Statistical Foundations for Business Forecasts

by H. W. Gearing

Summary: Economic progress in the second half of the twentieth century depends upon the investment of large amounts of capital in new factories, in new processes and on the exploitation of the world's natural resources. The future success of many businesses will depend largely on the ability of their managements to make reasonably accurate forecasts. This article discusses some of the possible uses of computers in the establishment and maintenance of records which will serve as foundations for business forecasting routines, and reviews some recently published details of work carried out at the U.S. Bureau of Census.

It is nearly eight years since the Anglo-American specialist team on management accounting (1950) reported that American management was conscious that for an industrial enterprise to be successful there must be reliable and detailed advance planning. It was said that the approach to forecasting must be flexible, with constant revision of forecasts in the light of changing conditions. A recent article in a popular American management magazine (M/M 1958) claimed that one-third of the firms in a sample had no advance planning of any kind and only a fifth had formal routines: presumably small businesses were weighted equally with large in this calculation. Some large British companies were operating systems of forecasting and budgetary control before the Second World War. A high level of capital investment is required for industries which use automatic control and transfer equipment, where a given plant layout is likely to be fixed for many years, and it is all the more desirable to make forecasts, and to make them for longer periods, than we have covered before. The statistical foundations of business forecasts are of four kinds:

(*a*) published statistics for the community as a whole,
(*b*) the records of the particular business—internal series,
(*c*) estimates of customers' intentions over the next few months and availability of materials,
(*d*) estimates of what is expected to happen to the market for the particular product in future years.

The application of statistical methods to biological experiments, industrial experimentation, quality control and survey work has advanced, because the mathematical statisticians who have developed the techniques of experiment and survey design have worked closely with those who required to know the answers to specific problems; they have been able to influence the systems of collection in field or factory and the subsequent classification and analysis of the collected data. The mathematical economist and economic statistician have been less fortunate: economic theories have been discussed in the abstract and it has not been possible to experiment (except with rates of interest). Much of the data collected by the central government, in response to the demand for better economic time series, has been bedevilled by roughness of classification, because firms called upon to make returns are often engaged in more than one trade. The headings of the sales analysis and stock records of individual firms are dictated by internal requirements: management structure thus influences the sales analysis much more than, say, an occasional census of production; only the larger companies with punched-card equipment are able to make a special analysis for census purposes. The most we can hope for, in published series, is reasonable comparability from period to period.

Apart from the difficulties of exact definition and classification in economic time series, which no amount of subsequent mathematical treatment can overcome, econometricians have experimented with advanced mathematics for twenty-five years without the normal business man or industrial accountant becoming aware that anyone was trying to apply the disciplines of actuarial science to the data which he helped to create. This is largely the fault of the econometricians, who have made little attempt to explain their science to those who cannot follow their symbols. The student of statistical methods has, however, been made aware of the techniques available for time-series analysis by standard texts such as Davis (1941), Yule and Kendall (1950), Croxton and Cowden (1955), and others.

There has recently been a slight change in these conditions, due mainly to the efforts of a few keen operational research statisticians and practising accountants. The experience of those who have practised the art of forecasting is available mainly in a few published articles (e.g. Friday 1952, Liston 1955, 1956). Tinbergen (1939) applied multiple regression analysis to the statistical testing of business cycle theories, but the experience of those who undertake business forecasting is that, over the majority of businesses, mathematical models are not applicable.

Where the business is dealing with the supply of some basic service or raw material, fuel or power, foodstuffs or water, there may be scope for fitting mathematical models to time series, with a view to their projection.* Provided that the total area of the investigation is analysed into relatively homogeneous strata and separate curves are fitted to each stratum, those engaged in the work of forecasting demand in such businesses may reasonably continue to rely upon the use of their mathematical models. As Kendall, Friday and others have

* Mr. Muir's paper on "Automatic Sales Forecasting," to be published in October, is an example of this technique.

said, in the great majority of businesses, even if a mathematical model has been found to summarize past experience and can be shown to pass a goodness-of-fit test when projected into recent data, the probability that the causative factors will continue to exert their influence in the same direction cannot be assumed. Where a model is based on a multiple regression analysis, any attempt at projection involves forecasting the "independent" variables in the equation and such forecasts will be rough by comparison with the precision of the calculations to which the data will be subjected. Even with basic raw materials, and services such as public transport, changes in fashion and the evolution of substitutes can alter substantially the causative factors in the total demand and these changes are usually difficult to anticipate.

The application of electronic computers to the work of business forecasting, or to the work of setting up foundations from which managements may make their forecasts, cannot proceed faster than the ability of economic statisticians and accountants to devise systematic models of a system and clerical recording procedures to provide the data required. The statistical methods to be used will vary from one industry to another, but in most cases it will be desirable to begin by rationalizing past experience, by analysing the time series into trends, periodic fluctuations, seasonal variations and random influences. The techniques involved use simple moving averages to extract linear trends or weighted moving averages to follow a cyclical pattern. When there is no limitation on output, arising from shortages of materials, labour, productive capacity, suitable packaging or transport, then as Dent (1935) has shown, the sales forecast becomes the basis of all other forecasts in the business. From it, the production managers or buyers derive their plans, the accountant calculates his financial forecasts and forecasts of capital expenditure are made. The work of forecasting the sales of a company, for varying periods of time ahead, changes, as Liston (1956) has said, as we go further into the future. Monthly forecasts for the next quarter or half year are partly a translation from outstanding orders, made in conjunction with those responsible for production planning. An annual sales forecast involves market research projections, and sales forecasts for longer periods ahead are based on researches into the business of the community as a whole. These long-term forecasts will be based on trends revealed by published and private statistics, but there will be many elements at work in the economy which cannot be measured.

Where random influences are at work affecting changes of demand, the innocent might assume that those factors, which it may be more convenient to ignore, will tend to offset each other, but such an assumption cannot be justified unless it can be demonstrated that past deviations from a calculated trend have been randomly distributed. Even then, fashions will change and demand may be affected significantly by the action taken by the larger company managements concerned: the routine analysis of past series and recent figures must highlight any significant deviations from expected trends and seasonal movements, so that they can be investigated.* The efficiency of this analysis as a tool for forecasting will depend upon the ability of those making the forecasts to detect any permanent changes and to alter the parameters of their model. All this may sound promising for an electronic computer, but there are many limitations on what is possible and there have been only a few examples reported. I review some of these in the hope that those specializing in this work will bring others to readers' notice.

(*a*) *Published statistics:* At a joint meeting of the American Statistical Association and the Econometric Society in New York at the end of 1955, two members of the staff of the U.S. Bureau of the Census gave a paper entitled "Seasonal Adjustments by Electronic Computer Methods." The revision of their paper, published two years later (Shiskin and Eisenpress, 1957), describes some of the computer programs for seasonally adjusting time series, which have been developed at the U.S. Bureau of Census and improved and extended at the U.S. National Bureau of Economic Research. These programs have been made available to other organizations and by mid-1957 more than 3,000 series had been adjusted.

The computer programs were written with the object of reducing the cost of making seasonal adjustments of time series. Two methods, developed at the Bureau of the Census, are described in the paper and the results compared. The first method was based on the ratio-to-moving-average method (Mills 1955, Croxton and Cowden 1955, and others); a simple 12-month moving average of the original observations was taken and the original observations were then divided by the moving average to give a series which contained the seasonal and irregular components. These components for each month in successive years were then averaged to provide 12 seasonal factors, one for each month. These seasonal factors were divided into the original observations to provide a series which included only the trend, cyclical and irregular elements: a five-month moving average of this was used to smooth out the irregular elements. It was appreciated that this method leaves a number of weaknesses; the final five-month average does not result in a smooth representation of the trend-cycle component and two observations at beginning and end have to be covered by extrapolation.

* In *Statistical Analysis of Stationary Time Series*, by Ulf Grenander and Murray Rosenblatt (New York, 1957: John Wiley and Sons), the results of new theoretical work carried out at Chicago and Stockholm are explained. The new techniques use concepts from the theory of stochastic processes, and they are applied to the study of random noise; turbulence and storm-generated ocean waves are examined. The authors say (p. 178) they would hesitate to apply these techniques to econometrics and the social sciences, because the assumptions of stationarity will in general be valid only for short time intervals. It is further suggested that appropriate means for analysis of such series may be found where there is more quantitative theory, but in most contexts the common-sense methods of practical economists will give more.

The second method begins in the same way as the first, by dividing the original observations by a simple moving average: the trend-cycle curve is obtained by means of a weighted 15-month moving average (Spencer formula, Kendall 1946). This technique, which has been used in actuarial work and which results in a curve resembling a third degree polynomial, is further improved upon by introducing a control chart procedure to identify "extreme" observations, which are then adjusted. Further adjustments are made to smooth the seasonal factors; alternative formulae are available for smoothing irregular elements of differing magnitude and the most appropriate of these is automatically selected. Adjustments for number of working days are carried out initially, provided the number of working days is punched at the same time as the original observations. The article includes a number of charts comparing original observations and seasonally adjusted series yielded by the two methods. The assessment of their validity is partly a question of relating the arithmetical procedures to a logical argument and partly a subjective measure of which seasonal series has the minimum of movement between years. It is also recognized that moving-average techniques tend to distort the incidence of turning-points in series.

The authors report that further revisions of the computer programs are now under consideration, including variable methods of graduation and extrapolation at the ends of the series and finer control limits for separating extreme ratios from normal ratios. The contribution of the Univac Computer thus includes a limited editing facility as well as its speed; a ten-year series of 120 monthly observations was analysed in 2·3 minutes and the computing and printing of 3,000 series could have been completed in a week. To the argument that such complicated procedures are out of tune with the crudeness of the original data to which they have been applied, the authors argue that these tools permit the rapid analysis of large amounts of data in the search by economists for uniformities in economic behaviour. In an appendix, they discuss two other computer methods for making seasonal adjustments, one of which makes use of a regression technique. The other uses a moving third-degree polynomial fitted by least squares to an eleven-term series of ratios, the program for which has been coded on an IBM 701 Computer.

The work of Shiskin, Eisenpress and others demonstrates what a computer can do in the way of carrying out mass computations on large amounts of data. They are fully aware, as is demonstrated in their paper, of the limitations as well as the advantages of computer programs of this kind. After several years of work they admit that there are "certain desirable adjustments that appear to be extremely difficult to make mechanically" (p. 437): examples of these are missing observations, data arising during strikes, and other extreme individual observations. Despite the control chart procedure built into the program, they think that some manual adjustment of the basic data prior to input would be simpler and more effective. Their economist friends suspect that the fascination of the new tool may divert econometricians from the analysis of real problems, but all would agree that analysis of available data must continue. Their published paper, and any subsequent reports which the National Bureau of Economic Research may issue on this research project, should be read in detail by all who contemplate the analysis of time series on a computer.

(*b*) *Internal series:* There is more hope for the use of computers on internal records and statistics, some of which could be analysed into seasonal and trend elements, or related to internal standards.

In a paper entitled "A Central Intelligence Program for Management" given to the Systems and Procedures Association of America in October 1955 (Workshop for Management 1956), Mr. Ray R. Eppert, an Executive Vice-President of Burroughs Corporation, stressed the importance of well-designed *pipelines* for reporting information from outlying branches to management at all levels. Reports to executives should be tailored to provide each management function with data arranged to give all relevant pieces of information in one statement. For a sales manager of office equipment, this would include recent sales figures and their proportion to past and budgeted sales, number of men selling, total selling expenses and expenses related to men and sales.

The production of such reports automatically would seem to require very considerable data storage capacity, if all the series are to be brought together from different accounting points of detailed origin, for collation and calculation in different ways for the various management functions. Some companies in America are moving towards such consolidated reports, when their punched-card equipment is extended by the addition of a computer. Montgomery Ward and Company (Chicago) were probably one of the first companies to embark on such work, covering 568 retail stores (Klingman 1955), and it is understood that the Burroughs Corporation has also carried out, on a computer, statistical analyses of sales leading to the setting of new targets. In England, J. Lyons and Company Limited have programmed LEO to produce management statistics as by-products of their integrated routines for organizing deliveries from their central bakeries to their retail teashops and retail agents, and of their leaf-tea inventory and stock control procedures (P.E.P. 1957). Other companies with comprehensive punched-card routines for sales analysis are currently considering how their detailed statistical returns can be improved, by the use of a computer. to make comparisons and extract significant information for management's attention: a section of a London study group of the British Computer Society is currently discussing this subject.

Provided that the various classifications required by the manager or forecaster are included in the coding system applied to the detail throughout the period, or when not so coded, if they can be derived by grouping combinations of other elements in the code, then there

should be no difficulty in recovering the information from the records to produce forms of analysis appropriate to the work of each manager or the business forecaster. The biggest problem will be the design of coding systems which are sufficiently detailed to yield the required finished analysis, but which are not a burden to the normal accounting and recording work of the business. There is, however, usually a high degree of redundancy in business coding systems (Oettinger 1957); any one customer may take only a limited range of products and his account code identification will be correlated with codes for representative, area, industrial classification, etc. A detailed coding system need not be a burden if it is added to the order for internal records and transmitted through an integrated system to each delivery against that order. Such a routine can be automatically carried out without manual labour, where there is an integrated system on punched cards, or a computer. But in this example of a sales analysis coded for a variety of attributes, combinations of which are of interest for different management functions, we meet a major difficulty when accumulating records covering all aspects of past business. If all codes are covered in cumulative summary cards, a very large volume of cumulative cards have to be updated in the last months of the year, particularly where there are a large number of small customers.

If the cumulative record is on magnetic tape, there is the further problem of which code should be used as the major sequence. As with manual systems, we find it difficult to preserve major sequences suitable to every inquirer, but computers will enable us to select specific classifications and perform a limited analysis on them more rapidly than hitherto, if programs are written with variable parameters to permit selection of code combinations by the inquirer. The use of separate tape searching equipment, such as has been demonstrated by one British manufacturer and several American companies, will simplify such programs by first selecting the records required for the survey and then processing the selected records through the computer.

(*c*) *Customers' intentions:* A company making heavy capital equipment, to the order of industrial buyers, will find that its ability to forecast sales over the immediate future will depend on the available analysis of uncompleted orders and the tendency, or otherwise, of its factory to keep to delivery promises. Industries engaged on the manufacture of consumer goods receive the demands of the customers at third hand, transmitted through retailer, wholesaler and packer. Marketing research companies have improved the knowledge of manufacturers by systems of retail shop audits, and the application of punched-card equipment and computers to this work is already well advanced in the offices of those concerned with such surveys. The vertically integrated company, which is both retailer and manufacturer (e.g. Lyons), can use the computer to accelerate and improve such information routines. The quality of the result and its cost depends on the collection routine at shop level, which has to be carefully considered so as not to inconvenience the work of retail distribution.

The application of electronic computers to the analysis of current demand will probably follow several lines of development according to the nature of the particular business, for example:—

(i) analysis of current orders into their expected periods of execution, and the establishment of a detailed programme from these, for each manufacturing operation;

(ii) analysis of customers' past records and their presentation to selected customers as a basis for future estimates; this technique may be valuable in seasonal trades.

The first, under such titles as *production scheduling and material control*, had already been made a matter of routine by the application of punched-card techniques. The introduction of computers into the final tabulating stages of these routines has reduced the time taken for the analysis and increased its scope. Some American companies are claiming considerable economies from the introduction of computers on these routines; others that the economies were made mainly when the system was organized originally for punched cards. Others are considering the use of linear programming and other statistical methods, to achieve optimum allocations of resources and the minimization of transport costs, change-over time, etc. (Kozmetsky and Kircher 1956). Mathematical models of ideal systems may demonstrate where corrective action should be fed back into a system, but the many discontinuities in the factors of production in real life are difficult to build into a computer program.

The second line of possible development, to provide representatives with detailed statements of customers' past requirements as a basis for discussing new delivery instructions, would not be difficult to organize, where there is a facility for mechanical tabulation, but it could be an exceedingly time-consuming operation for the machine room. Whether it would be justified would depend upon the circumstances of each particular business. It could be used to reduce any tendency to overstock, when there is the fear of a shortage. Such a tabulating service, if backed by reliable delivery promises, might assist customers to work with smaller stocks and, irrespective of the periods covered by their individual orders or contracts, it could help to smooth out the peaks and valleys of their delivery instructions.

(*d*) *Market forecasts:* The long-term sales forecast remains the hardest nut to crack, and it is difficult to see how the arrival on the scene of the electronic computer can be of much help in the great majority of industries. The dangers of projection from past series into the future are well known. For the individual firm there is an additional hazard, that the proportion of the total business which it gets will be partly dependent on the activities of others in the same trade. For long-term forecasts, a slide rule will usually give a sufficiently

accurate calculation of the components required to achieve any total target.

In routine market surveys of current demand, the analysis of the collected information on a computer will improve on conventional punched-card tabulating procedures in two ways:—

(i) to carry out in one run, counts of unit observations and accumulate totals which would normally require several processes of tabulation;
(ii) to perform limited editing operations by signalling, for closer scrutiny, any observations having extreme values in observed characteristics;

but this is a far cry from peering into the future.

Conclusion: The writer believes that one of the biggest problems of industry today is to make accurate forecasts. For a quarter of a century we have had econometricians working on theories and they have had little if any messages for those who run the world's commerce. The electronic computer may assist some managements to analyse the implications of customers' orders already in their order book at any time, in terms of load on available resources and optimum distribution over productive capacity. But the importance of this to any particular sales forecasting routine will depend on the length of the production process. By analysing past series into trend, seasonal elements, etc., the computer may permit us to examine series for each product, instead of total series only, and thereby establish probable limits for future fluctuations in demand and stock levels. This may contribute to the stability of the aggregates (Friday 1952) which we compile in building up our forecasts, by adding together the pieces of a future jigsaw puzzle as we see them take shape. It will not, however, make forecasting a science; good forecasting will remain very much of an art. As Lady Lovelace wrote in 1842, "the computer can do whatever we know how to order it to perform" (Bowden 1953).

References

Anglo-American Council on Productivity (1950). *Report of specialist team on Management Accounting.* London: November 1950.

Bowden, B. V. (Editor) (1953). *Faster than Thought; a symposium on digital computing machines.* London: Sir Isaac Pitman and Sons Ltd.

Croxton, F. E., and Cowden, D. J. (1955). *Applied General Statistics.* New York: Prentice-Hall Inc.; London: Sir Isaac Pitman and Sons Ltd.

Davis, H. T. (1941). *The Analysis of Economic Time Series.* Indiana: Bloomington Press.

Dent, A. G. H. (1935). *Management Planning and Control.* London: Gee and Co. (Publishers) Ltd.

Eppert, R. R. (1956). "A Central Intelligence Program for Management." Management Publishing Corporation, Greenwich, Connecticut: *Workshop for Management:* Systems and Procedures Assn. of America, Eighth (1955) Annual Meeting.

Friday, F. A. (1952). "Business Forecasting." *The Incorporated Statistician* (The Assn. of Incorporated Statisticians), **3–2**, 25.

Friday, F. A. (1952). "The Problem of Business Forecasting." *Journal of Industrial Economics* (Blackwell: Oxford), **1**, 55.

Kendall, M. G. (1946). *The Advanced Theory of Statistics.* London: Charles Griffin and Company Ltd.

Klingman, H. F. (Editor) (1955). *Electronics in Business; a descriptive reference guide.* New York: Controllership Foundation Inc.

Kozmetsky, G., and Kircher, P. (1956). *Electronic Computers and Management Control.* New York: McGraw-Hill Book Co. Inc.

Liston, D. J. (1955). "Sales Forecasting: an evaluation of some practical results." *Applied Statistics*, **4–3**, 162.

Liston, D. J. (1956). "The Statistician in Business Research." *The Incorporated Statistician*, **7–2**, 63.

M/M (1958). "How Management tackles Advance Planning, a survey of Company Practice." *Management Methods* (Greenwich, Connecticut), January 1958.

Mills, F. C. (1955). *Statistical Methods.* New York: Henry Holt.

Oettinger, A. G. (1957). "Account Identification for Automatic Data Processing." *J. Assoc. for Computing Machinery*, **4–3**, 245.

P.E.P. (1957). *Three Case Studies in Automation.* London: P.E.P.

Shiskin, J., and Eisenpress, H. (1957). "Seasonal Adjustments by Electronic Computer Methods." *J. American Statistical Assn.*, **280**, 415.

Tinbergen, J. (1939). *Testing Business Cycle Theories.* Geneva: League of Nations.

Yule, G. U., and Kendall, M. G. (1950). *An Introduction to the Theory of Statistics* (Ninth edn.). London: Charles Griffin and Co. Ltd.