TRANSACTIONS OF SOCIETY OF ACTUARIES
1973 VOL. 25 PT. 1 NO. 73

TIME SERIES ANALYSIS AND FORECASTING

ROBERT B. MILLER AND JAMES C. HICKMAN*

If you do not think about the future, you cannot have one.

—GALSWORTHY

ABSTRACT

Actuaries study the past experience of claims, expense, investment, and other economic processes developing over time in order to build models for insurance systems that will operate in the future. This paper outlines a method of parametric modeling that is applicable to a broad class of time series. The resulting models may be used to produce forecasts and forecast intervals that will convey a measure of the reliance that may be placed in the forecasts.

I. INTRODUCTION

FORECASTING is at the heart of actuarial science. Actuaries fix insurance prices not to match past costs but to balance a stream of future expected costs. Reserves reflect expectations about the insufficiency of future income to balance future claim payments.

In formulating these expectations, some kind of model for the future is required. This model can be built only on knowledge of the past. Consequently, actuaries study past claims, expense, and investment experience as a preliminary step to building a model of the future. In this analysis it is highly convenient to view each value of a particular process of interest as being a realization from an unchanging random generator of values. The actuary's task is then to use the assumed independent observed values of the process to estimate the distribution from which the stream of future values, which will determine the experience of the insurance system, will be selected. Many of the existing stock of statistical methods are directed toward estimating a distribution from a set of independent observations determined by the distribution.

Yet the concept of a single unchanged distribution generating independent insurance cost values is obviously not universally adequate. An example of distributions shifting over time is provided by annuity

* Mr. Miller, not a member of the Society, is associate professor of statistics and business, University of Wisconsin, Madison.

mortality. The experimental evidence of this century is very convincing that annuity mortality is not generated by an unchanging process. An alternative to the classical stationary model is the time-trend model brought into the mainstream of North American actuarial science by Jenkins and Lew [8]. They adopted the model ${}^{1950+k}q_x = {}^{1950}q_x(1 - s_x)^k$, where ${}^{1950+k}q_x$ is the mortality probability that will be expected to be experienced in calendar year $1950 + k$ at age x, and s_x is an annual improvement rate. Although we cannot push the analogy very far, this time-trend model, in its basic concept, has elements in common with the autoregressive time series model to be discussed later.

In their role as technical insurance managers, actuaries are involved with short-term as well as long-term forecasts. Cash flows, security prices, policy loans, sales, claim numbers, and claim amounts are examples of processes developing over time for which short-term forecasts may be very useful.

The purpose of this paper is to explain a technical method for analyzing the past experience of a process that is developing over time for the purpose of forecasting. Almost equally important, the method will provide indexes of the reliance that may be placed on the forecast and will also provide a warning when one is yielding to the temptation to read more structure into past experience than is really there. The justification for this review is that these methods are not covered in the Society's course of study on statistics, and forecasters in diverse fields have found them useful.

Since the basic ideas that are the building blocks for these methods are already covered in the Society's course of reading on statistics and numerical analysis, is it reasonable to ask why these methods were not made routine long ago? The answer is that these methods require extensive computation in their application and this was impractical before large-scale electronic computers became available. In addition, the reader will note that these methods involve ideas from several traditional fields in applied mathematics. Methods intersecting several existing fields seem to develop more slowly because there are relatively few generalists who can effectively integrate ideas from apparently independent areas.

Even before starting our technical exposition, it seems wise to make tentative responses to certain fundamental questions that will naturally arise. The methods reviewed in this paper are only one of several families of technical methods for analyzing numerical data for the purpose of forecasting [4, 7]. Some of the others, such as multiple regression, have already been used in actuarial science [10, 12]. The methods to be developed here will involve the analysis of a single variable time series. Analysis involving past values of several series often lead to excellent forecasts, but such models require more intricate mathematics. In addition, the values of the associated series may not in practice be available in time to permit forecasting. Frequently, the past values of the single variable time series under study capture the essence of the impact of associated series [9, 11].

II. INTUITIVE DISCUSSION

A common assumption made in many theoretical discussions of statistics is that one's observations are generated independently. However, particularly in the nonexperimental sciences, research workers are painfully aware of the fact that their observations are often necessarily correlated. Unfortunately, many statistical methods designed for use on independent observations are not effective when used on correlated observations. Hence it behooves practical research workers to include in their bags of statistical tools methods which are designed to handle such observations. Such tools fall under the rather vague title of "time series analysis" in the parlance of statisticians. Our purpose in this paper is to provide an introduction to time series methods for actuaries. We will deal mainly with the approach developed by Box and Jenkins [1]. Our justification is that this approach is fairly easy to understand and to implement and that this approach is being used successfully by a number of research workers faced with business and economic forecasting problems.

In this section we will look at several examples of time series data and some of the complications that arise when they are analyzed. These examples set the stage for our discussion of methods for handling such data.

Our first example is the hypothetical series plotted in Figure 1. The abscissa is a "time" axis, and time is assumed to be measured at equally spaced, discrete points. The ordinate is the observation axis. An observation at time t is denoted by Z_t. Even though Figure 1 displays only 100 observations, the points are connected by straight lines to aid the visual examination of the series. Casual observation reveals that the observations appear to vary about a fixed mean level (the line denoted by μ in Fig. 1) and within fixed bands about that level (the lines marked u and l in Fig. 1). Note further that the observations tend to fall above and below their mean level in an alternating pattern and that the pattern is quite consistent throughout the series. This suggests that the observations may be correlated. To check this, we constructed Figures 2 and 3. Figure 2 is a scatter diagram of the points (Z_{t-1}, Z_t) for $t = 2, 3, \ldots, 100$, while Figure 3 is a scatter diagram of the points (Z_{t-2}, Z_t) for $t =$

3, 4, . . . , 100. These figures indicate negative and positive correlation between observations one and two time periods apart, respectively. Observations separated by a specified number of time periods are said to be *lagged*. Thus Z_{t-1} and Z_t are one lag apart, Z_{t-2} and Z_t are two lags apart, and so forth. If lagged observations in a time series are correlated, then the series is said to be *autocorrelated*. A numerical measure of autocorrelation in an observed series at any lag $k = \ldots -2, -1, 0, 1, 2, \ldots$, is given by the sample autocorrelation coefficient

$$r_k = \frac{\sum_{t=k+1}^{n} (Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^{n} (Z_t - \bar{Z})^2}, \tag{1}$$

where n is the number of observations and

$$\bar{Z} = \frac{1}{n} \sum_{t=1}^{n} Z_t .$$

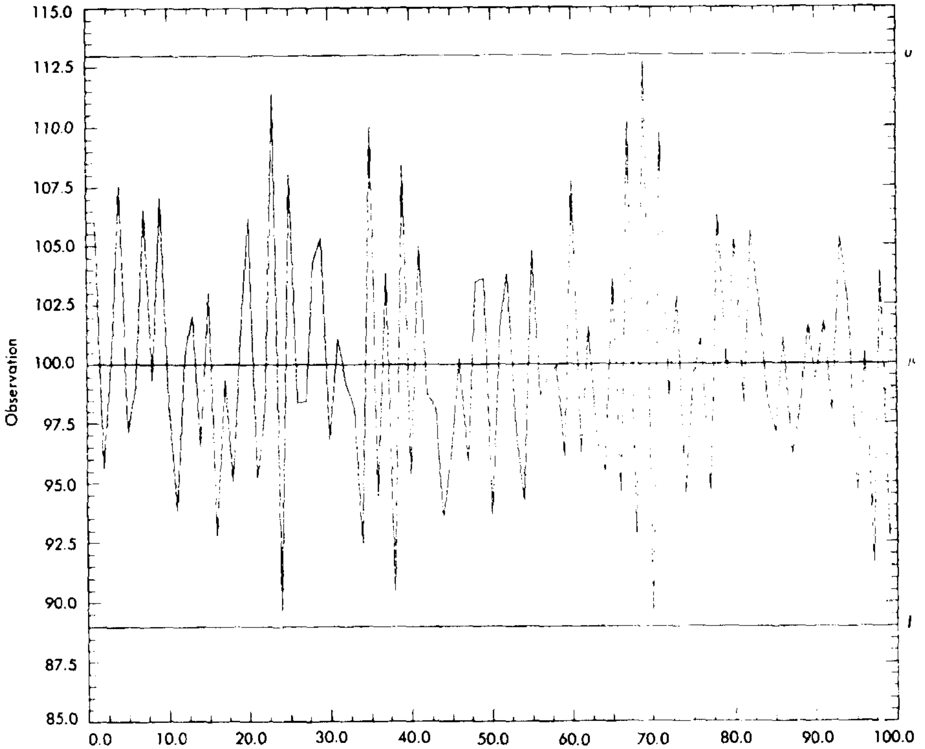

FIG. 1.—A discrete time series

For our hypothetical series in Figure 1, $\bar{Z} = 99.92$, and

$$r_0 = 1 ,$$

$$r_1 = \frac{\sum_{t=2}^{100}(Z_t - \bar{Z})(Z_{t-1} - \bar{Z})}{\sum_{t=1}^{100}(Z_t - \bar{Z})^2} = -0.55 ,$$

$$r_2 = \frac{\sum_{t=3}^{100}(Z_t - \bar{Z})(Z_{t-2} - \bar{Z})}{\sum_{t=1}^{100}(Z_t - \bar{Z})^2} = 0.30 ,$$

$$r_3 = -0.18 , \quad r_4 = 0.14 , \quad r_5 = -0.14 ,$$

$$r_6 = 0.03 , \quad r_7 = 0.00 .$$

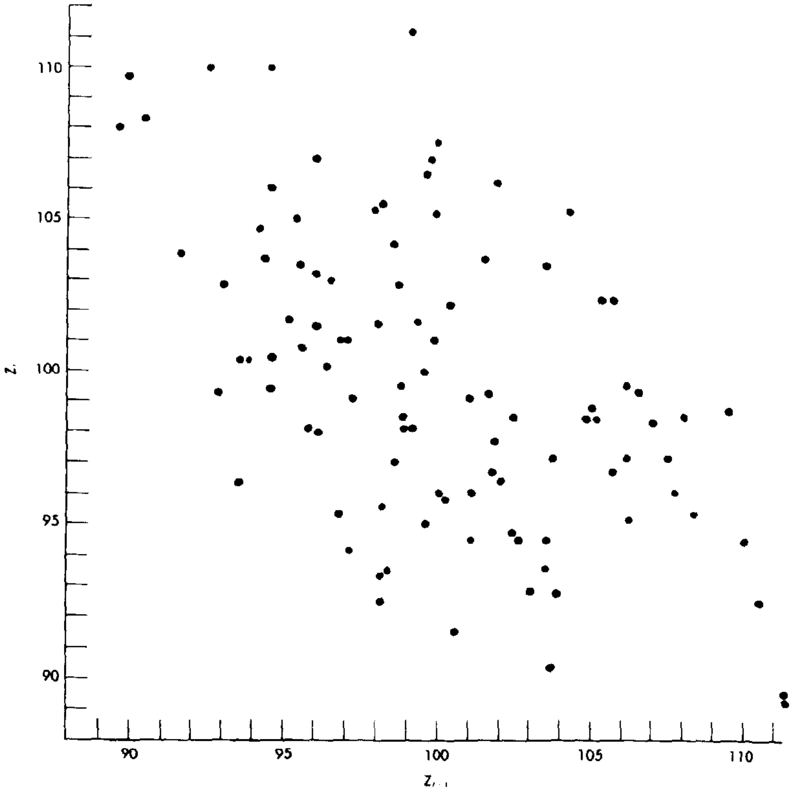

FIG. 2.—Scatter diagram of (Z_{t-1}, Z_t) for series in Fig. 1

(We need only consider $k \geq 0$ because $r_{-k} = r_k$.) The sample autocorrelation coefficients move rapidly toward zero, indicating that autocorrelations after lag 2 or lag 3 are not very significant compared to those of shorter lags.

The series in Figure 1 evidences properties of stationarity. The process generating a series is said to be stationary (in the wide sense) if the following facts are true:

a) $\mathcal{E}(Z_t)$ and Var $(Z_t) = \mathcal{E}[Z_t - \mathcal{E}(Z_t)]^2$ are constant functions of the index t.

b) Cov $(Z_t, Z_{t+k}) = \mathcal{E}(Z_t Z_{t+k}) - [\mathcal{E}(Z_t)]^2$ is some function of the lag k alone. (Note that $\mathcal{E}$ denotes mathematical expectation.)

A stricter form of stationarity is the requirement that the joint distributions of any two sets of random variables $\{Z_{t_1}, Z_{t_2}, \ldots, Z_{t_s}\}$ and $\{Z_{t_1+\tau}, Z_{t_2+\tau}, \ldots, Z_{t_s+\tau}\}$ be the same. For many practical applications wide-sense stationarity is all that need be assumed, although it should be pointed out that

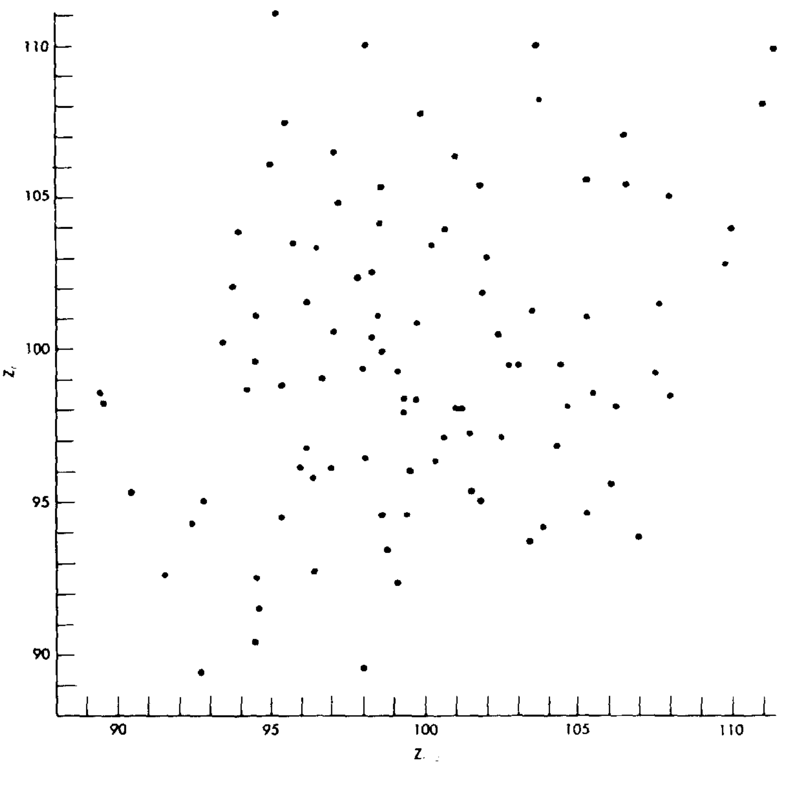

FIG. 3.—Scatter diagram of (Z_{t-2}, Z_t) for series in Fig. 1

the imposition of joint normal distributions on the variables implies the stricter form of stationarity. The series in Figure 1 appears to be generated by a stationary process because it varies about a fixed mean level, and it displays a consistent pattern of autocorrelation and variability throughout its length. A further useful characteristic of many stationary processes is that their autocorrelation functions (the theoretical analogue of the sample autocorrelation coefficients) decay quickly as k increases [1, p. 174].

Many series in practice are not generated by a stationary process. Often series fail to vary about a fixed mean level. Instead they "meander" about or exhibit strong upward or downward trends. Our next two example series exhibit these sorts of nonstationarity. They were selected not only because they illustrate practical and important models but also because they are series of direct concern to actuaries.

Figure 4 displays the series of nominal interest rates, expressed

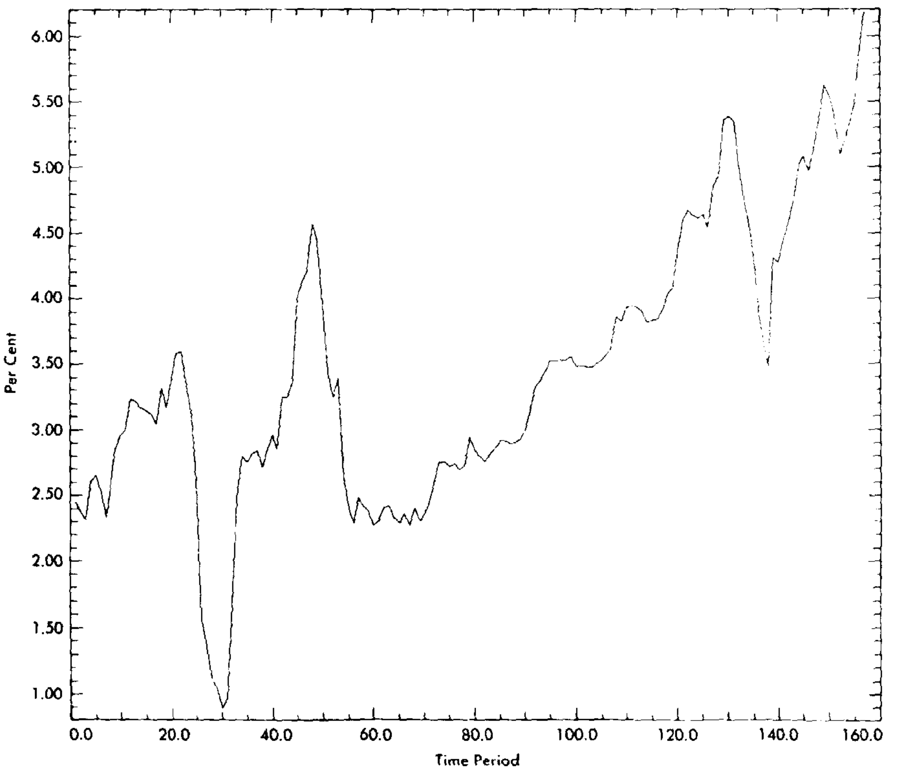

FIG. 4.—Nominal annual interest rates on three-month Treasury bills, 1956–69

as a per cent, on three-month United States Treasury bills for the period January, 1956—January, 1969. This is a meandering series which does not display consistent behavior about its mean value $\bar{Z} = 3.439$. Hence we cannot expect the autocorrelation coefficients to provide information about the autocorrelation which may exist in this series. Panel 1 of Table 1 shows the autocorrelation coefficients for lags 1–24.

TABLE 1

SAMPLE AUTOCORRELATION COEFFICIENTS FOR TREASURY BILL SERIES

Panel 1: Original Series		Panel 2: First Difference	
Lag	Coefficient	Lag	Coefficient
1	.95	1	.39
2	.89	2	.20
3	.82	3	.07
4	.76	4	.02
5	.69	5	−.10
6	.63	6	−.23
7	.58	7	−.30
8	.53	8	−.16
9	.50	9	−.07
10	.47	10	.01
11	.45	11	−.01
12	.43	12	.08
13	.40	13	.04
14	.37	14	−.00
15	.35	15	−.00
16	.33	16	−.10
17	.31	17	−.15
18	.31	18	−.18
19	.31	19	−.15
20	.34	20	−.08
21	.36	21	−.10
22	.39	22	−.16
23	.42	23	−.01
24	.44	24	.15

They do not decay toward zero very quickly, indicating the lack of stationarity that was obvious by visual inspection of the series as well.

In view of this difficulty it is natural to seek a transformation of the original interest rate series that is stationary and to try to analyze this transformed series. A reasonable quantity to look at here is the *change* in interest rates from month to month. In finite-difference notation familiar to actuaries, we want to look at the first backward difference $\nabla Z_t = (1 - E^{-1})Z_t = Z_t - Z_{t-1}$. Figure 5 is a plot of the ∇Z_t series, and the second panel of Table 1 displays some autocorrelation coefficients calcu-

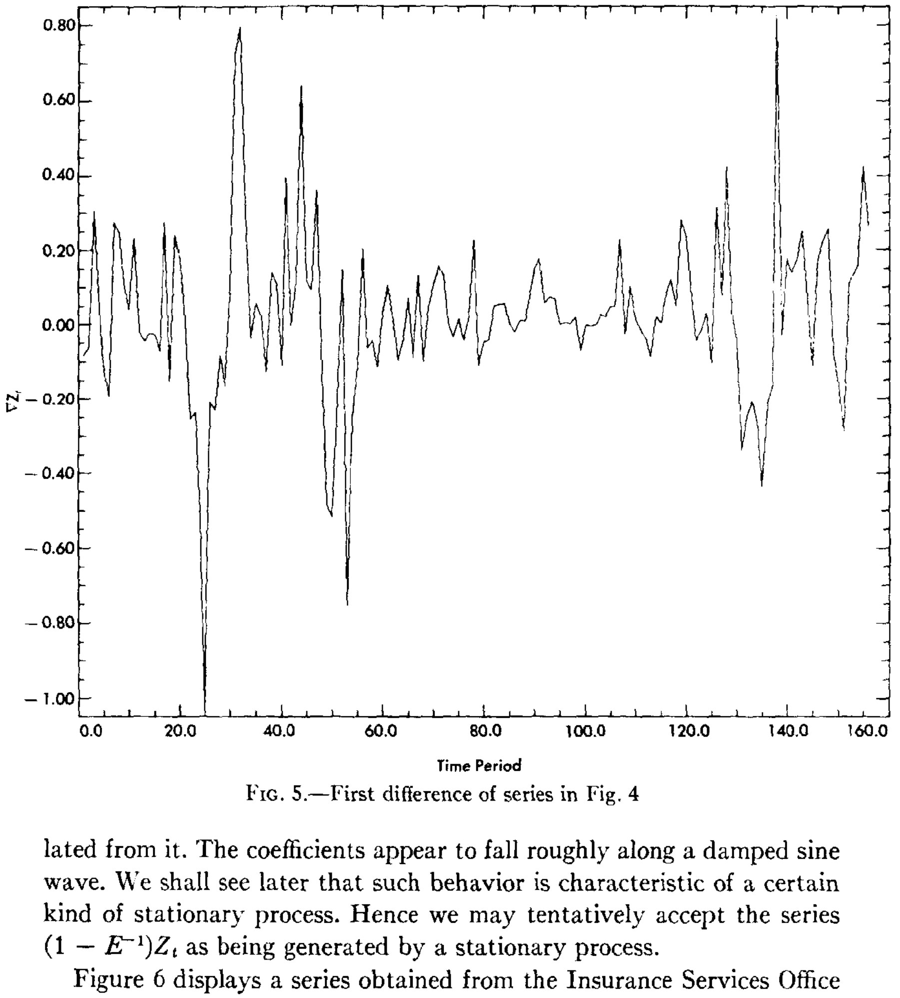

FIG. 5.—First difference of series in Fig. 4

lated from it. The coefficients appear to fall roughly along a damped sine wave. We shall see later that such behavior is characteristic of a certain kind of stationary process. Hence we may tentatively accept the series $(1 - E^{-1})Z_t$ as being generated by a stationary process.

Figure 6 displays a series obtained from the Insurance Services Office in New York. The observations are year-end quarterly indexes of automobile property damage paid claim costs, 1954–70. The series has a decided upward trend that appears somewhat quadratic. This nonstationary behavior is reflected in the autocorrelation coefficients displayed in panel 1 of Table 2. Again we must look for a transformation to reduce the series to stationarity. We might try the first backward difference again, but panel 2 of Table 2 shows that this series is also nonstationary. Such a result should not be too surprising given the quadratic-like trend exhibited in Figure 6, but we should expect the second backward differ-

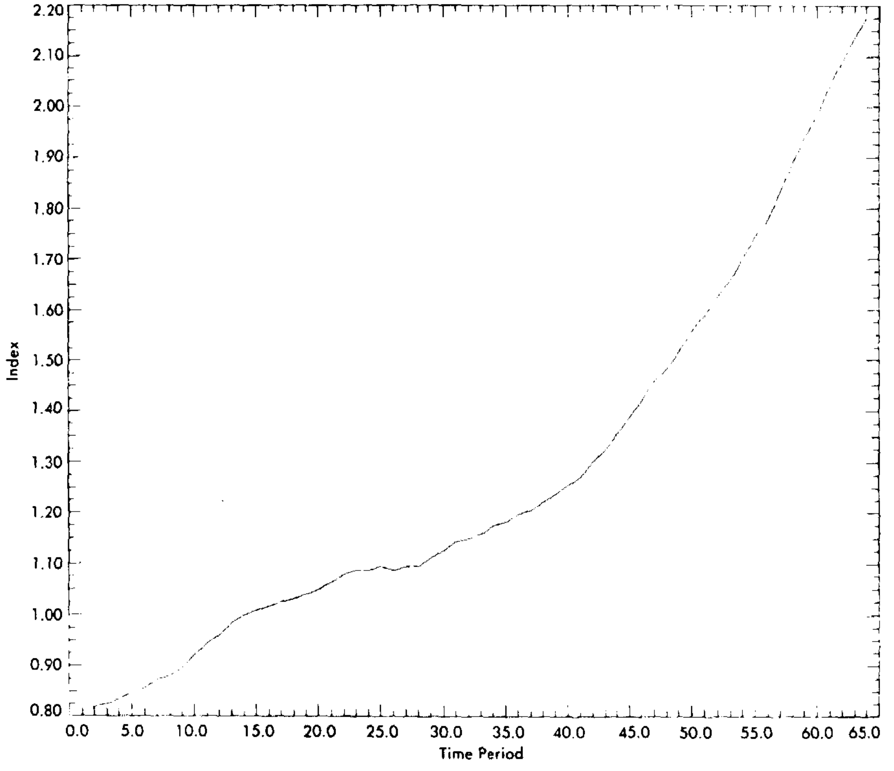

FIG. 6.—Quarterly indexes of automobile property damage paid claim costs, 1954–70

TABLE 2

AUTOCORRELATIONS FOR PROPERTY DAMAGE CLAIM COST SERIES

Panel 1: Original Series		Panel 2: First Difference		Panel 3: Second Difference	
Lag	Coefficient	Lag	Coefficient	Lag	Coefficient
1	.94	1	.84	1	−.54
2	.88	2	.82	2	.15
3	.81	3	.76	3	.02
4	.75	4	.68	4	−.05
5	.69	5	.63	5	.02
6	.64	6	.57	6	−.06
7	.58	7	.51	7	−.13
8	.53	8	.50	8	.24
9	.48	9	.45	9	−.05
10	.43	10	.40	10	−.15
11	.39	11	.38	11	.19
12	.34	12	.33	12	−.09

ence to appear stationary. Panel 3 of Table 2 is consistent with this expectation. The autocorrelation coefficients of $(1 - E^{-1})^2 Z_t$ die out very quickly.

Our final example illustrates a different sort of nonstationarity that is sometimes encountered in practice—nonstationarity of variance. Figure 7 shows daily stock market rates of return, approximately from 1963 to 1971. These data are based on Standard and Poor's index and were made available to us through the courtesy of the New York Life Insurance Company. Note that while the series appears to vary about a fixed mean level of zero, the pattern of variability changes with time. In technical terms, it appears as if occasionally the variance of the distribution shifts.

The ancillary benefit of the examples of this section is that they illustrate the insights that may be gained from simple visual displays of the original data and of transformations of the data.

III. LINEAR TIME SERIES MODELS

In this section we shall introduce a class of mathematical models that is often used to graduate time series data. The distinctive feature of these models is that they explicitly allow for correlation among the observations. The primary purpose of fitting models to time series data is to allow the researcher to forecast future observations in the series and to obtain a measure of the potential error in his forecasts.

We shall motivate the time series models by reference to the familiar linear multiple regression model

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + e_i \,, \qquad (2)$$

$i = 1, 2, \ldots, n$. Here y is considered the dependent variable whose value depends on the values of the independent variables $X_1, \ldots, X_p$ plus some observational error e. The β coefficients are unknown parameters, and the errors $e_1, e_2, \ldots, e_n$ are assumed to be uncorrelated, with means equal to 0 and variances equal to the common unknown value σ^2. Lewis has used a multiple regression model to calculate (forecast) expected profit as a function of withdrawal rates and amount of insurance [10].

The time series models we wish to consider are similar to equation (2), except that the "independent" variables are to be taken as *past observations* in the series. This is how autocorrelation is to be introduced. In order to distinguish the time series models from the model in equation (2), we shall introduce some new notation. The correspondences between the items in the time series models and equation (2) are given in the tabulation on page 279. Thus the time series models may be written in

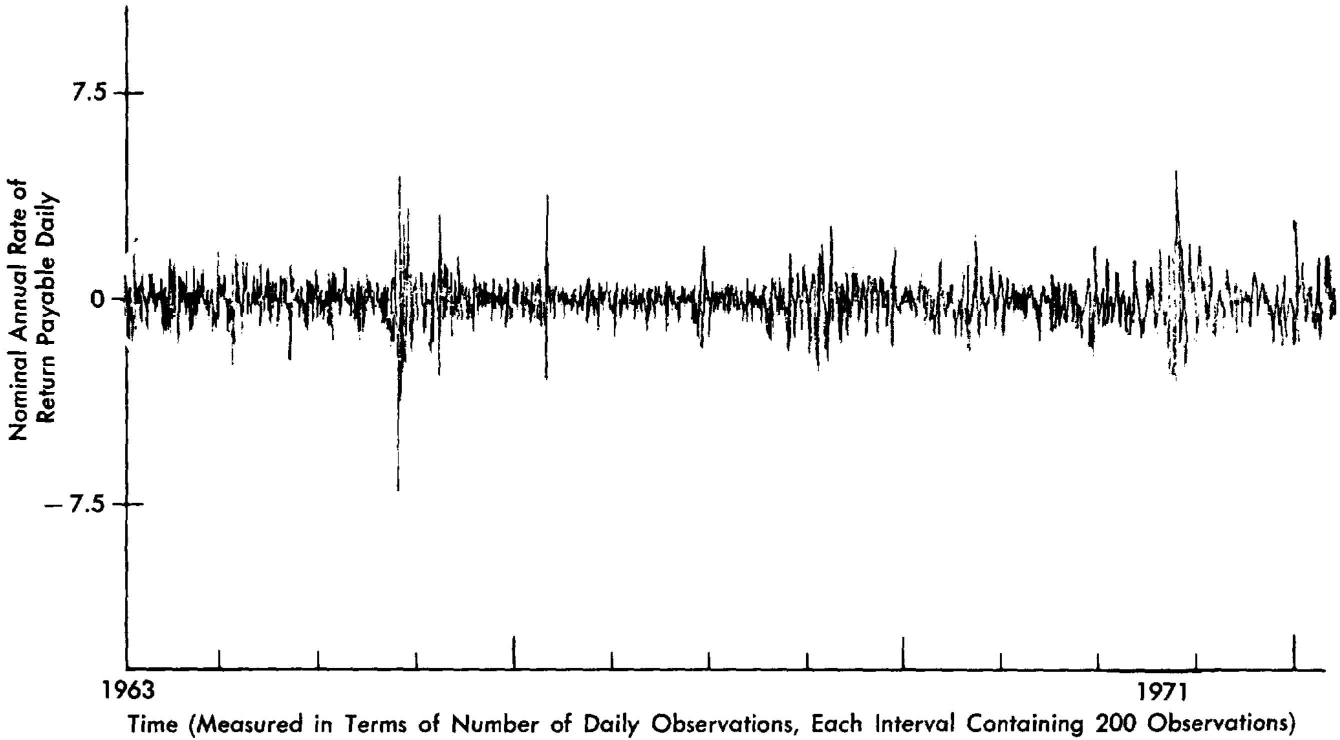

FIG. 7.—Daily stock market rates of return

Item	Regression Model	Time Series Models
Index	i	t
"Dependent" variable	y_i	$\dot{Z}_t$
"Independent" variable	X_{ki}	$\dot{Z}_{t-k}$
Coefficients	β_k	Π_k
Error	e_i	a_t

the general form

$$\dot{Z}_t = \Pi_1\dot{Z}_{t-1} + \Pi_2\dot{Z}_{t-2} + \ldots + \Pi_k\dot{Z}_{t-k} + \ldots + a_t , \tag{3}$$

where we assume that the a_t's are uncorrelated, with common mean 0 and common variance σ^2. We also assume that any a_t is uncorrelated with past observations $\dot{Z}_{t-1}, \dot{Z}_{t-2}, \ldots$. The dot over the Z's denotes the fact that if the Z's are stationary with fixed mean μ, then we consider $\dot{Z} = Z - \mu$ in equation (3). If the Z's are nonstationary, there may be no fixed mean level μ, so that $\dot{Z} = Z$ and we need not use the dot. Note that equation (3) allows us to enter an infinite number of past Z's into the model. In other words, theoretically speaking, the present observation may be a function of the entire past history of the series. We will now consider some specific examples of equation (3).

Example 1.—Suppose that $\Pi_1 = \phi$, $\Pi_k = 0$, $k \geq 2$, $-1 < \phi < 1$. Then

$$\dot{Z}_t = \phi\dot{Z}_{t-1} + a_t . \tag{4}$$

This is called the *autoregressive model* of order 1—abbreviated AR(1)—because it looks like a simple linear regression of the present observation on the previous observation of the series. The model can be interpreted as saying that the current observation $\dot{Z}_t$ is a fraction ϕ of the previous observation plus a random "shock" a_t.

The assumption $-1 < \phi < 1$ ensures that equation (4) is a stationary model. To prove that this assumption will ensure stationarity is a rather technical matter. However, some insight may be gained from the following development. Using successive substitution, we obtain

$$\begin{aligned} \dot{Z}_t &= \phi\dot{Z}_{t-1} + a_t = \phi(\phi\dot{Z}_{t-2} + a_{t-1}) + a_t \\ &= \phi^2\dot{Z}_{t-2} + \phi a_{t-1} + a_t \\ &= \phi^3\dot{Z}_{t-3} + \phi^2 a_{t-2} + \phi a_{t-1} + a_t \\ &\;\;\vdots \\ &= \phi^k\dot{Z}_{t-k} + \sum_{j=0}^{k-1}\phi^j a_{t-j} . \end{aligned} \tag{4'}$$

This shows that the dependence of $\dot{Z}_t$ on past $\dot{Z}_{t-k}$'s *and* a_{t-k}'s decreases, since $\phi^k \to 0$ as $k \to \infty$. Hence the observations are being drawn toward an equilibrium position. In fact, if the shocks, a_t, into the model were somehow "turned off" at some point in time, the observations, $\dot{Z}_t$, would eventually satisfy the equation $\dot{Z}_t = \phi^k \dot{Z}_{t-k}$, which tends to the equation $\dot{Z}_t = 0$, or $Z_t = \mu$, the equilibrium value.

Another way to express stationarity is to define the theoretical autocorrelation function

$$\rho_k = \frac{\mathcal{E}(Z_t - \mu)(Z_{t+k} - \mu)}{\mathcal{E}(Z_t - \mu)^2} = \frac{\mathcal{E}(\dot{Z}_t \dot{Z}_{t+k})}{\mathcal{E}(\dot{Z}_t^2)}, \tag{5}$$

$k = \ldots -2, -1, 0, 1, 2, \ldots$ [5]. This function is the theoretical analogue of the sample function defined in equation (1). The assumption of stationarity implies that the denominator in equation (5), the variance of the series, is a constant value, γ_0, say, while the numerator, the covariance, is a function of k alone, γ_k, say. To compute γ_k for the AR(1) model in equation (4), note that

$$\begin{aligned} \mathcal{E}(\dot{Z}_t \dot{Z}_{t+k}) &= \mathcal{E}[(\phi^k \dot{Z}_{t-k} + \phi^{k-1} a_{t-k+1} + \ldots + \phi a_{t-1} + a_t)\dot{Z}_{t-k}] \\ &= \phi^k \mathcal{E}(\dot{Z}^2_{t-k}) . \end{aligned} \tag{6}$$

Equation (6) follows because $a_t, a_{t-1}, \ldots, a_{t-k+1}$ are uncorrelated with $\dot{Z}_{t-k}$, that is, $\mathcal{E}(a_{t-j}\dot{Z}_{t-k}) = 0$ for $j = 0, 1, \ldots, k-1$. But equation (6) may be written

$$\gamma_k = \phi^k \gamma_0 ,$$

so that

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k , \quad k = \ldots -2, -1, 0, 1, 2, \ldots . \tag{7}$$

Thus it is clear that the theoretical autocorrelations tend to zero as $|k|$ becomes large. If we had taken $|\phi| \geq 1$, we can see from equation (4′) that the effect of past observations and shocks would have grown rather than diminished. In this case the model would not be stationary, as the observations would not have been drawn toward an equilibrium value. We could not have defined ρ_k as we did in equation (5), because the variance of the model depends on t when $|\phi| \geq 1$.

An important example of the nonstationary case occurs when $\phi = 1$, so that the model is

$$Z_t = Z_{t-1} + a_t . \tag{8}$$

This is a random walk model which has appeared in many models of stock price change. The present observation is the previous observation plus a random shock. If in equation (7) we let ϕ approach unity, we see that $\rho_k \to 1$, which suggests that if the sample autocorrelations were calculated for data generated by a random walk, they would tend to stay close to unity rather than die out. We have seen in our example that this sort of behavior is characteristic of other nonstationary series as well.

If we assume in equation (8) that $Z_t = 0$, $t \leq 0$, then

$$Z_1 = a_1 ,$$

$$Z_2 = Z_1 + a_2 = a_1 + a_2 ,$$

$$Z_t = Z_{t-1} + a_t = a_1 + a_2 + \ldots + a_{t-1} + a_t ,$$

so that $\mathcal{E}(Z_t) = 0$ and Var $(Z_t) = t\sigma^2$, $t > 0$. This result shows the dependence of the variance of Z_t on t and hence the absence of statistical equilibrium.

Example 2.—The autoregressive model of order p, abbreviated AR(p), is simply the model in equation (3) with $\Pi_p \neq 0$, and $\Pi_k = 0$, $k \geq p + 1$. The autoregressive model of infinite order is given by equation (3) with the condition that there is no $p' > 0$ such that $\Pi_k = 0$ for all $k \geq p'$. For these models to be stationary, certain restrictions must be placed on the coefficients. For example, the AR(2) model with $\Pi_1 = \phi_1$, $\Pi_2 = \phi_2$, $\Pi_k = 0$, $k \geq 3$, is stationary if and only if the following conditions are satisfied:

$$\phi_2 + \phi_1 < 1 , \quad \phi_2 - \phi_1 < 1 , \quad -1 < \phi_2 < 1$$

(see Box and Jenkins [1, Fig. 3.2, p. 59]). Figure 8 shows the various possible patterns that AR(2) autocorrelations can take. In general, for stationary AR models the autocorrelations decay toward zero as k increases. Notice in particular the pattern in Figure 8, *b*, which is reminiscent of a damped sine wave.

Now consider the stationary infinite autoregressive model with $\Pi_k = -\theta^k$, $-1 < \theta < 1$, that is,

$$\dot{Z}_t = -\theta \dot{Z}_{t-1} - \theta^2 \dot{Z}_{t-2} - \ldots + a_t , \tag{9}$$

or

$$a_t = \dot{Z}_t + \sum_{k=1}^{\infty} \theta^k \dot{Z}_{t-k} . \tag{9'}$$

Multiply equation (9′), with t replaced by $t - 1$, by θ, and subtract the resulting equation from equation (9′) to obtain

$$\dot{Z}_t = a_t - \theta a_{t-1} . \tag{10}$$

Thus an observation from the infinite autoregressive model in equation (9) can be thought of as the difference between two components: (*a*) the current shock and (*b*) a fraction θ of the previous shock. The model in

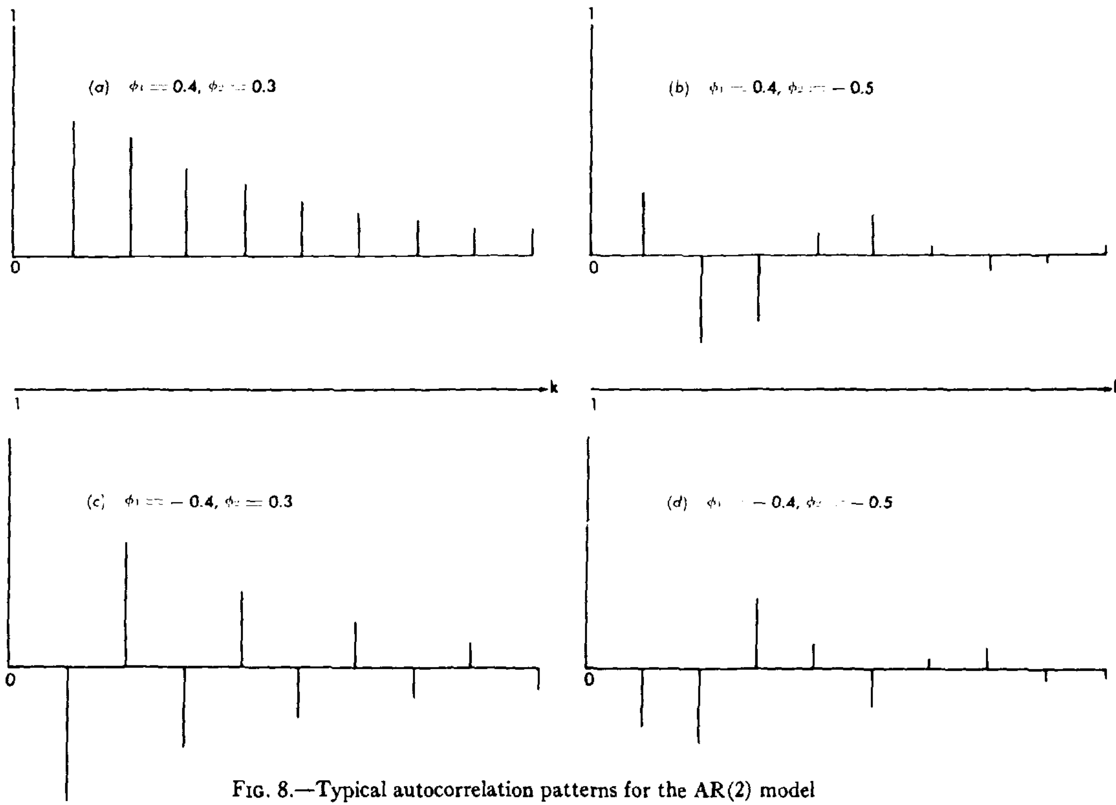

FIG. 8.—Typical autocorrelation patterns for the AR(2) model

equation (10) is called a *moving average* model of order 1—abbreviated MA(1). The advantage of the simple representation in equation (10) becomes clear when one contemplates calculating the theoretical autocorrelations of the model in equation (9). This appears to be a formidable task because of the infinite number of past observations present on the right-hand side. However, from equation (10) we obtain the simple calculations

$$\mathcal{E}(\dot{Z}_t^2) = \mathcal{E}(a_t - \theta a_{t-1})^2 = \mathcal{E}(a_t^2 - 2\theta a_t a_{t-1} + \theta^2 a_{t-1}^2) = (1 + \theta^2)\sigma^2 ,$$

$$\mathcal{E}(\dot{Z}_t\dot{Z}_{t-1}) = \mathcal{E}[(a_t - \theta a_{t-1})(a_{t-1} - \theta a_{t-2})] = -\theta\sigma^2 ,$$

$$\mathcal{E}(\dot{Z}_t\dot{Z}_{t-k}) = \mathcal{E}[(a_t - \theta a_{t-1})(a_{t-k} - \theta a_{t-k-1})] = 0 , \quad k \geq 2 .$$

Hence

$$\rho_0 = 1 , \quad \rho_1 = \frac{-\theta}{1 + \theta^2} , \quad \rho_k = 0, \quad k \geq 2 .$$

In other words, the MA(1) model implies nonzero autocorrelation only between observations one lag apart. This model has been discussed in the context of estimating future claims [6].

It is clear from this development that the MA(1) model is stationary no matter what the value of θ. However our restriction of θ to the interval $(-1, 1)$ was necessary to make this model fit into the general model in equation (3). To see this, consider equation (10) written in finite-difference notation:

$$\dot{Z}_t = (1 - \theta E^{-1})a_t .$$

Now if $|\theta| < 1$, we may invert the difference operator to obtain

$$\begin{aligned} a_t &= (1 - \theta E^{-1})^{-1}\dot{Z}_t \\ &= (1 + \theta E^{-1} + \theta^2 E^{-2} + \ldots)\dot{Z}_t \\ &= \dot{Z}_t + \sum_{k=1}^{\infty}\theta^k \dot{Z}_{t-k} , \end{aligned}$$

which is equation (9′). In fact the condition $|\theta| < 1$ represents no essential restriction, because, if we have an MA(1) model with $|\theta| > 1$, there is a unique MA(1) model with $|\theta| < 1$ which generates the same time series. This is because, if $\rho_1 = -\theta_1/(1 + \theta_1^2)$, then $\theta_2 = 1/\theta_1$, $|\theta_2| < 1$, produces the same ρ_1. Hence we may always consider MA(1) models to be invertible and therefore contained in the model in equation (3).

We may define moving average models of finite order q—abbreviated MA(q)—as follows:

$$\dot{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} . \tag{11}$$

We shall always assume that these models are invertible in the sense that $(1 - \theta_1 E^{-1} - \ldots - \theta_q E^{-q})^{-1}$ provides a convergent infinite expansion. That is, the roots of the polynomial equation of degree q, $P_q(E^{-1}) = (1 - \theta E^{-1} \ldots \theta_q E^{-q}) = 0$, are outside the unit circle in the complex plane. It should be clear from equation (11) that the general MA(q) model exhibits nonzero autocorrelation at most at the first q lags, but $\rho_k = 0$ for $k \geq q + 1$.

Example 3.—We may define a hybrid and very rich class of models which is a mixture of the autoregressive and moving average models, denoted by ARMA(p, q) and given by

$$\dot{Z}_t - \phi_1 \dot{Z}_{t-1} - \ldots - \phi_p \dot{Z}_{t-p} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} . \quad (12)$$

This is a flexible class of models which often allows one to fit rather complicated time series with a model containing a very small number of parameters, say p and q less than or equal to 2. This is a very advantageous situation from a statistical point of view because then one loses only a small number of degrees of freedom in estimating the parameters. See Box and Jenkins [1, Fig. 3.11, p. 78] for patterns of theoretical autocorrelation coefficients for the ARMA(1, 1) model.

Example 4.—Suppose that $\Pi_k = (1 - \theta)\theta^{k-1}$, $-1 < \theta < 1$, $k = 1, 2, \ldots$. Then

$$\dot{Z}_t = (1 - \theta)(\dot{Z}_{t-1} + \theta \dot{Z}_{t-2} + \theta^2 \dot{Z}_{t-3} + \ldots) + a_t , \quad (13)$$

and

$$\theta \dot{Z}_{t-1} = (1 - \theta)(\theta \dot{Z}_{t-2} + \theta^2 \dot{Z}_{t-3} + \ldots) + \theta a_{t-1} . \quad (14)$$

Subtracting equation (14) from equation (13) yields

$$\dot{Z}_t - \theta \dot{Z}_{t-1} = (1 - \theta)\dot{Z}_{t-1} + a_t - \theta a_{t-1} ,$$

or

$$Z_t - Z_{t-1} = a_t - \theta a_{t-1} . \quad (15)$$

Thus the model in equation (13) implies that the first difference of the Z series is stationary and follows an MA(1) model. This, of course, implies that the Z series itself is nonstationary with no fixed mean level. Hence we have not put dots on top of the Z's in equation (15).

Interestingly enough, equation (13) can be looked upon as a model for simple exponential smoothing, a popular tool of many forecasters [3]. This can be seen as follows. We adopt a rather conventional approach related to the idea of using next year's expected claims, given all past claims, as an estimate of next year's net or pure premium. That is, we define the smoothed value of Z_t to be

$$\begin{aligned} S_t(\theta) &= \mathcal{E}(Z_t \mid Z_{t-1}, Z_{t-2}, \ldots) \\ &= \mathcal{E}[(1 - \theta)(Z_{t-1} + \theta Z_{t-2} + \theta^2 Z_{t-3} + \ldots) + a_t \mid Z_{t-1}, Z_{t-2}, \ldots] \\ &= (1 - \theta)(Z_{t-1} + \theta Z_{t-2} + \theta^2 Z_{t-3} + \ldots) . \end{aligned}$$

In other words, the smoothed value of Z_t is its conditional expected value at time $t - 1$. Now

$$S_t(\theta) = (1 - \theta)Z_{t-1} + \theta(1 - \theta)(Z_{t-2} + \theta Z_{t-3} + \dots)$$
$$= (1 - \theta)Z_{t-1} + \theta S_{t-1}(\theta) ,$$

so that the smoothed value of Z_t is a weighted average of the observation at time $t - 1$ and its smoothed value. But this is just the familiar formula for exponential smoothing with smoothing constant $\alpha = 1 - \theta$.

A simple operational method of obtaining equation (13) from equation (15) is to write equation (15) in difference notation,

$$(1 - E^{-1})Z_t = (1 - \theta E^{-1})a_t ,$$

and invert the operator on the right-hand side to obtain

$$(1 - \theta E^{-1})^{-1}(1 - E^{-1})Z_t = a_t ,$$

or

$$[1 - (1 - \theta)(E^{-1} + \theta E^{-2} + \theta^2 E^{-3} + \dots)]Z_t = a_t$$

or

$$Z_t = (1 - \theta)(Z_{t-1} + \theta Z_{t-2} + \dots) + a_t ,$$

which is equation (13). This method can be used to show formally that any model of the form

$$(1 - \phi_1 E^{-1} - \dots - \phi_p E^{-p})(1 - E^{-1})^d \dot{Z}_t = (1 - \theta_1 E^{-1} - \dots - \theta_q E^{-q})a_t \quad (16)$$

is also of the form of equation (3), provided that the operator on the right-hand side of equation (16) is invertible. The model in equation (16) is called an *integrated autoregressive moving average* model, abbreviated ARIMA(p, d, q), where p, d, and q are nonnegative integers. The symbol d is called the order of differencing, and

$$\dot{Z}_t = \begin{cases} Z_t - \mu & \text{if} \quad d = 0 \\ Z_t & \text{if} \quad d > 0 . \end{cases}$$

The model for $W_t = (1 - E^{-1})^d Z_t$ will be stationary as long as suitable restrictions are placed on $\phi_1, \phi_2, \dots, \phi_p$.

A slight extension of the model in equation (16) is the addition of a trend term on the right-hand side, namely,

$$(1 - \phi_1 E^{-1} - \dots - \phi_p E^{-p})(1 - E^{-1})^d \dot{Z}_t = \theta_0 + (1 - \theta_1 E^{-1} - \dots - \theta_q E^{-q})a_t . \quad (17)$$

Such an extension occasionally proves useful in practical problems where differencing does not eliminate a deterministic trend.

Our review of time series models has culminated in the classes of models displayed in equations (16) and (17). These classes contain all our previous examples as special cases. We shall see in Section V that such models can often be fitted successfully to data that arise in practice. Consequently, they represent a potential basis for making forecasts. The question of forecasting is considered in the next section.

IV. FORECASTING

Once a satisfactory model has been obtained, it can be used to generate forecasts of future observations. "Forecasts" must be interpreted as values we would expect to observe if future observations followed the model we are using. But we can hardly hope that our mathematical model reflects all the forces influencing the empirical phenomenon of interest. Furthermore, we are using "expect" in the technical, statistical sense of expected value. Thus, because we are dealing with stochastic models, we could not predict future observations with certainty even if our model were exact. Consequently, our forecasts must be accompanied by a statement of the error inherent in them. In the examples below, forecasts will take the form of a "best guess," an upper limit, and a lower limit. These limits will reflect the effect of random error, but they cannot take into account the possibility of massive changes in the structure underlying the actual observations.

In summary, we conceive of the forecasting problem in the following way:

a) A portion of the past history of a series, $\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n$, is known.

b) Forecasts of future values of the series, $\dot{Z}_{n+1}, \dot{Z}_{n+2}, \ldots$, are needed.

c) In addition, a measure of the potential error in the forecasts is required.

If it can be assumed that the data came from a model of the form of equation (16), then it is possible to use this assumption to produce the required forecasts. The usual criterion for choosing forecasts is the minimization of the mean-square forecast error. Let $\hat{\dot{Z}}_n(l)$ denote the forecast of $\dot{Z}_{n+l}$ when $\dot{Z}_1, \ldots, \dot{Z}_n$ are known. We shall choose $\hat{\dot{Z}}_n(l)$ so that

$$\mathcal{E}_n[\dot{Z}_{n+l} - \hat{\dot{Z}}_n(l)]^2 = \min_a \mathcal{E}_n(\dot{Z}_{n+l} - a)^2 , \tag{18}$$

where $\mathcal{E}_n$ stands for conditional expectation, given $\dot{Z}_1, \ldots, \dot{Z}_n$. It is well known that the solution of this minimization problem is to choose $\hat{\dot{Z}}_n(l) = \mathcal{E}_n(\dot{Z}_{n+l})$, the conditional expected value of $\dot{Z}_{n+l}$. It follows that

$$\mathcal{E}_n[\dot{Z}_{n+l} - \hat{\dot{Z}}_n(l)]^2 \tag{19}$$

is the conditional variance of $\dot{Z}_{n+l}$, and this quantity is a natural measure of the degree of potential forecast error. Thus we must develop expressions for the two quantities $\mathcal{E}_n(\dot{Z}_{n+l})$ and $\mathcal{E}_n[\dot{Z}_{n+l} - \mathcal{E}_n(\dot{Z}_{n+l})]^2$. We shall begin with an example.

Example 5.—Suppose that our model is the AR(1) model

$$\dot{Z}_t - \phi\dot{Z}_{t-1} = a_t\,, \quad -1 < \phi < 1\,.$$

Then

$$\begin{aligned} \dot{Z}_{n+l} &= \phi\dot{Z}_{n+l-1} + a_t \\ &= \phi^k\dot{Z}_{n+l-k} + \sum_{j=0}^{k-1}\phi^j a_{n+l-j}\,, \end{aligned} \tag{20}$$

using equation (4′). Setting $k = l$ in equation (20), we obtain

$$\dot{Z}_{n+l} = \phi^l\dot{Z}_n + \sum_{j=0}^{l-1}\phi^j a_{n+l-j}\,. \tag{21}$$

Finally,

$$\hat{\dot{Z}}_n(l) = \mathcal{E}_n(\dot{Z}_{n+l}) = \phi^l\dot{Z}_n + \sum_{j=0}^{l-1}\phi^j\mathcal{E}_n(a_{n+l-j}) = \phi^l\dot{Z}_n\,, \tag{22}$$

because $\mathcal{E}_n(a_{n+l-j}) = 0$ for $j < l$. Furthermore,

$$\begin{aligned} \mathcal{E}_n[\dot{Z}_{n+l} - \hat{\dot{Z}}_n(l)]^2 &= \mathcal{E}_n\Bigl(\sum_{j=0}^{l-1}\phi^j a_{n+l-j}\Bigr)^2 \\ &= \Bigl(\sum_{j=0}^{l-1}\phi^{2j}\Bigr)\sigma^2 \\ &= \frac{1-\phi^{2l}}{1-\phi^2}\,\sigma^2\,. \end{aligned} \tag{23}$$

Thus a reasonable forecast interval for $\dot{Z}_{n+l}$ would be

$$\phi^l\dot{Z}_n \pm L\Bigl[\sqrt{\Bigl(\frac{1-\phi^{2l}}{1-\phi^2}\Bigr)}\Bigr]\sigma\,, \tag{24}$$

where L is a number which determines the number of standard errors one wishes to use. A common value of L is 2 because this provides approximate 95 per cent forecast intervals, assuming that the a_t's are normally distributed. Note that, as l increases, the forecast interval in equation (24) approaches

$$0 \pm \frac{L}{\sqrt{(1-\phi^2)}}\,\sigma\,. \tag{25}$$

Equations (21) and (23) are a special case of an interesting fact about the ARIMA class of models which is generally true, namely, that

$$\dot{Z}_{n+l} - \hat{\dot{Z}}_n(l) = \sum_{j=0}^{l-1}\psi_j a_{n+l-j}\,, \tag{26}$$

where $\psi_0 = 1$ and the other ψ_j's are functions of the parameters of the model. Furthermore, the ψ_j's can be determined by equating the coefficients of powers of E^{-1} in the equation

$$(1 + \psi_1 E^{-1} + \psi_2 E^{-2} + \ldots)(1 - \phi_1 E^{-1} - \ldots - \phi_p E^{-p})(1 - E^{-1})^d = (1 - \theta_1 E^{-1} - \ldots - \theta_q E^{-q}) . \tag{27}$$

The proof of this fact can be made by induction (see Box and Jenkins [1, pp. 114–19]). It follows from equation (26) that

$$\mathcal{E}_n[\dot{Z}_{n+l} - \hat{Z}_n(l)]^2 = \sigma^2 \sum_{j=0}^{l-1} \psi_j^2 , \tag{28}$$

and this provides a convenient method for calculating our measures of potential forecast error. The following example illustrates the procedure.

Example 6.—Suppose that the model is

$$Z_t - 2Z_{t-1} + Z_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} , \tag{29}$$

that is, the second difference of the Z_t's follows an MA(2) model. We have

$$\begin{aligned}
Z_{n+1} - 2Z_n + Z_{n-1} &= a_{n+1} - \theta_1 a_n - \theta_2 a_{n-1} , \\
Z_{n+2} - 2Z_{n+1} + Z_n &= a_{n+2} - \theta_1 a_{n+1} - \theta_2 a_n , \\
&\;\;\vdots \\
Z_{n+l} - 2Z_{n+l-1} + Z_{n+l-2} &= a_{n+l} - \theta_1 a_{n+l-1} - \theta_2 a_{n+l-2} , \\
&\;\;\vdots
\end{aligned}$$

From these equations it is clear that

$$\hat{Z}_n(1) - 2Z_n + Z_{n-1} = -\theta_1 a_n - \theta_2 a_{n-1} , \tag{30a}$$

$$\hat{Z}_n(2) - 2\hat{Z}_n(1) + Z_n = -\theta_2 a_n , \tag{30b}$$

$$\hat{Z}_n(l) - 2\hat{Z}_n(l-1) + \hat{Z}_n(l-2) = 0 \quad \text{for} \quad l \geq 3 . \tag{30c}$$

Equation (30c) is a homogeneous difference equation in l with two initial conditions (30a) and (30b) which specify the values of $\hat{Z}_n(1)$ and $\hat{Z}_n(2)$ in terms of the quantities θ_1, θ_2, Z_n, Z_{n-1}, a_n, and a_{n-1}. The general solution of the difference equation is

$$\hat{Z}_n(l) = A_0 + A_1 l , \tag{31}$$

where A_0 and A_1 are determined from the initial condition by the equations

$$\hat{Z}_n(1) = A_0 + A_1 , \quad \hat{Z}_n(2) = A_0 + 2A_1 . \tag{32}$$

In order to use equation (28) to obtain our measure of forecast error, we must determine the ψ_j's from the equation

$$(1 + \psi_1 E^{-1} + \psi_2 E^{-2} + \ldots)(1 - E^{-1})^2 = (1 - \theta_1 E^{-1} - \theta_2 E^{-2}) ,$$

or

$$1 + (\psi_1 - 2)E^{-1} + (\psi_2 - 2\psi_1 + 1)E^{-2} + (\psi_3 - 2\psi_2 + \psi_1)E^{-3} + \ldots$$
$$= 1 - \theta_1 E^{-1} - \theta_2 E^{-2} .$$

Equating coefficients, we have

$$\psi_1 - 2 = -\theta_1 ,$$
$$\psi_2 - 2\psi_1 + 1 = -\theta_2 ,$$
$$\psi_j - 2\psi_{j-1} + \psi_{j-2} = 0 , \quad j \geq 3 .$$

Solving this linear difference equation with initial conditions, we obtain

$$\psi_j = (1 + \theta_2) + (1 - \theta_1 - \theta_2)j , \quad j \geq 1 . \tag{33}$$

Thus a forecast interval for Z_{n+l} is

$$A_0 + A_1 l \pm L\sqrt{\left(\sigma^2 \sum_{j=0}^{l-1} \psi_j^2\right)} ,$$

where A_0 and A_1 are given by equation (32) and the ψ_j's are given by equation (33). It is evident that the forecasts lie on a line with slope A_1 and intercept A_0, while the standard errors are *increasing* functions of l. Thus the forecast intervals become wider as l increases. This behavior is quite different from that in equations (24) and (25). The reason is that the model in example 5 is stationary, while the model in the present example is nonstationary. (By observing eq. [29], you will note that the second difference of the Z_t's is a stationary MA(2) model, while the Z_t's are nonstationary.)

V. MODEL-BUILDING

We noted in Section IV that if one could assume that his time series came from a model in the ARIMA class, then he could obtain forecasts of future observations and a measure of the potential errors in his forecasts. But the question surely arises: How can one justify assuming an ARIMA model? In this section we will present a technique for determining an ARIMA model which best describes a given set of time series data. In effect, we are seeking to discover whether our data behaved *as if* they were generated by such a model. Our statistical procedure does not attempt to say why a series fits a particular model well, although it may be possible on intuitive grounds to explain certain types of behavior discovered during the modeling process.

We assume that we have in hand n observations from a time series $\dot{Z}_1, \dot{Z}_2, \ldots, \dot{Z}_n$. We wish to choose a model from the ARIMA class which well describes our data in the following ways:

a) If $\hat{\dot{Z}}_1, \hat{\dot{Z}}_2, \ldots, \hat{\dot{Z}}_n$ are the observations generated by the model, then the residuals $\dot{Z}_t - \hat{\dot{Z}}_t = \hat{a}_t$ should evidence no significant autocorrelation.

b) The model contains as few parameters as possible consistent with requirement *a*. This is a modern-day restatement of the dictum of Ockham ("Ockham's razor") that "entities should not be multiplied without necessity." The reason for requirement *b* is that we want our model-building procedure to be as efficient as possible. By adding a large number of parameters to the model, we can make the residuals essentially zero, but the data would contain so little information about these parameters that we could not estimate them with any degree of confidence. Our strategy is to use a small number of parameters, saving a relatively large number of degrees of freedom for estimating residual autocorrelation and residual variance. As soon as residual autocorrelation becomes negligible, we can be satisfied that our model is using the information contained in the fact that the observations are autocorrelated.

The strategy for finding a suitable model proceeds as follows. Because the ARIMA class of models is quite large, we first try to eliminate unrealistic candidates from the class before going to more detailed analysis. This can be done most simply by calculating the autocorrelation coefficients for the data and trying to match them to theoretical autocorrelation coefficients of ARIMA models. This procedure is illustrated in the following example.

Example 7.—Consider the data from the first example in Section II, displayed in Figure 1. The first ten sample autocorrelation coefficients are given below:

k	1	2	3	4	5	6	7	8	9	10
r_k	−.55	.30	−.18	.14	−.14	.03	.00	−.10	.06	−.01

The first ten autocorrelation coefficients of an AR(1) model with $\phi = -0.6$ are as follows:

k	1	2	3	4	5	6	7	8	9	10
ρ_k	−.6	.36	−.216	.129	−.078	.047	−.028	.017	−.010	.006

The correspondence is quite close, so that it would be worthwhile to consider an AR(1) model for the data of this example. We will discuss this model further in a moment.

Next let us consider the Treasury bill data in the second example of Section II. These data are pictured in Figure 4. There we found that the first-difference series appeared stationary. Furthermore, the sample autocorrelations appeared to follow a damped sine wave pattern of decay. We noted in Section III that the AR(2) model can produce such a pattern of theoretical autocorrelations. Hence it would be sensible to consider the model

$$(1 - \phi_1 E^{-1} - \phi_2 E^{-2})(1 - E^{-1})Z_t = a_t$$

for the Treasury bill data.

Finally, let us look at the claim index data in the third example of Section II. There we found that the second difference appeared stationary, and it is evident from panel 3 of Table 2 that at most two sample autocorrelations from the second-difference series are significant. This suggests an MA(2) model for the second difference, so that we might entertain the model

$$(1 - E^{-1})^2 Z_t = (1 - \theta_1 E^{-1} - \theta_2 E^{-2})a_t$$

for the claim index data of this example.

We have now tentatively chosen models for the data in the three examples of Section II. We next make a detailed study of the residuals from fitting these models to the data. We will use the method of least squares to fit the models. The procedure will be illustrated in the next two examples.

Example 8.—We will fit the AR(1) model

$$Z_t - \mu = \phi(Z_{t-1} - \mu) + a_t$$

to the data in the first example. Define

$$a_t(\mu, \phi) = (Z_t - \mu) - \phi(Z_{t-1} - \mu) .$$

We wish to choose μ^* and ϕ^* so that

$$\sum_{t=1}^{n} a_t^2(\mu^*, \phi^*) = \min_{\mu,\phi} \sum_{t=1}^{n} a_t^2(\mu, \phi) .$$

Unfortunately, $a_1(\mu, \phi)$ is a function of the unobserved quantity Z_0, so we set $a_1(\mu, \phi)$ equal to its expected value, 0, and choose $\hat{\mu}$ and $\hat{\phi}$ so that

$$\sum_{t=2}^{n} a_t^2(\hat{\mu}, \hat{\phi}) = \min_{\mu,\phi} \sum_{t=2}^{n} a_t^2(\mu, \phi) .$$

If n is reasonably large, $\hat{\mu}$ and $\hat{\phi}$ will be very close to μ^* and ϕ^*, so that there is very little loss of efficiency in setting $a_1(\mu, \phi) = 0$. For specified values of μ and ϕ the quantities

$$
\begin{aligned}
a_2(\mu, \phi) &= (Z_2 - \mu) - \phi(Z_1 - \mu) , \\
a_3(\mu, \phi) &= (Z_3 - \mu) - \phi(Z_2 - \mu) , \\
&\;\;\vdots \\
a_n(\mu, \phi) &= (Z_n - \mu) - \phi(Z_{n-1} - \mu) ,
\end{aligned}
$$

and hence the quantity

$$\sum_{t=2}^{n} a_t^2(\mu, \phi)$$

can be calculated. Figure 9 shows a contour map of this sum-of-squares function and indicates that $\hat{\mu} = 99.9$ and $\hat{\phi} = -0.56$. The values of $\hat{\mu}$ and $\hat{\phi}$ are most efficiently found by an iterative nonlinear least-squares routine on a computer [1, p. 495].[1]

The residuals $a_t(\hat{\mu}, \hat{\phi})$, $t = 2, \ldots, n$ may now be checked for autocorrelation by examining the autocorrelation coefficients

$$r_k(a) = \frac{\sum_{t=2+k}^{n} [a_t(\hat{\mu}, \hat{\phi}) - \bar{a}(\hat{\mu}, \hat{\phi})][a_{t-k}(\hat{\mu}, \hat{\phi}) - \bar{a}(\hat{\mu}, \hat{\phi})]}{\sum_{t=2}^{n} [a_t(\hat{\mu}, \hat{\phi}) - \bar{a}(\hat{\mu}, \hat{\phi})]^2} ,$$

$k = 1, 2, \ldots$, where

$$\bar{a}(\hat{\mu}, \hat{\phi}) = \frac{1}{n-1} \sum_{t=2}^{n} a_t(\hat{\mu}, \hat{\phi}) .$$

Naturally, if the autocorrelation in the residuals is negligible, the $r_k(a)$'s should be quite small. If, on the other hand, the $r_k(a)$'s display a distinct pattern, then one must modify the original model to account for the autocorrelation still present.

The first ten residual autocorrelations for the data above are given in the accompanying tabulation. A rough measure of their significance is

k	1	2	3	4	5	6	7	8	9	10
$r_k(a)$	.00	.01	.01	−.03	−.15	−.03	−.06	−.15	.01	.08

obtained by comparing them with their approximate standard error, assuming that the residuals are uncorrelated. This standard error is simply

[1] A discussion of the numerical methods used to solve this type of nonlinear least-squares problem would direct the exposition from its main goal. However, within the Society of Actuaries' current Syllabus of Examinations, these methods are outlined under the heading of Newton's iterative method of solving systems of nonlinear equations, p. 312 of *Theory and Problems of Numerical Analysis* by F. Scheid.

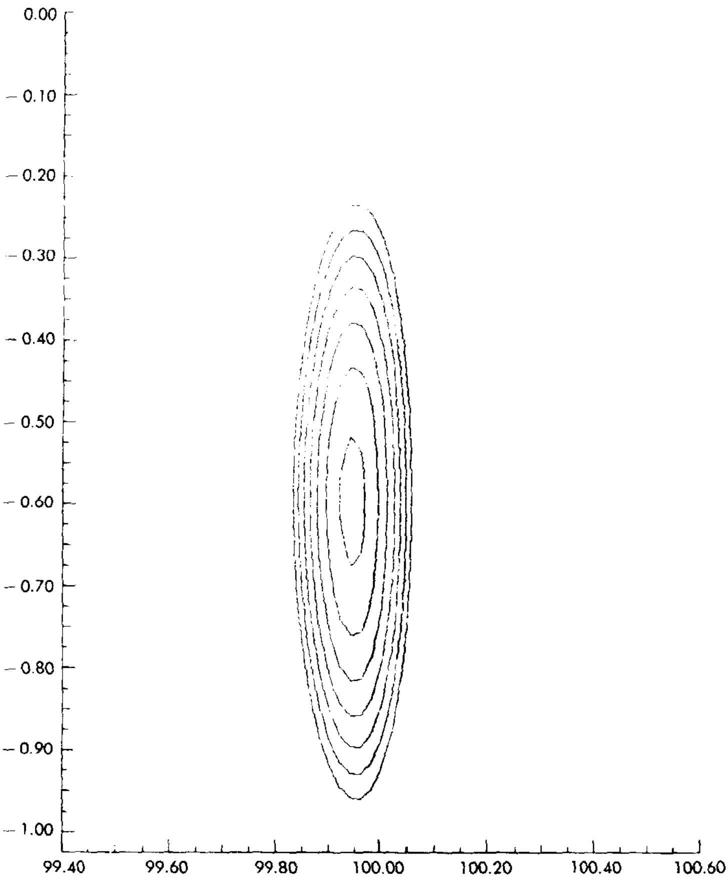

FIG. 9.—Contours of the sum-of-squares function for the series in Fig. 1

$1/\sqrt{(n-1)} = 1/\sqrt{99} \doteqdot 0.1$. It is seen that none of the coefficients is greater than twice its approximate standard error. An over-all test of the significance of the coefficients is given by comparing the quantity

$$99\sum_{k=1}^{10} r_k^2(a) = 5.6529$$

with the upper 100α per cent critical point in a chi-square distribution with $10 - 2 = 8$ degrees of freedom. (See Box and Jenkins [1, p. 291] or Box and Pierce [2].) The general rule is to compare

$$m\sum_{k=1}^{M} r_k^2(a) \tag{34}$$

with a chi-square value with degrees of freedom equal to M minus the number of parameters appearing in the sum-of-squares function. M is the number of autocorrelation coefficients considered, and m is the number of residuals computed after the least-squares estimation of the parameters of the ARIMA model. This result is valid if the residuals are assumed to be normally distributed. For our example, with $\alpha = 0.05$, we could compare 5.6529 with 15.507, which is the upper 5 per cent critical value of a chi-square distribution with 8 degrees of freedom. Clearly we need not be dissatisfied with the fit of our model here.

The unbiased least-squares estimate of σ^2 is given by

$$\frac{1}{97}\sum_{t=2}^{100} a_t^2(\hat{\mu}, \hat{\phi}) = 18.025 .$$

In general, the estimate is obtained by dividing the minimum sum-of-squares function by the number of residuals less the number of parameters estimated. An estimate of σ is provided by the square root of the variance estimate, which in our example is $\sqrt{18.025} \doteq 4.25$.

Example 9.—We will fit the model

$$Z_t - 2Z_{t-1} + Z_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

to the claim index data in the third example of Section II. Define

$$a_t(\theta_1, \theta_2) = Z_t - 2Z_{t-1} + Z_{t-2} + \theta_1 a_{t-1}(\theta_1, \theta_2) + \theta_2 a_{t-2}(\theta_1, \theta_2) ,$$

and note that $a_1(\theta_1, \theta_2)$ and $a_2(\theta_1, \theta_2)$ cannot be calculated because they depend on the unknown quantities Z_0 and Z_{-1}. Setting $a_1(\theta_1, \theta_2) = a_2(\theta_1, \theta_2) = 0$, however, it is possible to calculate

$$\sum_{t=3}^{n} a_t^2(\theta_1, \theta_2)$$

for given values of θ_1 and θ_2. We take as our estimates of θ_1 and θ_2 the values $\hat{\theta}_1 = 0.62$ and $\hat{\theta}_2 = -0.28$ that minimize this sum-of-squares function.

Contours of the sum-of-squares surface are shown in Figure 10. The estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{60}\sum_{t=3}^{64} a_t^2(\hat{\theta}_1, \hat{\theta}_2) = 0.49 \times 10^{-4} .$$

The estimate of σ is 0.007.

The first ten residual autocorrelation coefficients are given in the accompanying tabulation. Their estimated standard error, assuming that

k	1	2	3	4	5	6	7	8	9	10
r_k	−.03	−.05	−.02	−.06	−.12	−.21	−.06	.30	.01	−.11

the residuals are uncorrelated, is approximately $1/\sqrt{62} = 0.127$. This may be used to individually test the hypotheses that the autocorrelations are zero. The value of

$$62\sum_{k=1}^{10} r_k^2(a)$$

is 10.6454, and by locating the upper 5 per cent critical value for a chi-square distribution with 8 degrees of freedom, we see that it is not significant at the 5 per cent level. Hence we may be satisfied with our model here. In reaching this judgment, we observe that r_8 is somewhat greater than twice the standard error, assuming that the residuals are

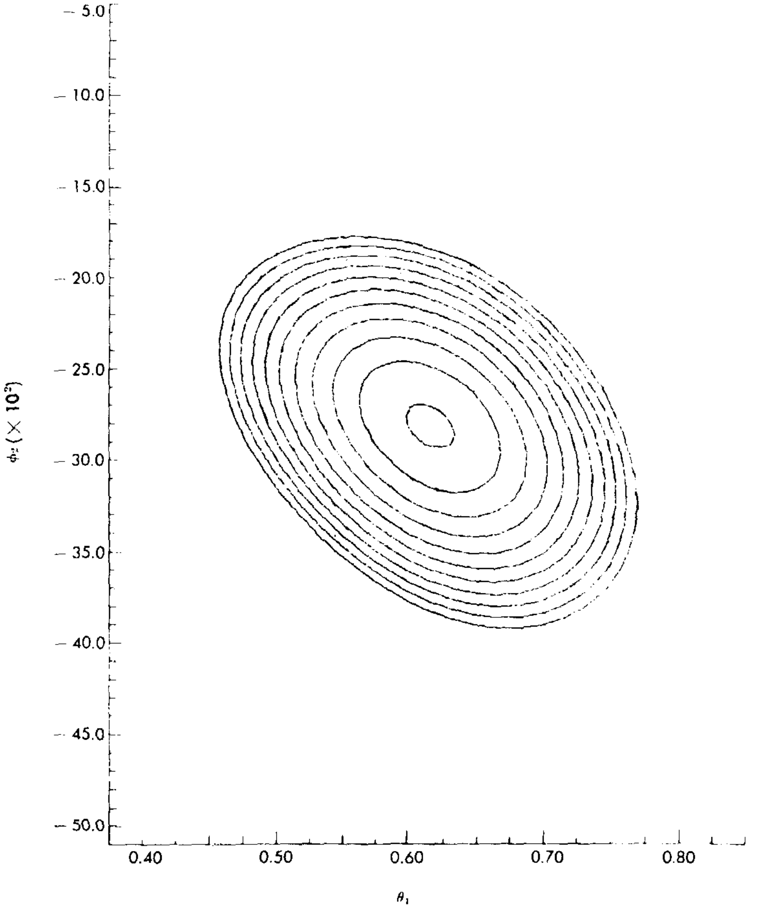

FIG. 10.—Contours of the sum-of-squares function for the MA(2) model

mutually independent. However, in making ten tests, this single deviation is not alarming, and in view of the acceptable value of the over-all test statistic, we decided to terminate model-building.

Example 10.—Without going into details, we note that the model

$$(1 - \phi_1 E^{-1} - \phi_2 E^{-2})(1 - E^{-1})Z_t = a_t$$

appears to be adequate for the Treasury bill data in the second example of Section II. We have $\hat{\phi}_1 = 0.38$, $\hat{\phi}_2 = 0.06$, and $\hat{\sigma} = 0.213$.

VI. APPLICATIONS TO FORECASTING

In this section we will apply the results of Sections IV and V to the forecasting of three of the series presented in Section II. Figure 11 displays the basic format of our results. There we have pictured the claim index data from the third example of Section II, two "future observations" not used in the modeling and held aside for subsequent forecast checking, (denoted by x in Fig. 11), the forecasts produced by our fitted model, and

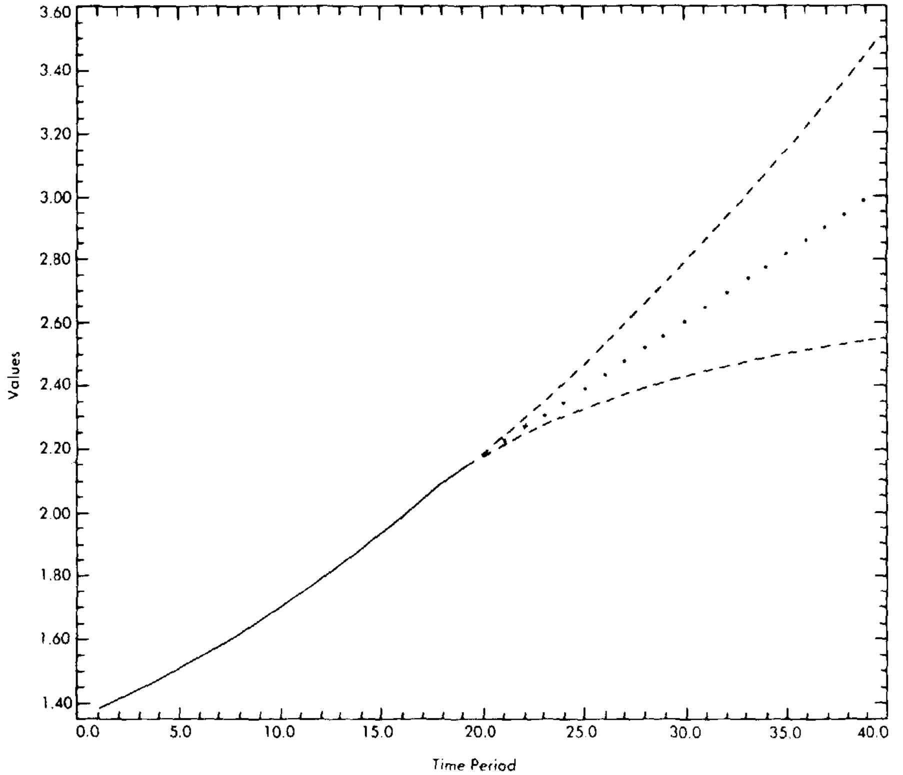

FIG. 11.—Twenty past observations, twenty forecasts with error bounds, and two "future" values of the series in Fig. 6.

approximate 95 per cent forecast error bounds. It is interesting to compare the forecasts with "future" values not used in the forecast. Table 3 shows comparisons made by means of (*a*) the root-mean-square forecast error and (*b*) Theil's coefficient [13]. The root-mean-square forecast error is defined by

$$\text{RMS} = \sqrt{\left\{\frac{1}{k}\sum_{l=1}^{k}[Z_{n+l} - \hat{Z}_n(l)]^2\right\}} ,$$

while Theil's coefficient is defined by

$$U = \sqrt{\left\{\frac{\sum_{l=1}^{k}[Z_{n+l} - \hat{Z}_n(l)]^2}{\sum_{l=1}^{k}(Z_{n+l} - Z_{n+l-1})^2}\right\}} ,$$

where k is the number of "future" values available.

TABLE 3

FORECAST INFORMATION FOR PROPERTY DAMAGE SERIES*

Forecast Lead Time l	Lower Forecast Boundary	Forecast	Upper Forecast Boundary	"Future" Values	ψ_l
1	2.208	2.222	2.236	2.225	1.380
2	2.242	2.265	2.289	2.272	2.043
3	2.271	2.308	2.345	....	2.706
4	2.299	2.351	2.403	....	3.368

* RMS = 0.059; U = 0.08.

Theil's coefficient is a comparison of the forecasts $\hat{Z}_n(l)$ with a forecasting procedure which forecasts the next period's observation with that of the present period. This forecast would be appropriate for a random walk model. A U value less than 1 indicates a rather efficient forecast procedure. The following example shows how the forecasts in Table 3 were obtained.

Example 11.—From equations (30) we see that the forecasts are given by the equations

$$\hat{Z}_{64}(1) = 2Z_{64} - Z_{63} - \theta_1 a_{64} - \theta_2 a_{63} ,$$

$$\hat{Z}_{64}(2) = 2\hat{Z}_{64}(1) - Z_{64} - \theta_2 a_{64} ,$$

$$\hat{Z}_{64}(l) = 2\hat{Z}_{64}(l-1) - \hat{Z}_{64}(l-2) , \quad l \geq 3 .$$

Thus, knowing Z_{64}, Z_{63}, θ_1, θ_2, a_{64}, and a_{63}, we can calculate the forecasts recursively. The estimation phase of model-building gives us estimates

$\hat{\theta}_1 = 0.62$, $\hat{\theta}_2 = -0.28$, $\hat{a}_{64} = -0.0083$, and $\hat{a}_{63} = -0.0031$, so that, using these estimates, we can actually generate forecasts. Thus

$$\begin{aligned}
\hat{Z}_{64}(1) &= 2(2.177) - 2.136 - (0.62)(-0.0083) - (-0.28)(-0.0031) \\
&= 2.222\,, \\
\hat{Z}_{64}(2) &= 2(2.222) - 2.177 - (-0.28)(-0.0083) \\
&= 2.265\,, \\
\hat{Z}_{64}(3) &= 2(2.265) - 2.222 = 2.308\,, \\
\hat{Z}_{64}(4) &= 2(2.308) - 2.265 = 2.351\,.
\end{aligned}$$

TABLE 4

FORECAST INFORMATION FOR AR(1) DATA*

Forecast Lead Time l	Lower Forecast Boundary	Forecast	Upper Forecast Boundary	"Future" Values	ψ_l
1	95.60	104.10	112.60	114.75	0.3136
2	87.96	97.70	107.44	92.84	0.0983
3	91.21	101.29	111.37	99.45	0.0308
4	89.04	99.28	109.52	92.58	0.0096
5	90.16	100.40	110.64	105.20	0.0030

* RMS = 6.48; U = 0.42.

Alternatively, we could use equations (31) and (32) as follows:

$$2.222 = A_0 + A_1\,, \qquad 2.265 = A_0 + 2A_1\,,$$

yielding $A_0 = 2.1793$ and $A_1 = 0.0429$. Thus

$$\hat{Z}_n(l) = 2.1793 + 0.0429l\,.$$

Using equation (33), we obtain $\psi_0 = 1$, $\psi_1 = 1.380$, $\psi_2 = 2.043$, $\psi_3 = 2.706$, and the approximate 95 per cent forecast intervals are

$$2.1793 + 0.0429l \pm 2\sqrt{\left[(0.000049)\sum_{j=0}^{l-1}\psi_j^2\right]}\,,$$

for $l = 1, 2, 3, 4$.

Calculations similar to these in example 11 produce forecasts for our other examples. Tables 4 and 5 summarize the results. The reader's attention is especially directed to the forecast boundaries in Table 5 which show a forecast range of almost one percentage point for the one-month-ahead forecast and of over three percentage points six months ahead. These figures indicate a high degree of uncertainty in forecasting short-term interest rates, a phenomenon familiar to many actuaries.

VII. CONCLUSIONS

Our purpose in this paper has been to illustrate the parametric modeling approach to time series analysis by presenting several simple examples. Guided by the reference list, which contains brief notes on the principal references, the interested reader may now construct an individualized tour through the expanding literature of this subject.

TABLE 5

FORECAST INFORMATION FOR TREASURY BILL DATA

Forecast Lead Time l	Lower Forecast Boundary	Forecast	Upper Forecast Boundary	"Future" Values*	ψ_l
1	5.83	6.30	6.72		1.3778
2	5.86	6.37	7.07		1.5839
3	5.43	6.39	7.37		1.6957
4	5.22	6.41	7.61		1.7372
5	5.03	6.42	7.82		1.7631
6	4.89	6.43	8.00		1.7761

* Not available.

The approach that we have illustrated involves the following steps:

1. Graph the data, searching for patterns that may reveal a shifting mean or a changing variance.
2. Obvious shifts in the mean may often be removed by differencing the data. Variance shifts are much harder to deal with and have not been discussed in this paper.
3. Tentatively identify a model for the series by examining the sample autocorrelation function, and compare it with the autocorrelation function of standard models. In this process keep in mind the necessity of economizing on parameters.
4. Estimate the parameters of the model using a nonlinear least-squares computer program.
5. Check the adequacy of the model by examining the residuals. If the model is adequate, the residuals should be approximately independently and identically distributed. Nothing is gained, and much foolishness may result, from going further with model-building after this step is finished.
6. Use the model to produce forecasts with associated approximate probability intervals.

The process that we have outlined is a dynamic process. That is, steps 3, 4, and 5 may be repeated several times before a satisfactory model is achieved. As new observations are obtained, the whole process may be repeated.

It would be unrealistic to assert that this recipe for time series analysis is universally applicable. We claim only that it is an orderly method for approaching many forecasting problems. It does not require extensive input or deep technical knowledge, and it may be easily carried out using modern computers. In many practical situations it will lead to greater insight into the process under study and to improved forecasts.

It might be useful to speculate on the likely role of parametric time series analysis in actuarial practice. There seems little doubt that the primary application will be in producing short-term forecasts. Such forecasts can serve to improve the management of cash flows arising from claims, investment, and sales operations. Experience in other fields seems to indicate that single variable time series models produce forecasts almost as accurate as more elaborate econometric models which incorporate many variables and fixed relationships among the variables. Time series models may also serve as components of comprehensive corporate models or management training games.

In this review we have stressed the forecasting role of time series models. The strategy has been to squeeze all the useful (nonrandom) information from the past history of the series for the purpose of forecasting. The models have not been explanatory in nature. The goal has not been to uncover new relationships among economic variables or to provide guides to the indirect control of such variables. The object has been forecasting, but that seems more than enough to justify the effort.

REFERENCES

1. Box, G. E. P., and Jenkins, G. W. *Time Series Analysis, Forecasting and Control.* San Francisco, Calif.: Holden-Day, 1970.

   A full-scale exposition of parametric time series modeling and forecasting.

2. Box, G. E. P., and Pierce, D. A. "Distribution of Residual Autocorrelations in Autoregressive-integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, LXV (1970), 1509–26.

   Tests of fit for ARIMA models are developed and illustrated.

3. Brown, R. G. "Less Risk in Inventory Estimates," *Harvard Business Review*, July–August, 1959.

   This review of time series analysis, specifically for inventory estimation, contains an excellent informal discussion of exponential smoothing as a forecasting method.

4. Chambers, J. C., Mullick, S. K., and Smith, D. D. "How to Choose the Right Forecasting Technique," *Harvard Business Review*, July–August, 1971.

This article is addressed to executives, but it contains an interesting overview of a variety of informal and mathematical techniques of forecasting.

5. GREVILLE, T. N. E. "Discussion of 'Six Term Formulas for Routine Actuarial Interpolation,'" *RAIA*, XXXIV (1945), 35–39.

   This contains a brief discussion of the autocorrelation between successive terms of a polynomial observed with random error. If $U_x = V_x + a_x$, where V_x is a polynomial, then the autocorrelation between $\Delta^{n+1}U_x$ and $\Delta^{n+1}U_{x+1}$ is $-(n+1)/(n+2)$. This fact is used to interpret the behavior of higher orders of differences.

6. HICKMAN, J. C. "Discussion of 'Credibility for Group Insurance Claim Experience,'" *TSA*, XXIII (1971), 254–57.

   An MA(1) model is suggested for the change in group claims from one year to the next. Unfortunately, the autocorrelation for lag 1 was written as $-\theta/(1+\theta)$ instead of the correct $-\theta/(1+\theta^2)$.

7. JENKINS, G. M., AND WATTS, D. G. *Spectral Analysis and Its Applications*. San Francisco, Calif.: Holden-Day, 1968.

   A stationary time series may be described either by its autocovariance function or by the Fourier transform of the autocovariance function, the spectrum. The book by Jenkins and Watts develops time series analysis using the spectrum rather than the autocovariance function as the primary identification tool.

8. JENKINS, W. A., AND LEW, E. A. "A New Mortality Basis for Annuities," *TSA*, I (1949), 369–466.

   A classic example of a nonstationary insurance process.

9. LEUTHOLD, R. M., MACCORMICK, A. J. A., SCHMITZ, A., AND WATTS, D. G. "Forecasting Daily Hog Prices and Quantities: A Study of Alternative Forecasting Techniques," *Journal of the American Statistical Association*, LXV (1970), 90–107.

   Forecasts derived from an econometric model and a noncausal time series model are compared using Theil's coefficient. The forecasts from the complex econometric model are only slightly better than those from the time series model.

10. LEWIS, J. L. "Expected Profit Formulas," *TSA*, XXIII (1971), 325–30.

    An application of regression to estimating profit in individual life insurance.

11. NELSON, C. R. "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," *American Economic Review*, LXII (1972), 902–17.

    Forecasts made using a sophisticated econometric model of the United States economy for fourteen economic variables are compared with those made using ARIMA models. The time series models, reproduced in the

Appendix of the paper, may be of interest to economic forecasters. The relative success of the naïve time series models is startling.

12. PARKER, G. G., AND SEGURA, E. L. "How to Get a Better Forecast," *Harvard Business Review*, March-April, 1971.
    Multiple regression for the executive.
13. THEIL, H. *Applied Economic Forecasting*. Amsterdam: North-Holland Publishing Co., 1966.
    An excellent guide to practical methods of evaluating forecasting methods.

DISCUSSION OF PRECEDING PAPER

FRANK G. REYNOLDS:

Messrs. Miller and Hickman are to be congratulated for making a significant contribution to actuarial literature. Since World War II the use of time series as a method of making forecasts has been developed and has come to be one of the standard forecasting tools of the statistician. Yet this paper is the subject's first real appearance in the Society's literature. Our gratitude is due Messrs. Miller and Hickman for rectifying this deficiency.

To the practical actuary involved in everyday problems in an insurance company or a consulting firm, the most important question about any new area of theory is, "How well does it work in practice?" A general conclusion is impossible, but the following three studies attempt to illustrate the use of time series as a tool in helping to solve practical problems and, hopefully, illustrate some of the pitfalls into which novices, such as myself, can fall.

I. *Cash-Flow Forecasting*

INTRODUCTION

In recent years one of the important problems facing the actuary has been to forecast his company's cash flow. Some of the components, such as salaries, mortgage repayments, individual policy premiums, and investment income, can be handled by computer inventories or relatively simple projections of preceding years' figures. Other components, such as bond maturities and tax payments, can be predicted with reasonable accuracy for a year or two in advance. Much more troublesome, however, are the items of policy loans and group annuity cash-outs. Both are related to changes in the money markets. High interest rates on long-term bonds and mortgages, coupled with a buoyant stock market, led to extremely large and, to a great extent, fluctuating cash outflows for several years in the late 1960's. As a result, the investment divisions of many companies experienced major disruptions in their operations and in some cases became net sellers of securities rather than investors. Most companies took steps to insulate themselves against these outflows and, in so doing, attempted to discover whether there was a relationship between the changes in interest rates and their subsequent cash outflows.

In looking for such a relationship, one of the difficult problems to be dealt with was a seemingly seasonal pattern in the cash flow. The remainder of this discussion is devoted to investigating how time series can remove this seasonal variation, leaving a series which then can be used in observing the effects of interest rate changes.

THE DATA

Ideally, the problem should be broken up into several components and each studied separately. Unfortunately, the data available for this example consisted of the net cash flow from all sources, with the consequence that the predictions were not quite as good as could be made with a finer breakdown of the figures.

TABLE 1

CASH-FLOW DATA: AUTOCORRELATION FUNCTION

Lag	Value	Lag	Value	Lag	Value
0	1.000	17	−0.193	34	−0.069
1	0.358	18	0.088	35	0.010
2	0.259	19	−0.193	36	0.424
3	0.136	20	−0.281	37	0.090
4	0.047	21	−0.286	38	0.016
5	0.106	22	−0.193	39	−0.062
6	0.402	23	−0.142	40	−0.094
7	0.071	24	0.333	41	−0.053
8	−0.036	25	−0.063	42	0.142
9	−0.056	26	−0.105	43	−0.040
10	0.002	27	−0.193	44	−0.146
11	0.064	28	−0.246	45	−0.169
12	0.539	29	−0.179	46	−0.066
13	0.053	30	0.080	47	−0.045
14	−0.015	31	−0.123	48	0.300
15	−0.158	32	−0.179	49	0.051
16	−0.251	33	−0.168	50	−0.051

A reasonable volume of data was obtained by using the monthly net cash flow in thousands of dollars for the period January, 1961—April, 1972. The monthly periods seemed to provide a reasonable balance between a large number of observations and the instability associated with short time intervals.

PRELIMINARY ANALYSIS

From Figure 1 the extreme variability and the wide range of possible values are apparent. As a first step, the autocorrelation function was calculated (Table 1) to determine whether there were any seasonal patterns, as general reasoning would suggest. The only values which are 0.300 or larger are r_1, r_6, r_{12}, r_{24}, r_{36}, and r_{48}, which suggests an annual pattern.

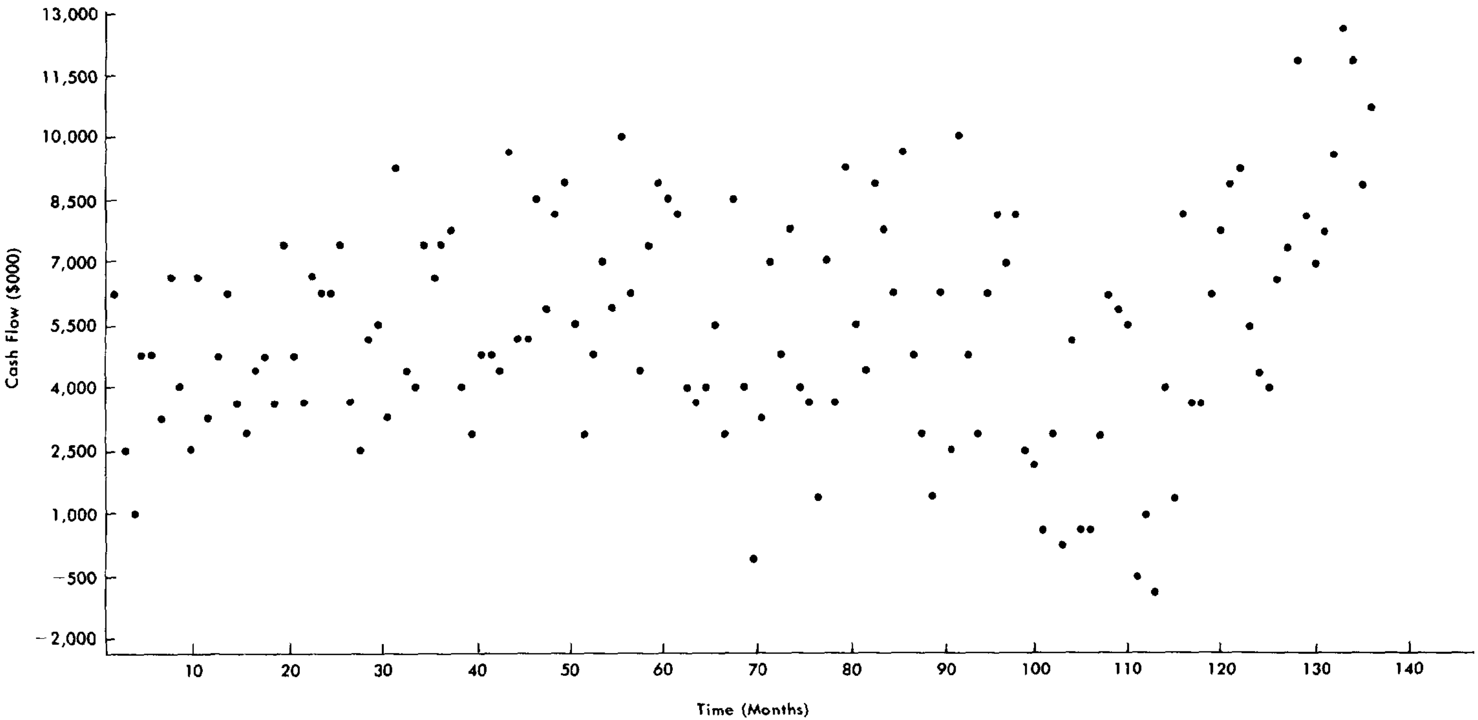

FIG. 1.—Cash-flow data: mixed autoregressive moving average models

Working under the assumption that the series of differences from the preceding year was stationary, the following prediction equation was obtained after a very considerable number of iterations:

$$(1-1.983B + 0.998B^2)(1 - B^{12})Z_t = 9.025 + (1 - 1.683B + 0.708B^2)(1 - 0.584B^{12})a_t,$$

where Z_t is the value of the monthly cash flow, B is an operator such that $BZ_t = Z_{t-1}$, and a_t is the random change which took place at time t. With the exception of the constant term, all the coefficients are very significantly different from zero.

Looking at this equation and remembering that the constant term is not significantly different from zero, it is easily seen that the following equation could be used almost equally well.

$$(1 - 2B + B^2)(1 - B^{12})Z_t = (1 - 1.7B + 0.7B^2)(1 - 0.584B^{12})a_t .$$

But this factors into

$$(1 - B)^2(1 - B^{12})Z_t = (1 - B)(1 - 0.7B)(1 - 0.584B^{12})a_t$$

and reduces to

$$(1 - B)(1 - B^{12})Z_t = (1 - 0.7B)(1 - 0.584B^{12})a_t .$$

This common factor was, of course, the reason why so many iterations were necessary in order to obtain the coefficients in the prediction equation.

This factoring leaves us with the information that the series of annual increases should be differenced from month to month. A look at the autocorrelation function for the series of annual differences (Fig. 2) shows that this problem might have been anticipated (since the autocorrelation function displays something of the typical "damped sine wave" shape usually associated with series which require that successive values be differenced before they are stationary).

Working with this new information, the prediction equation

$$\begin{aligned}(1 - B)(1 - B^{12})(1 + 0.0398B - 0.9499B^2)\\ \times (1 + 0.9284B^{12} + 0.3036B^{24})Z_t\\ = 62.2266 + (1 + 0.2718B - 0.5573B^2)\\ \times (1 + 0.3048B^{12} - 0.4985B^{24})a_t\end{aligned}$$

was obtained. With the exception of ϕ_1 and the constant term, all the coefficients are significantly different from zero.

A search for common factors was not rewarding, and accordingly this model, despite its complexity, was retained.

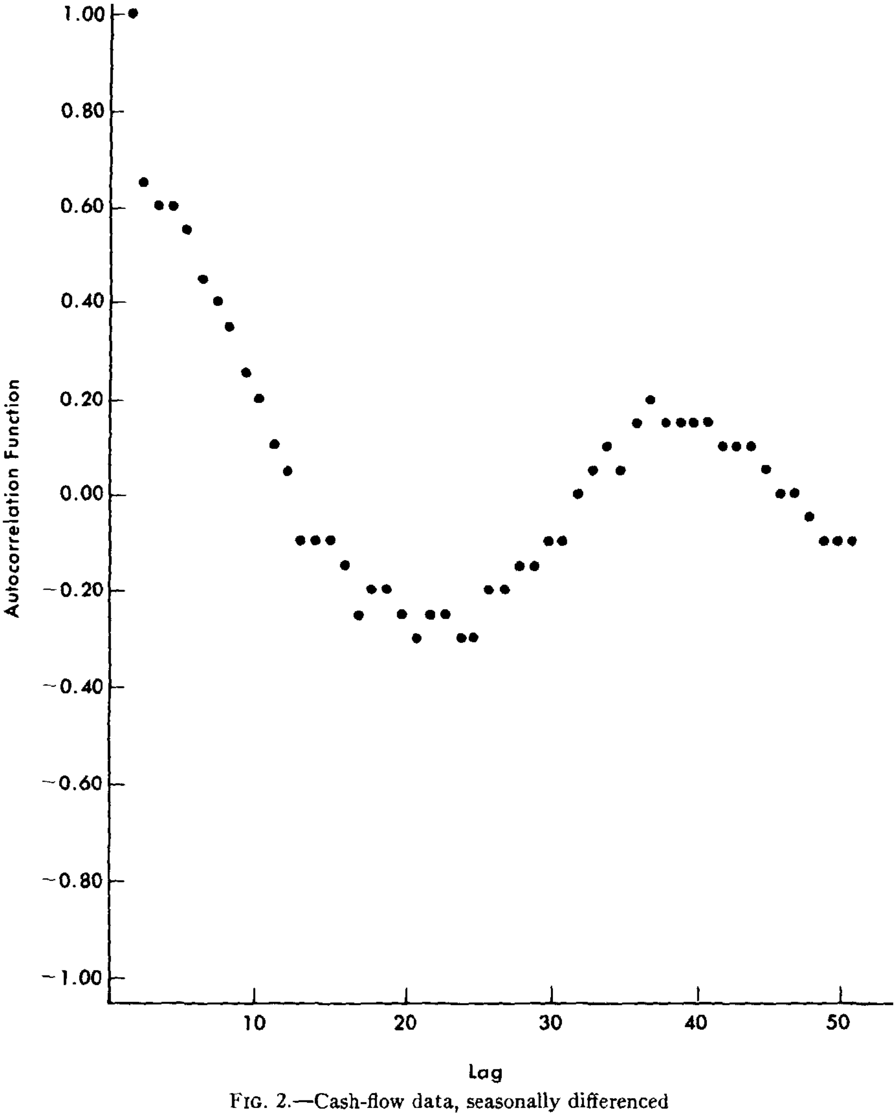

FIG. 2.—Cash-flow data, seasonally differenced

The residual standard deviation was 1,350. When compared with the data, this value, although in line with the fluctuations therein, is still much too large to permit one to have too much confidence in the prediction formula's ability to forecast the future.

PREDICTIONS

Using the prediction equation, the one-step-ahead forecasts for the last twelve times at which the values were known were calculated. They, and the associated random shocks, are displayed in Table 2.

TABLE 2

ONE-STEP-AHEAD FORECASTS FOR CASH-FLOW MODEL

(In Thousands of Dollars)

t	Z_t	Forecast	a_t
125	6,586	4,350	2,236
126	7,402	5,360	2,041
127	12,003	12,200	− 162
128	7,982	10,900	−2,936
129	6,836	9,740	−2,905
130	7,824	4,210	3,616
131	9,797	7,990	1,803
132	12,518	9,620	2,901
133	11,823	20,700	−8,868
134	8,745	9,060	− 311
135	10,870	4,030	6,845
136	2,012	9,370	−7,359

The first reaction of most people looking at the results is one of dismay. After all, the forecasts run from \$6,845,000 under to \$8,868,000 over. However, two points must be remembered. First, the original data varied in value from \$12,518,000 to −\$1,050,000 and did so quite erratically, suggesting that close adherence to the actual is not to be expected. Second, the purpose of this analysis is not so much to derive an accurate prediction equation as to ascertain whether the seasonal factors can be eliminated from the series so that a further analysis can be done to determine the effect of interest rate changes and changes in other variables on the experience. As a consequence, large residual variations which would be unacceptable in a pure prediction equation are more acceptable in this instance. The real problem, however, emerges in the next paragraph.

Table 3 shows the actual cash flows for the last two years and the monthly predictions for the next three years. The predictions show a very strong tendency to explode. A review of the original series (Fig. 1) shows why. During the 1960's the cash flows tended to increase only slowly. However, starting in the first half of 1970, they accelerated very rapidly. The prediction equation is a function primarily of the preceding two years' changes, and, when these show a new and distinct pattern, the forecasts will reflect this new direction.

This new trend explains some of the earlier problems. First, the reduction in the variance upon taking the first differences of the annual changes was not very great. It would appear that the need for this difference existed primarily in the last two years. Second, the predictions in Table 2 were disappointingly poor. Now, this can be attributed to the fact that the predictions were based in part on the new trend and in part upon the older data base, which confused the prediction equation.

TABLE 3

CASH-FLOW MODEL'S FORECASTS FOR THE FUTURE

(In Thousands of Dollars)

t	Z_{t-12}	Z_t	P_{t+12}	P_{t+24}	P_{t+36}
125	3,971	6,586	15,500	23,300	34,000
126	1,525	7,402	23,500	35,100	50,100
127	8,175	12,008	26,000	38,100	52,800
128	3,487	7,982	10,400	17,400	26,600
129	3,714	6,836	1,050	967	4,450
130	6,415	7,824	14,200	23,100	33,800
131	7,602	9,797	24,000	36,200	50,600
132	8,827	12,518	36,300	54,100	74,500
133	9,178	11,823	4,530	5,630	8,920
134	5,650	8,745	− 304	7,510	3,510
135	4,226	10,870	26,600	40,700	57,000
136	4,183	2,012	− 2,540	− 3,410	− 1,160

CONCLUSIONS

To the actuary looking for a total solution to his cash-flow prediction problems the foregoing is not too reassuring, but it does serve to illustrate the following points:

1. Time series can be used to eliminate many seasonal fluctuations and trends, so that one can obtain basic data which can be analyzed by other means—for example, regression analysis. (In support of this contention it is to be noted that the prediction equation accounts for 75 per cent of the variation in the data, leaving only 25 per cent attributable to chance—a very significant reduction.)
2. In performing a time series analysis, one should check the following:
   a) Whether data are plentiful. Even here, as the variance shows, the volume is, at best, marginal.
   b) Whether the prediction equation can be simplified by factorization. Otherwise, convergence problems arise.

c) Whether new trends have emerged in the last few values which would invalidate any analysis. The theory underlying time series depends upon the absence of major nonrandom changes. In the case of this particular company, time series would have been useful during the 1960's, but, with the changes of the early seventies, they will have to wait until the trend becomes more discernible mathematically.

3. This analysis required roughly 25 minutes of running time on an IBM 360/75 computer, a costly means of making the original estimate.
4. Despite the fact that the final prediction equation did not produce results that were in line with the most optimistic expectations, several interesting points did emerge.
   a) The current month's cash flow is related to the change that took place last year and the year before.
   b) It is important to look at the change which took place two months previously.

In summary, it would appear that the attempts to use time series alone to predict accurately an insurance company's total cash flow will not be too successful, but the analysis may provide important facts for use in making such forecasts.

II. *Forecasting Life Insurance Sales*

INTRODUCTION

One of the earliest successful applications of time series was the projection of sales for the large airlines. It is well known that these projections have led to both under- and overcapacity at times. Nevertheless, they have been reasonably accurate when averaged over the last decade.

Accordingly, it was felt that it might be of interest to attempt to fit a time series to life insurance sales. Two sets of data were available: the quarterly Canadian individual life insurance sales of all companies combined for the period January, 1947—September, 1972, and the monthly Canadian individual life insurance sales of one large company for the period January, 1962—December, 1971.

The two sets of data are portrayed in Figures 3 and 4. The all-companies data show a general tendency to increase with time and a seasonal pattern with high sales in the last quarter of each year and a much lower level of activity in the first and third quarters. The single-company data also show the increase with time and a seasonal pattern.

These observations lead one to anticipate a model containing either an ordinary or a seasonal difference and a seasonal prediction term.

THE ALL-COMPANIES MODEL

As a preliminary step, the residual variance for all models of the form $(p, d, q) \times (0, DS, 0)_4$ (where $p = 0, 1, 2, 3$; $d = 0, 1, 2$; $q = 0, 1, 2, 3$;

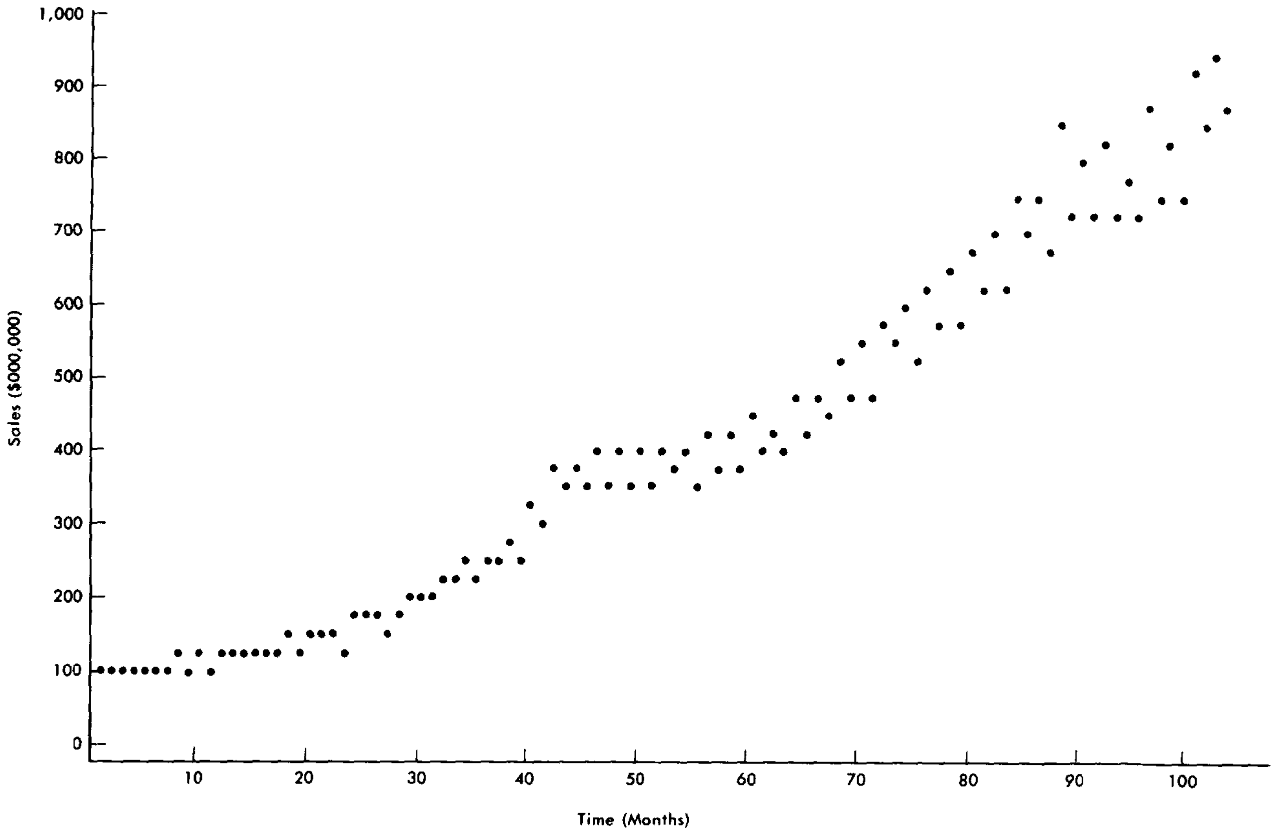

FIG. 3.—Life insurance sales for all companies

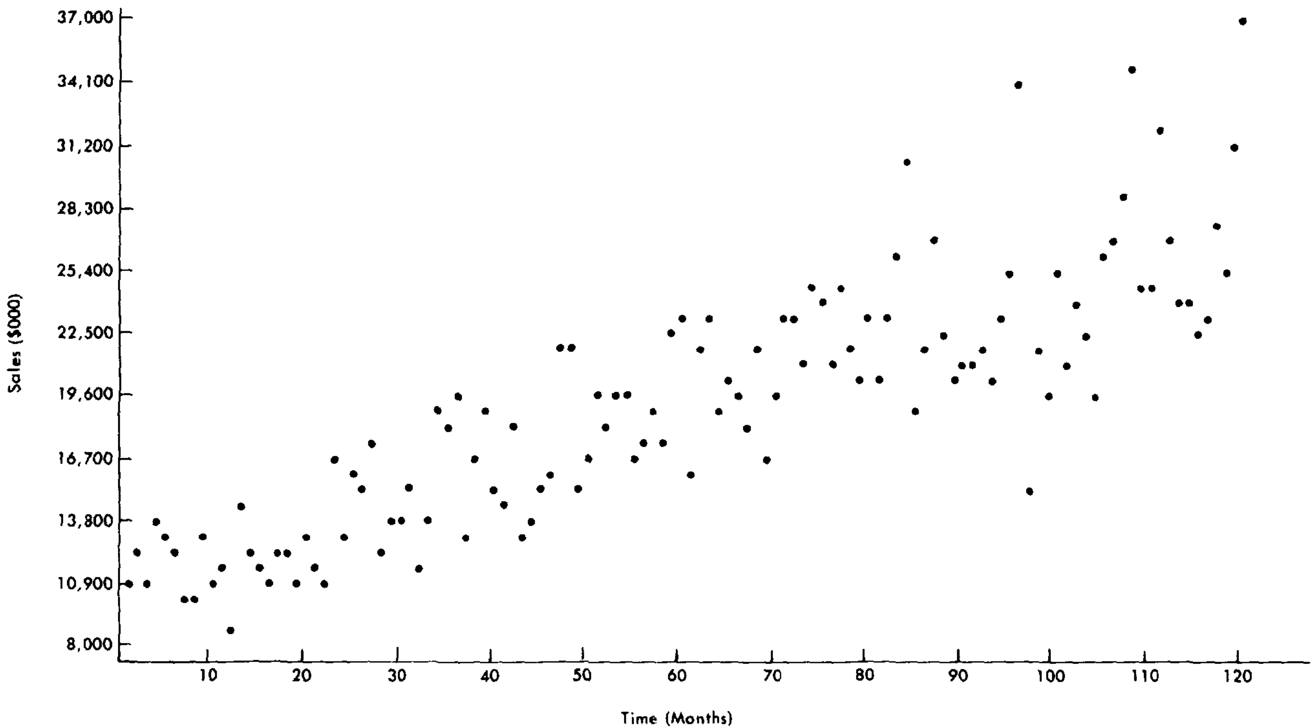

FIG. 4.—Sales data for a single Canadian company

$DS = 0, 1, 2$) was calculated. The $(0, 1, 3) \times (0, 1, 0)_4$ model, which involves approximating the results, after an ordinary and a seasonal difference are taken, by a three-term moving average formula, had the lowest residual variance. The partial autocorrelation function (Table 4) shows an interesting pattern. Most of the partial autocorrelations corresponding to a lag of $1 + 4n$ ($n = 0, 1, 2, 3, 4, \ldots, 10$) are considerably larger in absolute value than their neighbors, suggesting a strong annual pattern.

At this point, because of the increasing complexity of the model, it was decided to look at the series of increases from the same quarter in the

TABLE 4

LIFE INSURANCE SALES FOR ALL COMPANIES: PARTIAL AUTOCORRELATION FUNCTION

Lag	Value	Lag	Value	Lag	Value
1	−0.260	17	−0.170	33	−0.077
2	−0.076	18	0.104	34	−0.080
3	0.138	19	−0.085	35	0.037
4	−0.281	20	−0.021	36	0.001
5	−0.099	21	−0.151	37	0.057
6	0.232	22	0.035	38	−0.026
7	−0.008	23	−0.015	39	−0.000
8	−0.109	24	−0.132	40	0.032
9	−0.146	25	−0.032	41	−0.100
10	−0.056	26	−0.036	42	0.064
11	−0.054	27	0.017	43	0.059
12	−0.015	28	−0.011	44	−0.055
13	−0.172	29	−0.092	45	0.065
14	−0.001	30	0.050	46	0.105
15	0.069	31	−0.023	47	0.052
16	0.083	32	0.037	48	0.155

preceding year (i.e., the original series after an annual difference). This step decreased the volume of calculations considerably, and hence the time needed to obtain estimates.

The residual variances for models of the form (p, 1, q) with values of p and q ranging from 1 to 6 were calculated and compared. The (5, 1, 2) model had by far the smallest residual variance and was selected for further testing. The value of ϕ_5 was found to be not significant, and hence the model dropped to a (4, 1, 2) model. Calculations for this model gave the results shown in Table 5. Obviously, there is an annual component in the series, and probably a semiannual one as well (the change from a (5, 1, 2) model to a (4, 1, 2) model increased the residual variance by less than 1 per cent, another indication that ϕ_5 was not an important parameter). Thus the results obtained by working with the series of increases from the preceding year pointed to a $(2, 1, 2) \times (p, 1, 0)_4$ model for the series of actual quarterly sales.

Working once more with series of actual quarterly sales, an attempt was made to fit a $(2, 1, 2) \times (2, 1, 0)_4$ model to the data. The value of $S\phi_2$ was found to be nonsignificant, and the model was reduced to a $(2, 1, 2) \times (1, 1, 0)_4$ form. When the calculations were done, this model had the parameters shown in Table 6. The key values are all significantly different from zero, and all would seem well. However, one significant

TABLE 5

INITIAL PARAMETERS FOR ALL-COMPANIES MODEL

Parameter	Value	Standard Deviation	Parameter	Value	Standard Deviation
ϕ_1	−0.0158*	0.1378	ϕ_4	−0.4484	0.1124
ϕ_2	−0.5843	0.1496	θ_1	0.2575	0.1247
ϕ_3	−0.1641*	0.1142	θ_2	−0.7471	0.1241

* These two values are not significantly different from zero at the 95 per cent confidence level.

TABLE 6

TENTATIVE PARAMETERS FOR ALL-COMPANIES MODEL

Parameter	Value	Standard Deviation	Parameter	Value	Standard Deviation
Constant term	0.9330*	1.4139	$S\phi_1$	−0.3254	0.1232
ϕ_1	−0.0054*	0.0907	θ_1	0.1272	0.0462
ϕ_2	−0.6914	0.1018	θ_2	−0.9255	0.0421

* *Not significantly different from zero at the 95 per cent confidence level.*

change had occurred in moving from the series of increases from the preceding year to the series of actual quarterly sales—namely, the residual variance had risen from 281.988 to 300.749, despite the fact that the models were supposedly the same.

The reason for the change in the residual variance lay in the nature of the prediction equation. The model for the series of increases from the preceding year was of the form

$$(1 - B^4)(1 - B)(1 - \phi_2 B^2 - \phi_4 B^4)Z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$$

(showing only the parameters which are significantly different from zero), while the $(2, 1, 2) \times (1, 1, 0)_4$ model is of the form

$$(1 - B^4)(1 - B)(1 - \phi_2 B^2)(1 - S\phi_1 B^4)Z_t = (1 - \theta_1 B - \theta_2 B^2)a_t .$$

This later model involves a term of the form $S\phi_1\phi_2B^6$ which is not present in the earlier one and accounts for the difference in the residual variance. It provides an excellent illustration of the extreme care that must be used in making "small" changes in the model.

The (4, 1, 2) $\times$ $(1, 1, 0)_4$ model was investigated and the seasonal $S\phi_1$ term found to be nonsignificant.

Finally, the following (4, 1, 2) model with an annual difference was examined and found to have the form

$$\begin{aligned}(1 - B)(1 - B^4)& \\ \times (1 + 0.0188B + 0.5836B^2 + 0.1624B^3 + 0.4470B^4)Z_t& \\ = 1.0351 + (1 - 0.2539B + 0.7440B^2)a_t\,.&\end{aligned} \tag{1}$$

All the parameters except ϕ_1, ϕ_3, and the constant term are significantly different from zero. The residual standard deviation is only 16.81, which is small relative to the actual sales (which are up to 1,000). Furthermore, the model explains 99.53 per cent of the variation in the original series, a rather remarkably high percentage. The chi-square test for goodness of fit produces a value of 25.52 for 42 degrees of freedom, indicative of a good fit.

As a further test on the adequacy of the model, the one-step-ahead forecasts were calculated for the last twelve time periods for which actual values were available (October, 1969—September, 1972) and compared with the actual sales. The results are shown in Table 7.

TABLE 7

ONE-STEP-AHEAD FORECASTS FOR ALL-COMPANIES MODEL

Time Period (t)	Actual Sales (Z_t)	Prediction ($P_{(t-1)+1}$)	Residual Error (a_t)
1969:			
October–December (92)	822.77	868.00	−45.21
1970:			
January–March (93)	732.49	751.00	−18.23
April–June (94)	780.86	808.00	−27.56
July–September (95)	718.90	697.00	22.34
October–December (96)	887.22	837.00	50.12
1971:			
January–March (97)	753.61	785.00	−31.11
April–June (98)	812.63	818.00	− 5.54
July–September (99)	757.12	736.00	21.38
October–December (100)	932.52	890.00	42.77
1972:			
January–March (101)	841.93	820.00	21.94
April–June (102)	962.36	919.00	43.63
July–September (103)	869.22	884.00	−14.88

A visual comparison shows that the largest difference is only 50, while the smallest is a mere 5. More mathematically, the chi-square test gives a statistic of 1.2 with 6 degrees of freedom, a value which will be exceeded by random chance 97.5 per cent of the time even if the model is a perfect fit.

A final check was made by comparing the last four available sales figures with the projections of them that would have been made a year earlier (Table 8). The errors are much larger than those in Table 7. This is to be expected, since the projections extend much farther into the future. Also, a look at Figure 3 shows that sales leveled out in 1970 and 1971 and then surged ahead in 1972, catching the prediction equation flat-footed.

All things considered, the model in equation (1) appears to pass the reasonability tests.

TABLE 8

ONE-YEAR-AHEAD FORECASTS FOR ALL-COMPANIES MODEL

Time Period	Actual	$P_{(t-4)+4}$	Δ
October–December 1971	932.52	907.00	25.52
January–March 1972	841.93	775.00	66.93
April–June 1972	962.36	846.00	116.36
July–September 1972	869.22	795.00	74.22

THE SINGLE-COMPANY MODEL

As a preliminary step, the residual variance for all models of the form $(p, d, q) \times (0, DS, 0)_{12}$(where $p = 0, 1, 2, 3$; $d = 0, 1, 2$; $q = 0, 1, 2, 3$; $DS = 0, 1, 2$) was calculated. A comparison of the residual variances indicated that a $(2, 0, 3) \times (p, 1, q)_{12}$ model would be best. Both the autocorrelation function and the partial autocorrelation functions showed very high values for a lag of one year but did not show unusual values for lags of multiples of one year (Table 9). Accordingly, a $(2, 0, 3) \times (1, 1, 1)_{12}$ model was constructed. It was found that θ_3 was not significantly different from zero. Better results were obtained with a $(2, 0, 2) \times (1, 1, 1)_{12}$ model, as can be seen in Table 10. All parameters of the model are significantly different from zero. The residual standard deviation of 2,169 (or $2,169,000) is disappointingly high, since it is about 10 per cent of many of the sales values. Comparison of the variance in the original series with that in the residual series shows that the model accounts for about 86 per cent of the variation, a reasonably high proportion. A chi-square statistic of 22.05 with 43 degrees of freedom indicates that the fit is certainly reasonable.

TABLE 9

SALES DATA FOR A SINGLE CANADIAN COMPANY

AUTOCORRELATION FUNCTION

Lag	Value	Lag	Value	Lag	Value
0.	1.000	17.	−0.143	34.	0.023
1.	0.011	18.	−0.101	35.	0.003
2.	0.185	19.	−0.058	36.	−0.112
3.	0.060	20.	−0.118	37.	−0.067
4.	0.137	21.	0.061	38.	−0.048
5.	0.080	22.	−0.077	39.	0.032
6.	0.070	23.	0.003	40.	−0.035
7.	−0.060	24.	−0.040	41.	−0.064
8.	−0.061	25.	0.011	42.	−0.081
9.	−0.006	26.	0.054	43.	0.126
10.	−0.103	27.	0.053	44.	−0.032
11.	0.026	28.	0.154	45.	−0.042
12.	−0.324	29.	0.155	46.	−0.013
13.	−0.031	30.	0.023	47.	−0.015
14.	−0.071	31.	−0.098	48.	0.117
15.	−0.093	32.	0.107	49.	0.123
16.	−0.128	33.	−0.045	50.	0.019

PARTIAL AUTOCORRELATION FUNCTION

Lag	Value	Lag	Value	Lag	Value
1.	0.011	18.	−0.037	35.	0.038
2.	0.185	19.	−0.041	36.	−0.144
3.	0.059	20.	−0.132	37.	−0.126
4.	0.106	21.	0.147	38.	−0.005
5.	0.062	22.	−0.075	39.	0.028
6.	0.027	23.	0.020	40.	0.114
7.	−0.101	24.	−0.136	41.	−0.004
8.	−0.108	25.	−0.043	42.	−0.158
9.	−0.006	26.	0.094	43.	0.048
10.	−0.091	27.	−0.038	44.	−0.029
11.	0.049	28.	0.155	45.	0.013
12.	−0.282	29.	0.105	46.	−0.012
13.	−0.018	30.	−0.099	47.	0.003
14.	0.047	31.	−0.218	48.	0.070
15.	−0.069	32.	−0.094	49.	0.012
16.	−0.062	33.	0.033	50.	−0.037
17.	−0.107	34.	−0.086		

TABLE 10

INITIAL VALUES FOR SINGLE-COMPANY MODEL

Parameter	Value	Standard Deviation	Parameter	Value	Standard Deviation
Constant term. .	1,698.28	103.832	θ_1.	1.4124	0.0191
ϕ_1.	1.3336	0.0642	θ_2.	−0.9894	0.0140
ϕ_2.	−0.8296	0.0679	$S\theta_1$.	0.8771	0.0396
$S\phi_1$.	0.2398	0.1087			

Two reasons for the lack of fit can be given. First, and most important, one usually likes to have two to three hundred values of the series, particularly where the seasonal period is long. The all-companies model involves data for nearly twenty-five years, whereas this model involves only a ten-year period. Second, the data are monthly and for only one company. Naturally they are subject to more fluctuation than combined sales by quarter for all companies.

To test the adequacy of the model, one-step-ahead forecasts for 1971 (the last year actual sales figures were available) were calculated and compared with the actual sales figures (Table 11). A visual comparison

TABLE 11

ONE-STEP-AHEAD FORECASTS FOR SINGLE-COMPANY MODEL

(In Thousands of Dollars)

Period	Actual Sales (Z_t)	Predicted Sales ($P_{(t-1)+1}$)	Residual Error (a_t)
January	24,827	27,000	−2,186.3
February	24,727	30,900	−6,194.9
March	32,156	31,200	980.9
April	26,715	30,500	−3,818.8
May	24,300	29,200	−4,937.1
June	23,984	30,100	−6,088.4
July	22,820	28,900	−6,055.4
August	22,880	28,900	−5,972.1
September	27,686	30,100	−2,389.6
October	25,528	30,900	−5,352.2
November	31,221	34,500	−3,223.9
December	36,898	37,200	− 259.1

shows that the errors range in size from $302,000 to $6,173,000. These results look poor until they are analyzed. The graph of the actual sales (Fig. 4) is most illuminating. Beginning in 1969, sales began to fluctuate widely from month to month, particularly between December and January. Under such conditions, a mathematical analysis is bound to break down somewhat.

In summary, the model of Table 10 appears to be reasonable, all things considered.

FORECASTS OF THE FUTURE

For the all-companies model, predictions were made for 1973 and 1974 (Table 12). The increases appear fairly large, but, when compared as in Table 13 with the increases for the preceding two years, they are seen to fall reasonably in the range. A modifying influence seems to be at work.

Wherever the actual increases are largest, they exceed the forecasts, and wherever the actual increases are smallest, the forecasts are larger.

For the single-company model, similar projections were made. They are summarized by quarter in Table 14. It would appear that the prediction equation was able to foresee a better year in 1972 (which actually occurred).

CONCLUSIONS

The all-companies model is certainly adequate for prediction purposes and even for a single company, time series models seem to perform adequately.

TABLE 12

ALL-COMPANIES MODEL'S FORECASTS FOR THE FUTURE

Period	Actual 1971	Actual 1972	Predicted 1973	Predicted 1974
January–March	753.61	841.93	931	1,050
April–June	812.63	962.36	1,040	1,150
July–September	757.12	869.22	974	1,070
October–December	932.52	....	1,160	1,260

TABLE 13

ANALYSIS OF THE INCREASES FORECAST BY ALL-COMPANIES MODEL

Period	Penultimate Actual	Last Actual	1973 Forecast	1974 Forecast
January–March	21.12	88.32	91	119
April–June	31.77	149.73	78	110
July–September	38.22	112.10	105	96
October–December	64.45	45.40	120	100

TABLE 14

SINGLE-COMPANY MODEL'S FORECASTS FOR THE FUTURE

(In Thousands of Dollars)

Period	Actual 1970	Actual 1971	Predicted 1972	Predicted 1973
January–March	56,900	81,700	95,800	104,300
April–June	70,700	75,000	94,000	103,600
July–September	68,400	73,400	92,600	102,300
October–December	91,000	93,600	108,600	116,100

III. *Forecasting Stock Market Values*

INTRODUCTION

The introduction of variable annuities and other section 81(5) funds has brought many actuaries into closer contact with their company's investment division. Sooner or later, the actuary is asked whether there is not some way in which he can apply his mathematical knowledge to assist the investment officers in making stock market predictions. Most actuaries back off at this point; for the venturesome individual, however, time series can provide some interesting insights into the workings of the stock market.

The following six companies were chosen to illustrate a variety of common situations: C.I.L., a chemical manufacturer, whose stock has remained relatively stable over the period; Distillers Seagram, whose

TABLE 15

Company	Model	χ^2
C.I.L.	$(1-0.943B)Z_t=a_t+0.913$	17.97*
Distillers Seagram	$(1-B)Z_t=(1-0.103B)a_t+0.189$	44.35
Gomtar	$(1-B)Z_t=(1-0.280B)a_t+0.028$	9.67*
Doodyear	$(1-B)Z_t=(1+0.004B)a_t+0.382$	21.65*
Sherritt-Gordon	$(1-B)Z_t=(1-0.126B)a_t+0.073$	53.38
Stelco	$(1-B)Z_t=(1+0.097B)a_t+0.092$	27.79*

* Not significant.

stock rose rapidly during the early part of the period but which has been relatively stable for some years; Domtar, a primary iron and steel manufacturer, whose stock tends to be cyclical; Goodyear, a rubber and tire manufacturer, whose stock declined until very recently; Sherritt-Gordon, a mining firm, whose stock rose through most of the period; Steel Company of Canada, a primary iron and steel manufacturer, whose stock tends to be cyclical. Month-end stock prices for the period January, 1959—March, 1973, were available.

THE MODELS

The results of attempting to fit models to the data are summarized in Table 15. The value of χ^2_{23} at the 5 per cent level is 35.2. Thus the models fit reasonably well except in the cases of Distillers Seagram and Sherritt-Gordon. Despite some indications of a seasonal factor in the autocorrelation function, the introduction of a seasonal factor did not reduce the residual variance.

A further test on the adequacy of the models was made by calculating the one-step-ahead forecasts for the last twelve time periods for which actual values were available and comparing these with the actual prices by means of a chi-square test. For all six models the fit was excellent.

One interesting fact emerges from Table 15, namely, that each model is virtually of the form $(1 - B)Z_t = a_t$, that is, the month-to-month change is a completely random value—the so-called white noise.

CONCLUSIONS

The confidence intervals for the estimates are much too large to give one confidence that the method produces useful projections. However, the fact that the month-to-month changes in values are themselves virtually white noise is interesting and suggests that a fundamental review of a company's financial outlook by a good securities analyst may still be worth much more than a great deal of mathematical analysis.

IV. *General Conclusions*

The analysis of a number of series of data of the nature of those with which the actuary may have contact leads me to suspect that time series will be useful to the actuary in the same way that regression analysis often is, namely, in pointing out relevant factors that are too subtle for the eye to extract from the mass of data available.

RICHARD W. ZIOCK:

The purpose of this discussion is to present a modification of time series analysis, which sacrifices very little in accuracy to achieve a huge gain in ease of practical application. This modification should prove to be of special value when a large number of series must be forecast.

Performing a time series analysis is no mean task. There are many steps, and, at several points in the process, decisions have to be made, of which the ultimate effect on the final forecast is unknown. Generally speaking, the process one follows is to calculate autocorrelations on the data and several meaningful transformations of the data and then search for a simple and recognizable pattern of the autocorrelation function. Having done this, one has identified the process (i.e., model) and the transformation to be used. At this stage, unless the model is very simple, it is necessary to estimate the coefficients of the model. Then these initial estimates are used in a computer program which refines the estimates using a nonlinear least-squares method. Finally, when the model has been fitted, forecasts can be made. If the forecasts are not suitable in some respect, then probably an error has been made in the selection of the model and/or the transformation of the data.

I would be very much surprised if many actuaries take the time and trouble to go through this long process. It should also be pointed out that it is not a clear-cut process at each juncture. At the Waterloo Time Series Conference (sponsored by the Committee on Research of the Society of Actuaries) the experts in time series who were present spent considerable time on certain models and autoregression patterns, discussing whether the model was of this type or that type.

There is more evidence of this fact in a standard reference on time series analysis.[1] These authors show models for six data series, which they call Series A–F. Of the six series, four have two fitted models shown, indicating that they are unable to find a single model for every set of data.

The modification of time series analysis (TSA) which I call "stepwise elimination" (SWE) eliminates much of the arbitrary decision-making which must be done in the process of fitting time series models to the data.

With SWE it is only necessary, once the method is programmed on the computer, to choose the transformation of the data on which the model is to be built. The remainder of the model-building is performed entirely by mathematical routines within the computer. The basis of the SWE procedure is that, by giving up very slight amounts of accuracy in the fitting process, we can make a huge gain in terms of the elimination of decisions and work.

Consider the following general autoregressive model of order 7, AR(7):

$$Z_t = \Pi_0 + \Pi_1 Z_{t-1} + \Pi_2 Z_{t-2} + \ldots + \Pi_7 Z_{t-7} + a_t , \qquad (1)$$

where the Z_t's are values of the time series and a_t is a normal random deviate with zero mean and some specified nonzero variance. By a suitable choice of the coefficients in this model, letting coefficients equal zero for those variables to be eliminated from the model, one can create any kind of autoregressive model needed in practice. This model is about all one needs in practice, because moving average processes are very seldom encountered. However, on an approximate basis, this model also can handle adequately moving average processes as well. This is true because any moving average model can be transformed into an autoregressive model containing an infinite string of autoregressive parameters.

The above AR(7) model cuts off the autoregressive parameters after lag 7; however, for a moving average process with effective shocks at lags of 1, 2, or 3, the coefficients decline very rapidly in the rewritten autoregressive model. In fact, for a MA(1) model rewritten as an autoregres-

[1] G. E. P. Box and G. W. Jenkins, *Time Series Analysis: Forecasting and Control* (San Francisco, Calif.: Holden-Day, 1970), p. 239.

sive model with an infinite number of terms, if the coefficient of Z_{t-1} is Π_1, then the coefficient of Z_{t-7} will be Π_1 to the seventh power (Π_1^7). Since Π_1 is always absolutely less than unity (for stationary models), we have a quick die-off of the coefficients of the rewritten autoregressive model.[2] Hence, on an approximate basis, the above model handles moving average processes as well.

The above AR(7) model could be fitted to the data using a least-squares multiple regression method. The problem is that the least-squares procedure would assign a nonzero value to each Π_i whether or not the Z_{t-i}'s were significantly helpful as predictors. Thus the model would be overparameterized. This would cause the coefficients to move up and down with no regularity. The technical reason for this is known as multicollinearity. In laymen's terms, too many variables are moving parallel for the procedure to determine adequately the effect of each variable.

This difficulty is solved quite easily by a procedure commonly used by econometricians and statistical researchers—stepwise multiple regression. Basically, the stepwise multiple regression procedure uses only those variables in the final equation which it finds help significantly to explain the variance or changes in the dependent variable. All other coefficients are set at zero.

There are many technical variations in the stepwise multiple regression procedure. The one which is in the IBM computer package used by us is the Doolittle method. In this method all variables which predict at least 1 per cent (this can be varied) of the variance of the dependent variable are retained in the final multiple regression. All others are eliminated. In the Doolittle procedure one has the ability to force a variable, that is, a variable will be tried first even though it may not explain as much variance or have as much potential for explaining variance as another variable. When I apply stepwise multiple regression to the AR(7) model, I force the variable at lag 1 and at lag 4. The variable at lag 1 is forced to cause a pickup of trend factor, if such exists, and at lag 4 is forced to pick up a seasonal factor. (I am dealing always with quarterly data.) The other variables are allowed to enter freely. However, on occasion the forcing will delete lag 4 because it does not explain 1 per cent of the variance. The forcing simply provides a better chance to explain 1 per cent of the variance, since it is considered first.

SWE is a fitting to the general AR(7) model by a stepwise multiple regression technique. I think one can see how much simpler in practice

[2] This is also illustrated in eqs. (9) and (10) of the paper under discussion.

SWE is than TSA. In order to demonstrate how well SWE works, I carried it out on Series A and Series E that had already been fitted by conventional time series methods. Box and Jenkins provide data in the appendix to which I fit the generalized autoregressive model with the stepwise elimination method. Series A has 197 observations of "Chemical Process Concentration Readings: Every Two Hours." The model[3] of the differenced data is

$$\nabla Z_t = a_t - 0.70a_{t-1}\,.$$

This can be rewritten into the following general autoregressive formula by substituting in order to eliminate terms involving lagged a's. This was done, and the result is

$$\begin{aligned}\nabla Z_t = &-0.70\nabla Z_{t-1} - 0.49\nabla Z_{t-2} - 0.34\nabla Z_{t-3} - 0.24\nabla Z_{t-4} \\ &-0.17\nabla Z_{t-5} - 0.12\nabla Z_{t-6} - 0.08\nabla Z_{t-7} + R + a_t\,,\end{aligned} \tag{2}$$

where R involves terms of ∇Z_{t-k} with $k > 7$.

The equation I obtained through the SWE procedure is

$$\begin{aligned}\nabla Z_t = 0.01 &- 0.63\nabla Z_{t-1} - 0.41\nabla Z_{t-2} - 0.38\nabla Z_{t-3} - 0.33\nabla Z_{t-4} \\ &- 0.33\nabla Z_{t-5} - 0.32\nabla Z_{t-6} - 0.21\nabla Z_{t-7} + a_t\,.\end{aligned} \tag{3}$$

Here the coefficients do not die off as fast, and a small constant term is present; but the form is very similar to that of formula (2). The data fit was actually better, the residual variance of formula (3) being 93 per cent of the residual variance of formula (2). It should be pointed out that my computer program always eliminates the first seven data points to simplify the fitting process; thus we would expect the coefficients to be slightly different even if there were no other differences.

The model[4] for Series E, which is "Woelfer Sun Spot Numbers: Yearly," of which there are a hundred observations, is

$$Z_t = 11.31 + 1.57Z_{t-1} - 1.02Z_{t-2} + 0.21Z_{t-3} + a_t\,. \tag{4}$$

The model obtained through the SWE procedure was

$$Z_t = 10.44 + 1.62Z_{t-1} - 1.12Z_{t-2} + 0.28Z_{t-3} + a_t\,. \tag{5}$$

Here, there is only a small difference between the two models. (All of the difference in this case is due to eliminating the first seven data points.) This illustrates the fact that SWE works better for autoregressive processes.

[3] Box and Jenkins, *Time Series Analysis*, p. 239.

[4] *Ibid.*

Although SWE is good for pure autoregressive processes because of similarity and proved fair in one case for a pure moving average process, it was less successful with ARIMA (i.e., mixed) processes.

In order to test SWE's performance under stable conditions, 196 data values were generated (using random numbers) from this ARIMA model:

$$Z_t - 0.92Z_{t-1} = 1.45 + a_t - 0.58a_{t-1}\,. \tag{6}$$

This is the model which Box and Jenkins fit to the Series A undifferenced data. (The SWE fit to actual data was not good, but this may have been because of the consistency of eq. [6] and the data.) The AR(7) form of equation (6) is

$$\begin{aligned} Z_t = 3.59 &+ 0.34Z_{t-1} + 0.20Z_{t-2} + 0.11Z_{t-3} + 0.07Z_{t-4} \\ &+ 0.04Z_{t-5} + 0.02Z_{t-6} + 0.01Z_{t-7} + a_t\,. \end{aligned} \tag{7}$$

The SWE procedure yielded

$$Z_t = 7.58 + 0.23Z_{t-1} + 0.23Z_{t-2} + 0.13Z_{t-3} + a_t\,, \tag{8}$$

which is different from formula (7). The sizes of the coefficients at lags 1, 2, and 3 are in the same range. All coefficients less than 0.10 in formula (7) were eliminated, and because of that the constant term is larger to give the model the correct mean or expected value. Thus it appears that SWE tends to make simpler models out of complex ARIMA models.

What is the consequence of this? The residual variance of formula (6) was 0.097, whereas that of formula (8) was 0.099. Relative to the data (generated) variance of 0.1199, under formula (6) 81 per cent of the data variance is unexplained and under formula (8) 83 per cent of the data variance is unexplained. Viewed differently, formula (6) explains 19 per cent of the variance and formula (8) explains 17 per cent of the variance. The difference of 2 per cent does not seem large, especially in light of the gain in practicality.

I also generated data with formula (7). The SWE equation fitted to these data was practically identical with formula (8). This shows that the effect of limiting the number of terms to seven is almost nil.

What has been the practical experience with SWE? Each quarter, at my company, we update our forecasts for twenty-seven quarterly time series. Some of these time series are of cash-flow components, and others are of gain from operations by line of business.

Before fitting the general autoregressive model by SWE, the data may be transformed. Our computer program handles five transformations: first differences, differences over lag 4, logarithms, logarithms of differences, and logarithms of differences over lag 4. We have found that sometimes it is necessary to difference the data, and while usually the differences are over interval 1, on some strongly seasonal series we difference over interval 4 (first quarter from first quarter, etc.). For those series which increase geometrically or as a constant percentage, logarithms are taken before differencing. The most common transformation is the differences of logarithms. This means that the series is trending upward or downward by a constant percentage each period. Because of the trend toward term insurance and other factors, I felt that some of the series involving either assets or reserves would probably increase arithmetically rather than geometrically in the future. Thus "Interest Received on Bonds" (which involves assets multiplied by rates of interest) was only differenced before fitting. Only a few of the series were stationary without transformation. One is "New Flow from Changes in Other Noninvested Funds." The correct transformation is usually apparent from a graph of the raw data.

Most of the fitted models conform to expectations. In many of the models, only the first few coefficients are nonzero, indicating a basic autoregressive process. In others, all of the coefficients are nonzero, and they decline with increasing lag, indicating a moving average. A few appear to be combined ARIMA models. Of course, the textbook labeling of these models does not matter to us. What is important is the reasonableness of the forecasts. Since, with SWE, this is only a function of transformation, we have only to choose the correct transformation. A chi-square test on the residuals' autocorrelations helps to test the transformation. However, the resulting forecast is really the final test of the choice of transformation.

The resulting forecasts serve well in picking up seasonal patterns and in following the general trends.

(AUTHORS' REVIEW OF DISCUSSION)

ROBERT B. MILLER AND JAMES C. HICKMAN:

The two discussions of our paper are interesting but very different. Mr. Reynolds presents us with a set of case studies illustrating some of the problems that inevitably seem to arise in using time series analysis to produce forecasts. Mr. Ziock, on the other hand, has proposed a reduction in the size of the class of models used in routine time series analysis. He also outlines a semiautomatic method for selecting the model from within the reduced class and estimating the parameters of the model. His testimonial on the importance of using transformations as a preliminary step to modeling is welcome.

Mr. Ziock proposes that as a practical matter the class of models used to analyze business time series may be reduced from the fairly broad class of autoregressive integrated moving average (ARIMA) models examined in the paper to the subclass of autoregressive models with a maximum lag of 7, AR(7). In supporting his proposition, he examines three sets of time series data for which models within the ARIMA class are already available. For each of these sets of data he selects and fits a model from within the AR(7) class, using a stepwise regression program. In the comparisons of his model with those from within the ARIMA class he gives a fair appraisal of the results.

We would like to make a few comments on each of his examples. In the first example, the first differences of the observations have been fitted to a moving average model of order 1, MA(1). To the same data Mr. Ziock has fitted an AR(7) model. His model involves eight parameters, plus the variance of the residuals, as compared with one parameter, plus the variance of the residuals, in the original model. Despite the convenience of the semiautomatic way in which the model was selected and the parameters estimated, it is difficult to see why an AR(7) model is simpler than a MA(1) model.

On a somewhat more technical level, we acknowledge that we adopted a "build up" strategy rather than a "build down" strategy in selecting a model. That is, we advocated starting with the simplest model suggested by the sample autocorrelation function and a graph of the data, then adding as few parameters as possible consistent with the residuals appearing to be white noise. This was done not out of slavish devotion to the slogan "parsimony" but because we wanted to minimize our estimation problems and conserve degrees of freedom so that we could make sharper statements in testing hypotheses about the parameters and the residuals. Perhaps we are overconservative, but we want to avoid reading more structure into the data than is incontrovertibly there.

We are mildly distressed by the constant term in Mr. Ziock's equation (3). It is easy to see that the forecasts made from the original model will be a horizontal line. (See the discussion on MA(1) models and exponential smoothing in Sec. III of the paper.) The forecasts made from equation (3) will edge upward as a result of the constant term. If forecasts do not seem reasonable, given the data, there is good reason to stop and examine the model.

Mr. Ziock's second example involves fitting a model from within the class of AR(7) models to data to which a satisfactory AR(3) model has already been fitted. Since the actual parameter-fitting process differs at most in the handling of the starting values and the details of the least-squares method used, it is not surprising that the two models are practically identical.

The third and final example involves a simulation experiment in which a mixed autoregressive moving average model is used to generate data to which a member of the AR(7) class is fitted. The resulting model is interesting, but the experiment was not highly successful.

In summary, we agree that there is a degree of arbitrariness in starting with the large class of ARIMA models. There are certainly even more all-encompassing and more complicated classes that could be considered initially. If prior evidence rather conclusively indicates that the data may be analyzed satisfactorily within a smaller class of models, such as the AR(7) class, the process of model selection and parameter estimation may proceed more rapidly. We would differ with Mr. Ziock primarily in that we are not convinced that mixed models are rare. We are also somewhat concerned with the apparent concentration on the reduction in the variance of residuals in the stepwise regression program. We are apprehensive that this emphasis may result in overparameterization and may discourage a detailed examination of the residuals. The ultimate test of the adequacy of the model is the behavior of the residuals.

If a semiautomatic analysis is desired, we suggest the X-11 program developed by the United States Bureau of the Census and described in somewhat extravagant terms in reference [4]. We tend to the view that X-11, exponential smoothing, or any forecasting technique that involves automatic analysis and does not compel examination of the data and multiple tests of the residuals to check the adequacy of the model may lead the forecaster to lose touch with reality.

Instead of suggesting a reduction in the class of models, Mr. Reynolds uses an even broader class of models in some of his case studies. In selecting a model he goes beyond the sample autocorrelation function and displays partial autocorrelations. In both these instances he augments our paper.

Mr. Reynolds' first case study involves cash-flow forecasting. He provides the reader with an excellent account of his problems in modeling these data. In his comments he suggests that perhaps the process generating these data shifted around 1970. A study of his Figure 1 would lead one to speculate that the observations beyond observation 100 might produce a model different from that which was selected for the entire series. The analysis of blocks of economic data with end points identified with known economic events (imposition of price controls, devaluations, etc.) can often lead to insights.

The second case study involves the use of seasonal models and the use of the partial autocorrelation function as an identification tool. We decided to omit these topics from our expository paper, but the interested reader will find them explained in Box and Jenkins (ref. [1] of the paper).

The third case study involves forecasting common stock prices. Mr. Reynolds adds yet another piece of evidence that, for all practical purposes, stock prices behave as a random walk. There does not appear to be enough structure in past observations to permit predictions that would lead to consistent gains.

Our purpose in writing this paper was to bring an exposition of time series analysis to actuaries. The two discussions have served to advance this project.