Industrial and business forecasting methods

A practical guide to exponential smoothing and curve fitting

COLIN D LEWIS

Industrial and business forecasting methods

A practical guide to exponential smoothing and curve fitting

COLIN D LEWIS

Professor of Operations Management
University of Aston

Butterworth Scientific
London Boston Sydney Wellington Durban Toronto

First published 1982

British Library Cataloguing in Publication Data

Lewis, C.D.
Industrial and business forecasting methods.
1. Forecasting
I. Title
003 CB158

ISBN 0-408-00559-9

Typeset by Scribe Design, Gillingham, Kent
Printed in England by The Camelot Press Ltd., Southampton

Preface

This book is aimed quite simply at those who possess time-series data which they wish to forecast.

Forecasts are interpreted as being the result of extrapolating the past into the future. It is assumed that forecasts are derived from an objective series of calculations or computations involving data, whereas subjective estimates of future values are termed 'predictions'. Forecasts are, therefore, unbiased estimates of future data values. As such they can and should be modified if subjective predictions confidently indicate that an unbiased estimate is unlikely to be an effective estimate of what will occur. For example, given that it is known that a major customer is going to suffer a strike next month, it is obviously quite pointless to assume that a forecast based on the last nine months of that customer's demand is likely to be at all accurate.

Since the early sixties, the proportion of industrial and business organizations using forecasting techniques has increased steadily, such that in 1977[1] it was established that forecasting techniques were used by 88% of the 500 largest industrial companies in the USA. Moreover, it was also established that no other single family of techniques was used as much as forecasting.

Along with the growth of the use of forecasting techniques, there has been a parallel growth in the variety of forecasting models on offer from the theorists. In spite of this increasingly wide range of available forecasting models, it has also become apparent that the choice of forecasting model actually used follow a typical *Pareto* relationship, such that 20% of the models available are used in 80% of practical forecasting applications and the remaining 80% of models are used in only 20% of applications. It is with the former which this book is exclusively concerned.

The two principal forecasting model types (and their derivatives) which feature in the vast majority of forecasting applications are exponential smoothing and regression (including curve fitting). The main reasons for the undoubted popularity of these methods are:

(i) their relative simplicity;
(ii) their economy in computational and storage terms;

(iii) the fact they are automatic, in model identification terms, such that forecasts can be produced without subjective intervention;
(iv) they have been extensively used for over twenty years.

The first section of this book deals with exponential smoothing methods which are generally recognized as falling into the *short-term* forecasting area. These forecasting techniques are associated mainly with data based on a time period of less than one year, i.e. demand per month, sales per quarter, etc.

The second section of the book deals with regression and curve fitting methods which are generally recognized as falling into the *medium-term* forecasting area and, although more usually associated with yearly-based data, can often be used for data based on shorter time periods if no seasonality is present.

As, in practice, most forecasting is done using a computer (or programmable calculator) the text of this book is supported by flow diagrams and worked tables to assist the reader in programming and debugging forecasting programs designed to suit his or her own situations.

C.D. Lewis
Solihull

1. LEDBETTER, W.N. and COX, J.F. (1977) 'Operations research in production management: an investigation of past and present utilization', *Production and Inventory Management*, **18**, 84

Contents

Section 1

Short-term forecasting (exponential smoothing methods)

Introduction

Short-term forecasting is generally concerned with:

(i) data associated with a time period of less than one year (i.e. calendar months, accounting periods in a year – usually 12 or 13, quarters, weeks, etc.);

(ii) situations where forecasts are associated with a particular item and for which forecasts are updated every time period;

(iii) situations where forecasts are required for a large number of items;

(iv) situations where the forecasts produced for a particular item or product and are used on a period to period basis to: (a) analyse demand and to assess appropriate inventory levels and production schedules, and (b) analyse sales to help assess cash flows and to ascertain marketing procedures.

It is apparent that for these types of application the most appropriate forecasting model (or series of models) needs to be:

(i) cheap to operate – in terms of implementation, routine updating and storage requirement costs;

(ii) flexible and hence able to offer a variety of different, but closely related, model types suitable for a wide variety of items and situations for which forecasts could be required;

(iii) largely automatic, such that a minimum of manual interruption is necessary;

(iv) well proven, and hence, readily available in both the literature and computer software.

The range of forecasting models based on the exponentially weighted average (collectively referred to as exponential smoothing methods), which were introduced in the early 1960's, has been shown to fulfil most of the above requirements.

The superiority of exponential smoothing methods over the traditional moving average concept is such that today most manufacturing organizations of any size use them and no industrial or business computer software is complete without them.

In Section 1, Chapter 1 develops the concept of the simple exponentially weighted average on which all smoothing models are based and

Chapter 2 is concerned with the more sophisticated variants of the basic model, which are necessary for growth and seasonal situations. Chapter 3 examines various measures of dispersion of forecasting errors and Chapter 4 the smoothing techniques developed specifically for the automatic monitoring of short-term forecasts. Chapter 5 examines the possibilities of adaptive forecasting models and Chapter 6, although not specifically based on exponential smoothing methods, examines the use of autocorrelation analysis in identifying data characteristics, a necessary pre-requisite for determining the appropriate type of forecasting model.

The forecasting methods included within this book must necessarily be expressed in mathematical terms. In order not to 'frighten off' the layman this Introduction will be used to present the ideas of symbolic representation of variables and the use of algebraic equations. Anyone able to understand these simple concepts will have no difficulty in following the material presented.

As an example, let us take a naive forecasting scheme which says 'Let the forecast for next month's expected demand be equal to the demand that occurred this month'. This could be written as:

$$\text{FORECAST}_{\text{next month}} \text{ EQUALS DEMAND}_{\text{this month}}$$

Using the first letters of *forecast* and *demand* as symbols representing those variables respectively, and introducing an equals (=) sign, this forecasting scheme could be rewritten as:

$$f_{\text{next month}} = d_{\text{this month}}$$

To complete the 'tidying up' of this equation we need to simplify and generalize the subscripts *next month* and *this month*. The simplest way of doing this is to work with months (or, in other situations, accounting periods, weeks or even days) as our period of time, and to refer all time to current (or present) time. Thus if we regard present time as t, then future time can be considered positive with respect to t (i.e. as $t + 1$, $t + 2$, etc.) and past time can be considered as negative (i.e. $t - 1$, $t - 2$, etc.). Having developed a method of subscripting to indicate time we can now rewrite our naive forecasting scheme as a generalized algebraic equation of the form:

$$f_{t+1} = d_t \tag{0.1}$$

Evidence of how much simpler is such symbolic representation, compared with a written statement, can be seen by considering a more practical forecasting scheme which predicts that 'Next month's expected demand will be equal to the arithmetic average of the last six months' demand or sales figures'. This lengthy statement can be written very briefly in equation form as:

$$f_{t+1} = {}^1/_6 \sum_{i=t}^{t-5} d_i \tag{0.2}$$

where the summation sign

$$\sum_{i=t}^{t-5}$$

means in this situation the sum of the values of d_i from i equals t up to and including $t-5$ (i.e. $t, t-1, t-2, t-3, t-4$ and $t-5$). Expanding the summation sign this equation would appear as:

$$f_{t+1} = {}^1/_6(d_t + d_{t-1} + d_{t-2} + d_{t-3} + d_{t-4} + d_{t-5}) \tag{0.3}$$

which would be an alternative form of equation (0.2).

This concludes the introduction to the concepts of symbolic representation and algebraic equations. The material presented in this book is deliberately chosen not to be more complicated than this.

Forecasting for stationary situations

A stationary situation is one in which, although observed values fluctuate from one time period to the next, the average value remains steady over a reasonably long period of time. To illustrate this, Fig. 1.1 shows a series of demand values plotted against time. It can be seen that the average value per month over the one-year period is about 100 items per month, and that this average figure is neither increasing nor declining significantly with time. This is a typical stationary situation,

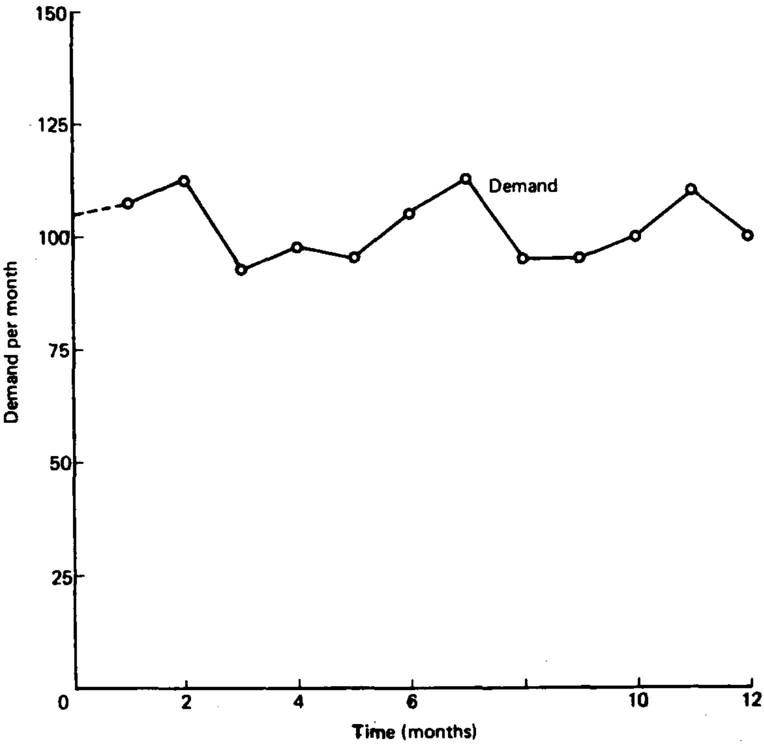

Fig. 1.1 Typical stationary situation

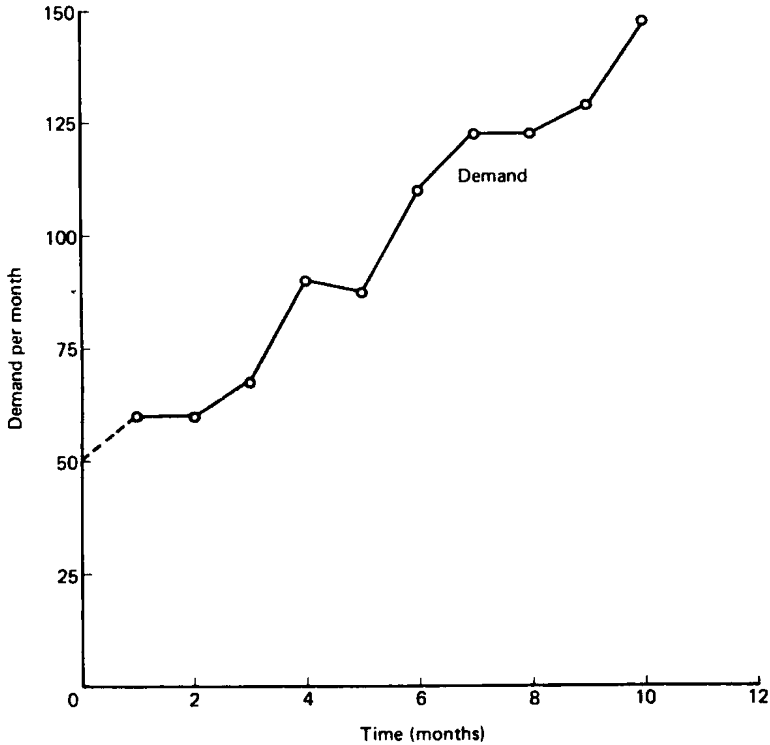

Fig. 1.2 Non-stationary situation

with individual values fluctuating above and below a reasonably steady average figure.

Examining Fig. 1.2, however, reveals a completely different situation. Here the average value is definitely not stationary but increasing with time. Such a situation could arise with a product maintaining its share of a rapidly expanding market or a product gaining an increasing share of a static (or even declining) market. Exactly what types of market condition have caused such values can only be identified by sales and marketing intelligence, but whatever the reason, this forecasting situation is obviously a more complicated one than that depicted in Fig. 1.1 and will be dealt with separately in Chapter 2 as one of two possible types of linear, non-stationary situations.

Before discussing the mathematical techniques involved in forecasting, it might be best to discuss initially just what one should expect of a forecasting system. First, it should be appreciated that because one's forecasts are based on past information, there will always be some degree of forecasting error. Accepting this, it is apparent that the most logical aim of a forecasting system should be to minimize such errors over a longish period of time. Because forecasting errors can be both positive and negative (i.e. the forecast can be below or above the actual value that occurs) a simple cumulative sum of the errors alone will not be a good indication of whether forecasting errors are being minimized or not, as this sum will tend towards zero irrespective of the forecasting system used (see discussion of CUSUM techniques, page 122). A more effective measure is the cumulative sum of the *squared* errors since squaring always produces a positive result irrespective of whether the original figure was positive or negative. Thus all errors contribute towards the cumulative sum of squared errors.

Once it has been accepted that some degree of error is present in any forecasting system, it is evident that the forecast can be only an average value of what is expected to occur, with errors distributed evenly either side of that average. In practice it is generally assumed that these errors are distributed according to a probability distribution known as the Gaussian or Normal distribution. This point will not be laboured or

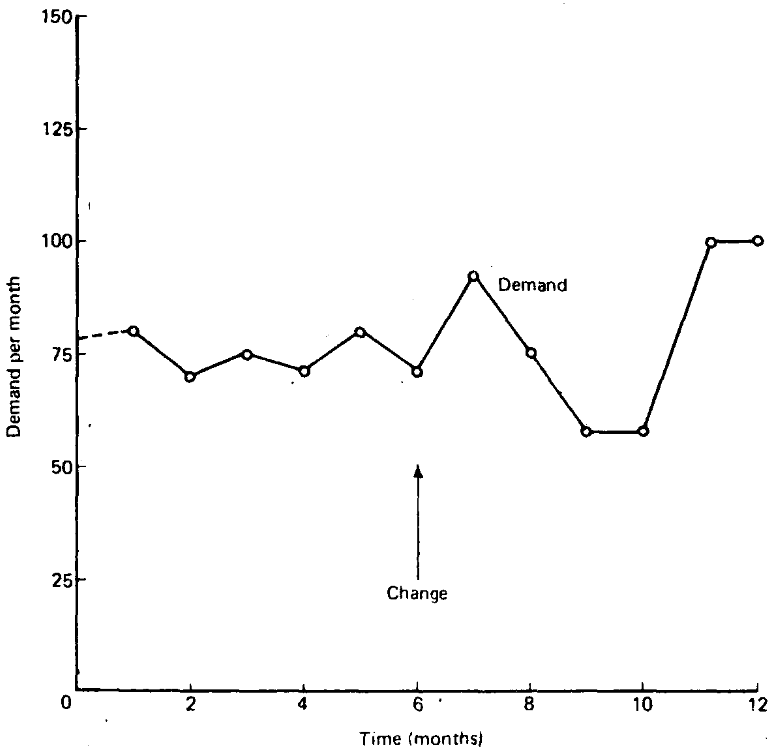

Fig. 1.3 Situation in which the standard deviation suddenly increases

explained further here except to say that the assumption is reasonably true so long as the value of the average per time period is not too low, which is generally true of the items we are concerned with in short-term forecasting if this average value is generally greater than ten.

Assuming that errors are distributed normally, we need some measure of the spread or the degree of dispersion of errors around the average. The usual measure of dispersion is known as the *standard deviation,* represented universally by the Greek symbol σ, and Fig. 1.3 indicates a situation in which the standard deviation suddenly increases owing to a change in the underlying data pattern.

For this situation the forecast would not change appreciably as the average value remains at approximately 75 per month, even after the change. However, it is evident that the spread of values, or variation, has changed after the sixth month and this change, in an efficient forecasting system, would be noted by a change in the value of the standard deviation σ.

Thus there are *two* basic parameters we wish to estimate in any forecasting system. The first is the actual forecast which predicts what the expected or average value in the future is likely to be. The second is the standard deviation which measures the spread or dispersion of individual values about that average. Chapters 1 and 2 present methods of forecasting the expected value in different types of situation and Chapter 3 describes how the standard deviation can be estimated together with the methods of evaluating other measures of forecasting accuracy.

Forecasting time period

A short-term forecasting system treats the total of individual values in each time period as a single item of data, i.e. demand per day, sales per week, production per month, etc. Increasing the length of the time period increases the sample size and hence reduces the variability of successive individual values per period thus enabling more accurate forecasts to be made. At the same time, however, the speed of response of the forecasting system to real or actual changes in the data is obviously reduced with longer time periods. A balance between these two effects can be achieved only by the selection of a suitable forecast time period.

It can be shown that the minimum value of the forecast time period should be of a duration to ensure that at least one non-zero value occurs in two time periods, that is that there is a 50 per cent probability that a non-zero value will occur during one time period. This is a minimum; overall considerations may require a longer interval.

This criterion of having one non-zero value occurring within two periods is normally met for industrial and business data if the time period is greater than or equal to one week, and a calendar month (or more practically a planning or accounting period of that order) is most typical.

The number of time periods ahead for which forecasts are produced is known as the forecast *horizon*.

The moving or rolling average*

One traditional method of forecasting the future average expected value per time period is to average the past individual values over the last n time periods. Such a moving average (m_t) has already been discussed in the Introduction see page 5, and could be defined as

$$m_t = \frac{1}{n} \sum_{i=t}^{t-n+1} d_i \tag{1.1}$$

or, alternatively,

$$m_t = m_{t-1} + \frac{1}{n}(d_t - d_{t-n}) \tag{1.2}$$

This latter version simply means that we put the current value of the moving average as being equal to the immediate past value plus $1/n$ times the current value less the value now n periods old.

Having calculated m_t for a stationary situation this then becomes the forecast of what one expects to occur, not only in the next time period but for any future time period.

This does not mean, however, that if one makes a forecast, say, for six months hence, this estimate cannot be modified next month, when an additional month's information can be used to improve the estimate, which is now only five months away. This concept may be slightly difficult to follow but corresponds very much with planning schedules, which are definite for the first and second months but only tentative for the third and fourth months. The tentative plans then become definite at the next two-monthly review in the same way that a vague forecast can become more definite as further information is received as time progresses.

The moving average does, however, in practice have several drawbacks which are discussed here:

*Known to latter-day, purist statisticians as an autoregressive process.

(i) Starting a moving average. When beginning the calculation of a moving average from new data, because it is necessary to have individual values available for the previous $n-1$ periods, no true forecast can be made until at least n periods have passed.

(ii) With a moving average, all the data included within the average are 'weighted' equally and data too old to be included are obviously given zero weighting. The weight of an item indicates the proportion of its value that an item contributes to an average, which in the case of the moving average is $1/n$ for all items included in the average, and zero for those not included.

A criticism of a method of equal weighting is that more recent data should be more important than older data and should, therefore, be weighted more highly.

A method using unequal weights could be proposed to resolve this feature, and two possible weighted averages based on either fractional or decimal concepts are shown here as equations (1.3) and (1.4) respectively. Note that in both the sum of the weights is *one*; this, by definition, is always necessary of a true average.

$$m_t = \tfrac{1}{2}d_t + \tfrac{1}{4}d_{t-1} + \tfrac{3}{16}d_{t-2} + \tfrac{1}{16}d_{t-3} \tag{1.3}$$

or

$$m_t = 0.4d_t + 0.3d_{t-1} + 0.2d_{t-2} + 0.1d_{t-3} \tag{1.4}$$

(iii) With a moving average, the amount of past data that must be retained can become excessive. A six-period moving average has been discussed here but, in practice, to obtain an average which is not excessively sensitive, one can require data from up to 20 periods.

(iv) As the sensitivity or the speed of response of a moving average is inversely proportional to n, the number of time periods included in the average, it is difficult to change this sensitivity since it is also most difficult to change the value of n, as already illustrated by the initializing-average situation.

Most of the disadvantages presented by the moving average can be overcome by a moving average of a special type in which the weighting series is exponential.

The exponentially weighted average*

Suppose that, instead of one of the weighting systems used previously, it is proposed to use a series of weights whose values decrease *exponentially* with time.

*Referred to in more recent, advanced statistical texts – most confusingly – as a 'moving average' process.