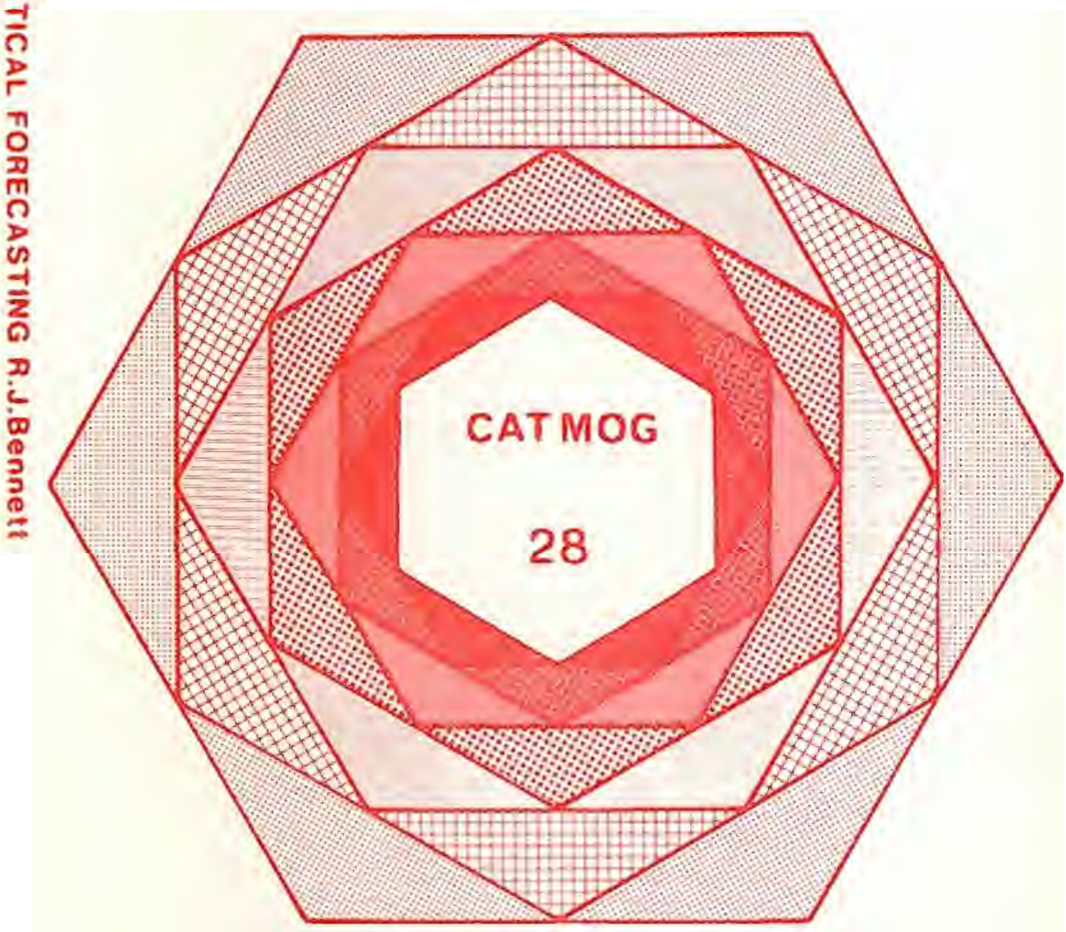
STATISTICAL FORECASTING

R.J.Bennett



STATISTICAL FORECASTING R.J.Bennett

ISBN 0 86094 C64 0
ISSN 0306 – 6142

CATMOG

(Concepts and Techniques in Modern Geography)

CATMOG has been created to fill a teaching need in the field of quantitative methods in undergraduate geography courses. These texts are admirable guides for the teachers, yet cheap enough for student purchase as the basis of classwork. Each book is written by an author currently working with the technique or concept he describes.

1. An introduction to Markov chain analysis - L. Collins
2. Distance decay in spatial interactions - P.J. Taylor
3. Understanding canonical correlation analysis - D. Clark
4. Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5. An introduction to trend surface analysis - D. Unwin
6. Classification in geography - R.J. Johnston
7. An introduction to factor analytical techniques - J.B.Goddard & A.Kirby
8. Principal components analysis - S. Daultrey
9. Causal inferences from dichotomous variables - N. Davidson
10. Introduction to the use of logit models in geography - N. Wrigley
11. Linear programming: elementary geographical applications of the transportation problem - A. Hay
12. An introduction to quadrat analysis - R.W. Thomas
13. An introduction to time-geography - N.J. Thrift
14. An introduction to graph theoretical methods in geography - K.J. Tinkler
15. Linear regression in geography - R. Ferguson
16. Probability surface mapping. An introduction with examples and Fortran programs - N. Wrigley
17. Sampling methods for geographical research - C. Dixon & B. Leach
18. Questionnaires and interviews in geographical research - C. Dixon & B. Leach
19. Analysis of frequency distributions - V. Gardiner & G. Gardiner
20. Analysis of covariance and comparison of regression lines - J. Silk
21. An introduction to the use of simultaneous-equation regression analysis in geography - D. Todd
22. Transfer function modelling: relationship between time series variables - Pong-wai Lai
23. Stochastic processes in one-dimensional series: an introduction - K.S. Richards
24. Linear programming: the Simplex method with geographical applications - J. E. Killen

Continued on inside back cover

CONCEPTS AND TECHNIQUES IN MODERN GEOGRAPHY No. 28

STATISTICAL FORECASTING

by

R.J. Bennett

(University of Cambridge)

CONTENTS

Notation (Major terms only).

X_t	independent variable at time t
Y_t, Y_i	dependent variable to be forecast at time t or time i
$X_{t,1}$	independent variable 1 at time t
Y_{t+k}	forecast variable at time t+k
$\hat{Y}_t$	estimate of forecast variable at time t
var (Y_t)	variance of Y variable
prob (Y_t)	probability of variable Y occuring with a given value
e_t, e_i	errors in forecasting at times t or i
$\hat{e}_t$	estimate of forecasting error at time t
$\underline{X}_t$	matrix or vector with values of X_t
$\underline{K}_t$	Kalman gain matrix
$\underline{P}_t$	covariance matrix between independent variables
I	information theoretic criterion
a,b,c,d,	coefficients or parameters in model
b_{ji}	coefficient for lag j related to independent variable i
α	weight used in exponential data weighting
Δ	difference operator
Δ^2	second order difference operator
Δ^s	seasonal difference operator
σ_e	standard deviation of errors
σ_Y	standard deviation of forecast variable
μ	mean
μ_j	mean for season j
$\underline{\theta}_t$	vector of parameters or coefficients at time t
t	the subscript for time
^	superscript used to denote estimate in all cases

1 INTRODUCTION

(i) Motivation for forecasting

Forecasting derives from a wish to foresee the future course of events. As such, forecasting is something we all undertake all the time, from the level of making an appointment in our diaries, listening to weather forecasts, or the timetable of our lecture courses, to the level of choosing a holiday or even a career. In each case we try to anticipate where we and others will be at some future time. Such anticipation is based upon a range of assumptions. Depending upon the validity of these assumptions we will place a considerable degree of confidence in our forecast or alternatively our anticipation may be so uncertain that we place very little confidence in it. Writing in our diaries an appointment or lecture timetable represents the former; choosing a career, perhaps, represents the latter. Making a forecast, then, involves the interplay of two sets of components: first, the assumptions from which our forecast derives: these represent in some sense a model of the way in which events occur; second, and deriving from the form of our forecasting assumptions, the certainty or level of confidence about our resulting anticipations: these represent forecasting confidence intervals.

The pre-requisities assumed in this monograph for understanding these two components are an elementary understanding of regression analysis and statistical hypothesis testing. Hence, this CATMOG should be read after an elementary introductory statistics course (Ferguson, CATMOG 15, should also be helpful).

In geographical forecasting two things should be made clear at the outset. First, it is normal to distinguish geographical prediction from geographical forecasting. Prediction is usually taken to mean the writing of a single, one-off scenario for the future. Forecasting, on the other hand, is usually reserved as a term to denote anticipations of the future which are held with less confidence, are based on an explicit model, and are accepted as largely a statistical exercise undertaken on a series of occasions. The recurrent nature of forecasting is important since this leads us, in assessing the accuracy of predictions, to evaluate not the single event, but the sum of events: hence to assess the sum of the forecasting errors, and the variance of the forecasting errors.

A second feature of geographical forecasting especially distinguishes it from other disciplines. This is the degree to which spatial components enter into predictions. Whereas the economist frequently uses one variable or event as an indicator to anticipate a change in another variable or events, the geographer often applies his knowledge of events in one region to anticipate what will happen in another region. The economist's model of forecasting is usually termed that of using a leading indicator. The geographer, in addition to constructing forecasts based on leading indicators, is frequently concerned with determining which are leading regions.

(ii) Forecasts and Planning

Forecasting is seldom a purely academic exercise. In fact is is usually undertaken with a particular policy or business goal in mind. Knowledge of the future course of events confers value, power, and in some cases, survival. A prediction of what will happen naturally elicits the question of what should we do? Hence forecasting is naturally linked with questions of public policy and control. Forecasting highlights problem areas and exposes the need for a policy response, it suggests the ranges and limitations of actions, and it gives aid in the formulation and achievement of social and economic goals. In particular it allows the simulation of policy alternatives. Thus for many writers, forecasting is a major element in the monitoring and planning process (see for example, Bennett and Chorley, 1978, chapter 6).

(iii) Methods of forecasting

Forecasting can be undertaken in a great variety of ways ranging from pure intuition to the 1,000-equation mathematical models favoured by some econometricians. Over this range, four main categories can be distinguished.

(a) Behavioural or causal forecasting, in which the mechanisms are sought which cause changes in the behaviour of events. Good understanding of models, of behavioural relationships allows good forecasts of changes. Hence the forecasts are as good as the models, but the major aim is to obtain forecasts which can be explained.

(b) Curve fitting and extrapolation, in which various forms of curve fitting are used e.g. linear, polynomial, Fourier and other mathematical functions. In each case the pattern of variation of relevant variables is examined and a mathematical function is fitted to the systematic component of that variation. Hence the aim of these forecasts is a good mathematical fit.

(c) Survey methods. In socio-economic systems where the course of events is determined by the actions of people, an obvious way of finding out what changes are likely to occur is to ask the people concerned. Good examples of such forecasts are consumer survey and opinion polls. More refined methods are used in government forecasting of the economy with consultations with employers and unions.

(d) Judgemental methods. Where survey methods cannot be adopted, or in systems where the interactions of large numbers of individuals make forecasting based on surveys very difficult, judgemental methods provide an alternative. These methods assemble the subjective assessments of groups of experts, the Delphi method being a common example. This method is particularly powerful in that it allows the incorporation of subjective elements which can never be included in mathematical models.

Each of these methods is interrelated, and rarely will a forecasting study involve only one technique. In this guide we shall be mainly concerned with statistical forecasting using behavioural and extrapolative models of categories (a) and (b) above.

(iv) Statistical vs. other forecasting methods

Many discussions of forecasting involve an assessment of the relative merits of statistical as opposed to more intuitive survey and judgemental methods. To some extent this amounts to assessing how far forecasting is an art or a science. The methods (c) and (d) are usually simpler and more easily understood, whereas statistical forecasting using methods (a) and (b) is usually felt to be more abstract and complex, it lends a false sense of certainty, and it generates remoteness of the forecaster from the assumptions involved. However, against these disadvantages, statistical methods do possess an important number of advantages which suggest that, where possible, they should be used. Five of the most important of these advantages can be summarised as follows:

(a) All forecasts involve assumptions, and statistical forecasts lay bare the assumptions required more clearly than intuitive methods for which it is often not at all clear precisely what assumptions are involved.

(b) Statistical models identified and fitted to the past behaviour of events provide important information of system behaviour and to ignore this information would be foolhardy.

(c) It is known that statistical models, and hence the forecasts derived from them, are approximations to system behaviour of the previous sample period. With more intuitive methods, because they are in a very real sense arbitrary, it is not possible even to know if they are correct over the previous period.

(d) Following from previous property, statistical forecasts allow the calculation of forecasting confidence intervals. These permit the range and probability of outcomes to be assessed. More intuitive methods allow, at best, comparisons between ranked alternatives deriving from different expert opinions or surveys and there is no clear way in which to rank these alternatives on a scale of confidence.

(e) Statistical models can be easily automated, readily updated, can take account of a large number of factors, and can incorporate complex models of system structure. In particular such models can often allow the treatment of what has been sometimes termed 'counter-intuitive' behaviour; that is, behaviour which would not be predictable on intuitive grounds because of the complexity of system interrelationships.

(v) Statistical forecasting vs. statistical estimation

It is highly desirable that a model used for forecasting is one which also incorporates understood behavioural relationships. But in many circumstances it may be sufficient to have merely a good forecasting equation. In this latter case, many of the criteria which assume such importance in statistical inference become unimportant (see Ferguson, CATMOG 15). In particular it is useful to note that three major assumptions of least squares are of little importance in forecasting, unless statistical inference is also required. -First, multicollinearity in regression analysis with a high degree of correlation between the independent variables leads to unstable and uncertain estimates of the regression coefficients (see Ferguson, CATMOG 15, p.28). However, multicollinear models are good forecasting devices, and are usually better (in terms of the error variance of the forecasts) than models in which the multicollinearity has been removed, or the number of independent variables reduced. A second major problem in regression analysis is autocorrelation of errors. For large samples the regression coefficients have large variance, whilst for small samples they become very unstable. If the autocorrelation pattern is controlled by moving-average terms, then the estimates of the coefficients will also be biased. For forecasting alone, the most important effect of autocorrelation of errors in biasing estimates of the t and F statistics is not important, and indeed many forecasting models with autocorrelated residuals should have lower error variance. However, forecasting confidence intervals can no longer be reliably constructed. Apart from the effects of residual autocorrelation, other factors may also lead to biased estimates of the regression coefficients. Most important are errors in measurement of the independent variables, specification errors, correlation between the independent variables and the residuals, non-normal residuals, and non-linearity. However, for forecasting, most of the causes of bias have no effect on forecasting accuracy, and biased models usually provide better forecasts.

Despite the convenience presented by forecasting models the problems discussed above do give rise to difficulty when statistical inference or calculation of forecasting confidence levels is also involved. However, it is important to bear in mind the distinction between the aims of forecasting and inference, since when only for former is required, many of the complexities of complicated estimation techniques can be avoided. In fact, for forecasting purposes alone, estimates of models deriving from least-squares regression (OLS) are optimal in the sense that they are minimum mean squares error (MMSE). This property of least-squares estimates discussed further in section IV.

II THE DEVELOPMENT OF FORECASTS IN THE SIMPLEST CASE (CONSTANT MEAN MODEL)

The development of forecasting equations for all possible statistical models is clearly beyond the scope of this short CATMOG. Hence this discussion concentrates attention on forecasting in the simplest case of a constant mean model. For more complex models good summaries are given by Box and Jenkins (19/0), Chatfield (1975), Anderson (1976), Gilchrist (1976), Makridakis and Wheelwright (1979) and Bennett (1979). In addition, a number of papers in the Journal or the Royal Statistical Society make very useful reading because of the detailed discussion which they include which emphasises the difference in viewpoints of analysts on the relative efficacy of different methods. The papers in that journal by Coen, Gomme and Kendall (1971), Box and Newbold (1971), dray (1971), Kendall (1971), and Chatfield and Prothero (1973) are particularly commended.

In the discussion which follows, the detailed derivation of the forecasting equations is given for the simplest model, that based upon assuming that the time series in question is characterised by a constant mean. Then, in the following pages, the discussion of the more complex forecasting models is more brief, emphasising the equations used, rather than the details of the derivation which can be followed in the references given.

The constant mean model. This is an extrapolative model falling into the second category of model given in section I (iii). Forecasting with any statistical model requires five components to be defined: the model assumed, the forecasting equation, the method of estimation, the forecasting confidence interval, and the estimate of the error variance. Each of these is discussed in turn below for the constant mean model.

(i) Assumed model. The model equation upon which the forecasts are based incoporates the a priori assumptions regarding the nature and behaviour of events, the system, or the variables with which we are concerned. For the constant mean model, the assumed model equation is given by the equation:

$$Y_t = \mu + e_t \qquad (1)$$

Here μ is a constant mean and e_t is a value for the error in the forecast at time t. Y_t is the value, at time t, of the variable which it is sought to forecast.

(ii) Forecasting equation. The aim of forecasting is, given a set of past data $Y_1, Y_2, \ldots, Y_t$ up to the present time t, what is the value of the variable Y at some future instant? If we are interested in the forecast of Y at k instants ahead in the future (k-step ahead forecast), this will give the term Y_{t+k}. Then the forecasting formula can be given simply by changing subscripts in the model equation (1) to give a forecasting equation:

$$Y_{t+k} = \mu + e_{t+k} \qquad (2)$$

(iii) Forecasting estimates. The forecasting equation is as yet incomplete since we have no way of knowing the values of either μ or e_{t+k} on the right hand side of the equation. To derive our final forecasts we require an estimation equation which gives us the values of these two terms. For the constant mean model this can be very simply accomplished by following two steps. First, if the forecasting errors e_{t+k} can be assumed to be a random normal sequence, our best estimate of these is given by setting them equal to their mean, which is zero (i.e. $e_{t+k} = 0$). Second, the estimate of the constant mean μ can be derived, as in the calculation of any mean, by summing the past observations of Y_t and dividing by the number of observations, i.e.

$$\mu = (Y_1 + Y_2 + \ldots + Y_t)/t = (\sum_{i=1}^{t} Y_t)/t$$

Incorporating these two steps into the forecasting equation (2) gives the final equation used for deriving forecast estimates. This gives us:

$$\hat{Y}_{t+k} = (Y_1 + Y_2 + \ldots + Y_t)/t \qquad (3)$$

In this equation the forecast is denoted by $\hat{Y}_{t+k}$ with the hat being added to distinguish equation (3), which incorporates estimates of the mean, from equation (2) which assumes that the mean μ is known. Of course, other estimates of the constant mean can be developed, and methods using 'local' and recursive forms will be discussed in later sections of this paper.

(iv) Forecasting confidence interval. If we are interested merely in forecasting the level of a variable at some future instant, then equation (3) allows us to accomplish this task. However, one major advantage of statistical over intuitive forecasting techniques is that they allow us to develop a further set of equations which permit measurement of how close our forecast will be to the true value of the variable, with a given probability. The result will then be a measure which suggests that at a given level of statistical significance, we expect the forecast to be within plus or minus a given amount of the true value.

The generation of such forecasting confidence intervals requires three decisions to be made. First, it must be established that the forecast is an unbiased estimate. (Hence to establish confidence intervals, the normal assumptions of regression are important and must be satisfied, whereas for the pure forecasting equation (3) they need not be satisfied; see the discussion in section I (v)). Second, it must be established that the errors e_t have a zero mean. Finally, if these two previous assumptions are met, a decision must be made as to the probability range within which we wish to know the accuracy of the forecasts. This last decision requires a tradeoff between desired accuracy of the forecasts and the veracity of the model itself, that is, there are the usual problems of trading off between making a type I and type II error of inference.

Let us take as an example the requirement that the forecasting confidence intervals be at the 95% level. This gives the range within which our forecasts will be no more than 5% incorrect when averaged over a large number of cases. If we can assume that the forecasting errors are normally distributed, then from the standard normal distribution at the 95% significance level, the forecasting errors will cover the range of $\pm$ 1.96a. of the true value of the forecast variable, where σ_e is the standard deviation of the errors e_t. A two-tailed distribution is usually assumed since we are normally interested in both over- and under- estimation. The position should be clear from figure 1. For the standard normal distribution, graphed in the upper part of the figure, the area under the curve in the two tails corresponding to the 5% chance of error is given by standard tabulations as 1.96, i.e.

$$\text{Prob } (-1.96\sigma_e \leq e_t \leq 1.96\sigma_e) = 0.95 \qquad (4)$$

Relating this result for the standard normal distribution to the scale of measurement of our example requires rescaling by the standard deviation of the errors in the model, σ_e. Hence in the figure, the probability values for the standard normal distribution are rescaled by σ_e. When related to the forecast level of $\hat{Y}_t$, this permits the forecasting confidence interval to be calculated as shown in the lower part of the figure. The forecast $\hat{Y}_t$ refers to the centre of the distribution, and we expect the real value Y_t to be within $\pm$ 1.96σ_e of occasions, on average. Hence the forecasting confidence level for $\hat{Y}_t$ is given at the 95% significance level by using equation (3) as :

$$\text{Prob } (\hat{Y}_t - 1.96\sigma_e < Y_t < \hat{Y}_t + 1.96\sigma_e) = 0.95$$

In evaluating this equation, it remains to obtain an estimate of the standard deviation of the errors σ_e.

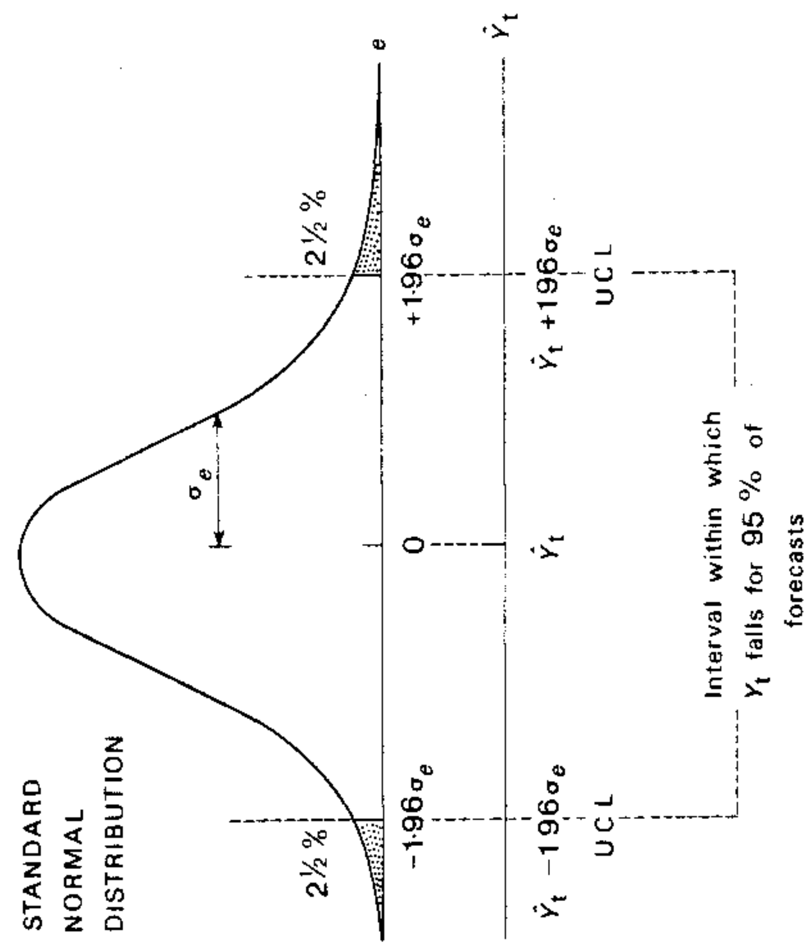


Fig. 1. Confidence interval for evaluation of forecasts at a given probability level

(v) Estimate of the error variance. For the constant mean model an equation which can be used to provide estimates of the variance of the errors is very easy to obtain. However, at first reading the reader may wish to omit this derivation and move directly to the result given in equation (6). The estimate of the forecast is given by equation (3) as:

$$\hat{Y}_t = \sum_{i=1}^{t} Y_i/t$$

or, substituting the forecasting model (1) we have:

$$\hat{Y}_t = \sum_{i=1}^{t} (\mu + e_i)/t$$

or,

$$\hat{Y}_t = \mu + (\sum_{i=1}^{t} e_i)/t$$

From this equation, if we take the variance of each term, represented by the term var(.), we obtain,

$$\text{var } (\hat{Y}_t) = \text{var } (\sum e_i/t)$$

$$= \sum \text{var } (e_i)/t^2$$

This uses the result that var(μ) = 0, and that the error terms are each independent. Now the term on the right hand side can be simplified by recognising that var(e_i) = σ_Y^2, where σ_Y^2 is the variance of the series to be forecast. Using this relation gives:

$$\text{var } (\hat{Y}_t) = \sigma_Y^2/t \qquad (5)$$

This result can now be used to give estimates of the forecasting errors. Remembering that

$$e_t = Y_t - \hat{Y}_t$$

again, find the variance on both sides:

$$\text{var } (e_t) = \text{var } (Y_t - \hat{Y}_t)$$

$$= \text{var } (Y_t) + \text{var } (\hat{Y}_t)$$

Using the result (5) for var ($\hat{Y}_t$), this gives

$$\sigma_e^2 = \sigma_Y^2 + \sigma_Y^2/t$$

or

$$\sigma_e^2 = \frac{t+1}{t} \sigma_Y^2 \qquad (6)$$

Hence, the estimate of the forecasting errors can be expressed in terms of the variance of the original data, but weighted by a term (t+1)/t. This weighting will always be smaller than one, and tends to the value of one as the sample size increases.

Using the expression for the error variance of the forecasts (6), the forecasting confidence intervals can now be obtained directly from equation (4) as:

$$\text{Prob } (\hat{Y}_t - 1.96\,\sigma_Y \sqrt{\frac{t+1}{t}} \leq Y_t \leq \hat{Y}_t + 1.96\,\sigma_Y \sqrt{\frac{t+1}{t}}) = 0.95 \qquad (7)$$

An example.

To obtain a clear grasp of the preceding discussion, it is useful to work through an example using actual data. the example used below will also be used throughout the ensuing discussion of more complex forecasting models. It uses a set of rainfall and runoff measurements for a small drainage catchment in Kenya. This catchment, the Lagan River, has been discussed by Blackie (1972) and the data have been previously analysed for forecasting purposes by Bennett (1979). The reader is directed to these two sources for a more detailed discussion. The data are reported in figure 2 and tabulated in Appendix 1 which allow the reader to follow and recalculate the results.

Using just the last 12 months of runoff data, we have the series of observations (60.7, 36.2, ... , 66.5, 86.5). Applying the estimating equation (3) we obtain:

$$\hat{Y}_{t+1} = (60.7 + 36.2 + \ldots + 66.5 + 86.5)/12$$

$$= 97.9$$

This forecast is shown in figure 3. Clearly the series of Lagan runoff data has a marked seasonality with a summer maximum and winter minumum, and thus the resulting forecast using a constant mean model will not be very accurate. The more complex seasonal models discussed in section III are required. However, for illustrative purposes we may go on and calculate confidence levels for the constant mean forecast. Again, using only the last 12 observations, we obtain an estimate of the variance of the runoff series of:

$$\sigma_Y^2 = 1906.0, \text{ and } \sigma_Y = 43.7$$

Entering this result into equation (7), we obtain the forecasting confidence intervals at the 95% significance level as:

$$\text{Prob } (\hat{Y}_t - 1.96\sigma_Y \sqrt{\frac{t+1}{t}} \leq Y_t \leq \hat{Y}_t + 1.96\sigma_Y \sqrt{\frac{t+1}{t}}) = 0.95$$

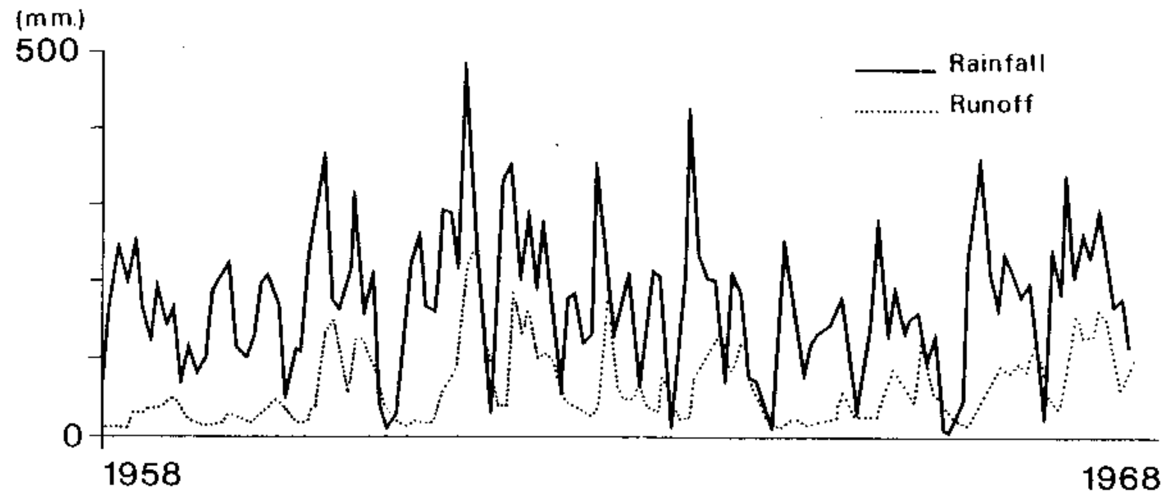


Fig. 2. Rainfall and runoff for Lagan catchment, monthly data. (Source: Blackie, 1972, table 3)

Fig. 3. Forecast of runoff for Lagan catchment together with 95% confidence interval

The resulting confidence interval of $\pm$ 1.96 (43.7) $\sqrt{(12+1)/12}$ = $\pm$ 89.15 is entered into figure 3, and shown by the shaded area. Within this area, using the results of the constant mean forecasting model alone, we expect to find the real value of Y_t with a 95% chance of being correct. As expected from a model with no seasonal factors, the forecasts are not particularly good, having wide confidence intervals indicating a high level of uncertainty in the forecasts produced.

III TYPES OF FORECASTING MODELS

Even within the class of statistical forecasting techniques to which this CATMOG is restricted, there is a wide range of available methods. These are grouped into six categories below and each is summarised in turn. In practice, a given forecast may require elements of two or more of these different forecasting models to be combined, e.g. as a constant mean, plus trend, plus seasonal model. In the discussion which follows it will not be possible to give the detail adopted for the constant mean model. In each case the equations for the forecasting estimates are given, but the equations for the forecasting confidence intervals are not derived, and are given instead in Appendix 2.

(i) Extrapolation (naive) models.

These models are very simple to apply and have been extensively used in market and business forecasting. The following references detail some examples of such applications: Fildes(1979), Brown (1962), Winters (1960), Theil and Wage (1964), and Theil (1966). The simplest extrapolation model is the no change model.

$$\hat{Y}_{t+1} = Y_t \qquad (8)$$

that is, there is no change from the present time period to the future at one step ahead e.g. streamflow now will be the same as when last measured. Two other simple models incorporate the effect of changes. The absolute change model is given by:

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) \qquad (9)$$

and the rate of change model is given by:

$$\hat{Y}_{t+1} = Y_t \left(\frac{Y_t}{Y_{t-1}} \right) \qquad (10)$$

Equation (9) states, for example, that streamflow will be the same as when measured plus a term to allow for changes observed over the last two measurements. Both equations (9) and (10) can be extended to (n+1) terms and an average taken over various periods of change. This gives the moving-average absolute change model:

$$\hat{Y}_{t+1} = Y_t + \frac{\sum_{k=0}^{n} (Y_{t-k} - Y_{t-k-1})}{n+1} \qquad (11)$$

and the moving-average rate of change model:

$$\hat{Y}_{t+1} = Y_t \left(\sum_{k=0}^{n} \frac{Y_{t-k}}{Y_{t-k-1}} \right) \frac{1}{n+1} \qquad (12)$$

Both of these models give equal weight to each period in the past. A more complicated model allows differential weighting of the contribution of each period of the past. One major example is the exponentially-weighted moving-average model:

$$\hat{Y}_{t+1} = \frac{1}{1+(1-\alpha)+(1-\alpha)^2+ \ldots +(1-\alpha)^{n-1}} \times \left(Y_t + (1-\alpha)Y_{t-1} + \ldots + (1-\alpha)^{n-1} Y_{t-n+1}\right) \qquad (13)$$

The weighting term α can be estimated by special estimation methods, or it can be chosen using the experience of the investigator. It is usually in the range $(0.1 \leq \alpha \leq 0.3)$.

The results of applying each of these forecasting models to the Lagan runoff series from September 1967 to December 1968 are given in Table 1. In this case forecasting has been applied to postdicting known outcomes. This enables us to see how good the forecasts have been. This results in the surprising finding that the no change model is the best of the simple predictors employed!

(ii) Trend and growth curve models.

Many temporal and spatial series exhibit trends of various forms, and these may be estimated and incorporated into forecasting equations. The simplest such model is a direct extension of the familiar regression equation :

$$\hat{Y}_{t+1} = a + bt + e_t \qquad (14)$$

This is a linear trend model which can be used for forecasting by changing the time origin as follows:

$$\hat{Y}_{t+1} = a + b(t+1)$$

This can be estimated by the normal least-squares equations with the independent variable being replaced by time values indexed from zero onwards, i.e. independent variable values will read as 0,1,2,3, ..., t, ...}

More complicated trend models can be used to describe complex curves of growth or decline. Simple extensions of the linear trend model are polynomial trend models:

TABLE 1. Extrapolation forecasts for the Lagan discharge data

Discharge in Lagan (Sept. 1967 - Dec. 1968)		No change one-step forecast model (8)	Absolute change model (9)	Rate of change model (10)	Two-term[a] moving-average rate of change model (12)	Two-term[b] exponential model (13)
September	86.9					
October	84.3	86.9				
November	80.5	84.3	81.7	81.8		
December	126.7	80.5	76.7	76.9	77.3	82.5
January	60.7	126.7	172.9	199.4	159.6	107.4
February	36.2	60.7	-5.3	29.1	62.2	82.3
March	35.5	36.2	11.7	21.6	19.4	56.1
April	77.4	35.5	34.8	34.8	27.9	39.3
May	162.3	77.4	119.3	168.8	122.3	59.5
June	130.0	162.3	247.2	340.3	346.5	119.9
July	133.7	130.0	247.7	104.1	187.8	131.7
August	160.5	133.7	137.4	137.5	122.3	136.8
September	139.4	160.5	187.3	192.7	178.9	148.5
October	85.6	139.4	118.3	121.1	144.3	144.6
November	66.5	85.6	31.8	52.6	63.3	111.7
December	86.5	66.5	47.4	51.7	45.9	82.4
Mean absolute forecasting error		29.8	49.9	51.1	52.2	32.3

Notes: a. model is given by:

$$\hat{Y}_{t+1} = Y_t\left(\frac{Y_{t-1}}{Y_{t-2}} + \frac{Y_{t-2}}{Y_{t-3}}\right)/3$$

b. model is given by:

$$\hat{Y}_{t+1} = (Y_t + 0.5Y_{t-1} + 0.25Y_{t-2})/(1 + 0.5 + 0.25)$$

quadratic $\hat{Y}_t = a + b_1 t + b_2 t^2$ (15)

cubic $\hat{Y}_t = a + b_1 t + b_2 t^2 + b_3 t^3$ (16)

etc.

These, and more complex, growth curves frequently used in forecasting are shown in figure 4. The Gompertz and logistic curves, for example, have frequently been used in UK traffic forecasts (see Tanner, 1978; Bennett, 1978), whilst species of the polynomial models have been applied to trend surface analysis of spatial distributions (see Haggett, 1964).

Although more complex, the growth curves shown in figure 4 can be estimated using normal least squares estimation equations by the simple device of linearizing the equations. The linearized version of each of the models is shown in figure 4, where the two devices of rearrangement and taking logs have been sufficient to reduce multiplicative models to additive ones. Using these equations, each of the models shown in figure 4 can be estimated using standard least-squares regression computer packages.

In other circumstances, very complex or inconsistent trend patterns may preclude description by mathematical functions such as those shown in figure 4. In these cases the following alternative procedures may also be considered:

(i) section the data and estimate different forecasting models for different periods of time,

(ii) construct models for the slope and intercept parameters which allow them to differ at different time periods. This requires the parameter change models discussed in section III(vi) of this CATMOG.

(iii) use nonlinear estimation techniques which can provide parameter estimates of series with complex trend functions which are not readily linearized (see Box and Jenkins, 1970; Bennett, 1979).

(iii) Regression models.

Regression models differ from the trend and extrapolation forecasting models discussed above in that knowledge of the behavioural structure is used to improve forecasting. Whereas extrapolation models involve assumptions as to trend or other elements, regression models develop optimal estimates (in some minimum error sense) of a mathematical model which best represents the underlying relationships. At best these models can be based on confirmed and well-understood relationships. At worst they may be mere curve-fitting exercises which add little to extrapolation models.

The simplest regression forecasting model is a straigthforward extension of the familiar linear regression equation, i.e.

$$Y_t = a + b X_t + e_t \quad (17)$$

In this equation X_t is an independent variable, Y_t is the variable to be forecast, and e_t is the error or residual sequence. The terms a and b are coefficients and the data are sampled over time instants t. This form

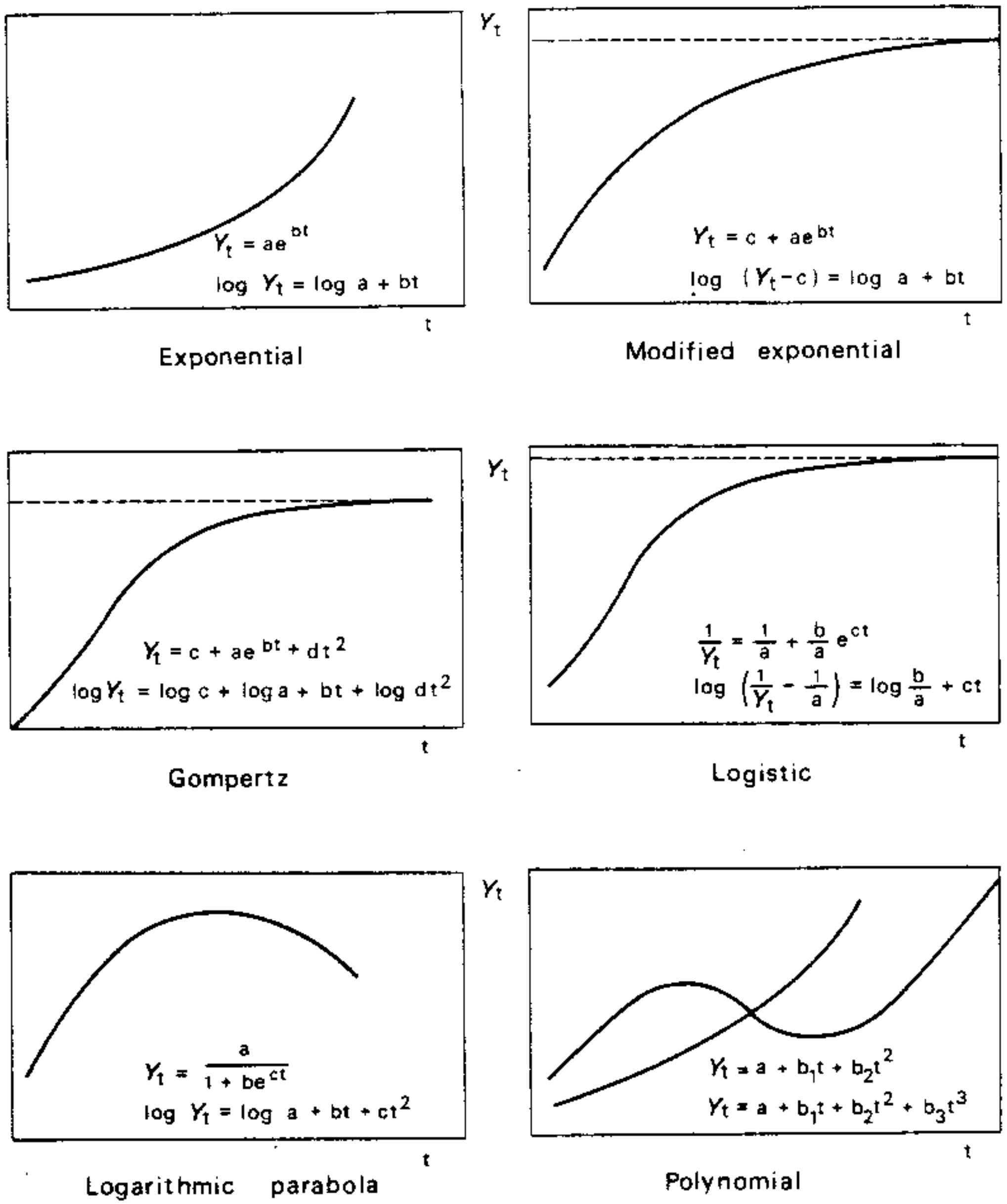


Fig. 4. Common trend and growth models used in forecasting. The upper equation in each case refers to the forecasting model, and the lower equation to the linearised versions after taking logs. The linearised version can then be estimated using least-squares regression. (note linearisation is not required for the polynomial models)

of linear regression can be recast as a forecasting equation by merely shifting the base of the time index t;

$$\hat{Y}_{t+k} = a + b\,X_{t+k} \tag{18}$$

Now a forecast at lead time k is given and the model can be estimated by the normal methods of least-squares regression. One major problem should be apparent, however. To forecast $\hat{Y}_{t+k}$ at k steps ahead, we require knowledge of X_{t+k} also at k instants ahead. Since this will usually require a further forecasting model to give values of $\hat{X}_{t+k}$, we may have not advanced in solving the forecasting problem. However, it is a frequent occurence that the independent variable X_t anticipates changes in Y_t; i.e. X_t is a leading indicator of Y_t. In this situation we may rewrite the forecasting equation (18) to give:

$$\hat{Y}_{t+k} = a + b\,X_t \tag{19}$$

Such leading indicators are common in practice: some stock prices are known to anticipate changes in the general level of the stock exchange indexes; changes in GNP anticipate changes in unemployment levels (Bray, 1971); changes in unemployment anticipate changes in migration in a region (Bennett, 1974); changes in rainfall into a catchment lead changes in runoff at its mouth; changes to the number of prey in an ecosystem anticipate changes in the number of predators; and so on. In many situations such lead-lag interrelationships can be represented in terms of systems as outputs that respond to inputs after some delay period. In other situations there may be no behavioural structure of input-output or stimulus-response which can be inferred, and the leading indicator model must be taken as merely a black-box forecasting device which 'works'. Clearly, the former circumstance is preferable to the latter because it allows forecasts to be meaningfully interpreted, more readily applied to planning, and which can be better justified.

The regression model can be extended to a whole set of lags of the leading indicator:

$$\hat{Y}_{t+k} = a + b_0 X_t + b_1 X_{t-1} + b_2 X_{t-2} + \ldots + b_q X_{t-q} \tag{20}$$

This is a multiple regression equation where the set of independent variables are in fact lags of the same independent variable. The normal least-squares equations can still be applied to estimating the coefficients. In this case there are q coefficients defining the maximum order of the equation (the maximum lag). The equation can also be extended to a set of lags and a set of independent variables, e.g.

$$\hat{Y}_{t+k} = a + b_{0,1} X_{t,1} + b_{1,1} X_{t-1,1} + \ldots + b_{q_1,1} X_{t-q_1,1} + b_{0,2} X_{t,2} + b_{1,2} X_{t-1,2} + \ldots + b_{q_2,2} X_{t-q_2,2} + e_t \tag{21}$$

Where $X_{t-j,i}$ is the ith independent variable and b_{ji} the coefficient at lag j for variable i. Models such as that given in equation (20) and (21) are discussed in more detail by Lai (CATMOG 22) who adopts the more general systems terminology of transfer function models to refer to them.

As an example of the application of regression forecasting models, consider again the Lagan runoff data. If rainfall is used as a leading indicator, the following regression model can be estimated:

$$\text{Runoff } Y_t = \underset{(0.034)}{0.207\, Y_{t-1}} + \underset{(0.035)}{0.199\, Y_{t-2}} + \underset{(0.011)}{0.075\, X_t} - \underset{(0.015)}{0.094\, X_{t-1}} \qquad (22)$$

This expresses runoff as a dependent variable on previous runoff levels and values of the rainfall as an input. The terms in brackets are the standard errors of the coefficients, and the detailed identification and estimation of this model using least-squares regression is discussed by Bennett (1979, pp 142-3, 235-6). Applying this model to the Lagan runoff series to give forecast values, we obtain the results shown in figure 5 for the forecast and confidence interval for the one-step ahead forecast.

(v) Stochastic models

Stochastic models arise in situations in which the error term and random components of the system play a more dominant part. There are two main classes of such models: autoregressive, and moving-average together with various 'mixed' models incorporating elements from both autoregresive and moving-average formats.

The general form of stochastic models has already been introduced in discussion of the extrapolation models. However, stochastic models differ from extrapolation models in using optimal estimation of the forecasting coefficients, whilst extrapolation models use ad hoc assumptions or external prior knowledge.

The various classes of stochastic models may be listed as follows:

(a) autoregressive (AR):

first order: $Y_t = a_1 Y_{t+1} + e_t$

second order: $Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + e_t$ (23)

....

p^{th} order: $Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + e_t$

(b) Moving-average (MA):

first order: $Y_t = e_t - c_1 e_{t-1}$

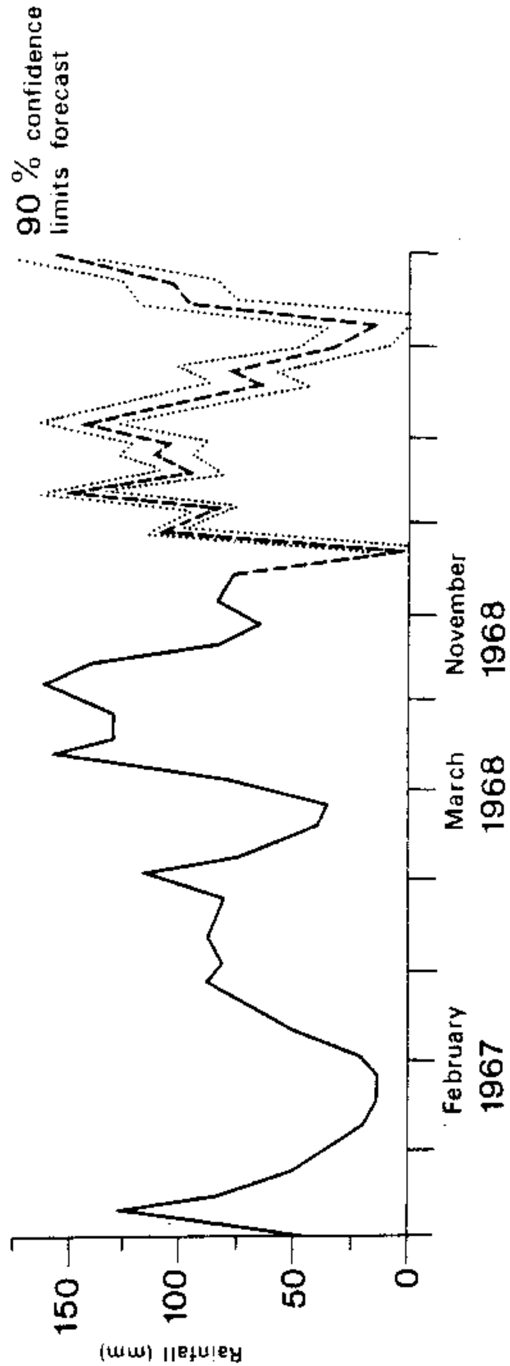


Fig. 5. Forecasts of Lagan runoff together with 90% confidence intervals using a stochastic transfer function model

second order: $Y_t = e_t - c_1 e_{t-1} - c_2 e_{t-2}$

....

v^{th} order: $Y_t = e_t - c_1 e_{t-1} - c_2 e_{t-2} - \ldots - c_v e_v e_{t-v}$ (24)

(c) Autoregressive moving-average (ARMA):

order p,v: $Y_t = a_1 Y_{t-1} + \ldots + a_p Y_{t-p}$

$+ e_t - c_1 e_{t-1} - \ldots - c_v e_{t-v}$ (25)

In these models the a_i and c_j terms are referred to respectively as the autoregressive and moving-average coefficients and the order p or v defines the number of lags involved.

Note that in each of the stochastic models there is no intercept term since it is assumed that the forecast variable Y_t has no trend (i.e. it is stationary), or that the trend has been removed prior to analysis, e.g. by trend and extrapolation models. A second family of forecasting models arises however in the special, but common, case in which the trend in the forecast variable is also stochastic, this gives rise to so-called integrated or cumulative processes. The simplest case of such an integrated model can be illustrated with no change extrapolation model (8). With a random error term added, this now becomes a random walk model:

$$Y_t = Y_{t-1} + e_t \quad (26)$$

Note that this can be rearranged to give:

$$Y_t - Y_{t-1} = e_t$$

or

$$\Delta Y_t = e_t$$

where

$$\Delta Y_t = (Y_t - Y_{t-1}) \quad (27)$$

It can be seen that this model can be represented by transforming the original data Y_t by taking the first differences defined as ΔY_t. Similarly, second differences yield:

$$\Delta^2 Y_t = e_t \quad (28)$$

or

$$(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = e_t$$

or

$$Y_t - 2Y_{t-1} + Y_{t-2} = e_t$$

Thus, when equation (26) represents a first order model, we obtain the equivalent second order model:

$$Y_t = 2Y_{t-1} - Y_{t-2} + e_t$$

Similar forms of differencing can be applied successively to give n^{th} order differences of a series, if this is desired. These can also be applied to the stochastic processes discussed above and this gives rise to a family of integrated autoregressive processes such as (26), to integrated moving-average models, and to the general family of autoregressive integrated moving-average models (ARIMA) to which Box and Jenkins (1970) devote primary attention.

The use of stochastic models for forecasting is very straightforward since simple shifting of the time origin transforms them from descriptive to predictive equations. Thus, the first order autoregressive and moving-average models become, respectively:

$$\hat{Y}_{t+1} = a_1 Y_t + e_{t+1} \quad (29)$$

and

$$\hat{Y}_{t+1} = e_{t+1} - c_1 e_t \quad (30)$$

In each case the e_{t+1} error term, which is unknown, can be replaced by setting it equal to its assumed mean value of zero. This yields the final forecasting equations as:

$$\hat{Y}_{t+1} = a_{1Y} Y_t \quad (31)$$

and

$$\hat{Y}_{t+1} = - c_1 e_t \quad (32)$$

A voluminous literature has developed around the two problems of determining the correct model (autoregressive, moving-average, ARMA, or ARIMA) to use in forecasting (the so-called' identification problem) and estimating the coefficients in each model. This literature is summarised by Anderson (1976), Box and Jenkins (1970), Richards (CATMOG 23) and Bennett (1979).

An example of stochastic forecasting can again be developed for the Lagan catchment, but for this particular problem a seasonal model is required and discussion of the resulting forecasting equations is left until this new species of model has been introduced in the next paragraphs.

(v) Seasonal models.

Many spatial and temporal series in both physical and economic problems contain marked cyclical patterns; e.g. unemployment, rainfall, insolation, etc. These are frequently referred to by the general title seasonal series even though in many cases seasonality is only one possible component together with other causes of cyclicity.

The simplest seasonal model is an adaptation of the constant mean model (1). This is termed the seasonal mean model (sometimes termed the Holt-Winters model; see Bennett, 1979). It expresses the forecast value at any instant as a function of the mean of the season (or other period), i.e.

$$Y_{tj} = \mu_j + e_{tj} \qquad (j=1,2, \ldots , r) \qquad (33)$$

Here j is used as an index of the season (or period) concerned, and there are r such seasons. In other respects this model is identical with equation (1). This equation can be thought of as giving a set of r separate forecasting models, one for each season, each with a different mean. By analogy with equation (3) estimates are given by:

$$\hat{\mu}_j = (r \sum_{i=1}^{t} Y_{ij})/t \qquad (34)$$

where r represents the length of seasonal period relative to the original data.

This simple form of seasonal model can be combined with extrapolation, trend, stochastic and each of the other forecasting models discussed in this handbook. The special case of trend-plus-seasonal model, which is frequently used in business forecasting, is often referred to as the Theil-Wage model (Bennett, 1979). Table 2 and figure 6 give examples of the use of three of the simplest seasonal models applied to the Lagan catchment discharge series:

(a) a previous season forecast

$$\hat{Y}_{t+1} = Y_{t+1-12} \qquad (35)$$

where 12 is used to refer to the annual cycle in these monthly data.

(b) a previous season plus absolute data change model

$$\hat{Y}_{t+1} = Y_{t+1-12} + (Y_{t+1-12} - Y_{t+1-13}) \qquad (36)$$

where an annual cycle is combined with first differences.

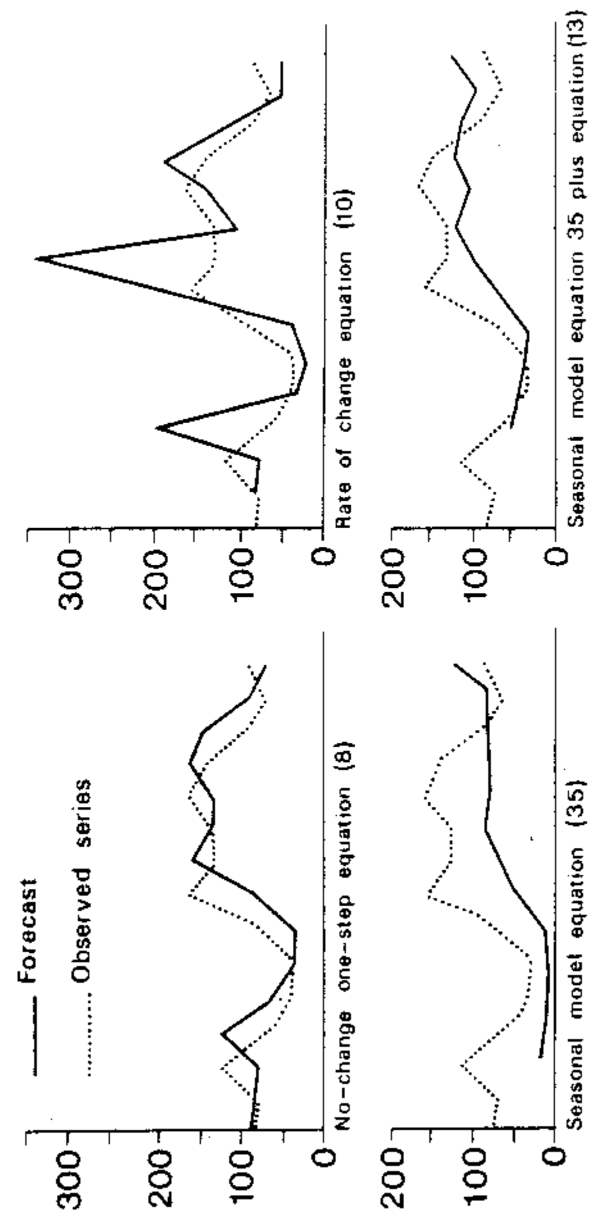


Fig. 6. Forecasts of Lagan runoff using various forecasting models

TABLE 2. Comparison of three seasonal forecasting models for the Lagan discharge series

Discharge in Lagan (Dec 1966 to Dec 1968)		Seasonal model (35)	Seasonal model (36)	Seasonal model (37)
December	27.6			
January	17.9			
February	12.0			
March	13.3			
April	18.4			
May	47.7			
June	64.2			
July	87.1			
August	82.3			
September	86.9			
October	84.3			
November	80.5			
December	126.7	27.6		
January	60.7	17.9	8.2	57.8
February	36.2	12.0	6.1	44.2
March	35.5	13.3	14.6	35.4
April	77.4	18.4	23.5	31.4
May	162.3	47.7	77.0	68.3
June	130.0	64.2	80.7	100.3
July	133.7	87.1	110.0	120.9
August	160.5	82.3	77.5	107.2
September	139.4	86.9	91.5	120.0
October	85.6	84.3	81.7	113.2
November	66.5	80.5	76.7	94.2
December	86.5	126.7	172.9	127.7
Mean absolute forecasting		50.8	35.6	29.2

(c) A previous season plus exponential data weighting forecasting

$$\hat{Y}_{t+1} = (Y_{t+1-12} + 0.5Y_{t+1-1} + 0.25Y_{t+1-2})/(1+0.5+0.25) \qquad (37)$$

The results of applying these various seasonal models demonstrate that neither of the purely seasonal models (35) and (36) produces very adequate forecasts, although both are better than some of the models shown in Table 1. The errors in these seasonal models are due to the large variation in cyclic amplitude from one year to the next (refer to figure 1.). This suggests that a combined

forecasting model might be more appropriate. Using a seasonal model to forecast from one year to the next plus exponential data weighting from recent months, model (37) gives the lowest error in Table 2, and it is likely that further refinements will produce further improvements.

There are two other important seasonal models apart from those introduced above. The first of these, the Fourier forecasting model is a form of deterministic mathematical function, similar to the trend models discussed previously, but giving a recurrent, cyclical structure. The simplest way of introducing this model is to describe it as a special mean model with estimates of the seasonal mean defined as:

$$\mu_j = 1 + \sum_{k=1}^{F} \left[a_k \cos \frac{2\pi kj}{F} + b_k \sin \frac{2\pi kj}{F} \right] Y_k \qquad (38)$$

where F is equal to one half of the length of the series seasonal periodicity, e.g. 6 for monthly data and an annual cycle. The seasonal mean for r seasons is described as a function of the sum of the sine and cosine contributions at various frequencies and a_k and b_k are coefficients, akin to regression coefficients, expressing the weight of each contribution to the overall sum. This model can be estimated by least-squares regression and there are various packages available for this task.

The stochastic seasonal forecasting model is a second important more complex seasonal model. this is a direct development from the stochastic trend model of integrated processes (27). They can be estimated by seasonal differencing, e.g.

$$\Delta^s Y_t = Y_t - Y_{t-s} \qquad (39)$$

where s is equal to a lag corresponding to the period of the seasonal cycle present in the original data. Using this concept, each of the models discussed earlier can be turned into seasonal models. For example, the stochasti random walk model (26) can be combined with a stochastic seasonal model to give:

$$\Delta^s \Delta Y_t = e_t \qquad (40)$$

or

$$(Y_t - Y_{t-1}) - (Y_{t-s} - Y_{t-s-1}) = e_t$$

or

$$Y_t = Y_{t-1} + Y_{t-s} - Y_{t-s-1} + e_t$$

Stochastic seasonal models are often preferable to Fourier models because they allow seasonality which has randomly displaced periods and amplitudes to be modelled. Such cases arise frequently, as for example when seasonal peaks of rainfall occur earlier greater or smaller than the average run of events (displaced amplitudes).

An example of a stochastic seasonal model can be developed with the Lagan catchment data for the rainfall series alone. An appropriate model to forecast this series is:

$$\text{rainfall } \Delta\Delta^{12}Y_t = \underset{(0.135)}{-0.62\,Y_{t-1}} - \underset{(0.068)}{0.46\,Y_{t-2}} - \underset{(0.036)}{0.25\,Y_{t-3}} \qquad (41)$$

This is a third order autoregressive model which has been estimated from data which has been differenced both as simple and seasonal differences to remove both stochastic trend and stochastic seasonality. The process by which this model is estimated and identified is described by Bennett (1979, pp 157-8, 237). The results of applying this model to forecast the Lagan rainfall are shown in figure 7 together with the forecasting confidence intervals. When the differencing operator is expanded the final model is given by:

$$(1-B-B^{12}-B^{13})\ (1-0.62B-0.46B^2-0.25B^3)$$

$$= 1 - 1.62B + 0.12B^2 + 0.25B^3 + 0.25B^4 - B^{12}$$
$$+ 1.62B^{13} - 0.12B^{14} - 0.46B^{15}$$

or

$$Y_{t+1} = Y_t - 1.62Y_{t-1} + 0.12Y_{t-2} + 0.21Y_{t-3} + 0.25Y_{t-4}$$
$$- Y_{t-12} + 1.62Y_{t-13} - 0.12Y_{t-14} - 0.46Y_{t-15}$$

where in each case B is the lag operator used to define $Y_{t-1} = BY_t$

(vi) Parameter change models

The forecasting models discussed above have all assumed that the most appropriate model is constant over time and space. Many situations arise, however, in which the most appropriate model will differ at different points in time or space. To cope with such situations, various parameter change models have been developed. The use of such models is referred to as adaptive forecasting, or nonstationary parameter forecasting. The use of these models is a relatively new field and a complicated one, and it will be possible here only to give a short summary. The reader is referred to more detailed discussion in Gilchrist (1976, chapter 4) and Bennett (1979, chapters 5 and 8). Various approaches to adaptive forecasting are discussed in these two books, but since each can be expressed as a special case of recursive least-squares, only this approach will be discussed here.

Parameter change forecasting requires, in addition to a forecasting model for Y_t (which we will refer to subsequently as the system model), also a forecasting model for (a vector of parameters) $\underline{\theta}_t$ (which we will refer to as the parameter model). The discussion is again most simply initiated by use of the constant mean model (equation (1)). Estimates of the mean, which is the single parameter of this forecasting model, are derived at any point in time using equation (3):

$$\hat{Y}_{t+1} = (Y_1 + Y_2 + \ldots + Y_t)/t \qquad (42)$$

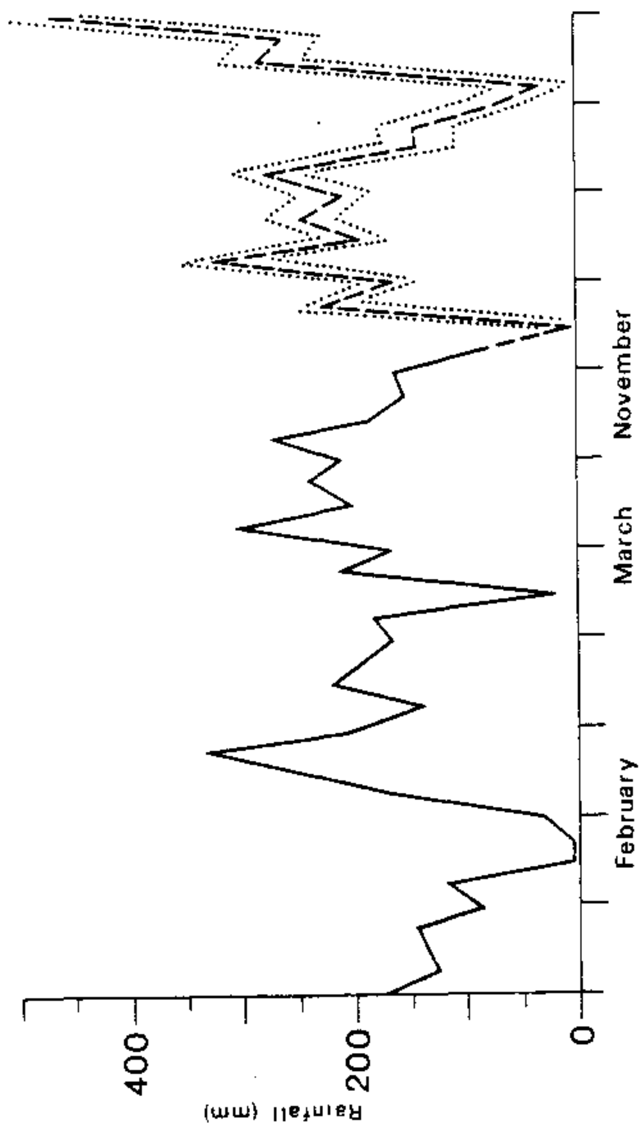


Fig. 7. Forecasts of Lagan rainfall together with confidence intervals using an autoregressive moving-average model

Now consider, at the next sampling instant t+2, how the forecast might be updated to include the new item of data Y_{t+1} which then becomes available. This updating rule can be obtained very simply from equation (42) as:

$$\hat{Y}_{t+2} = \left[(Y_1 + Y_2 + \ldots + Y_t) + Y_{t+1}\right] / (t+1)$$

This can be rearranged to give:

$$\hat{Y}_{t+2} = \left[t\,\hat{Y}_{t+1} + Y_{t+1}\right] / (t+1) \qquad (43)$$

This is an important relationship since it demonstrates a very simple way of generating forecasts recursively. The forecasting result from the previous step yields the old forecast $\hat{Y}_{t+1}$. When new information is available at the next instant t+1, this can be incorporated to give the new forecast $\hat{Y}_{t+2}$, but without the lengthy process of recalculating the mean: instead the new datum is added and the denominator (t+1) increased. In a very real sense, then, the previous forecast $\hat{Y}_{t+1}$ contains the previous calculation of the mean, and (for the purposes of this simple forecasting model) this value contains all the information from the data set (Y_o, Y_1, ... , Y_t) up to time t that we need to know. Hence equation (43) contains two elements: first, the term $t\hat{Y}_{t+1}$ which is a form of memory of previous data; and second, the term Y_{t+1} which is an update of this memory, adjusting it in the light of new information. Equation (43) can clearly be applied afresh to the next time instant yielding

$$\hat{Y}_{t+3} = (t\hat{Y}_{t+1} + Y_{t+2})/(t+2) \qquad (44)$$

and so on recursively.

The potential of this approach is somewhat limited for the constant mean model, but in more complex models, the recursive structure of equation (43) is especially significant. Two features should be noted. First, considerable economies are made in the amount of calculation required. Second, the recursive structure lends itself to methods which permit estimates of the changing model coefficients to be calculated. For the constant mean model, for example, equation (43) gives an adaptive forecast and an adaptive estimate of the mean as a single term.

In providing adaptive models for more complex forecasting equations it is easiest to proceed from the constant mean model to the regression model. For this and other parameter change models it is necessary to adopt a vector and matrix notation for the forecasting equations. Let us first assume that least-squares estimates of the regression parameters in our forecasting model can be employed (none of the assumptions of least-squares regression are undermined). The estimates of the regression parameters for the general case of a multiple regression forecasting model can then be written as follows:

$$\begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_h \end{bmatrix} = \begin{bmatrix} \sum_{t=o}^{T} X_t^1 X_t^1 & \cdots & \sum_{t=o}^{T} X_t^1 X_t^h \\ & \cdots & \\ \sum_{t=o}^{T} X_t^h X_t^1 & \cdots & \sum_{t=o}^{T} X_t^h X_t^h \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=o}^{T} X_t^1 Y_t \\ \vdots \\ \sum_{t=o}^{T} X_t^h Y_t \end{bmatrix} \qquad (45)$$

In this equation $\hat{\theta}_i$ are estimates of each regression parameter, $\sum_{t=o}^{T} X_t^i X_t^j$ is the unnormalised covariance of independent variable i with independent variable j (and lags thereof), and $\sum_{t=o}^{T} X_t^i Y_t$ is the covariance of independent variable i with the dependent variable Y_t. Note that although equation (45) is complex for the multiple regression case, it is identical in structure to the normal equations for least-squares regression, (see Ferguson, CATMOG 15). Equation (45) can be rewritten in a shorthand form as:

$$\hat{\underline{\theta}}_t = \left[\underline{X}_t' \underline{X}_t\right]^{-1} \underline{X}_t' \underline{Y}_t \qquad (46)$$

where the prime denotes the vector transpose, and where:

$$\underline{X}_t' = \begin{bmatrix} X_o^1 & X_1^1 & \cdots & X_t^1 \\ X_o^2 & X_1^2 & \cdots & X_t^2 \\ \cdots & & \cdots & \cdots \\ X_o^h & X_1^h & \cdots & X_t^h \end{bmatrix}$$

$$\underline{Y}_t = \begin{bmatrix} Y_o \\ Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_t \end{bmatrix}$$

There are assumed to be h parameters in all, and each vector and matrix has been explicitly subscripted by t to show its dependence upon the data available up to time t. Our problem is to find a recursive equivalent of the manipulation of the constant mean model which yielded equation (42). This recursive solution for equation (45) was first given by Plackett (1950). The derivation is long, but not complex, and can be followed in Bennett (1979, pp. 279-80). It results in what is normally termed the recursive least-squares estimator given by:

model equation

$$\hat{Y}_t = \underline{X}_t' \hat{\underline{\theta}}_t + e_t \qquad (47)$$

parameter equation

$$\hat{\underline{\theta}}_t = \hat{\underline{\theta}}_{t-1} - \underline{K}_t \left[\underline{X}_t' \hat{\underline{\theta}}_{t-1} - Y_t \right] \qquad (48)$$

covariance equation

$$\underline{P}_t = \underline{P}_{t-1} - \underline{K}_t \underline{X}_t' \underline{P}_{t-1} \qquad (49)$$

'gain' equation

$$\underline{K}_t = \underline{P}_{t-1} \underline{X}_t \left[1 - \underline{X}_t' \underline{P}_{t-1} \underline{X}_{t-1} \right]^{-1} \qquad (50)$$

and

$$\underline{P}_{t-1} = \left[\underline{X}_{t-1}' \underline{X}_{t-1} \right]^{-1} \qquad (51)$$

The P_t term is related to the standard errors of the regression coefficients, and $\underline{K}_t$ is usually termed the Kalman gain. This recursive least-squares solution has the important property that it is also a form of the Kalman filter (see Kalman, 1960; Willems, 1978; Bennett, 1979). This property does not concern us directly here, but is very significant in research developments, especially where policy models are concerned.

The operation of the parameter change forecasting model can be readily understood from the Lagan catchment model. The last three observations for the rainfall and runoff data (after second order and seasonal differencing) are:

rainfall:	-195.2	14.9	66.7
runoff:	-5.4	0.7	-5.7

The basic form of the regression model for this forecasting problem is given by equations (47) to (51), with coefficients as used in equation (22). This gives:

$$\hat{Y}_{t+1} = \underline{X}_{t+1}' \hat{\underline{\theta}}_{t+1} = \begin{bmatrix} 14.9 & 66.7 & 0.7 & -5.4 \end{bmatrix} \begin{bmatrix} 0.075 \\ -0.094 \\ 0.207 \\ 0.199 \end{bmatrix} = -6.082$$

Forecast values for each of the other terms in equations (48) to (51) can be derived as follows:

$$\hat{\underline{\theta}}_{t+1} = \begin{bmatrix} 0.075 \\ -0.094 \\ 0.207 \\ 0.199 \end{bmatrix} - \underline{K}_{t+1} \left\{ \begin{bmatrix} 14.9 \\ 66.7 \\ 0.7 \\ -5.4 \end{bmatrix}' \begin{bmatrix} 0.075 \\ -0.094 \\ 0.207 \\ 0.199 \end{bmatrix} + 5.7 \right\}$$

$$\underline{P}_{t+1} = \begin{bmatrix} 0.011 & 0 & 0 & 0 \\ 0 & 0.015 & 0 & 0 \\ 0 & 0 & 0.034 & 0 \\ 0 & 0 & 0 & 0.035 \end{bmatrix} - \underline{K}_{t+1} \begin{bmatrix} 14.9 \\ 66.7 \\ 0.7 \\ -5.4 \end{bmatrix}' \begin{bmatrix} 0.011 & 0 & 0 & 0 \\ 0 & 0.015 & 0 & 0 \\ 0 & 0 & 0.034 & 0 \\ 0 & 0 & 0 & 0.035 \end{bmatrix}$$

$$\underline{K}_{t+1} = \begin{bmatrix} 0.011 & 0 & 0 & 0 \\ 0 & 0.015 & 0 & 0 \\ 0 & 0 & 0.034 & 0 \\ 0 & 0 & 0 & 0.035 \end{bmatrix} \begin{bmatrix} 14.9 \\ 66.7 \\ 0.7 \\ -5.4 \end{bmatrix}$$

$$\times \left[1 + \begin{bmatrix} 14.9 \\ 66.7 \\ 0.7 \\ -5.4 \end{bmatrix}' \begin{bmatrix} 0.011 & 0 & 0 & 0 \\ 0 & 0.015 & 0 & 0 \\ 0 & 0 & 0.034 & 0 \\ 0 & 0 & 0 & 0.035 \end{bmatrix} \begin{bmatrix} 14.9 \\ 66.7 \\ 0.7 \\ -5.4 \end{bmatrix} \right] = \begin{bmatrix} 0.0023 \\ 0.0140 \\ 0.0003 \\ -0.0026 \end{bmatrix}$$

These give the following results for the parameters and the covariance matrix as:

$$\hat{\underline{\theta}}_{t+1} = \begin{bmatrix} 0.076 \\ -0.089 \\ 0.207 \\ 0.200 \end{bmatrix} ; \quad \underline{P}_{t+1} = \begin{bmatrix} 0.011 & 0 & 0 & 0 \\ 0 & 0.001 & 0 & 0 \\ 0 & 0 & 0.034 & 0 \\ 0 & 0 & 0 & 0.034 \end{bmatrix}$$

The inverse matrix $[1 + \underline{X}_t' \underline{P}_{t+1} \underline{X}_t]^{-1}$ for this simple case reduces to a scalar 0.014, and the error $e_{t+1} = 0.38$

The solution can now move to the next time period t+2 and the procedure can be repeated with the new parameters and covariance estimates, and so on.

A major problem with the direct application of equations (47) to (51) is that as time progresses the estimates weight the past data too highly, and give too little weight to new data. This occurs because P_t and Kt become very small. If it is required to keep the equations open to new information and hence better track any parameter shifts which occur, then old data must be weighted to a smaller extent. Various methods for changing the weighting of old and new data are available, and the three main species of techniques, ad hoc methods, window function and adaptive estimation, are discussed by Bennett (1979, chapter 5 and 8).

(vii) Spatial forecasting models

There is a large and growing literature on the application of forecasting techniques to geographical problems which are explicitly spatial Ind for such problems there have been four major approaches. First, the simplest approach is the so-called leading-indicator method by which changes in one region provide a time series to predict changes in the time series measured in another region. Examples of this type of model usually employ forms of the regression models discussed above. They were applied first by King et al. (1969) and Haggett (1971) to regional unemployment levels. The unemployment levels in one region were used to forecast the likely pattern of change in another region. There are now many examples available of studies which attempt to find lead-lag relationships between regions which are sufficiently stable for forecasting purposes and these are summarised by Cliff et al. (1975), and Haggett et al. (1977).

A second method is the so-called weights matrix approach in which the leading indicator which is used to forecast is the weighted sum of forecasting variables in other regions. This method has been developed by Cliff and Ord (1973, 1975), Cliff et al. (1975) and Haggett et al. (1977), and has been applied to forecasting the spatial diffusion of epidemics, unemployment cycles and other phenomena.

A third approach is the so-called spatial system structure in which weighted sums of regional forecasting variables are abandoned and instead the full set of interactions between all regions in the area of interest are used to forecast future changes in the region of interest. This approach, advocated by Bennett (1979), has been applied to environmental problems and to diffusion of relative economic changes.

A fourth approach is applied to forecasting where only spatial and no temporal data are available, e.g. where it is attempted to forecast from one part of a map to another, and there is no explicit time dimension present. Most commonly this arises with data interpolation (e.g. 'Kriging' methods- used in Geology). In addition, specialised Markov field models may be relevant to some empirical cases. Discussion of these methods is complex and beyond the scope of this CATMOG. The reader is referred to Besag (1974) and Bennett (1979, pp. 499-531).

IV EVALUATION OF FORECASTING MODELS

Evaluation of forecasts concerns the important stage of comparing the forecast with the actual value of the forecast variable when this becomes available. This can be accomplished in two ways:

(i) Ex Post: A model fitted to previous data is used to predict the past values, so-called post-diction. In some cases some of the past data may be saved to forecast from the other past data: model fitting is undertaken with the T-L data points, and then this model is used to forecast the last L data points.

(ii) Ex Ante: A model fitted to past data is used to predict the unknown future values, then the forecasts are compared with the actual realisation when with the passage of time the previously unknown future data become available.

Whichever forecasting situation is involved, a wide range of evaluation methods can be employed to assess the quality of the forecasts produced. The major categories are discussed below.

(a) Ad hoc loss function: the forecasts are compared against some criterion of the utility of the forecasts to the analyst, chosen with relation to the attributes of the problem in hand.

(b) Mean square prediction error: This is the most widely employed criterion and requires that the forecasts should have smaller variance, in the long run, than any other possible forecasts. This minimum mean square (MMSE) criterion measures the aggregate deviation of the forecasts from the actual realisation. Hence, an optimal forecasting method will be that which minimises the criterion:

$$\sum_{t=0}^{T} (Y_{t+k} - \hat{Y}_{t+k})^2 = \text{minimum} \tag{52}$$

The similarity of this criterion to that of the least-squares criterion used in regression estimation should be noted. In fact the estimation problem (to determine the regression coefficients of a model) and the forecasting problem (to determine the optimal forecasts using a given model) are identical in structure, and an optimal model (MMSE model) will yield optimal forecasts (MMSE forecasts). This is very important since it yields the result, due to Gauss, but developed by Wiener and Kolmogorov in the 1940's, that the minimum variance unbiased estimator of a regression model also yields minimum variance (MMSE) forecasts. Proof of this statement is available in Gilchrist, 1976 and Johnston, 1972.

(c) Inequality coefficients: This is a measure, resembling the correlation coefficient, of the forecasting error given by:

$$Q = \left[\frac{1}{T}\sum (Y_t - \hat{Y}_t)^2\right]^{\frac{1}{2}} \Big/ \left(\frac{1}{T}\sum Y_t\right)^{\frac{1}{2}} \left(\frac{1}{T}\sum \hat{Y}_t\right)^{\frac{1}{2}} \tag{53}$$

where the summations run over the forecasting lead time employed. The values of Q are bounded by zero (for perfect forecasting), and one (for forecasts which give no information).

(d) Symmetric inequality proportion: The equation for the forecasting errors (53) can be decomposed into three parts:

$$\frac{1}{T} \sum_{t=0}^{T} (Y_t - \hat{Y}_t)^2 = Q_1 + Q_2 + Q_3 \tag{54}$$

where.

$$Q_1 = \frac{T\,(\bar{Y} - \bar{\hat{Y}})^2}{\sum (Y_t - \hat{Y}_t)^2} \tag{55}$$

$$Q_2 = \frac{T\,(\sigma_y - \sigma_{\hat{y}})}{\sum (Y_t - \hat{Y}_t)^2} \tag{56}$$

$$Q_3 = \frac{2T\,(1-R)\,\sigma_y\,\sigma_{\hat{y}}}{\sum (Y_t - \hat{Y}_t)^2} \tag{57}$$

R is the correlation coefficient between Y_t and $\hat{Y}_t$ and $\bar{\hat{Y}}_t$ is the mean of the forecasts. The equations read, respectively, as the differences in the predicted means from the actual values, the variance of the forecasts and actual values, and the correlation of the forecasts and actual values. We wish Q_1 and Q_2 to be minimised and Q_3 to be maximised.

(e) Residual autocorrelation: Various tests of residual autocorrelation can be applied to the forecasting errors. If there are systematic errors detected by a significant value of the residual autocorrelation, then the model must either be modified, or a second model to reproduce the stochastic error correlation properties must be adjoined.

(f) Prediction-realisation diagram: This is a simple graphical plot suggested by Theil (1966) and shown in figure 8(a). A consistent correspondence between forecasts and actual realisations will show a direct linear relation in the figure. It is normal to rotate the figure through 45 , as shown in figure 8(b). After this manipulation, figure 8(c) shows the various categories of forecasting errors: overestimation, underestimation, and turning point errors.

(g) Information inaccuracy: This statistic derives from information theory and gives a measure of the information given by a new observation Y_t over its forecast value $\hat{Y}_t$. This is measured by the information statistic:

$$I(Y_t/\hat{Y}_t) = \sum_{t=0}^{T} \hat{Y}_t \log \frac{1}{\hat{Y}} - \sum_{t=0}^{T} Y_t \log \frac{1}{Y_t} \qquad (58)$$

This statistic ranges from zero, for perfect forecasting, to one with forecasts which give no information.

(h) Information gain: This is a measure of the information gained from new observations and is based on Shannon's measure of information gain:

$$I = \log \frac{1}{\hat{Y}} - \log \frac{1}{Y_t} \qquad (59)$$

V CONCLUSION: CHOICE OF FORECASTING MODELS

The preceding discussion reviews each forecasting model in detail without commenting on which is most appropriate in a particular context. From the experience of previous empirical studies, and of research using simulations of real problems, it is possible to give some general rules of guidance which should be followed in choosing a particular approach to forecasting.

In many forecasting situations, stochastic models have proved themselves preferable to exponential data weighting, seasonal models, arid Fourier models. Reid (1973) found such models preferable three out of four times. Much, however, depends upon the circumstances of the particular study. Stochastic models such as Box-Jenkins techniques are generally better for shorter forecasting lead times, and for more frequently observed data (i.e. monthly rather than annual) where more noise is present. Again, in parameter change situations, stochastic models, when combined with recursive least-squares estimation, are usually preferable to other methods

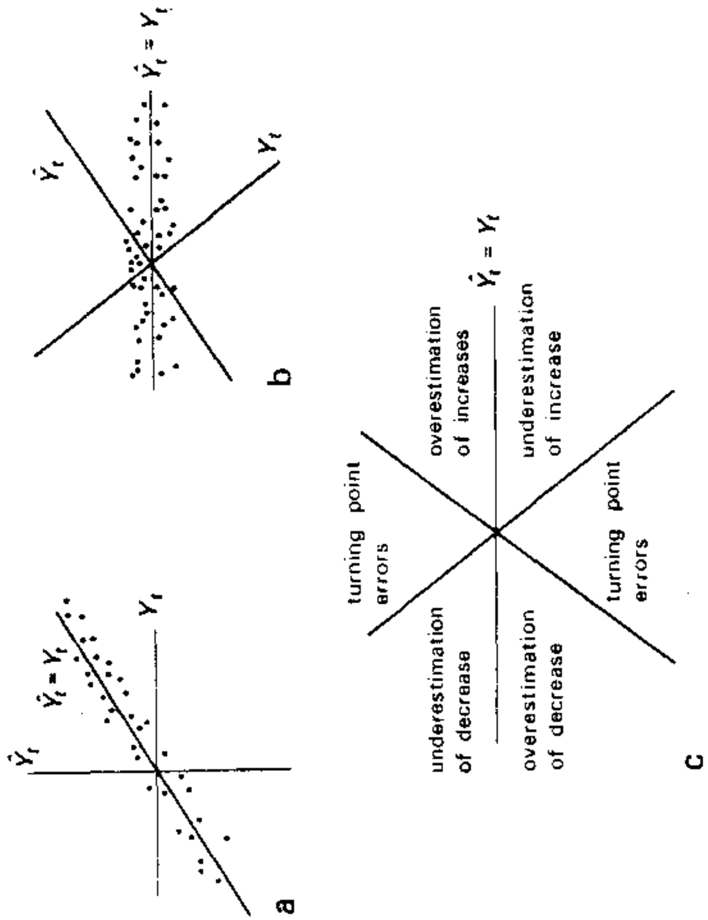


Fig. 8. Prediction-realisation diagram for forecasts (after Theil, 1966, figure 24)

(Kendall, 1973). In fact, there is usually the surprising result that stochastic models are usually better than regression models. This poses a considerable dilemma. Stochastic models are essentially black box forecasting devices for which it is usually difficult to give behavioural interpretations. Regression models, however, do give the opportunity for 'causal' or structural explanation, except where the leading indicators are merely used as black box forecasters. It may be, therefore, that in some circumstances a regression model should be preferred even if it gives larger mean square error than a stochastic model; the forecasts are more credible and interpretable.

To conclude, then, no hard and fast criteria for choice of forecasting model can be given. Much depends upon the experience and judgement of the analyst; as in all statistical methods this factor cannot be removed. Still more depends upon the nature of the data itself: in many cases the presence of trends, cycles and other factors indicates clearly the type of models which should be explored. In seeking the necessary_ experience the reader is strongly urged to experiment with different data sets, and is referred particularly to the guidance given in sources such as Chisholm and Whittaker (1971), Chatfield (1975), Gilchrist (1976), and Bennett (1979).

BIBLIOGRAPHY

GENERAL

Anderson, O.D., (1976), ***Time series analysis - The Box-Jenkins approach.,*** (London, Butterworths).

Bennett, R.J., (1979), ***Spatial time series: analysis - forecasting-control.*** (Pion, London).

Bennett, R.J. & Chorley, R.J., (1978), ***Environmental systems: philosophy, analysis and control.*** (Methuen, London).

Box, G.E.P. & Jenkins, G.M., (1970), ***Time series analysis, forecasting and control.*** (Holden-Day, San Francisco).

Chatfield, C., (1975), ***The analysis of time series: theory and practice.*** (Chapman & Hall, London).

Chisholm, R.K. & Whittaker, G.R., (1971), ***Forecasting methods.*** (R.D. Irwin, Homewood, Illinois).

Cliff, A.D., Haggett, P., Ord, J.K., Bassett, K. & Davies, R., (1975), ***Elements of spatial structure.*** (Cambridge University Press, Cambridge).

Cliff, A.D. & Ord, J.K., (1973). ***Spatial autocorrelation.*** (Pion, London).

Cliff, A.D. & Ord, J.K., (1975). Model building and the analysis of spatial pattern in human geography. ***Journal Royal Statistical Society.*** 8, 37, 297-384.

Ferguson, R.I., (undated). ***Linear regression in geography: CATMOG 15.*** (Geo Abstracts, Norwich).

Fildes, R., (1979), Quantitative forecasting - the state of the art: extrapolative models. *Journal Operations Research,* 30, 691-710.

Gilchrist, W., (1976). *Statistical forecasting.* (Wiley, London).

Haggett, P., Cliff, A.D. & Frey, A., (1977). *Locational analysis in human geography, second edition.* (Arnold, London).

Kendall, M.G., (1971). Review of Box, G.E.P. and Jenkins, G.M. ***Journal*** *Royal Statistical Society,* A, 134, 450-53.

Kendall, M.G., (1973). *Time series.* (Charles Griffen, London).

Makridakis, S. & Wheelright, S.C., (eds). (1979). Special issue on forecasting. *Management science,* vol. 25, 1175-1300.

THEORY

Besag, J.E., (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal Royal Statistical Society,* B, 36, 61-73.

Brown, R.G., (1962). *Smoothing, forecasting and prediction of discrete time series.* (Prentice-Hall, London).

Johnston, J., (1972). *Econometric methods, second edition.* (McGraw-Hill, London).

Kalman, R.E., (1960). A new approach to linear filtering and prediction problems. *Transactions of the American Society of Mechanical Engineers, Journal of Basic Engineering,* D, 82, 35-45.

Plackett, R.L., (1950). On some theorems in least squares. *Biometrika,* 37, 149-157.

Reid, D.J., (1973). A review of short term projection techniques. Paper presented at *Royal Statistical Society Conference on Practical Aspects of Forecasting.* (Imperial College London, Mimeo).

Theil, H., (1966). *Applied economic forecasting.* (North-Holland, Amsterdam).

Theil, H. & Wage, S., (1964). Some observations on adaptive forecasting. *Management Science,* 10, 198-206.

Willems, J.C., (1978). Recursive filtering. *Statistics Neerlandica,* 32, 1-40.

APPLICATIONS

Bennett, R.J., (1974). Process identification for time series modelling in urban and regional planning. *Regional Studies,* 8, 157-174.-

Bennett, R.J., (1978). Forecasting in urban and regional planning closed loops, the examples of road and air traffic forecasts. *Environment and Planning,* A, 10, 145-162.

Blackie, J.R., (1972). Hydrological effects of a change in land use from rain forest to tea plantation in Kenya. *Proceedings of Wellington Symposium. Results of Research on Representative and Experimental Basins, Volume 2.* (International Association of Scientific Hydrology, UNESCO, Paris), pp. 312-329.

Box, G.E.P. & Newbold, P., (1971). Some comments on a paper of Coen, Gomme and Kendall. *Journal Royal Statistical Society,* A, 134, 229-240.

Bray, J., (1971). Dynamic equations for economic forecasting with GDP-unemployment relation and the growth of GDP in the United Kingdom as an example. *Journal Royal Statistical Society,* A, 134, 167-227.

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

Coen, P.J., Gomme, E.D. & Kendall, M.G., (1969). Lagged relationships in economic forecasting. *Journal Royal Statistical Society,* A, 132, 133-152.

Duesenberry, J.S., Froom, G., Klein, L.R. & Kuh, E., (1965). *The Brookings Quarterly Econometric model of the United States.* (North-Holland, Amsterdam).

Haggett, P., (1971). Leads and lags in inter-regional systems: a study of the cyclic fluctuations in the south west economy. In *Spatial policy problems of the British economy,* eds. M. Chisholm and G. Manners. (Cambridge University Press, Cambridge).

Haggett, P., (1964). Regional and local components in the distribution of forested areas in Southeast Brazil: a multivariate approach. *Geographical Journal,* 130, 365-380.

King, L.J., Casetti, E. & Jeffrey, D., (1969). Economic impulse in a regional system of cities: a study of spatial interactions. *Regional Studies,* 3, 213-218.

Lai, P-W., (undated). *Transfer function modelling: relationship between time series variables: CATMOG 22.* (Geo Abstracts, Norwich).

Tanner, J.L., (1978). Long-term forecasting of vehicle ownership and road traffic. *Journal Royal Statistical Society,* 141, 14-62.

Winters, P.R., (1960). Forecasting sales by exponentially weighted moving averages. *Management Science,* 6, 324-342.

APPENDIX 1

Rainfall and runoff for the Lagan catchment, 1958-1968. (Source: Blackie, 1972, table 3.)

Month	Year 1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968
	Rainfall (mm)										
January	52.3	77.5	111.5	10.6	135.7	181.1	13.8	33.1	26.2	7.7	18.4
February	171.7	103.0	105.6	26.5	38.7	115.5	81.8	8.2	105.1	30.6	242.1
March	244.1	190.3	233.4	127.3	166.7	127.1	176.8	143.3	150.5	172.8	174.6
April	192.4	205.3	302.6	217.4	330.0	347.4	417.5	250.4	275.8	251.2	333.0
May	251.1	222.3	364.0	263.5	360.7	233.0	227.0	134.9	132.3	352.0	209.1
June	161.3	115.5	175.9	164.0	206.9	118.6	197.0	76.5	192.8	213.3	263.1
July	123.5	101.2	161.5	161.7	284.9	174.4	199.5	115.8	132.4	157.4	226.1
August	192.9	136.6	208.5	293.9	197.4	211.3	76.1	128.0	147.6	238.5	296.5
September	142.6	198.3	305.6	288.4	274.4	67.7	211.0	132.7	155.8	208.5	217.5
October	166.2	204.1	162.0	217.2	199.1	117.5	182.7	140.7	90.2	171.0	165.9
November	67.9	167.8	212.2	482.8	53.8	212.5	76.0	177.1	129.9	198.9	176.7
December	113.6	50.5	40.1	269.8	177.2	205.4	74.7	93.3	2.4	113.1	103.1
	Runoff (mm)										
January	9.6	13.1	23.0	26.1	133.7	37.3	45.7	23.2	21.7	17.9	60.7
February	10.7	8.8	14.4	12.8	52.5	28.2	24.3	12.6	20.2	12.0	36.2
March	12.6	11.4	17.7	10.7	34.2	25.5	20.6	13.7	22.7	13.3	35.5
April	9.4	13.5	47.0	12.0	37.7	44.5	68.1	17.7	46.4	18.4	77.4
May	31.8	29.2	132.4	20.0	183.5	179.4	87.3	20.0	84.2	47.7	162.3
June	29.1	22.8	145.4	14.7	131.4	78.8	97.5	13.3	72.1	64.2	130.0
July	35.4	18.2	88.9	18.9	164.3	44.0	127.2	13.8	59.0	87.1	133.7
August	32.3	18.2	57.4	62.6	103.6	47.0	108.5	18.4	50.9	82.3	160.5
September	41.6	29.4	128.7	76.6	107.8	60.4	80.2	19.9	137.1	86.9	139.4
October	46.7	36.6	120.7	130.2	98.1	37.2	129.9	22.2	73.1	84.3	85.6
November	28.7	50.1	95.5	220.0	58.4	29.1	73.8	57.4	41.3	80.5	66.5
December	19.2	33.3	57.5	245.3	38.7	83.5	37.3	36.2	27.6	126.7	86.5

APPENDIX 2

Equations for calculation of forecasting confidence intervals

This appendix gives the equations for calculation of the forecasting confidence intervals for each forecasting model. In each case only the estimate of the error variance is given; the confidence levels can then be derived by substituting these estimates into equation (4) with the chosen significance level.

No change model (8)

$$\sigma_e^2 = 2\ \sigma_Y^2$$

Absolute change model and rate of change model (9) and (10)

$$\sigma_e^2 = \frac{3}{2}\ \sigma_Y^2$$

Moving-average change and rate of change model (11) and (12)

$$\sigma_e^2 = \frac{n+1}{n}\ \sigma_Y^2$$

where n is the length of moving-average

Exponentially-weighted moving-average model (13)

$$\sigma_e^2 = \frac{\sum_{i=1}^{n+1} i\ \alpha^i}{\sum_{i=1}^{n} i\ \alpha^i}\ \sigma_Y^2$$

Trend models (14) to (16)

e.g. linear trend

$$\sigma_e^2 = \sigma_Y^2\ \left(\frac{1}{t} + \frac{(t+k)^2}{t^2} + 1 \right)$$

Regression models (17) to (21)

$$\sigma_Y^2 = \sigma_e^2 \sum_{j=g}^{k-1} V_j^2 + \sigma_e^2 \sum_{j=0}^{k-1} W_j^2$$

where V_j and W_j are derived from the coefficients of models describing the respective structure of autocorrelation in the X_t data and the relationship of Y_t to X_t. σ_e^2 is the residual variance of the model describing the X_t data, and when this sequence is uncorrelated, then the V_j terms equal zero. (See Box and Jenkins, 1970; Bennett, 1979, for further discussion).

Stochastic models (22) to (25)

The confidence intervals are given by:

$$\sigma_Y^2 = \left(1 + \sum_{j=1}^{k-1} W_j^2\right)^{\frac{1}{2}}\ \sigma_e$$

where

$$W_1 = a_1 - c_1$$

$$W_2 = a_1 W_1 + a_2 - c_2$$

...

$$W_j = a_j W_{j-i} + a_2 W_{j-2} + \ldots + a_{j-1} W_1 + a_j - c_j$$

and the a_j, c_j terms are the estimates of the model coefficients, with k the order of the model.

This series, Concepts and Techniques in Modern Geography is produced by the Study Group in Quantitative Methods, of the Institute of British Geographers.
For details of membership of the Study Group, write to the Institute of British Geographers, 1 Kensington Gore, London, S.W.7.
The series is published by Geo Abstracts, University of East Anglia, Norwich, NR4 7TJ, to whom all other enquiries should be addressed.