - t_k$ and the possibility that this covariance function fails to exist. These three types of non-stationarity will be referred to as non-stationarity in the mean, in the covariance, and in the variance, respectively.

Within the class of weakly stationary time series, further characterization is possible in terms of linear representations. The first type of linear representation for a weakly stationary time series is called a moving average. A weakly stationary time series $\{X(t), t \varepsilon T\}$ possessing an expected value $M = E\{X(t)\}$ is said to be a moving average process of order q if there exists a sequence of real numbers $b_0 = 1, b_1, b_2, \ldots, b_q$ such that

$$(1) \quad X(t) = M + \sum_{j=0}^{q} b_j \varepsilon(t-j) \quad ; t \varepsilon T,$$

where $\{\varepsilon(t), t\varepsilon T\}$ is a weakly stationary time series with zero expected value

and a covariance $\gamma_\varepsilon(h)$ which is zero for all $h\neq o$. Usually M is set equal to zero

without loss of generality. A second type of linear representation for a weakly

stationary time series is known as an autoregressive process. A weakly station-

ary time series $\{X(t), t\varepsilon T\}$ possessing an expected value $M = E\{X(t)\}$ is said to

be an <u>autoregressive process of order p</u> if there exists a sequence of real numbers

$a_o = 1$, a_1, a_2, ..., a_p such that

$$(2) \quad \sum_{j=o}^{p} a_j [X(t-j)-M] = \varepsilon(t) \quad ; t\varepsilon T ,$$

where $\{\varepsilon(t), t\varepsilon T\}$ has the same properties as given above. A time series $\{X(t), t\varepsilon T\}$

whose covariance $\gamma_x(t_j, t_k)$ is a function only of $t_j - t_k$ but yet whose expected

value is a function of t is non-stationary in the mean as we said before. A time

series exhibiting such non-stationarity may still, however, possess a representa-

tion given by (1) or (2). In this event, we shall still refer to these representa-

tions as a moving average process and an autoregressive process, respectively.

In examining representations (1) and (2) with M = 0, the theory of

ordinary linear difference equations plays a major role. Since (2) is (apart

from its stochastic nature) merely a linear difference equation of order p, a

general solution for X(t) as a linear combination of $\varepsilon(t)$, $\varepsilon(t-1)$, ... can be

found. This suggests, then, that any representation of the form (2) can be in-

verted into a representation of the form (1). Since this suggests the possible

equivalence of a moving average representation and an autoregressive representation,

it is worthwhile to examine this possible equivalence in detail. Such a detailed

examination is beyond the scope of this paper, but we certainly should state those

conditions under which a moving average process permits an equivalent representation

as an autoregressive process. A sufficient condition for this equivalence is that

the variance of X(t), given by $\gamma_x(o)$, be finite.[4] From (1), this requires

---

[4]<u>Ibid.</u>, 95.

(3) $\quad \gamma_\varepsilon(o) \sum_{j=o}^{q} b_j^2 < \infty$,

where $\gamma_\varepsilon(o)$ is simply the variance of $\varepsilon(t)$. While for finite values of q this may appear to be a trivial condition, one last property of the representations (1) and (2) indicates the non-triviality of condition (3). This property is that a moving average process of finite order q is equivalent to an autoregressive process of infinite order and that an autoregressive process of finite order p is equivalent to a moving average process of infinite order. Depending on whether p is finite, then, we may have to let q approach infinity in (3) and convergence of the sum in (3) must be verified before deciding on the equivalence of the representations (1) and (2).

Given an equivalence between the representations (1) and (2) when $\{X(t), t\varepsilon T\}$ is weakly stationary and where either p or q (or both) must be infinite, it is apparent that the number of parameters needed to characterize the time series of interest $(a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q)$ may be large, depending upon which representation is of direct interest. Therefore, in the interest of parsimony, a third linear representation for a weakly stationary time series has received an increasing amount of attention in recent years. Specifically, a time series $\{X(t), t\varepsilon T\}$ is said to possess a mixed autoregressive moving average representation of order (p,q) if there exist sequences of real numbers $a_o = 1$, a_1, ..., a_p, $b_o = 1$, b_1, ..., b_q such that

(4) $\quad \sum_{j=o}^{p} a_j X(t-j) = \sum_{j=o}^{q} b_j \varepsilon(t-j)$; $t\varepsilon T$,

where $\{\varepsilon(t), t\varepsilon T\}$ is a weakly stationary time series with zero expected value and a covariance $\gamma_\varepsilon(h)$ which is zero for all $h \neq o$. Representation (4) therefore includes representations (1) and (2) as special cases while avoiding the practical problems associated with a proliferation of parameters.

Representation (4) can be simplified by the introduction of linear operations. Specifically, let B be the backward shift operator and let ∇ be the backward difference operator so that

$$BX(t) = X(t-1) \text{ and } \nabla X(t) = X(t) - X(t-1) = (1-B)X(t).$$

Then if

$$(5) \quad \Phi_p(B) = \sum_{j=0}^{p} a_j B^j \text{ and } \Theta_q(B) = \sum_{j=0}^{q} b_j B^j,$$

the representation (4) may be rewritten in the form

$$(6) \quad \Phi_p(B)X(t) = \Theta_q(B)\varepsilon(t) \quad ; t\varepsilon T.$$

Keeping the representations (4) and (6) in mind, it is helpful to recall that we are still considering time series $\{X(t), t\varepsilon T\}$ that are weakly stationary. For some time series, it is clearly unrealistic to assume, a priori, that the property of weak stationarity holds. Box and Jenkins,[5] in a series of papers beginning in 1962, have suggested generalized forms of (4) and (6) for the linear representation of a broad class of time series exhibiting non-stationarity. To motivate their generalization,[6] consider the autoregressive process of order one given by setting p = 1 and q = o in (4):

$$X(t) + a_1 X(t-1) = \varepsilon(t) \quad ; t\varepsilon T$$

The equivalent moving average representation is then

$$X(t) = \varepsilon(t) - a_1\varepsilon(t-1) + a_1^2\varepsilon(t-2) - a_1^3\varepsilon(t-3) + \ldots$$

In order for $\{X(t), t\varepsilon T\}$ to be weakly stationary, it is necessary that the variance of X(t) be finite. This requires that $|a_1| < 1$. If $|a_1| > 1$, the process $\{X(t), t\varepsilon T\}$ is said to diverge and is clearly non-stationary in variance. If $|a_1| = 1$, the

[5]Box, G. E. P., and Jenkins, G. M. Models for Prediction and Control III Linear Non-Stationary Models, Technical Report No. 79, University of Wisconsin Department of Statistics.

[6]Watts, D. G. An Introduction to Linear Modeling of Stationary and Non-Stationary Time Series, Technical Report No. 129, August 1967, University of Wisconsin Department of Statistics.

process is still non-stationary but in an interesting way. For, if $a_1 = +1$, the

process oscillates with an unbounded range. A more important case is when $a_1 = -1$.

In this case,

$$X(t) = \varepsilon(t) + \varepsilon(t-1) + \varepsilon(t-2) + \varepsilon(t-3) + \ldots$$

is the realization of what is termed a random walk. In fact, it is seen that the

first backward difference given by

$$X(t) - X(t-1) = \varepsilon(t) = (1-B)X(t)$$

is weakly stationary and, in fact, equivalent to a process with a covariance $\gamma_\varepsilon(h)$

which is zero for all $h \neq o$. While the original process $\{X(t), t\varepsilon T\}$ has no variance

(or higher central moments in general) its behavior is not divergent when $a_1 = -1$,

but is said to be "homogeneously" non-stationary. This terminology was first

introduced by Box and Jenkins.[7]

The above motivation suggests a generalization of (6) to a more com-

plete representation given by

(7) $\Phi_p(B)(1-B)^d X(t) = \Theta_q(B)\varepsilon(t)$; $t\varepsilon T$,

which of course includes both stationary and non-stationary time series. Repre-

sentation (7), following Box and Jenkins, is called an <u>integrated autoregressive</u>

moving average of order (p,d,q).[8] If $d = o$, (7) represents "well-behaved",

weakly stationary time series. If $d > o$, (7) represents "homogeneously" non-

stationary time series possessing infinite variances. This latter possibility

is one which has frequently caused concern in the analysis of economic time series.

It has been argued that time series possessing infinite variances cannot be poss-

ible in the economic sector. If this is indeed the case, then careful analysis

of any given time series should reveal whether it possesses a representation in

(7) with $d = o$. The truth of this argument, therefore, does not prohibit us from

---

[7] Box and Jenkins, <u>Op. cit.</u>

[8] Box and Jenkins, <u>Op. cit.</u>

at least considering the possibility that d > o in (7). Alternatively, it has

been shown in some recent work by Harrison[9] and Cogger[10] that several time series

from the economic sector exhibit behavior typical of time series represented by

(7) with d > o.

Given that a time series, either stationary or non-stationary, can be

represented by (7) for some set of parameters a_1, a_2, ..., a_p, b_1, b_2, ..., b_q, p, d,

and q, it is necessary to estimate these parameters. Specifically, a sequence of

realizations X(1), ..., X(N) of this time series must be utilized to estimate

p, d, q, a_1, a_2, ..., a_p, b_1, b_2, ..., b_q before forecasts of future realizations

X(N+1), X(N+2), ... of this time series can be calculated. The logical approach

to take at this point might therefore be to discuss estimation procedures before

examining forecasting procedures. It is convenient for the purposes of the auto-

mobile study, however, to reverse our approach and discuss forecasting procedures

first. This approach is taken in light of the evolutionary manner in which

economic simulation models are usually constructed. A modular approach is usually

taken in constructing such models, beginning with a relatively simple model and

then adding sophistication to the model as familiarity with its workings is

attained. This incremental or modular approach is ideally suited to the incor-

poration of a time series analysis program within the simulation model, since the

problems of forecasting and estimation can be totally segregated. Initially, a

simulation model will have a forecasting segment in which the time series model (7)

is specified a priori. As the simulation model attains a higher level of sophisti-

cation, the appropriate model will no longer be specified a priori but on the

basis of information flowing from a separate estimation segment within the simulation

[9]Harrison, P. J. "Exponential Smoothing and Short Term Sales Forecasting," Management Science, 13 (1967), 821-842.

[10]Cogger, K. O. "Statistical Foundations of General Order Exponential Smoothing with Implications for Applied Time Series Analysis" Unpublished Ph.D. dissertation, The University of Michigan, to appear.

model. Forecasting procedures will therefore be the subject of Section III and estimation procedures will be developed in Section IV.

III. Forecasting

Initially, the auto study simulation model will utilize a time series forecasting technique based upon a single model from (7). Regardless of the time series under consideration, the parameters p, d, q, a_1, a_2, ..., a_p, b_1, ..., b_q will remain fixed. For this particular time series representation, the expected squared forecast error will be a minimum only when the time series under consideration possesses the chosen representation. It is important, then, to choose a representation which will achieve reasonably accurate forecasts regardless of the actual time series representation. It has been shown by Harrison[11] and Cogger[12] that model (7) when $p = o$, $d = q = 1$, $b_1 = -\beta$, $|\beta| < 1$ is one such representation that achieves this type of "robustness" under misspecification.

With these parameter values, model (7) may be written in the form

$$(1-B)X(t) = (1-\beta B)\varepsilon(t),$$

which can be written as

$$(8) \quad X(t) - X(t-1) = \varepsilon(t) - \beta\varepsilon(t-1)$$

after removing the backward shift operator B. The basic problem is to find a forecast, given by $\hat{X}_t(\ell)$, of some future value $X(t+\ell)$ given only the realizations $X(t)$, $X(t-1)$, ... such that the mean squared forecast error given by

$$E[X(t+\ell) - \hat{X}_t(\ell)]^2$$

is a minimum for any $\ell > o$. While this problem seems formidable, it is actually quite easy to solve given (8). From (8), which is an ordinary difference equation, the time series $\{X(t), t\varepsilon T\}$ has an alternative representation of the form

$$(9) \quad X(t) = M + \varepsilon(t) + (1-\beta) \sum_{j=1}^{\infty} \varepsilon(t-j),$$

---

[11] Harrison, P. J., Op. cit.

[12] Cogger, K. O., Op. cit.

where M is a constant. Representation (9) is simply the sum of the solution

(given by M) of the homogeneous difference equation (X(t)-X(t-1)=0) and the

particular solution of the non-homogeneous difference equation (X(t)-X(t-1) =

$\varepsilon(t)-\beta\varepsilon(t-1)$). The forecast of X(t+ℓ) which achieves minimum mean squared error

is known to be simply the conditional expectation of X(t+ℓ) given X(t), X(t-1), ...

which is given by

$$(10) \quad \hat{X}_t(\ell) = E\{X(t+\ell)|X(t),X(t-1),...\} = M + (1-\beta) \sum_{j=0}^{\infty} \varepsilon(t-j).$$

For a complete discussion of this well known property of conditional expectations,

the reader is referred to Doob[13] for an advanced discussion and Malinvaud[14] for

a less theoretical discussion. From (9) and (10), it is seen that

$$(11) \quad X(t+\ell) - \hat{X}_t(\ell) = \begin{cases} \varepsilon(t+\ell) + (1-\beta) \sum_{j=1}^{\ell-1} \varepsilon(t+\ell-j) & ; \ell \geq 2 \\ \varepsilon(t+\ell) & ; \ell = 1. \end{cases}$$

The mean squared forecast error for any ℓ ≥ 1 is given by

$$(12) \quad E[X(t+\ell)-\hat{X}_t(\ell)]^2 = \ell_\varepsilon(o)[1+(\ell-1)(1-\beta)^2] \quad ; \ell \geq 1,$$

since the time series {ε(t),t∈T} has been assumed to be uncorrelated with mean

zero.

The form of the ℓ-step ahead forecast given by (10) is not utilized

in practice. Instead, such a forecast is usually given in terms of X(t), X(t-1), ...

since these are the observable random variables. To calculate this more useful

form of a forecast, we start with the original representation (8) in the form

$$(13) \quad X(t) = X(t-1) + \varepsilon(t) - \beta\varepsilon(t-1).$$

The observation to be forecast can then be expressed as

$$(14) \quad X(t+\ell) = X(t+\ell-1) + \varepsilon(t+\ell) - \beta\varepsilon(t+\ell-1) \quad ; \ell \geq 1.$$

The forecast of X(t+ℓ) made at time t, which we have denoted by $\hat{X}_t(\ell)$, is now

found by taking conditional expectations at time t on both sides of (14) using

the relations

---

[13]Doob, J. L. Op. cit.

[14]Malinvaud, E. Statistical Methods of Econometrics. Chicago; Rand McNally and Co., 1966.

$$E\{X(+\ell)\,|\,X(t),X(t-1),\ldots\} = \hat{X}_t(\ell) \quad ; \quad \ell \geqslant 1$$

$$E\{X(t+\ell-1)\,|\,X(t),X(t-1),\ldots\} = \hat{X}_t(\ell-1) \quad ; \quad \ell \geqslant 2$$

$$E\{X(t+\ell-1)\,|\,X(t),X(t-1),\ldots\} = X(t) \quad ; \quad \ell = 1$$

(15)

$$E\{\varepsilon(t+\ell)\,|\,X(t),X(t-1),\ldots\} = o \quad ; \quad \ell \geqslant 1$$

$$E\{\varepsilon(t+\ell-1)\,|\,X(t),X(t-1),\ldots\} = o \quad ; \quad \ell \geqslant 2$$

$$E\{\varepsilon(t+\ell-1)\,|\,X(t),X(t-1),\ldots\} = \varepsilon(t) \quad ; \quad \ell = 1$$

Whence we obtain the forecast $\hat{X}_t(\ell)$ entirely in terms of previous forecasts and of known values of the series, since from (11) it is seen that $\varepsilon(t) = X(t) - \hat{X}_{t-1}(1)$. Specifically, (14) and (15) reveal that

(16)
$$\hat{X}_t(1) = X(t) - \beta[X(t) - \hat{X}_{t-1}(1)] = (1-\beta)X(t) + \beta \hat{X}_{t-1}(1) \quad ; \quad \ell = 1$$

$$\hat{X}_t(\ell) = \hat{X}_t(\ell-1) \quad ; \quad \ell \geqslant 2.$$

The form of (16) reveals several important properties of time series forecasts based upon representation (8). First, the forecasting procedure is (computationally) extremely efficient. It is recursive in nature, allowing forecasts to be quickly revised in the light of new observations on the time series of interest. A new forecast is obtained simply as a weighted average (depending on the value of β) of the most recent observation and the most recent one step ahead forecast. Second, the forecast is constant for all future time periods. This second property is a necessary consequence of representation (8) since the underlying time series is assumed to have a constant expected value M. Third, the forecast implicitly assumes that an infinite number of past realizations of the time series are available for forecasting purposes. This property follows from the fact that the conditional expectations in (15) are based upon $X(t)$, $X(t-1),\ldots$ ad inf.

This last property deserves special attention, since in the automobile study's simulation model any time series will be observed for only a finite number of time periods. To apply the forecasting procedure (16), suppose that an initial forecast is used to determine a forecast at time $t = 1$. Then

$$\hat{X}_1(1) = X(1)(1-\beta) + \beta \hat{X}_0(1)$$

where $\hat{X}_0(1)$ is determined in some arbitrary manner. By recursively updating the forecast as new observations become available, (16) finally reveals that

$$(17) \quad \hat{X}_N(1) = (1-\beta) \sum_{j=0}^{N-1} \beta^j X(N-j) + \beta^N \hat{X}_0(1)$$

is the forecast of $X(N+1)$ based on the available realizations $X(1)$, $X(2)$, ..., $X(N)$ and on the initial forecast $\hat{X}_0(1)$. As the number, N, of observations increases, the effect of the arbitrarily specified value for $\hat{X}_0(1)$ becomes negligible (since $|\beta|<1$, $\lim_{N\to\infty} \beta^N = 0$). The effect of $\hat{X}_0(1)$ is significant if N is small, however, and it is important to eliminate its effect upon $\hat{X}_N(1)$ in this case.[15] Taking expected values on both sides of (17), it is seen that

$$E\{\hat{X}_N(1)\} = \frac{(1-\beta)(1-\beta^N)M}{(1-\beta)} + \beta^N \hat{X}_0(1)$$

where M is the expected value of $X(N-j)$ for all $j \geq 0$. But then

$$E\left\{ \frac{\hat{X}_N(1) - \beta^N \hat{X}_0(1)}{(1-\beta^N)} \right\} = M$$

Therefore for any value of N and any initial forecast $\hat{X}_0(1)$, we may correct the forecast $\hat{X}_N(1)$ in such a way that it will be an unbiased estimate of M for the time series $\{X(t), t \epsilon T\}$. Specifically, it is recommended that $\hat{X}_0(1)$ be equated to zero in the forecasting procedure described by (16) and that $\hat{X}_t(\ell)$ be adjusted by a factor of $(1-\beta^t)$ for all $t \geq 1$. For large values of t, this correction will not be needed but for small values of t its use will insure unbiasedness in the resulting forecast.

To summarize the institution of this forecasting procedure in the automobile simulation model, we now recall the pertinent steps that must be carried out. It is assumed that N realizations of a time series have been observed and denoted $X(1)$, $X(2)$, ..., $X(N)$. It is further assumed that a forecast is required for any future realization $X(N+\ell)$ where $\ell > 0$. We have denoted this

---

[15] Wade, R. C. "A Technique for Initializing Exponential Smoothing Forecasts," *Management Science*, 13 (1967), 601-2.

forecast by $\hat{X}_N(\ell)$. Recursively, the following steps are carried out.

Step 1: $\hat{X}_1(1) = (1-\beta)X(1)$

Step 2: $\hat{X}_t(1) = (1-\beta)\hat{X}(t) + \beta\hat{X}_{t-1}(1)$; $t = 2, 3, \ldots, N$

Step 3: $\hat{X}_N(\ell) = \dfrac{\hat{X}_N(1)}{(1-\beta^N)}$; $\ell \geqslant 1$

Step 3, of course, is not necessary for large N but should always be made for

the small sample sizes (N≤20) extant in the automobile study.

Finally, it is necessary to choose a value for the parameter β. Since

we wish the forecasts obtained by this forecasting procedure to be as accurate

as possible (in the sense of minimizing the mean squared forecast error), the

choice of a value for β is important. A value of $\beta=.5$ is recommended for several

reasons. First, the above forecasting procedure is based on the model (7) when

$(p,d,q) = (o,1,1)$. If the time series under consideration is truly represented

by this model with $\beta=.5$, the forecast is known to possess minimum mean squared

error. If the true value of β is not .5, this forecast will possess a larger

mean squared error than is theoretically possible. The increase in mean squared

error will be, however, less than 33-1/3% for any true value of β in the range

$o \leqslant \beta \leqslant 1$, a considerably wide interval. Second, suppose the time series under con-

sideration cannot be represented by model (7) with $(p,d,q) = (o,1,1)$ for any

value of β. Then our choice of a value for β (given that we still insist upon

using the above forecasting procedure) should be based on the principle of provid-

ing a reasonable amount of protection against the existence of alternative repre-

sentations for the time series of interest. One commonly assumed time series

model is given by the condition $(p,d,q) = (1,0,0)$ with $-1 \leqslant a_1 \leqslant 0$. Such a time

series representation is called a first order Markov process and the above fore-

casting procedure will not achieve the minimum mean squared error that is theo-

retically possible. For a value of $\beta=.5$, however, the mean squared error of the

above forecast will be increased over this minimum by less than 30%, however.

The above properties of the chosen forecasting procedure are discussed in detail in Cogger.[16] The robustness of this forecasting procedure under misspecification of the true time series representation renders it quite valuable when this true representation is unknown as it will be in the automobile study.

The simple forecasting procedure described in steps 1-3 is called <u>first order exponential smoothing</u>. It is called "first order" because the expected value of the underlying time series is constant. The term "exponential" refers to the geometrically declining weights attached to previous observations in the forecast (17). This particular forecasting procedure, of course, is only one of many that can be derived from the integrated autoregressive moving average representation (7). Accordingly, we shall now consider more general time series forecasting procedures derived from (7).

The first generalization we wish to discuss is the time series forecasting procedure known as <u>general order exponential smoothing</u>. To motivate our discussion, it is convenient to reconsider steps 1-3 for first order exponential smoothing in the form:

Step 1': $S_1^{[1]} = (1-\beta)X(1)$

Step 2': $S_t^{[1]} = (1-\beta)X(t) + \beta S_{t-1}^{[1]}$; $t = 2, 3, \ldots, N$

Step 3': $\hat{X}_N(\ell) = S_N^{[1]}/[1-\beta^N]$; $\ell \geqslant 1$

This notation is more suitable for the discussion which follows, and we shall always refer to $S_t^{[1]}$ as the first order exponentially smoothed value of the data $X(1), \ldots, X(t)$.

These steps suggest an immediate generalization, however. Suppose we define $S_t^{[p]}$ as the p-th order exponentially smoothed value of the data

---

[16]Cogger, K. O. <u>Op. cit.</u>

X(1), ..., X(t) given by applying steps 1'-3' not to the original data but to the (p-1)-st order exponentially smoothed value of the data. More precisely, let $S_t^{[p]}$ be defined by

$$(18) \quad S_t^{[p]} = \begin{cases} (1-\beta)X(t) + \beta S_{t-1}^{[1]} & ; \ p=1 \ ; \ t \geq 1 \\ (1-\beta) S_t^{[p-1]} + \beta S_{t-1}^{[p]} & ; \ p \geq 2 \ ; \ t \geq 1 \\ o & ; \ p \geq 1 \ ; \ t=o. \end{cases}$$

To see that this definition can lead to useful results, (18) can be solved recursively for $t = 1, 2, \ldots, N$ to obtain[17]

$$(19) \quad S_N^{[p]} = (1-\beta)^p \sum_{j=o}^{N-1} \beta^j \binom{j+p-1}{p-1} X(N-j) \quad ; \ p \geq 1 \ ; \ N \geq 1.$$

(Note that if p=1, (19) reduces to the previously defined first order smoothed value of the data.) We now assume that the expected value of X(N-j) is a polynomial in j of degree n-1. Hence, this expected value may be expanded in a Taylor series about the point N in the form

$$(20) \quad E\{X(N-j)\} = \sum_{k=o}^{n-1} (-j)^k \ A_N^{[k]} /k!,$$

where the $A_N^{[k]}$, for any fixed value of N, are unknown constants. The important question to be answered at this point is whether the exponentially smoothed values of the data in (19) can be used to estimate the unknown parameters $A_N^{[k]}$ for k = 0, 1, ..., n-1. The answer to this question is in the affirmative, as shown in Brown[18] for the asymptotic case (N tending in the limit to infinite) and in Cogger[19] for any value of N. To show this, we take expected values on both sides of (19) and utilize (20) to get

---

[17] Brown, R. G., Op. cit.

[18] Ibid.

[19] Cogger, K. O., Op. cit.

(21) $E\{S_N^{[p]}\} = \sum_{k=0}^{n-1} \left\{ \frac{(-1)^k(1-\beta)^p}{k!(p-1)!} \sum_{j=0}^{N-1} \frac{j^k \beta^j (j+p-1)!}{j!} \right\} A_N^{[k]}$; $p \geqslant 1$; $N \geqslant 1$.

Defining the nx1 vector $\underline{S}_N$ in the form

(22) $\underline{S}_N = \begin{bmatrix} S_N^{[1]} \\ S_N^{[2]} \\ \vdots \\ S_N^{[n]} \end{bmatrix}$,

and the nx1 vector $\underline{A}_N$ in the form

(23) $\underline{A}_N = \begin{bmatrix} A_N^{[o]} \\ A_N^{[1]} \\ \vdots \\ A_N^{[n-1]} \end{bmatrix}$,

and the nxn matrix M(N) with elements in the p-th row and (k+1)-st column given

by

(24) $M_{p,k+1} = \frac{(-1)^k(1-\beta)^p}{k!(p-1)!} \sum_{j=0}^{N-1} \frac{j^k \beta^j (j+p-1)!}{j!}$; $1 \leqslant p \leqslant n$; $o \leqslant k \leqslant n-1$,

we now rewrite (21) in the matrix form

(25) $E\{\underline{S}_N\} = M(N)\underline{A}_N$; $N \geqslant 1$.

But then an unbiased estimate of the elements of $\underline{A}_N$ is given by

(26) $\hat{\underline{A}}_N = M(N)^{-1}\underline{S}_N$; $N \geqslant 1$,

where $M(N)^{-1}$ is merely the inverse of the matrix M(N). Defining

(27) $\underline{\ell} = \begin{bmatrix} 1 \\ \ell \\ \vdots \\ \ell^{n-1}/(n-1)! \end{bmatrix}$

as an nx1 vector, it is clear from (20) that

$$E\{X(N+\ell)\} = \sum_{k=0}^{N-1} \frac{\ell^k A_N^{[k]}}{k!} = \underline{\ell}' \, \underline{A}_N$$

and therefore an unbiased estimate of the mean of the time series at time $N+\ell$

is given by

(28) $\quad \hat{X}_N(\ell) = \underline{\ell}'\hat{\underline{A}}_N = \underline{\ell}'M(N)^{-1}\underline{S}_N \quad ; \quad N \geq 1.$

Alternatively, (28) describes the ℓ-step ahead forecast of the time series based

on the observations $X(1)$, $X(2)$, ..., $X(N)$. The above forecasting procedure is

called exponential smoothing of order n (or, sometimes, simply general order ex-

ponential smoothing).

We have yet to show that general order exponential smoothing minimizes the

mean squared forecast error for any time series represented in (7). The time series

representation in (7) for which this property holds is given by setting p=o, d=q=n,

and $b_j = \binom{n}{j}(-\beta)^j$; j=o, 1, ..., n. That is, the forecast given by exponential

smoothing of order n achieves minimum mean squared error for a time series possess-

ing the linear representation

(29) $\quad \nabla^n X(t) = \sum_{j=o}^{n} \binom{n}{j}(-\beta)^j \, \varepsilon(t-j)$

where the $\varepsilon(t)$ are uncorrelated with mean zero. A detailed proof of this result

is beyond the scope of this paper but may be found in Cogger.[20]

Exponential smoothing of order n is perhaps one of the most widely used

forecasting procedures based on (7). Its wide use has been promoted by the ease

with which time series forecasts may be made (cf. (18), (22), (24), (27), and (28)).

The only apparent difficulty with the procedure is the complexity of the matrix

M(N) for large values of n. This problem has been recently resolved.[21] A recur-

sive method of evaluating this matrix for any set of values for n, N, and β is given

by

---

[20] Cogger, K. O., _Op. cit._

[21] _Ibid._

(30) $M_{p,k+1}(N) = \begin{cases} 1-\beta^N \\ M_{p-1,1}(N) - (1-\beta)^{p-1}\beta^N \binom{p+N-2}{N-1} \quad ; p \geq 2 \; ; k=o \\ (P/k)[M_{p,k}(N)-M_{p+1,k}(N)/(1-\beta)] \quad ; p \geq 1 \; ; k \geq 1. \end{cases}$

The relationships in (30) are easily verified by substitution from (24).

We now consider representation (7) in its completely general form given by

(31) $\sum_{j=o}^{p} a_j \nabla^d X(t-j) = \sum_{j=o}^{q} b_j \varepsilon(t-j).$

For given values of p, d, q, a_1, ..., a_p, b_1, ..., b_q the representation (31) may be written alternatively in the form

(32) $X(t) = m(t) + \sum_{k=o}^{\infty} \Psi_k \varepsilon(t-k)$

where m(t) is the solution of the homogeneous difference equation given by

(33) $\sum_{j=o}^{p} a_j \nabla^d X(t-j) = o,$

and where the first few cofficients $\Psi_o=1$, Ψ_1, Ψ_2, ... which are needed in what follows may be obtained by equating coefficients of $B^o=1$, B, B^2, ... in the expression

(34) $\sum_{k=o}^{\infty} \Psi_k B^k = \sum_{j=o}^{q} b_j B^j / [(1-B)^d \sum_{j=o}^{p} a_j B^j].$

This follows from (7) by noting that if

$\Phi_p(B)(1-B)^d X(t) = \theta_q(B) \varepsilon(t),$

then

$X(t) = [\frac{\theta_q(B)}{\Phi_p(B)(1-B)^d}] \varepsilon(t) = \sum_{k=o}^{\infty} \Psi_k B^k.$

But then from (32), it is seen that

(35) $E\{X(t+\ell)|X(t),X(t-1), ...\} = m(t+\ell) + \sum_{k=\ell}^{\infty} \Psi_k \varepsilon(t+\ell-k) = \hat{X}_t(\ell).$

Combining (32) and (35), the minimum mean squared error of the ℓ-step ahead forecast will be given by

$$(36) \quad E\{X(t+\ell) - \hat{X}_t(\ell)\}^2 = \gamma_\varepsilon(o) \sum_{k=o}^{\ell-1} \psi_k^2 \quad ; \quad \ell \geqslant 1.$$

The above results enable us to determine the variance of the ℓ-step ahead forecast. The form of the forecast itself is most easily found in a different manner. Accordingly, we rewrite (31) in the form

$$(37) \quad \sum_{j=o}^{p} a_j \nabla^d X(t-j) = X(t) - \sum_{j=1}^{p+d} \Phi_j X(t-j) = \sum_{j=o}^{q} b_j \varepsilon(t-j).$$

This form assumes that $p+d \geqslant 1$, an assumption that will be maintained throughout our discussion. But then

$$(38) \quad X(t+\ell) = \sum_{j=1}^{p+d} \Phi_j X(t+\ell-j) + \sum_{j=o}^{q} b_j \varepsilon(t+\ell-j).$$

We now take expectations, conditional on $X(t)$, $X(t-1)$, ..., on both sides of (38) (recalling (15)) to get

$$(39) \quad \hat{X}_t(\ell) = E\{\sum_{j=1}^{p+d} \Phi_j X(t+\ell-j) | X(t), X(t-1), \ldots\} + E\{\sum_{j=o}^{q} b_j \varepsilon(t+\ell-j) | X(t), X(t-1) \ldots\}$$

where

$$(40) \quad E\{\sum_{j=1}^{p+d} \Phi_j X(t+\ell-j) | \ldots\} = \begin{cases} \sum_{j=1}^{\ell-1} \Phi_j \hat{X}_t(\ell-j) + \sum_{j=\ell}^{p+d} \Phi_j X(t+\ell-j) & ; \ p+d \geqslant \ell \geqslant 2 \\[2em] \sum_{j=1}^{p+d} \Phi_j X(t+1-j) & ; \ p+d \geqslant \ell = 1 \\[2em] \sum_{j=1}^{p+d} \Phi_j \hat{X}_t(\ell-j) & ; \ p+d < \ell ; \ell \geqslant 2 \end{cases}$$

and where

$$(41) \quad E\{\sum_{j=o}^{q} b_j \varepsilon(t+\ell-j) | \ldots\} = \begin{cases} \sum_{j=\ell}^{q} b_j \varepsilon(t+\ell-j) & ; \ 1 \leqslant \ell \leqslant q \\[1.5em] o & ; \ \ell > q \end{cases} \quad .$$

Since, from (32) and (35), $\varepsilon(t+\ell-j) = X(t+\ell-j) - \hat{X}_{t+\ell-j-1}$ (1) for $j \geqslant \ell$, (39)-(41)

completely determine the minimum mean squared error forecast $\hat{X}_t(\ell)$ in terms of

previous observations $X(t)$, $X(t-1)$, ... and of previous forecasts $\hat{X}_{t-1}$ (1),

$\hat{X}_{t-2}$ (1), ... for any values of p, d, and q such that $p+d \geqslant 1$.

For the case $p+d=o$, (31) may be written as

$$(42) \quad X(t) = \sum_{j=o}^{q} b_j \varepsilon(t-j).$$

But then

$$(43) \quad X(t+\ell) = \sum_{j=o}^{q} b_j \varepsilon(t+\ell-j).$$

Taking conditional expectations on both sides of (43) it is seen that

$$(44) \quad \hat{X}_t(\ell) = \begin{cases} \displaystyle\sum_{j=\ell}^{q} b_j \varepsilon(t+\ell-j) & ; \ 1 \leqslant \ell \leqslant q \\ o & ; \ \ell > q, \end{cases}$$

which is the same as the expression in (41)

Hence, equations (38)-(41) and (43)-(44) enable forecasts to be made

for any set of values for p, d, and q. One last point to be considered is the

need to calculate "initial conditions" for the forecast errors

$\varepsilon(t+\ell-j) = X(t+\ell-j) - \hat{X}_{t+\ell-j-1}$ (1) for $j \geqslant \ell$ found in (41) and (44). The recommended

procedure is to set these equal to zero or, equivalently, equal to their expected

values. For a large number of observations, of course, these "initial conditions"

will have a negligible effect upon the forecast $\hat{X}_t(\ell)$, and for any value of $t \geqslant 1$,

the forecast will be an unbiased estimate of $X(t+\ell)$. If the particular representa-

tion in (7) corresponds to general order exponential smoothing, of course, these

initial conditions may be eliminated in a more rigorous manner, and the necessary

forecasts may be calculated, as before, without the use of equations (38)-(44).

IV. Estimation of Forecasting Parameters

Suppose that N consecutive realizations $X(1), X(2), \ldots, X(N)$ of a time

series $\{X(t), t \varepsilon T\}$ represented by the model (7) have been observed. In order to

implement the forecasting procedures described in Section III, appropriate estimates of the unknown parameters $a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q$ must be obtained. In addition, the order (p, d, q) of the process must be determined.

Each of the three estimation procedures described in this section is based on the assumption that (p, d, q) is known. For this given order, each procedure yields estimates of the unknown parameters as well as a statistical measure of the goodness-of-fit of the available realizations to the model (7). These estimation procedures do, however, permit the determination of the proper order (p, d, q) in the following manner. First, estimation procedures are instituted for a large number of possible orders of the process. Empirical evidence has suggested that most economic time series can be modeled reasonably well by one of the following orders: $d = 0$, 1, or 2 and $p = q = 0$, 1, or 2. These nine different orders are therefore recommended for use in economic simulation models. Second, a statistical measure of goodness-of-fit is calculated for each of these nine different orders. Finally, the order (p, d, q) which yields the best fit of the available realizations to the model (7) is specified as the appropriate order of the process, and the corresponding estimates of the parameters $a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q$ are then used to carry out the forecasting procedures of Section III.

Each of the three estimation procedures to be presented are based only upon realizations of $\nabla^d X(t)$ where d, as indicated previously, has been specified. For notational convenience, we therefore define

$$Y(t) = \nabla^d X(t)$$

as the d-th backward difference of $X(t)$, and rewrite the model (7) in the form

$$(45) \quad \sum_{j=0}^{p} a_j Y(t-j) = \sum_{j=0}^{q} b_j \varepsilon(t-j).$$

Estimation Procedure A(q=o)[22]

The first estimation procedure to be discussed was first derived by Mann

and Wald for q=o in (45). In this case, (45) may be rewritten in the form

(46) $Y(t) = -a_1 Y(t-1) - a_2 Y(t-2) - \ldots - a_p Y(t-p) + \varepsilon(t)$.

But (46) is similar in appearance to equations appearing in ordinary multiple

regression theory and it might therefore be suggested that appropriate estimates

of the parameters a_j could be obtained by treating (46) as a classical least-squares

regression problem where $Y(t)$ is the dependent variable and $Y(t-1)$, $Y(t-2),\ldots,$

$Y(t-p)$ are the independent variables. Estimates of a_j obtained in this manner do not

possess the usual statistical properties of least squares estimates, however, since

the random variables $Y(t)$ are not independent.

Mann and Wald, however, showed rigorously for the first time that

asymptotically (large N) the statistical properties of least squares estimates of

a_j in (46) are the same as those of least squares regression estimates in multi-

variate normal systems. Summarizing Mann and Wald's results, we note first that

if N realizations of $X(t)$ are available, then only $(N-d)$ realizations of $Y(t)$ are

available, and these are denoted by $Y(1),Y(2),\ldots,Y(N-d)$. Second, estimates

$\hat{a}_1,\hat{a}_2,\ldots,\hat{a}_p$ of the parameters $a_1,a_2,\ldots,a_p$ are given by those values which

minimize the sum of squares

(47) $S = \displaystyle\sum_{t=p+1}^{N-d} [Y(t) + a_1 Y(t-1) + \ldots + a_p Y(t-p)]^2$.

Differentiating S with respect to $a_1,a_2,\ldots,a_p$ and equating to zero yields the

following matrix formulation of the normal equations:

(48) $(Y'Y)\hat{\underline{a}} = Y'y$,

---

[22]Mann, H. B., and Wald, A., "On the Statistical Treatment of Linear Stochastic Difference Equations," _Econometrica_, 11 (1943), 173-220.

where

$$(Y'Y) = \begin{bmatrix} \sum Y^2(t-1) & \sum Y(t-1)Y(t-2) & \dots & \sum Y(t-1)Y(t-p) \\ \sum Y(t-1)Y(t-2) & \sum Y^2(t-2) & \dots & \sum Y(t-2)Y(t-p) \\ \vdots & \vdots & & \vdots \\ \sum Y(t-1)Y(t-p) & \sum Y(t-2)Y(t-p) & \dots & \sum Y^2(t-p) \end{bmatrix} ,$$

$$\hat{\underline{a}} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} , \text{ and } Y'\underline{y} = \begin{bmatrix} -\sum Y(t)Y(t-1) \\ -\sum Y(t)Y(t-2) \\ \vdots \\ -\sum Y(t)Y(t-p) \end{bmatrix} , \text{ where}$$

all summations are from t=p+1 to t=N-d.

After obtaining the estimates $\hat{\underline{a}}$ from (48), the estimate of the variance of $\varepsilon(t)$ in (46) is given by

$$(49) \quad \hat{\gamma}_\varepsilon(o) = \frac{\displaystyle\sum_{t=p+1}^{N-d} [Y(t) + \hat{a}_1 Y(t-1) + \dots + \hat{a}_p Y(t-p)]^2}{(N-d-p)} = \frac{\displaystyle\sum_{t=p+1}^{N-d} Y^2(t) - \hat{\underline{a}}'Y'Y\hat{\underline{a}}}{(N-d-p)}$$

For purposes of testing hypotheses and setting confidence intervals, Mann and Wald show that the asymptotic covariance matrix of $\hat{\underline{a}}$ is given by

$$(50) \quad \text{Var } \hat{\underline{a}} = (Y'Y^{-1})\gamma_\varepsilon(o),$$

and that $\hat{\underline{a}}$ has a limiting multivariate normal distribution. That is, if the element in the i-th row and i-th column of $(Y'Y^{-1})$ is given by C_{ii}, then

$$z = \frac{\hat{a}_i - a_i}{\sqrt{\gamma_\varepsilon(o)C_{ii}}}$$

is asymptotically distributed as a standard normal random variate.

Estimation Procedure B (p=o)[23]

  The second estimation procedure to be discussed was first derived by Durbin for the case p=o in (45). In this case, (45) may be rewritten in the form

$$(51) \quad Y(t) = \sum_{j=0}^{q} b_j \varepsilon(t-j).$$

Durbin noted that since (51) describes a moving average process of finite order q, the time series $\{Y(t), t\varepsilon T\}$ must also possess an autoregressive representation of infinite order. If this autoregressive representation were of finite order, say k, then Estimation Procedure A would be directly applicable. Durbin showed that if k were taken to be sufficiently large, then this approach could be taken.

  Following Durbin, we first approximate the autoregressive representation of $\{Y(t), t\varepsilon T\}$ by

$$(52) \quad \sum_{j=0}^{k} a_j Y(t-j) = \varepsilon(t)$$

<u>for large k</u>. Let $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_k$ be the least squares estimates of $a_1, a_2, \ldots, a_k$ in (52) given by a direct application of Estimation Procedure A for p=k (cf. (48)). Durbin then showed that these estimates could be used to obtain consistent and asymptotically normal estimates of $b_1, b_2, \ldots, b_q$ in the following manner. Let the qxq matrix A be defined by

---

[23]Durbin, J., "Efficient Estimation of Parameters in Moving Average Models," <u>Biometrika</u>, 46 (1959), 306-316.

$$A = \begin{bmatrix} \sum_{j=0}^{k} \hat{a}_j^2 & \sum_{j=0}^{k-1} \hat{a}_j \hat{a}_{j+1} & \cdots & \sum_{j=0}^{k-q+1} \hat{a}_j \hat{a}_{j+q-1} \\ \\ \sum_{j=0}^{k-1} \hat{a}_j \hat{a}_{j+1} & \sum_{j=0}^{k} \hat{a}_j^2 & \cdots & \sum_{j=0}^{k-q+2} \hat{a}_j \hat{a}_{j+q-2} \\ \\ \vdots & \vdots & & \vdots \\ \\ \sum_{j=0}^{k-q+1} \hat{a}_j \hat{a}_{j+q-1} & \sum_{h=0}^{k-q+2} \hat{a}_j \hat{a}_{j+q-2} & \cdots & \sum_{j=0}^{k} \hat{a}_j^2 \end{bmatrix}.$$

Let $\hat{\underline{b}} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_q \end{bmatrix}$ and $\underline{a}^* = \begin{bmatrix} -\sum_{j=0}^{k-1} \hat{a}_j \hat{a}_{j+1} \\ -\sum_{j=0}^{k-2} \hat{a}_j \hat{a}_{j+2} \\ -\sum_{j=0}^{k-q} \hat{a}_j \hat{a}_{j+q} \end{bmatrix}$. Then $\hat{\underline{b}}$ is

given as the solution to the equation

(53) $A\hat{\underline{b}} = \underline{a}^*$

and the estimates $\hat{\underline{b}}$ possess the statistical properties previously stated.

The variance of these estimators is not as easily obtained as in Procedure A. However, when $q \leq 2$ (as will be the case according to our recommendations made at the beginning of this section), these variances are easily found. Durbin showed that if $q=1$, the variance of $\hat{b}_1$ is given by

$$\text{var } \hat{b}_1 = (1-b_1^2)/(N-d-k)$$

If $q=2$, the variance of $\hat{b}_1$ or $\hat{b}_2$ is given by

$$\text{var } \hat{b}_1 = \text{var } \hat{b}_2 = (1-b_2^2)/(N-d-k).$$

Hypothesis tests and confidence intervals may therefore be constructed by noting that

$$z = \sqrt{N-d-k} \; (\hat{b}_1 - b_1)/\sqrt{1-b_1^2}$$

is a standard normal random variable when $q=1$ and that

$$z_j = \sqrt{N-d-k}\ (\hat{b}_j - b_j)/\sqrt{1-b_2^2}\ ; \ j=1\ \text{or}\ 2$$

is a standard normal random variable when q=2. Finally, since this estimation

procedure utilizes Procedure A explicitly, an estimate of the variance of $\varepsilon(t)$ may

be obtained from (49) after setting p=k.

Estimation Procedure C (Any value for p, Any value for q)[24]

The last estimation procedure to be discussed was first derived by

Durbin for general values of p and q in (45). The procedure is _iterative_ in nature,

and each iteration involves a number of steps.

The _first step_ suggested by Durbin is to approximate the representation

for $\{Y(t), t \varepsilon T\}$ in (45) by an autoregressive representation

$$(54) \quad \sum_{j=o}^{k} a_j' Y(t-j) = \varepsilon(t)$$

for _large k_, and to apply Procedure A to obtain estimates $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_k$ of the

parameters in (54).

The _second step_ taken by Durbin is to calculate

$$(55) \quad e(t) = \sum_{j=o}^{k} \hat{a}_j' Y(t-j),$$

which is simply the residual error from the regression of Procedure A. Estimates

$\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p, \hat{b}_1, \hat{b}_2, \ldots, \hat{b}_q$ of the unknown parameters in (45) are then obtained by

calculating those values which minimize the sum of squares

$$(56) \quad S = \sum_{t=p+k+1}^{N-d} [Y(t)+a_1 Y(t-1)+\ldots+a_p Y(t-p)-b_1 e(t-1)-\ldots-b_q e(t-q)]^2.$$

(The limits on the summation are easily obtained by noting which terms in the

summand are available.) The appropriate estimates are given by solving

---

[24]Durbin, J., "The Fitting of Time Series Models," _Revue Inst. Int. de Stat._, 28 (1960), 233-243.

$$(57) \quad \begin{bmatrix} M_{11} & \vdots & M_{12} \\ \cline{1-3} M_{21} & \vdots & M_{22} \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hline \hat{b} \end{bmatrix} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix} \quad , \text{ where}$$

$$M_{11} = \begin{bmatrix} \sum Y^2(t-1) & \sum Y(t-2)Y(t-1) & \cdots & \sum Y(t-p)Y(t-1) \\ \sum Y(t-1)Y(t-2) & \sum Y^2(t-2) & \cdots & \sum Y(t-p)Y(t-2) \\ \vdots & \vdots & & \vdots \\ \sum Y(t-1)Y(t-p) & \sum Y(t-2)Y(t-p) & \cdots & \sum Y^2(t-p) \end{bmatrix} ,$$

$$M_{12} = \begin{bmatrix} -\sum e(t-1)Y(t-1) & -\sum e(t-2)Y(t-1) & \cdots & -\sum e(t-q)Y(t-1) \\ -\sum e(t-1)Y(t-2) & -\sum e(t-2)Y(t-2) & \cdots & -\sum e(t-q)Y(t-2) \\ \vdots & \vdots & & \vdots \\ -\sum e(t-1)Y(t-p) & -\sum e(t-2)Y(t-p) & \cdots & -\sum e(t-q)Y(t-p) \end{bmatrix}$$

$$M_{21} = M_{12}',$$

$$M_{22} = \begin{bmatrix} \sum e^2(t-1) & \sum e(t-2)e(t-1) & \cdots & \sum e(t-q)e(t-1) \\ \sum e(t-1)e(t-2) & \sum e^2(t-2) & \cdots & \sum e(t-q)e(t-2) \\ \vdots & \vdots & & \vdots \\ \sum e(t-1)e(t-q) & \sum e(t-2)e(t-q) & \cdots & \sum e^2(t-q) \end{bmatrix}$$

$$\underline{x}_1 = \begin{bmatrix} -\sum Y(t)Y(t-1) \\ -\sum Y(t)Y(t-2) \\ \vdots \\ -\sum Y(t)Y(t-p) \end{bmatrix} \quad , \quad \underline{x}_2 = \begin{bmatrix} \sum Y(t)e(t-1) \\ \sum Y(t)e(t-2) \\ \vdots \\ \sum Y(t)e(t-q) \end{bmatrix} \quad ,$$

$$\hat{\underline{a}} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} \quad , \text{ and } \quad \hat{\underline{b}} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_q \end{bmatrix} \quad , \text{ where all summations}$$

are from $t = p+k+1$ to $N-d$.

The <u>third step</u> suggested by Durbin is the following. Suppose that the true values of $a_1, a_2, \ldots, a_p$ are known in (45). Let

$$Z(t) = Y(t) + a_1 Y(t-1) + \ldots + a_p Y(t-p)$$

to obtain

(58) $Z(t) = \varepsilon(t) + b_1 \varepsilon(t-1) + b_2 \varepsilon(t-2) + \ldots + b_q \varepsilon(t-q)$

for $t = p+1, p+2, \ldots, N-d$. Then Procedure B ($p=o$) can be directly applied to obtain estimates $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_q$ by simply letting $Y(t) = Z(t)$ in the indicated procedure.

The <u>fourth step</u> suggested by Durbin is the following. Suppose that the true values of $b_1, b_2, \ldots, b_q$ are known in (45). Utilizing the backward shift operator B, (45) may be written in the form

(59) $\displaystyle\sum_{j=o}^{p} a_j B^j Y(t) = \sum_{j=o}^{q} b_j B^j \varepsilon(t).$

But then

(60) $\displaystyle\sum_{j=o}^{p} a_j B^j W(t) = \varepsilon(t) = \sum_{j=o}^{p} a_j W(t-j)$

when we define

(61) $\displaystyle W(t) = \sum_{j=o}^{\infty} \gamma_j B^j Y(t)$

such that

(62) $\displaystyle\sum_{j=o}^{\infty} \gamma_j B^j = [\sum_{j=o}^{q} b_j B^j]^{-1}.$

Approximating (61) and (62) for <u>large k,</u> we write

$$(63) \quad W(t) = \sum_{j=o}^{k} \gamma_j Y(t-j)$$

and

$$(64) \quad [\sum_{j=o}^{k} \gamma_j B^j] [\sum_{j=o}^{q} b_j B^j] = 1.$$

Knowing the values of b_j, the values of γ_j may easily be found by equating co-efficients of like powers of B in (64). For the values of q that are of concern, the following results are easily obtained.

q=1: $\gamma_j = (-b_1)^j$; $j \geq o$.

q=2: $\gamma_o = 1$; $\gamma_1 = -b_1$; $\gamma_j = -b_1 \gamma_{j-1} - b_2 \gamma_{j-2}$ for $j \geq 2$.

After making the indicated transformation (63) for <u>large k</u>, the parameters a_1, a_2,..., a_p may be estimated by a direct application of Procedure A, substituting W(t) for Y(t) in the normal equations (48). Thus, knowing $b_1, b_2, ..., b_q$, we obtain estimates $\hat{a}_1, \hat{a}_2, ..., \hat{a}_p$.

The above steps are related to each other in the following manner. First, preliminary (and usually quite accurate) estimates of $a_1, a_2, ..., a_p$, $b_1, b_2, ..., b_q$ are obtained by carrying out the first two steps described above. Given the resultant estimates $\hat{a}_1, \hat{a}_2, ..., \hat{a}_p$, the third step is carried out under the assumption that these estimates are equal to the actual values $a_1, a_2, ..., a_p$. This third step, of course, will yield new estimates of the unknown parameters $b_1, b_2, ..., b_q$. Given these latter estimates, the fourth step is then carried out under the assumption that these estimates are equal to the actual values $b_1, b_2, ..., b_q$. This fourth step, of course, will yield new estimates of the unknown parameters $a_1, a_2, ..., a_p$. This sequence of estimation procedures constitutes the <u>first iteration</u>. Additional iterations are then carried out <u>beginning with the third step</u>.

This estimation procedure clearly requires the establishment of a non-

arbitrary rule which can be used to terminate the iterative process above. Durbin

has indicated that the preliminary estimates obtained with steps one and two above

are usually quite good, thereby reducing the number of iterations that might be

required if poor "starting values" were utilized. Accordingly, it is recommended

that an upper limit of 5 iterations (including the first pass through the indicated

four steps) be established to reduce computational effort. Further, it is recommended

that each time the fourth step is carried out, the estimate of the variance of

$\varepsilon(t)$ given in (49) be calculated. The iterative procedure should be terminated

whenever $\hat{\gamma}_\varepsilon(o)$ changes by less than 10% from the previous iteration. This additional

termination condition should further reduce the expected number of iterations

without adversely affecting any forecasts that are subsequently made.

Finally, the only remaining ambiguities are the number of lags (k) to be

taken in some of the steps above where k was simply taken to be large, and the

statistical measure of goodness of fit that will be used to determine the proper

order of the process. It has been shown in recent Monte Carlo studies that when

both the sample size (N-d-p) and the number of lags (p) are assumed to be large, a

reasonable rule of thumb is to take $N \geq 50$ and p to be approximately N/10.[25] When

these conditions cannot be satisfied, the only recourse is to rely upon the proven

robustness of time series forecasts derived from the models discussed in this

paper. The statistical measure of goodness-of-fit that should be used in deciding

upon the order (p,d,q) of the process is given by the estimate of the variance

of the time series $\{\varepsilon(t), t \varepsilon T\}$ in (49). This measure is a logical one to utilize

in the context of time series forecasting, since the variance of a one-step ahead

forecast will be equal to the variance of the time series $\{\varepsilon(t), t \varepsilon T\}$. We would

prefer to choose, as a model of the time series of interest, that particular model

---

[25]Cogger, K. O., _op. cit._

(and order) which minimizes the residual variance given by (49). Moreover, this measure of goodness-of-fit can be shown to be extremely sensitive to misspecification of the correct value of d. It is less sensitive to misspecification of p and q but, fortunately, it can be shown that this latter type of misspecification does not usually entail very large increases in the variance of a forecast.[26]

---

[26]Ibid.

OTHER WORKING PAPERS

Working Paper
Number

1   "Reflections on Evolving Competitive Aspects in Major Industries, " by Sidney C. Sufrin

2   "A Scoring System to Aid in Evaluating a Large Number of Proposed Public Works Projects by a Federal Agency, " by M. Lynn Spruill

3   "The Delphi Method: A Systems Approach to the Utilization of Experts in Technological and Environmental Forecasting, " by John D. Ludlow

4   "What Consumers of Fashion Want to Know, " by Claude R. Martin, Jr. --Out of print. To be published in a forthcoming issue of the Journal of Retailing.

5   "Performance Issues of the Automobile Industry, " by Sidney C. Sufrin, H. Paul Root, Roger L. Wright, and Fred R. Kaen

6   "Management Experience with Applications of Multidimensional Scaling Methods, " by James R. Taylor

7   "Profitability and Industry Concentration, " by Daryl Winn

8   "Why Differences in Buying Time? A Multivariate Approach, " by Joseph W. Newman and Richard Staelin--Out of print. To be published in a forthcoming issue of the Journal of Marketing Research.

9   "The Contribution of the Professional Buyer to the Success or Failure of a Store, " by Claude R. Martin

10   "An Empirical Comparison of Similarity and Preference Judgments in a Unidimensional Context, " by James R. Taylor