FORECASTING AND RISK ANALYSIS APPLIED TO MANAGEMENT PLANNING AND CONTROL

JAN BERNY

Doctor of Philosophy

THE UNIVERSITY OF ASTON IN BIRMINGHAM

June 1988

THE UNIVERSITY OF ASTON IN BIRMINGHAM

FORECASTING AND RISK ANALYSIS APPLIED TO MANAGEMENT PLANNING AND CONTROL

JAN BERNY
Doctor of Philosophy

1988

THESIS SUMMARY

The aim of this research was to improve the quantitative support to project planning and control principally through the use of more accurate forecasting for which new techniques were developed.

This study arose from the observation that in most cases construction project forecasts were based on a methodology (c1980) which relied on the DHSS cumulative cubic cost model and network based risk analysis (PERT). The former of these, in particular, imposes severe limitations which this study overcomes.

Three areas of study were identified, namely growth curve forecasting, risk analysis and the interface of these quantitative techniques with project management. These fields have been used as a basis for the research programme.

In order to give a sound basis for the research, industrial support was sought. This resulted in both the acquisition of cost profiles for a large number of projects and the opportunity to validate practical implementation.

The outcome of this research project was deëmed successful both in theory and practice. The new forecasting theory was shown to give major reductions in projection errors. The integration of the new predictive and risk analysis technologies with management principles, allowed the development of a viable software management aid which fills an acknowledged gap in current technology.

KEY WORDS :-

Deterministic growth (S) curve
Forecasting
Risk analysis
Project management
Management software systems

ACKNOWLEDGMENTS

First and foremost my heart felt thanks to my wife Wilma for her patience over many years.

There are many members of industry who have contributed to the practical implementation of the theories which were developed in particular Hugh Davies and Paul Townsend.

Without the support of both SERC and South Bank Polytechnic this work could not have been completed, particular mention should be given to Rod Howes for his support in the earlier phases.

Last but not least I wish to thank Colin Lewis for his patience and the clear headed and frank way in which he has guided me.

CONTENTS

MAIN TEXT

APPENDIX

FIGURES

MAIN TEXT

FIGURES

APPENDIX

FIGURES

APPENDIX

TABLES

MAIN TEXT

TABLES

APPENDIX

MAIN TEXT

CHAPTER 1

INTRODUCTION

The introduction describes the background setting, which stimulated the author to embark on research into upgrading the quality of information feedback for management decision-makers. This led to the establishment of the key issues and consequently the plan of research which forms the structure and thrust of the thesis.

1.0 Overview

The well being of a large organisation is strongly dependent on the success of management planning and control. The overall aim of this research was to enhance these facilities, particularly with respect to projects and corporate strategies.

The major research effort was aimed at the integration of the quantitative aspects of project management. The use of the developed techniques was seen to apply to projects in the wide sense of the term. For instance, construction industry projects apply to estate development as well as building.

In manufacturing, one may be concerned with a marketing project, which in turn interfaces with production and/or research and development. All these may in varying degrees be planned and controlled, indicating a broad spectrum of potential applications.

The currently used project and corporate management techniques fall into two major areas namely network analysis and spreadsheet 'what if?' routines; and to a much lesser degree risk analysis and least of all forecasting. It was observed by Tan (1987) that currently little or no integration existed between these methodologies in computer applications. Furthermore, it was found that no attention was paid to the relationship of risk analysis with forecasting. For instance the construction industry has been making occasional use of risk methods since 1970, Yates (1986), but only currently is this subject receiving any prominence. Of equal importance is the largely ignored area of systematic use of historical project information as a direct support to future planning.

The major source of information for this research was derived from construction industry project cost breakdown data. Analysis showed a notable limitation in the quality of current forecasting methods and their practical use Berny, (1984).

Limitations to risk analysis methods, in particular PERT/RISK, have been frequently discussed e.g. Miller (1962). This point was well summed up in Battersby (1970), nevertheless this outdated technique still prevails in the few current software packages aimed at analysing risks of both time and cost overruns in projects.

The overall purpose of this research was, in the first instance, to improve the key technologies; namely forecasting and risk analysis. Having procured improved tools for planning and control, it was considered prudent to complete this research by establishing an integrated methodology to use these techniques and successfully implement them in industry.

1.1 Origins of the general area of research

This research originated as a result of the author's involvement in a refurbishment project for York County Hospital, Berny et al (1978). The author was charged with the responsibility of producing a portfolio for the purpose of acquiring large funds to alter the use of this listed building from a Hospital to a Hotel and Conference centre.

It was observed, from discussions with quantity surveyors, that the construction industry in general depended on a job contingency worth about 10% of the value of the work. This contingency was necessary to compensate for possible variations in cost and time and calculated in this ad hoc manner, rather than by any more scientifically accurate method. This rule of thumb was not historically based but enquiries indicated that it was thought to apply to most new building projects.

The DHSS (Department of Health and Social Services) cubic forecasting method, Hudson (1978), of calculating the potential cost breakdown was used as part of the York project, but was found to be inappropriate for this refurbishment project. The DHSS model required estimates to be made of parameters which related to the type of cost breakdown the construction experts, Shepherds of Portakabin fame, anticipated. This requirement was particularly exacerbated due, to the fact that it had to reflect the problems relating to work on buildings dating from 1740 to 1910. Furthermore, due to the fact that the DHSS model is a cubic equation the derived parameters generated projections which passed the curve maximum and hence were non-feasible. The combination of both the age of the building and properties of the cubic, caused the author to abandon this model for the York project.

An attempt to use PERT/RISK in a relevant way, Miller (1962), failed due to the lack of detailed planning information. The only methodology which made a positive contribution was discounted cashflow, which was used to assess the break-even point in time between cost and income.

Discounted cashflow, is accepted as a useful but limited appraisal technique, particularly in the case of large scale investment. This unsatisfactory financial prognosis led the author to an in depth inspection of the DHSS cubic forecasting model.

The author generalised and tested this new cubic model against cost breakdown data, Berny and Howes (1983) but the resulting investigation of a wide range of construction industry data showed the inadequacy of polynomial models, Berny (1984) and classical growth curve models such as the logistic.

It was concluded that there was a need for research into better means of forecasting, in particular for construction industry data. Further this project showed the necessity to obtain a simple means of assessing risk.

Reconsideration of the York County Hospital project, Berny et al (1978), showed the need to integrate planning and control methodologies with the support of better forecasting and risk assessment techniques.

Finally there was seen to be a considerable need to improve the promotion of the use of better planning and control techniques. The attitude of clients was seen as the main source of the dearth of application of better techniques. It was commonly found that in the case of large-scale projects, which generally have political strings attached, that overruns in cost and time terms were not blamed on project management. Extra costs remained hidden and were paid for by the unsuspecting taxpayer.

While such attitudes prevail, the extra effort required to organise work efficiently was seen by many in the profession as unnecessary and it appeared that more lip-service rather than action was being paid to the need to consider risk assessment.

This attitude was also reflected in the area of forecasting. Sparks and McHugh (1984) whose work surveyed the use of forecasting showed that 73% of 'extensive' influence stemmed from executive assessments. This author believes that the forecasting syndrome cited above also parallels the situation appertaining to risk analysis.

1.2 Key issues

This research is based on the hypothesis that to successfully promote better planning and control aids for project management, two separate aspects had to be considered; namely :- fundamental issues and key pragmatic problems.

1.20 Fundamental issues

The fundamental issues were in the first instance to improve medium-term forecasting and secondly to find a risk distribution which could be fully implemented by 'naive users'.

1.20a Forecasting

The forecasting methodology required specific attention be paid to growth curves, as the nature of a project was that in general it involves the life cycle of birth, maturity and death. This means that any representative model has to be monotonically increasing. The current approach for this type of prediction uses stochastic methods. However, as such procedures were dependent on deterministic models, it was argued that improvement in deterministic growth curve models was the first necessary step to enhance the quality of prediction.

The ability to progressively interpret from practical data considerations the nature of a model, further emphasized the need for deterministic, rather than the currently preferred stochastic, approach to medium-term forecasting.

1.20b Risk analysis

The most fundamental issue here was the shortfall in practical risk analysis methodologies. The major observation made by this author was that PERT/RISK, which is based on the beta distribution and requires four estimates from a user. These are pessimistic, optimistic, most likely (mode) and expected (mean) or standard deviation times or costs of an activity or project, Farnum and Stanton (1987).

It is quite impractical to expect the user to estimate four out of the five pieces of information cited above. The current method of using the beta distribution in PERT and other time-risk analysis procedures require the user to state only three quantities namely :- pessimistic, most likely and optimistic activity durations. As four parameters were required to satisfy the beta distribution the standard deviation was set at the value of $^1/_6$ from a range of 0 to $^1/_2$ on a normalized scale of activity duration from 0 to 1. An approximate equation was used to estimate the mean duration, thus further limiting the accuracy of the estimated expected value. Other consequences of these simplifications were discussed in detail by Battersby (1970) who highlights the limitations and their affect on the success of this technique. The same drawbacks apply to cost risk analysis.

These relatively complex demands on users probably influenced the lack of widespread implementation by industry of this family of techniques.

To sum up, the fundamental issues were to find :-

(1) An improved deterministic growth curve model, with the facility that practical interpretation was implementable.

(2) A new risk distribution which did not overtax the user and had no hidden assumptions such that probability 'guesstimates' were fully translated into risk distribution parameters.

1.21 Pragmatic issues

This aspect of the research is more qualitative in nature, hence it was not possible to measure success in a clear and quantitative manner. Nevertheless, acceptance of the techniques by industry, their continued interest and feedback were the main means of assessing the value of this part of the research.

The pragmatic aspects of forecasting have been reviewed by Fildes (1978).

The discussion about the problems emerging from the York Hospital project, Berny et al (1978), helped clarify areas of study, namely the importance of plan and control techniques and, secondly a means to successfully implement such methods. It was found necessary to improve planning accuracy, feedback from monitoring, translation of control decisions to achieve more effective and economic project completion.

The implementation requirements for successful completion of major projects primarily hinge on good communications between the user and the information which may be accessed, generated and highlighted by appropriate techniques. Practical feedback by "sympathetic" members of industry, their field trials and assessment of the success or practical advantages gained from this work was seen as a means of evaluating the solutions generated from considering the pragmatic issues.

The key pragmatic problems addressed in this work were :-

(1) More accurate and flexible planning.

(2) More valuable feedback from monitoring information.

(3) Greater flexibility in assessing the potential outcome of control decisions.

(4) Improvement to the communication interface between management and the feedback of accessible quantitative information.

CHAPTER 2

RESEARCH OBJECTIVES AND PLAN

The research structure required the key issues in chapter 1 to be set in a satisfactory framework of study, which reflected the management aim and was balanced by adequate theoretical progress. The need to stress limitations on the research was apparent from the outset as the potential literature base was known to be very extensive (European Space Agency Data Retrieval). This chapter analyses and sets out the research plan.

2.0 Plan basis

The key issues described in 1.2 required to be set into a systematic framework. The plan used to conduct the research required that the major goals were placed in an order that led logically from one phase to another. The nature of the subject was highly open ended. For this reason it required that limitations be set, both in the depth and breadth of the different areas of work, in such a manner that they would not detract from the value of the research project as a whole. In order to create an implementable final product of the results from this thesis, it was perceived as essential to create a balance of effort between the key theoretical and pragmatic issues; this entailed taking careful account of the research limitations. Due cognizance was paid to the need to ensure that each section of the research has the stature of a viable contribution to knowledge in its own right.

The research strategy was influenced by the project management needs of the construction industry, which was the major source of practical information.

2.1 Priorities

The need to establish a better forecasting methodology was seen as the critical factor in this research. Bromillow (1973) showed that improved project management within the construction field required that most attention be paid to planning project cost-breakdowns. Pilcher (1981) further showed the need to study S-curves, hence curve fitting procedures had to be a consideration. These fall into the growth curve forecasting field, hence these were selected as the top priority area of study.

Growth curve studies have been primarily in areas such as sales projection for a product or growth of cells Stone (1980). These studies consider a homogeneous 'item' or 'activity'. In contrast a construction project consists of many activities generating a heteregenous entity. Evidence from collected data showed that construction activity characteristics are that of a homogeneous entity.

Due to the monotonic nature of a growth curve the cumulation of many growth activities also generate a monotonic 'growth' pattern. For this reason a heteregenous system which is constituted from homogeneous growth entities will have growth properties but will be distinguished by showing much more noise (small variations) and hence more difficult to describe in model terms. It is for this reason that the modelling requirements for projects are much more stringent and hence highlighted the need to rethink the underlying concepts for generating deterministic models.

The natural continuation from growth curve forecasting was to explore the practical uses of new models. The importance of devising a means of parameter generation from project cost information was previously stressed and explored by Berny (1984).

The next priority was the exploration of the means by which project management could utilize interpretable curve fitting and forecasting data. It was decided that the plan, monitoring and control elements should be considered in this context.

A further aspect of curve fitting was the need to establish the statistical reliability of the chosen model and determine the associated confidence limits. This information was seen as the entry point to risk analysis studies.

Risk assessment led to the study of the next quantitative key issue, namely to establish a means of resolving the aforementioned problems (see 1.20b) associated with PERT/RISK, Battersby (1970). Work in this area, and its interface with project management, was seen as critical to the enhancement of progress in the area of project planning and control.

The amalgam, in a suite of computer programmes, of the techniques generated above was viewed as the key to an integrated approach to project management planning and control.

2.2 Priority list and limitation of research framework

PRIORITY LIST

(1) Find a better growth curve forecasting model which has properties interpretable by management.

(2) Utilize the model for planning, monitoring and control aspects of project management.

(3) Consider uses of the statistical reliability aspect of the model vis-a-vis risk assessment.

(4) Find a new risk distribution which overcomes the PERT/RISK drawbacks and employ it in a pragmatic manner.

(5) Devise the means by which the devices above can be used in an integrated manner by management.

LIMITATIONS

It was clearly not possible to explore all possible ramifications which would emerge from new growth curve models or probability distributions for risk. For this reason it was seen as important to find a general basis for formulating these functions and select one model by appropriate testing.

Limitations in the management interface and integration section were less easy to define. In order to produce a practical tool, a consensus of the feedback from industry was used to establish the practicability of the technology. Limitations in this work were of a qualitative nature. A plausible approach was exemplified by Eilon (1979) who stressed that planning and control should be highly adaptable - i.e 'what if?' In nature.

Summary of limitations :-

(1)	Theoretical :-	Detailed exploration of only one model for forecasting and one for risk analysis.
(2)	Pragmatic :-	The interaction with industry was controlled by field trials, which in turn set the limits of "industrial approval', measured by the purchase of the resulting software.

CHAPTER 3

FORECASTING

This chapter is central to the major theoretical key issue: forecasting. The historical background, its consequences in the light of pragmatic needs and the resulting development of new concepts for building deterministic growth curve models are fully described. Some models which result from the new principles are derived and one selected for further analysis.

The bulk of this chapter and further developments have been refereed and accepted for publication :-

"New concepts in deterministic growth curve forecasting" Journal of Applied Statistics: anticipated publication date March 1989.

3.0 Introduction

The majority of current studies of growth curves consider stochastic methods applied to deterministic models and are most frequently tested against marketing problems. Construction Industry project management applications have been found to require more robust deterministic growth curves. It is for this reason that this chapter is primarily concerned with improving non-stochastic models.

Tests on fifty construction project cumulative cost curves have shown that 16% were of an exponential nature and a large number showed that the non-cumulative costs have bimodal profiles. It was found that currently used deterministic models e.g.:- Stone (1980) and those using stochastic applications e.g. autocorrelation based on Mar-Molinero (1980) gave poor curve fits for these project cost profiles. It was further contended that the better the deterministic model the earlier one can start reasonable forecasting. It is fully accepted that stochastic methods enhance prediction, but a good deterministic basis was seen as equally important.

Based on the above background, it was found that further progress in analysing these cumulative cost curves required new concepts for building deterministic models.

The key properties of new models to be considered were :-

a) To include descriptions for non 'S' shaped curves which also pay attention to the early history of the process being described.

b) The need for models to pass through the origin.

c) Uphold cases which have no apparent saturation level.

d) The non-cumulative profile of the model should also follow the non-cumulative smoothed trend of the data.

Decay is the inverse of growth and hence should be considered as part of this study. However, due to lack of data and extent of the current work, decay models have been excluded.

A by-product of the database of the large spectrum of curves, required that alternative statistical measures had to be developed which permit comparison between widely differing sets of data. The results of this work summed up in Berny (1987c) have shown that multi-component models, describing growth as the sum of several functions, consistently reduce by some 50% testable forecast errors as well as giving better curve fits. The comparison with the current work has been made on the basis of like for like i.e. using the equivalent number of parameters to those used in the best quoted results from other studies.

3.00 Area of study

The major bias of this research has been to advance the use of growth-curves by project management. In this guise the applicable use of 'S-curves' by the Construction Industry has served as the major guide to the potential use of growth curves. This industry's cooperation is appreciated as a source of cost data relating to some 200 projects of which 50 have been analysed in depth, in addition to 12 problems from other industries. A concern for the need to forecast, control and plan budgets for construction projects with a higher degree of precision, has been highlighted by Pilcher (1973), Trimble (1974) and many others.

This issue is by no means confined to the construction industry, hence it was perceived that a useful practical outcome of this study would be to enable managers to simplify the means of creating budgets, thence to update them and aid the control of projects. An alternative use is that these facilities also create a means of checking the reasonableness of a budget constructed in the traditional way by considering individual activities.

3.01 Project growth properties

A budget can be interpreted as a prediction of project costs with respect to time, based on knowledge from past experience. Conversely it can be regarded as an extrapolation of past experience to a point in the future. The end point could be defined as the value at an infinite point in time, or more practically it can be identified as a known end-time and cost through which the growth curve must pass.

Growth can display many different profiles according to the phenomenon under study. These differences are particularly well demonstrated in non-cumulative or period value profiles. Some growth curves have regular features, whilst others can embody pronounced irregularities. In order to advance growth curve modelling the study of data which displays the irregular features of growth was necessary. The analysis of project expenditure data collected from the construction industry, appears to be ideal for this purpose.

From the projects which were considered, their project data were collected in the form of activity cost. These data were derived by a logical analysis of the project using networking or by bar chart methods. The expenditure associated with each activity was subsequently allocated to its time span on a period basis. The summation of all the activity expenditure occurring in each period produced a non-cumulative cost profile for the total project duration. Hence, it proved possible to identify the activities having the greatest influence on expenditure, both in value and timing. In all instances it was observed that cumulative activity expenditure profiles follow a variety of growth patterns. The above analysis showed that a project cost profile could be generated by cumulating the superimposed simulated cost breakdown of activities using growth curve models. This procedure was deemed to give a more precise total cost plan of a project than one which current methods generate, as they rely on using only a straight line model.

Another means of investigation adopted was to study actual overall costs from different projects of a similar type in order to establish any similarities in their expenditure patterns. Whilst some consistency was observed there proved to be no refined pattern of generality. This finding conforms with the unique characteristics associated with most construction projects, coupled with the wide variety of types. Whilst this has the obvious disadvantage of being responsible for the wide degree of variability, it offers the advantage of supplying a wide variety of growth curve profiles for study.

3.02 Selection of case studies

A recent study by Meade (1984) surveyed earlier work in marketing. Meade selected four classical case studies, dating back to 1967, for detailed analysis. It was decided that the same four should be used in this research.

These are :-

a) U.K. consumption of PVC, Hutchesson (1967)

b) U.K. consumption of titanium dioxide, Harrison and Pearce (1972)

c) Tractors in Spain, Mar-Molinero (1980)

d) Adoption of colour TV's in U.K., Bagot (1978)

In addition to the above four, two projects were used as further case studies. These latter were selected from the Construction Industry. The choice of construction projects was based on specially selected growth curve types, in order to demonstrate both the degree of variability which exists and the ability of the models developed to fit less well known types of curve profile.

It was anticipated that this approach would provide a direct comparison of performance with 'the state of art' relating to growth curve modelling. The wide variety of projects considered led the author to believe that a new approach should be attempted to produce a model which has the dual properties of flexibility and practical interpretation, whilst embodying sufficient accuracy to render the model useful in real life circumstances. A key factor in the development of this part of the research programme was the realisation that the consideration of cumulative growth curves did not provide adequate sensitivity for analysis. It was therefore decided that an extensive study should be made of non-cumulative curve profiles, which subsequently provided an ideal means for examining the effects of smoothing and curve fitting the data when testing models. However, it was established that the cumulative curve provided the best means of assessing the predictive properties.

In view of the importance and hence frequent reference to non-cumulative values/costs these will be referred to as periodic values/costs and cumulative values/costs as values/costs (to date).

3.1 Survey of growth curve modelling

This section deals with both the practical uses of growth curve models and the range of model types investigated.

3.10 Polynomial and error function models

The study of growth curves commenced from an evaluation of the DHSS cubic model for costing hospital construction Hudson (1978). Berny and Howes (1983) generalised and simplified the cubic model for cumulative project costs and developed a means of practical interpretation by evaluating its parameters from data related to the point of inflection i.e. the peak activity period (in terms of expenditure per unit time) and its cost to date.

A major limitation of the DHSS model was that when projects exhibited an early rapid expenditure towards the peak activity point, then the maximum value of the fitted cubic curve (which represents the cumulative costs,) was reached in many instances prior to the end of the project. This meant the period costs after the maximum were negative rendering the solution impractical. Other forms of polynomial equations, in particular the quartic, were investigated Berny (1984) but these were also found to be of little practical value.

A major step forward in the resolution of this problem was made as a result of discussions with A. J. White, who suggested the use of either the logistic or the differential of the error function (to be abbreviated as the DEF), a form of which is referred to in computer software literature as the Rayleigh - Norden model; Putnam and Fitzsimmons (1979). Neither of these models have the aforementioned "maximum" problem of the cubic. Investigation by this author of these two suggested models gave preference to the DEF model, as a basis for modelling construction project costs to date, because it gives a zero value at commencement (time=0), which is never the case with the family of modified exponential curves of which the logistic is one. Furthermore the relationship with the peak activity point in time was mathematically more straightforward to use.

Further support for the DEF model comes from Norden (1960) who compared the cubic logistic with a DEF model and stated that the former was found "to closely approximate a considerable number of projects,.. but it was found to be inadequate for extrapolation". Norden continues further stating that "the latter (DEF) was found to describe more realistically engineering research and development project phases".

The issues generated by comparing the potential use of the cubic with the DEF models was addressed by Berny (1984) who agreed with Norden (1960) that in spite of the DEF model's predictive attributes, used alone it was, in some cases, unable to adequately describe project costs. For this reason a model was formulated combining the best features of the cubic and DEF models as a cubic exponential spline. The combined (cubic exponential spline) model outlined by Berny and Howes (1983) proved to be easier to use from the point of view of interpretation, following principles indicated earlier.

However, even this combined model was found to be unsatisfactory for some construction projects; furthermore it was found difficult to give the model a sound statistical basis. Berny (1984) outlined a series of new models which appear to overcome the cited shortfalls of the DEF, cubic and logistic type models.

3.11 Classical growth curve models

The major area of work in this field has concentrated on variants of the logistic and Gompertz equations related to the generalized modified exponential and its variants. These types of models are explored and surveyed in detail by Stone (1980), who also considers other types of models excluding those in section 3.10.

Current "vogue" inclines to the stochastic methodologies e.g. Oliver (1987). However all these techniques rely in varying degrees on deterministic models. The observation made by Berny (1984) and implied by Tucker (1986) is that, in order to make a further advance in this field, there is seen to be a need to reconsider the basis of generating deterministic models prior to utilising stochastic methods. In addition models of the type cited in 3.10 have apparently been overlooked.

3.12 Literature survey

Prior to further developments and full evaluation of possible new models, a wide literature survey was conducted by the author, including a search of abstracts using the European Space Agency Information Retrieval Service (1985).

The following observations were made :-

1) The great majority of fundamental research was concerned with variants of the family of modified exponential models.

2) The source data of recent research emanated primarily from the areas of marketing/economics and biology.

In the author's opinion, the above observations indicate that previous areas of research have been limited in their scope when many other areas of application exist. A simple example, which demonstrated the spectrum of the potential impact on relevant problems in the scientific world, was the study of nuclear isotope decay. The decay process may be treated as the inverse of growth e.g. growth models could aid in determining the half life measurement of an isotope and its intermediate daughter products.

The conclusion drawn from the literature search was felt to confirm the need for a fresh look at the underlying concepts of growth curve modelling.

3.2 Growth curve concepts

In this section the basic definition of a growth curve will be discussed and reformulated. The term 'S-curve' has been specifically avoided as growth may be represented by curves showing no 'S-shape' and hence the generality of definition would be lost.

3.20 Earlier definitions relating to growth curves

Resulting from the investigation of past and current literature, it was concluded that no clear and unambiguous definition of a growth curve was readily available. Two examples of definitions are given below.

3.21 Dictionary definition

The McGraw-hill dictionary of Scientific and Technical terms refers to a microbiological and nuclear definition :-

"A curve showing how some quantity associated with 'an induced nuclear reaction' increases with time."

3.22 Technical definition

Rektorys (1969) in the book 'Survey of applicable mathematics' gives the following definition in the section on growth curves.

Using the notation of Rektorys :-

"A solution $x=F(t)$ of the differential equation $dx/dt=f(x)$, (where t=time and x represents the growth curve) is called the 'law of growth' assigned to any phenomenon observed to satisfy the equation."

Rektorys then makes three assumptions :-

a) the necessary parameters involved in $f(x)$ are to be established for the phenomenon on the basis of statistical data.

b) the growth of the quantity x in time takes place without external intervention

c) the initial condition $t=0$, $x=x_0$ holds

Rektorys's definition can be used to describe a major traditional group of models.

By way of example two models are derived :-

The logistic may be derived as follows :-

given :-

$$dx/dt = cx(1-x/a)$$

then $x = a/(1 + b\exp(-ct))$ ...eq.1

where a,b,c >0 and b is derived from the constant of integration

The differential form of the error function (DEF) is derived as follows :-

given :-

$$dx/dt = 2gt(K-x)$$

then $x = K(1 - \exp(-gt^2))$ ...eq.2

where K,g >0

Both the above definitions omit the finer details of growth curves which have been observed. For example some projects, after smoothing to remove 'random variations', exhibit a bimodal period cost profile. The significance of omitting bimodality is that the quality of forecasts deteriorates.

Multimodality generated by cyclical effects was considered and has been shown Berny (1985) that it may be modelled. However, as only data which exhibits bimodality could be found the cyclical cases were excluded from the bulk of the thesis.

3.23 Definition of a growth process and its curve

In the absence of a known general definition of growth it was perceived necessary to establish some clear terms of reference to enhance the understanding of this research.

Growth curve

It is this Author's contention that in general a growth curve describes how some quantity, associated with a growth process, increases with time.

This normally implies that the measured quantity is given in cumulative form e.g. total production or project expenditure to date.

> A growth process is further defined as depicting the life cycle of a phenomenon described by either a normal growth pattern or a combination of normal growth and other secondary growth processes, in such a way as to always exhibit a growth curve.

> Normal growth is assumed to consist of a birth, maturity, death cycle which excludes other processes of growth and has the property of always increasing with time.

Secondary growth processes are described as combinations of other processes which may or may not have the properties of normal growth. Thus the secondary processes consist of a series of internal and or external elements which combine to influence the overall pattern of normal growth.

These concepts expand on the current basic model formulation, which only uses the normal growth definition as originated by Gompertz in the early 1800's and summarised by Rektorys (1969), see 3.22. The main thrust of current theoretical work is primarily concerned with stochastic techniques and not in the methodology of building the underlying deterministic models.

3.24 Mathematical formulation of a growth curve

The above definitions lead to the following mathematical formulation :-

F^*_t = measured value at time t

Y^*_t = model estimate of value at time t

If Y^*_t is described by a function $Y^*(t)$ then :-

$Y^*_t = Y^*(t)$

must be such that for $t >= 0$, during a life-cycle, at time t' greater than t the value of $Y^*(t')$ must be greater than $Y^*(t)$.

Assumptions attached to the definition of $Y^*(t)$:-

3.24a Parameters of the model

The necessary parameters involved in $Y^*(t)$ have been established based on data describing the growth phenomenon.

3.24b Initial conditions of the growth process

The start of a growth process will be assumed to begin at time=0, thus the growth curve will always start at the origin. As the precise start of some growth processes is unknown, it is proposed that in such circumstances the initial value is set equal to the first measured value F^*_0. This convention does not prevent one from using a model to backtrack and find the time when $F^*_0=0$, should this be required.

Based on the above it is convenient to define :-

$$Y_t = Y(t) \text{ such that}$$

$$Y^*_t = Y^*(t) = Y_t + F^*_0 = Y(t) + F^*_0$$

with $Y(t)$ having the growth property defined for $Y^*(t)$ and $F^*_0 >= 0$

Note :-

For comparative studies account is taken of the fact that the traditional models start with time unity i.e. $t = 1$

3.24c Growth curve components

Following the definition that a growth process consists of a normal growth component and may also have a secondary process, it is necessary to define each of these components by separate general functions as below.

$$Y = Y(t) = N(t) + S(t),$$

where $N(t)$ is the normal growth component

and $S(t)$ describes the secondary processes.

It should be noted that this formulation does permit these two functions to have common parameters. Nevertheless, as a secondary process is not seen to be an essential part of all growth processes, only that the normal growth component is required to always increase with time during the life-cycle. However, it is necessary that the secondary process must be such that the growth nature of the total process is maintained see figure 3.1. This results in the following :-

at time $t'>t$ both

$Y(t')>Y(t)$ and

$N(t')>N(t)$ during the life-cycle

The graph in figure 3.1 and 3.2 (see end of 3.42a) shows modified data from a 'shop' project. It consists of three components : acceleration / deceleration (circle), a growth curve labelled DEF (box) and initial slope curve (diamond); these are shown in cumulative form in the upper curve (dot).

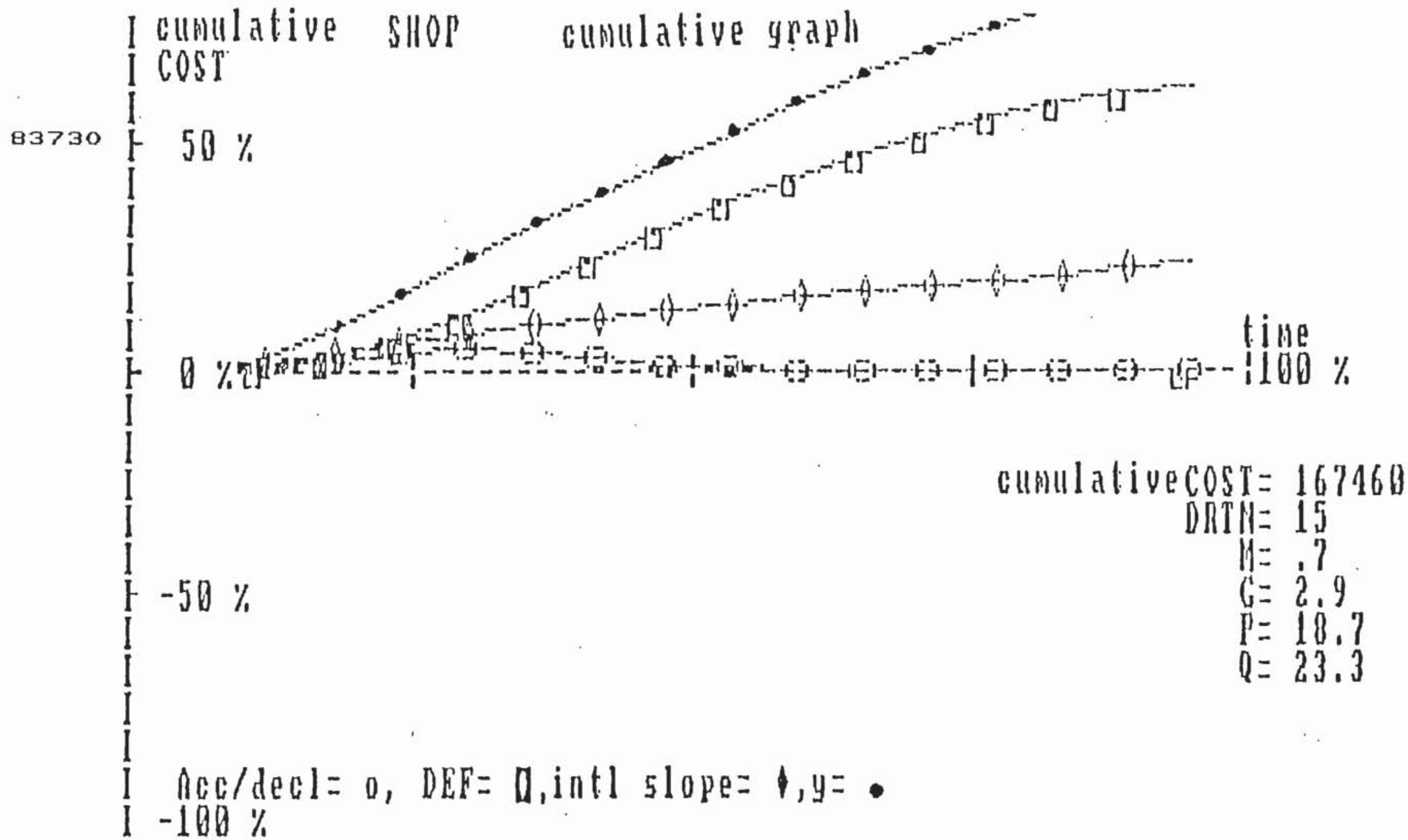


FIGURE 3.1 *Cumulative graph of 'shop' project showing the normal and secondary growth curve components*

3.24d Period value properties

The term "period" is used in the open-ended sense that it is a time unit whose value is specified by the user, but intervals do not have to be equally spaced if information is missing. The period values may be represented by the rate of change of Y(t) providing this is a continuous function. Following the properties set down in (c) above the rates of change of both Y(t) and N(t) must always be non-negative. The rate of change of the secondary process may be negative providing, that during the life cycle, the total growth model function Y(t) has a positive (i.e. non-negative) rate of change.

This is shown below :-

$dY/dt >= 0$ and likewise

$dN/dt >= 0$ for all $t >= 0$,

during the life cycle

3.24e A specifiable point on the growth curve

In growth phenomena it is possible to specify some point through which the growth curve may pass, apart from the point given by the initial conditions (see 3.31). If this point is specified to occur at a time t=D then one may associate with it a value Y^*_D. For most forecasting problems it has been found that it is best to select this time point, such that its value is close to or equal to the last known measured value of the growth data.

If it is desired to identify the cycle end it is suggested that, providing the growth curve has a finite upper bound (saturation level), the growth can be defined as ceasing when 99% of the saturation level is reached or some other value close to 100% specified by the forecaster on practical grounds. The only exception to this is when the end of the cycle can be estimated as a time and value, e.g. the contractual end of a project. In such a case, e.g. when considering budgets, it has been found useful to set this specified point, through which the model curve will pass, to have the value and time of the perceived cycle end.

In the traditional models no point through which the growth curve passes is specified. Although a marginal loss of accuracy was observed in some curve fitting by the imposition of specifying such a point, this was considerably outweighed by the many advantages gained.

Thus one may define :-

A specified point occurs at time t=D and has a value Y^*_D

or $Y_D = Y^*_D - F^*_0$

The definition of a growth process and its curve (see3.23) and its mathematical expression given in this section leads to several observations which can be made. The following notes consider the manner in which these concepts remove potential restrictions which may be found to exist in traditional models such as the family of modified exponential.

Observations on the general model

The above definitions do not require as a prerequisite :-

a) that the growth should necessarily end in an asymptotic manner.

b) a specification of points of inflection, thus the growth curve does not necessarily have to be S-shaped. Analysis of 50 projects has demonstrated that the 'S' curve is not always valid, in this instance it was valid in five out of six cases.

c) that the nature of the secondary process, if it exists, has properties similar to normal growth.

d) the measured value of the phenomenon say v should be the same as F e.g. F=ln(v) may be advantageous if growth is very rapid.

3.25 Comparison with Rektorys's definition

The above definition deviates from that of Rektorys in two major respects :-

i) It is not seen as a prerequisite for formulating models of the components of the growth process N(t) and S(t) that these should be derivable only from a differential equation as opposed to being empirically deduced.

ii) the assumption that growth occurs without external intervention has been omitted. For example, if one considers delays in project site work, these are frequently accounted for by external influences e.g. the weather, which thus constitutes part of the growth pattern. In the study of the tractor economy of Spain Mar-Molinero (1980) showed the politico-economic influence on the production level of Spanish tractors.

3.3 New growth curve models

The models which can be derived from the definition of a growth curve developed in this study cover a large number of possibilities. In order to retain a viable basis for curve fitting, it was decided to only consider functions with normally four but not more than five parameters. The objective was to find models exhibiting relatively straightforward mathematical properties which also incorporated a means of interpretation. As a result of studying many alternative models, it was found that the DEF gave the most promising basis for model building. From this standpoint one of its important merits was that it only has two parameters.

3.30 Model structure

When considering the formulation of a model, the detailed implications resulting from the definitions given above need to be explored more fully. This section considers in depth the consequences of the proposal that the growth curve passes through a defined point and questions the need to specify one point of inflection. It also suggests a means of reducing the effects of discontinuities.

3.31 Specified point of a growth curve and its consequences

If a growth curve must pass through a point at time $t=D$, this has the result that all models of the form $Y= Y(t) = kg(t)$ may be simplified as follows :-

$$\text{if } Y = Y_D \text{ at } t = D \text{ then}$$
$$k = Y_D/g(D)$$

Under these conditions k is no longer a parameter of the model, but a function of the specified value Y_D at time D and the remaining parameters of the function g(t). This implies that all models which can be reformulated as above, lose one parameter.

The above reformulation has been tested on the logistic curve. It was found, in common with the DEF, that the sum of squared errors (SSE) resulting from curve fitting, using the calculated value of k, was compatible with the SSE values obtained when the value of k was estimated by conventional least square methods.

The parameter k is the scaling constant of the model or saturation level for 'S' curves. It is conventionally evaluated by an algorithmic search process (minimising the SSE) used to derive growth curve parameters.

The new types of models may be formulated with several linear constants linking the components of the growth function. For this reason an alternative means of deriving these constants was considered.

Inspection of the method of least squares showed that a general simplification to algorithmic procedures could be formalised for the purpose of evaluating linear constants.

The remainder of this sub-section considers both the evaluation of linear constants and the special case of the saturation level and its properties.

3.31a Statistical derivation of linear constants

Consider an expression of the form :-

$$Y^*_t = a_1 f_1(t) + a_2 f_2(t) + \dots + F^*_0$$

$$\text{if } SSE = \sum (F^*_t - Y^*_t)^2$$

then if all parameters in the functions $f_i(t)$ are held constant the minimum of SSE is given by :-

$$dSSE/da_i = -2 \sum F^*_t f_i(t) + 2 \sum f_i(t) Y^*_t = 0$$

Thus a set of simultaneous equations is formed, which can be solved by the usual matrix methods to evaluate all the linear constants in the above generalised expression of Y^*_t. Hence the construction of algorithms is simplified, the main task being to find nonlinear parameters.

The solution is simplest for the case of $Y(t) = kg(t)$ giving a solution for k as shown below.

$$k = \sum\{F_t g(t)\} / \sum\{g(t)^2\}$$

where $F_t = F^*_t - F^*_0$,

F^*_t being the quantity associated with the measured value of the growth curve at time t.

This procedure was found to consistently give marginally better results including comparisons with those published by Meade (1984).

3.31b Consequences of deriving the saturation level

As can be seen, the above developments not only produce a simplification in terms of standard procedures for parameter estimation such as that of Marquardt used by Madison Academic Computing Centre at the University of Wisconsin (MACC 1984), but more importantly show that :-

a) The majority of traditional models can be reduced by one parameter.

b) Algorithmic fitting procedures can be simplified.

c) The saturation level (Y^* at time infinity) becomes a function of all the remaining parameters i.e. only of those in g(t). For this reason comparison with known research, which considers in its analysis of results the saturation level, is rendered less meaningful in this context.

d) The procedure for calculating the variance-covariance or information matrix has to be altered, described in appendix 1.

e) Due to the altered nature of the saturation level, an alternative procedure had to be formulated for evaluating the potential outcome of forecasts. This was resolved by estimating the standard error (SE) of predicted results. An approximate estimate was calculated based on Gregg et al (1964) and Williams (1959) (see appendix 1) and SE values calculated 10%, 30% and 50% into the future i.e. after the last measured value.

This procedure had a further consequence as the SE values depend on the variance-covariance matrix and the SSE, hence they are an embodiment of the latter data and thus permit one to more simply evaluate this statistical information.

f) The final consequence of this parameter loss was that the value of k is not rigidly determined. As has been stated earlier Y_D and D may be set by the forecaster at a point which is seen to be most appropriate for the problem in hand, thus one is left with an element of control over the value of this constant.

The facility to set a point through which the curve may pass, is advantageous in projects which usually have contractual requirements. This feature enables the forecaster to give estimates of costs, which bear a realistic relationship with the project cost and duration.

3.32 Points of inflection

The majority of research on S-shaped curves pays attention to a single point of inflection, because this is a unique property of traditional models and gives them the 'S' form. Experience has shown that this emphasis needed to be reassessed since some of the construction industry projects have shown the need to consider bimodal patterns of periodic growth. Bimodality requires that three points of inflection need to be considered, that is two peaks and a dip between them. In these circumstances there is no way of knowing which peak is the most significant, if indeed it is valid to ask such a question.

For the above reasons, models were formulated with and without specifying conditions for a dominant peak. All the models in this category showed merit, when undergoing initial curve fitting tests.

3.33 Discontinuities

It is not uncommon to find discontinuities in growth curve data. If these are known to exist then it seems appropriate to artificially suppress these features of the data. For instance in the Construction Industry, projects frequently show discontinuities due to holidays when the period costs are very small or zero and occasionally due to extra high period costs when expensive plant is being installed.

The artificial suppression of discontinuities was based on the fact that these events were already known and hence could be excluded from model considerations. The periods when these discontinuities occurred were suppressed from analysis and reinstated in the final growth curve. Furthermore, it seems appropriate that this methodology should also be adopted when using data to find parameters for the purpose of forecasting.

3.34 General form of the new model

The final part of this section considers the most general form of the proposed model. Experience has shown that at least two significant secondary elements exist. Discussions with industry support the view that many construction projects start with high period costs which slow down and subsequently speed up. These circumstances may be the cause of bimodality. A less common case of bimodality occurs when a project appears to be conducted in two major phases. This effect is observed some time after the project starts and is seen as a deceleration in period costs followed by acceleration.

3.34a Normal growth component

While the two parameter form of the logistic or other traditional models could have been used for N(t), the four parameter limit imposed (ref 3.3) would have restricted the nature of the functions which could be selected for the secondary process. For this reason the DEF, which can be treated as a one parameter growth model, was used. This then released three parameters for models of the secondary processes.

The form of N(t) used was :-

$$N(t) = H(1-\exp(-Gt^2)) \qquad \text{...eq.3}$$

such that $H,G > 0$, $N(0) = 0$ and at infinity $N(\infty) = H$

Its differential is :-

$$dN/dt = 2GHt\exp(-Gt^2) \qquad \text{...eq.3a}$$

thus $dN/dt >= 0$ for $t>=0$ and hence obeys the stipulated conditions in the definitions.

In the next section a form of modified exponential will be examined. As this function has the defined properties of growth it could validly replace the DEF as a normal growth component of the growth curve. The general form of the model being described does allow for this feature. However all tests to date but one (2%) show that the DEF describes the normal growth component. It is furthermore generally superior to, for instance, the logistic. The DEF's peak properties are more variable over a wider span of peak times when the DEF rather than the logistic is combined with secondary processes.

3.34b <u>Secondary Process</u>

In order to adequately account for the difficulties associated with project cost profiles e.g. high initial cost outlay, it has been found necessary to introduce three secondary elements, see figure 3.1. Tests comparing traditional models with those based on these concepts have shown an improvement in all the case studies, in some cases this improvement is quite significant.

(i) Initial slope process

In the introduction to this section it was noted that most construction projects start with a high capital outlay. It was further observed that these initial costs appeared to fall off exponentially. The later periods of the projects were generally found to show a period cost profile typified by the normal growth component.

It has been suggested by professionals within the construction industry that these early high period costs are due to the expenses associated with installing the site setup. Within the area of marketing Lewis (1982) points out that many traditional models cannot account for high product sales that occur when the item is launched with a well timed advertising campaign.

This process also has the effect of giving the growth curve a high initial slope, it is represented by the function, based on the modified exponential, shown below.

$$S_1(t) = J(1-\exp(-Mt)) \qquad \text{...eq.4}$$

such that $J, M >= 0$

When J or M are zero this term is lost i.e. the growth process has no initial slope thus growth starts very slowly.

At infinity $S_1(\infty)=J$, $S_1(0)=0$ and its initial slope=JM as

$$dS_1(t)/dt = JM \exp(-Mt)$$

As $dS_1/dt > 0$ for $t>0$ it has the required properties of growth.

However, as stated earlier, based on experience this function behaves generally as a secondary element.

An important result of using this term is that it increases the saturation levels, it may be shown that Y=J+H for t>>0. This property answers, to some degree, the criticism made about traditional models, namely that they are frequently criticised for under-estimating reality e.g. Meade (1984).

(ii) Accelerator/decelerator processes

Apart from the aforementioned need to describe the effects of bimodality, there is also a need to control the rate of climb to the peak activity point in the unimodal case. This latter effect overcomes a weaker aspect of the DEF.

The period value profile of the DEF was used to describe this process as this particular curve (see equation 3a) would only give a viable increase of growth over a limited time span during which this curve shows acceleration. Deceleration may then be described by the negative form of this curve (-dN/dt) as this is the reflection in the X-axis of the acceleration component. The combination of these elements then describes this secondary process as shown below.

$$S_2(t) = Rt \exp(-Pt^2) - St \exp(-Qt^2) \qquad \text{...eq.5}$$

such that R,P,S,Q >=0

This function has many properties and the ones which have been found most useful will be discussed in section (3.4). It can also be removed from the model, since for instance when R=S and P=Q the function vanishes. A minor drawback is that this function and its differential can have negative values.

3.35 Further simplifications

The present methodology shows it is advantageous to specify a period and its corresponding value through which the curve passes. This can lead to a simplification in that, if this point is set to have a value of unity and that the time expended to this point is also unity, all data may then be expressed as a proportion of the time and value at the "unity point" ($t=D$ and Y_D). A further consequence of this simplification is that the SE (Standard Error) of curve fitting can be standardised to a percentage form resulting in a measure which is best described as the Standard Percentage Error (SPE). The SPE has the added advantage that results of fitting growth curves can be more easily compared, regardless of the field of study to which they relate.

This proposal leads to the need to give some further definitions :-

The last measured value is :-

F^*_d at time $t=d$ or $F_d=F^*_d-F^*_0$,

thus there are normally $n=d+1$ observations.

It was found necessary to define F_d separately from the unity point value at $t=D$ as the latter is alterable and would thus vary the SPE values. Usually, in practice, $D=d$ and $F_d=Y_d$ with the exceptions of the use of D as the project duration when $D>d$ and when testing forecasting ability when $D<d$.

The proportionate measured value of the growth curve (f_t), with the starting value excluded is :-

$$f = f_t = (F^*_t - F^*_0)/Y_D$$

and occurs at proportionate time $x=t/D$

The proportionate predicted value y_t is :-

$$y = y_t = (Y^*_t - F^*_0)/Y_D = Y_t/Y_D$$

Thus the SPE is given by 100 times SE divided by the value of the last data point :-

$$SPE = 100\ SE/F_d$$

or

$$SPE = 100\ Y_D\ sqr\{ \sum_1^1 (f-y)^2/n \}/F_d$$

$$= 100\ sqr\{ \sum_1^1 (F^*-Y^*)^2/n \}/F_d \qquad \ldots eq.6$$

Similarly the mean error (ME) may be replaced by the Mean Percentage Error (MPE) :-

$$MPE = 100\ ME/F_d$$

or

$$MPE = 100\ Y_D\{ \sum_1^1 (f-y)/n\}/F_d \qquad \ldots eq.6a$$

3.36 General formulation of the model

The general form of the proposed model is derived by combining the normal growth component with the secondary process. This may then be simplified into the proportionate form.

The simplest formulation combining equations 3 to 5 is :-

$$Y^*_t = F^*_0 + H(1-\exp(-Gt^2)) + J(1-\exp(-Mt)) + t(R\exp(-Pt^2) - S\exp(-Qt^2))$$

or omitting the initial condition :-

$$Y_t = H(1-\exp(-Gt^2)) + J(1-\exp(-Mt)) + t(R\exp(-Pt^2) - S\exp(-Qt^2)) \quad \ldots eq.7$$

In the above equation the functional representations are in sequential order as follows (see figure 3.1) :-

$H(1-\exp(-Gt^2))$ is the DEF normal growth component

$J(1-\exp(-Mt)$ is a modified exponential type function which represents the initial slope, but may occasionally behave as the normal growth component in exchange for the function above.

$tR\exp(-Pt^2)$ is the accelerator element and the last function is its complement the deceleration element.

Based on the proportionate definitions one obtains from equation 7 :-

$$y = Y/Y_D = h(1-\exp(-gx^2)) + j(1-\exp(-mx)) + x(r\exp(-px^2) - s\exp(-qx^2)) \quad \ldots eq.8$$

where :-

$h = H/Y_D$, $j = J/Y_D$, $r = RD/Y_D$ and $s = SD/Y_D$

and as $x = t/D$

then $g = GD^2$, $m = MD$, $p = PD^2$ and $q = QD^2$

It is this formulation (equation 8) which has been explored in this study.

In equations 7 and 8 all parameters are non-negative, with the exception that :-

if G or g = 0 then M or m>0 i.e. the modified exponential is the normal growth component

if M or m = 0 then G or g>0 i.e. the DEF is the normal growth component

A mathematical drawback of these types of models is that some parameter combinations give nonfeasible curves which do not exhibit growth. But these types of models retain a high degree of variability when compared with the cubic and the non-feasibility only occurs when parameters p and q are used. For this reason all solutions using the accelerator/decelerator process have to be checked for feasibility during the growth lifespan, this has been found adequately catered for by simulating 300% into the future.

3.4 Derivation of viable models

The proportionate formulation (equation 8) will be used as the basis, for the remainder of this thesis for the proposed models. Four models will be presented below, and are presented in two pairs.

3.40 Model types

a) Dominant peak specified

Each one of the first pair of models has one dominant peak specified (see earlier explanation 3.32).

b) Peak unspecified

The next two models have unspecified peaks permitting critical freedom which allows such models to adapt more naturally to a phenomenon. Under these circumstances either the normal or a secondary process may be dominant, leading to more favourable results.

In order to fully utilise the unity point, which allows one parameter to be removed from the models, the proportionate formulation (equation 8) will be restated.

Equation 8 was given as below :-

$$y = h(1-\exp(-gx^2)) + j(1-\exp(-mx)) + x(r\exp(-px^2)-s\exp(-qx^2))$$

This model may be restated in the form 'y = kg(x)' equivalent to Y = kg(t) (see 3.31)

$$y = k\{h(1-\exp(-gx^2)) + j(1-\exp(-mx)) + x(r\ \exp(-px^2)-s\ \exp(-qx^2))\} \quad \text{...eq.9}$$

Thus at the unity point where y=1 when x=1 the constant k is given by :-

$$k = 1/\{h(1- \exp(-g)) + j(1-\exp(-m)) + r\ \exp(-p) - s\ \exp(-q)\} \quad \text{...eq.10}$$

3.41 Models with a specified peak

These models require the dominant peak to be specified, which adds one parameter to the model. This is compensated for by one of the linear constants being replaced by an expression which is a function of the remaining parameters. This expression is deduced by solving, at the point of inflection, the equation derived by setting the second differential to zero.

3.41a Three parameter model

A three parameter model may be formulated as a function of m and g and a dominant peak. This model has only been rudimentarily tested with particularly sensitive data which gave promising curve fit results. As it caters for the most commonly found secondary process, the initial slope element, it may prove worthy of further exploration.

In order to remove one parameter, the nature of the initial slope process was considered further. It could be argued that the initial slope element, which controls high initial cost outlays, depends not only on (m) the increase it gives to the costs generated by the normal growth process, (see 3.34b) but that the rate of decline of this element is controlled by this same quantity i.e j=m. Thus the initial slope $(dy/dt)_{t=0}$ is proportional to m^2.

A further four parameters are lost by setting r=s=0, which causes the accelerator/decelerator element to be suppressed. Finally the linear constant h is replaced as a function of m,g and the point of in time when the dominant peak (t_{pk}) occurs.

Thus equation 9 may be stated as below :-

$$y = k\{m(1-\exp(-mx)) + h(1-\exp(-gx^2))\} \quad \text{...eq.11}$$

If the dominant peak is at t_{pk} or $x_{pk} = t_{pk}/D$ then solving at $x = x_{pk}$:-

$$d^2y/dx^2 = k\{-m^3\exp(-mx) + 2hg(1-2gx^2)\exp(-gx^2)\} = 0$$

gives $h = m^3\exp(-mx_{pk}+gx_{pk}^2)/(2g(1-2gx_{pk}^2))$

such that $m,g > 0$ and $x_{pk} >= 0$

Note :- m and g in this model cannot be zero but x_{pk} could have the unlikely value of zero and thus $h = 0$.

The above expression for h and the earlier one for k (equation 10) renders $y = f(m,g,x_{pk})$ i.e, equation 11 has only three parameters.

As can be seen, the above expressions are rather complex. This situation becomes more severe when considering the four parameter model. For this reason the following simplification has been found useful.

If the parameters are replaced by upper case letters and x replaced by $X = x/x_{pk}$ then, the point of inflection occurs at X=1. Equation 9 and its parameters, shown in 3.4, are transformed as follows :-

$$y = K\{H(1-\exp(-GX^2)) + J(1-\exp(-MX)) + X(R\exp(-PX^2) - S\exp(-QX^2))\} \quad \text{...eq.12}$$

where $m = M/x_{pk}$, $g = G/x_{pk}^2$, $p = P/x_{pk}^2$, $q = Q/x_{pk}^2$

and $y = 1$ at $x = 1$ or $X = 1/x_{pk}$, for the case t=D

The format of H,R and S have been left unspecified as these parameters are separately definable.

The parameter J may also be left unspecified, but if j=m then $J=M/x_{pk}$ for the above model and R=S=0, then H is re-evaluated.

Thus the three parameter equation 12 is transformed to :-

$$y = K\{M(1-\exp(-MX))/x_{pk} + H(1-\exp(-GX^2))\}$$

...eq.13

with $X = x/x_{pk} = t/t_{pk}$

giving $H = M^3\exp(G-M)/(2Gx_{pk}^2(1-2G))$

and $K = 1/\{M(1-\exp(-M/x_{pk}))/x_{pk} + H(1-\exp(-G/x_{pk}^2))\}$

3.41b Four parameter model

This formulation was devised in order to test the validity of using a dominant peak. Evidence appears to indicate that superior results are achieved by not seeking to find a dominant peak, however this question is by no means fully explored hence its inclusion in this report.

This model is generated by the following simplifications:-

$X = x/x_{pk} = t/t_{pk}$, $J = M$, $P = G+Q$, $R = S$

such that R is evaluated by solving $d^2y/dx^2 = 0$

when $X = 1$ or $x = x_{pk}$

The simplest form of the resulting expression is :-

$$y = K\{M(1-\exp(-MX))/x_{pk} + (1-\exp(-GX^2))(1-RX\exp(-QX^2))\}$$

...eq.14

and at the unity point ($X=1/x_{pk}$) :-

$$K = 1/\{M(1-\exp(-M/x_{pk}))/x_{pk} + (1-\exp(-G/x_{pk}^2))(1-R\exp(-Q/x_{pk}^2)/x_{pk})\}$$

with

$$R = \{M^3\exp(Q+G-M)/2x_{pk}-G(1-2G)\exp(Q)\} / \{(Q+G)(2Q+2G-3)-Q(2Q-3)\exp(G)\}$$

such that $M,Q >= 0$ and $G,x_{pk} > 0$

A special case arises, when G=.5 then y is independent of the accelerator/decelerator process.

The fact that the M and/or Q may be zero allows the model to have two to four parameters, thus one may generate the DEF if M=Q=0. The fact that there are three possible solutions requires the need to create a selection process. This requirement is taken into account in the algorithmic search procedure.

3.42 Models without peak dominance specified

The formulation is simpler when peak dominance is unspecified and thus more easy to interpret, since the earlier form of proportionate model (equation 9) may be used. The simplification of the initial slope constant holds i.e. j=m. As the normal growth pattern is the DEF it was seen as reasonable to set h=1. The premises upon which r and s are simplified will be left to later. The difference between the two forms of these four parameter models is that the accelerator and decelerator processes are treated independently in the first model and dependently in the second.

The potential drawback of the retaining the independence of these processes is that they affect the initial slope thus clouding the principle of the j=m simplification see 3.41. For this reason the last model allows a clear distinction to be made between the two secondary components. In the absence of hard evidence to clearly distinguishes between the three "four parameter" models, for the planned evaluation of the practicality of the concepts given, the last model was chosen on the theoretical premises outlined.

Various other simplifications of r and s have been considered, a quite effective substitution was to set them both equal to unity. Closer inspection of the properties of the accelerator/decelerator elements, discussed below, led to the formulation of the last two models.

Accelerator and decelerator elements

In order to establish a simplification of the linear constants r and s of the accelerator and decelerator elements, it is helpful to assume that the initial slope is zero (m=0). Under these conditions a direct comparison may be made between the relative values generated during a life cycle by the normal growth (DEF) component and the remaining secondary process.

The greatest impact that the secondary process can have on the DEF component is at the points when its elements have extreme relative values, the DEF having a maximum of k. The accelerator maximum occurs at $x_p=1/sqr(2p)$ with a value of $krx_p exp(-.5)$ approximately equal to $.6krx_p$, thus the ratio of this maximum to that of the DEF is approximately $.6rx_p:1$.

Experience has shown that good results are obtained by setting :-

$$r = 1/2x_p = sqr(p/2) \quad (\text{as } p=1/2x_p^2)$$

$$\text{hence } rx_p = 1/2 \text{ giving a ratio of } .3:1.$$

This means that the maximum of the DEF curve exceeds this element by up to 70% and this result is independent of the value of x_p. Clearly this condition may contradict the growth curve definition, however it must be seen in contrast with the decelerator element. This element has a minimum at $x_q=1/sqr(2q)$ and if one accedes to the analysis used for the accelerator then :-

$$s = sqr(q/2) = 1/2x_q \text{ and the ratio } = -.3:1.$$

Thus if the extrema of the elements of this secondary process are close together the overall deviation may be quite small. Furthermore as all parameter sets are checked for feasibility the implied contradiction is only academic.

3.42a Independence of accelerator and decelerator elements

This model required the following substitutions for linear constants :-

$h = 1$, $j = m$ for the DEF

and initial slope components (modified exponential)

$r = 1/2x_p = sqr(p/2)$ where $x_p = 1/sqr(2p)$

which is the time when the accelerator element is at its maximum

$s = 1/2x_q = sqr(q/2)$ where $x_q = 1/sqr(2q)$

which is the time when the decelerator element reaches its minimum

In addition to the above it has been found advantageous to derive the time (x_g) when the DEF point of inflection occurs. This may be shown to be at $x_g = 1/sqr(2g)$.

The above then gives the following two forms of this model :-

$$y = k\{1-exp(-gx^2) + m(1-exp(-mx)) + x((p/2)^{.5}exp(-px^2)-(q/2)^{.5}exp(-qx^2))\} \quad ...eq.15$$

$$\text{and } k = 1/\{1-exp(-g) + m(1-exp(-m)) + (p/2)^{.5}exp(-p)-(q/2)^{.5}exp(-q)\}$$

the second form replaces the values of g,p,q by :-

$g = 1/2x_g^2$, $p = 1/2x_p^2$ and $q = 1/2x_q^2$

or if $m=0$ then $g>0$, the latter being the much more common case.

The second form of this equation has the merit that the properties of this model can be directly related to the aforementioned extrema and the DEF point of inflection. Thus when studying changes to parameter values these can be more readily interpreted.

However there is interaction between the component parts, hence a bimodal period curve does not always visually accord with parameter changes, but the growth components may be studied separately as indicated in figure 3.2. see 3.24c (page 33) for annotation.

The special feature of this model, as defined by equation 15 is that each element is separately controlled by its own parameter. This has the drawback that the premises upon which the linear constant of the initial slope were based are invalidated see 3.41a. The value of the slope at $x=0$ was assumed to be proportional to m, however in this case this value at $x=0$ is given by :-

$$dy/dx = k(m^2 + sqr(p/2)-sqr(q/2))$$

Thus the initial slope is a function of both secondary elements and may create difficulties for interpretation in particular, because it can theoretically be negative. The final model may be marginally outweighed in curve fitting terms in some instances, but no adequate evidence is available to suggest that complete element independence describes growth with a sufficient improvement in accuracy to outweigh the use of the next model.

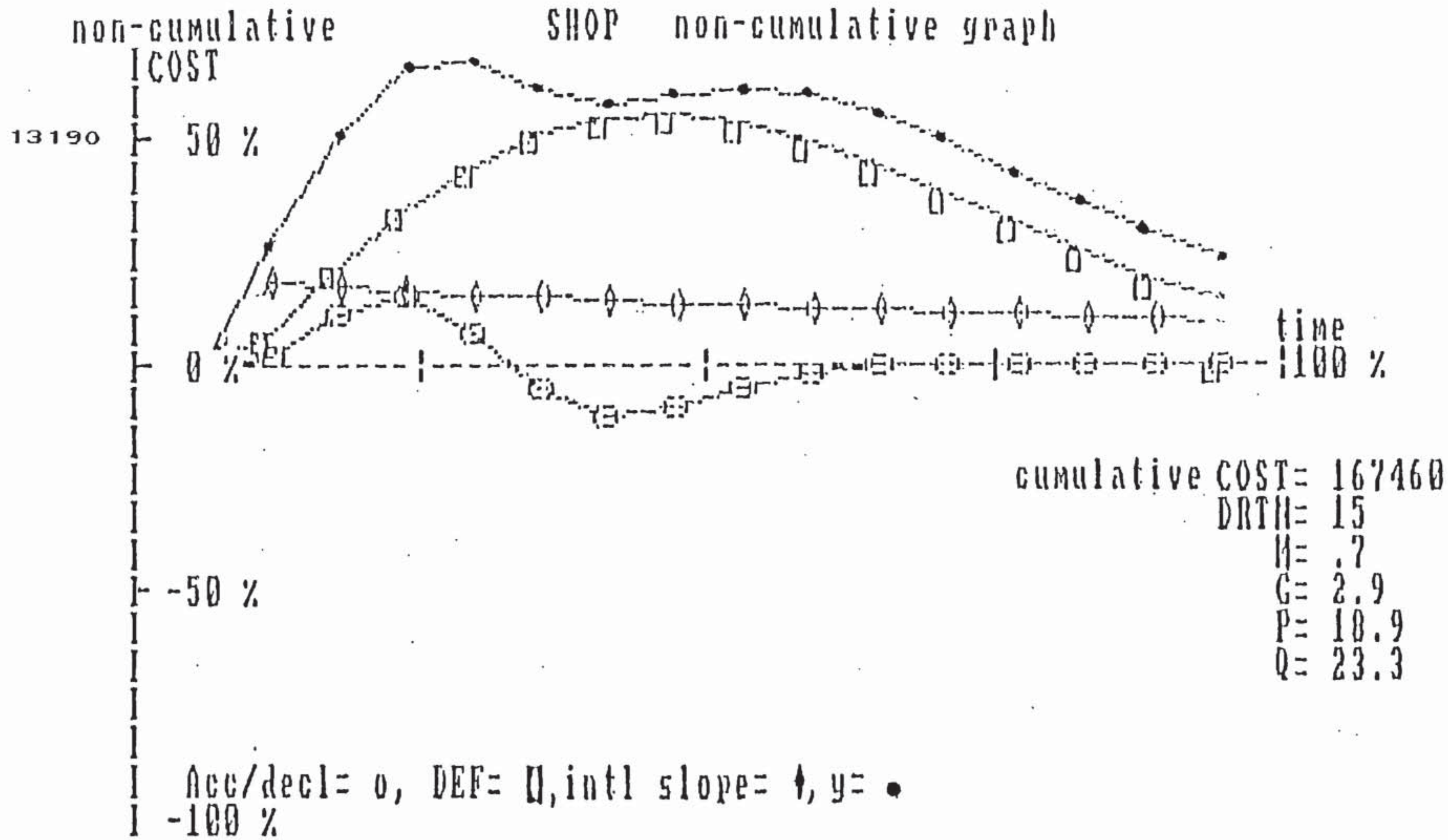


FIGURE 3.2 *Non-cumulative graph of 'shop' project showing the normal and secondary growth curve components*

3.42b Accelerator dependence on decelerator

This model differs from the previous one in that the linear constant of the accelerator is identical to the decelerator constant i.e. r=s=sqr(q/2). This makes the accelerator element dependent on the decelerator.

The effect of this modification is to make the accelerator maximum proportional to the ratio of x_p/x_q. If extrema are considered and that of acceleration (governed by x_p) occurs before deceleration this ratio is less than one. The result will be that bimodality is less pronounced. Since the major cause of nonfeasibilty is the deceleration element, the likelihood that it will occur during the lifecycle will diminish as x_q increases, hence a wider range of solutions should exist. It can be further argued that these processes are likely to work in tandem lending more credibility to this model option.

Thus :-

$$y = k\{1-\exp(-gx^2) + m(1-\exp(-mx)) + x(q/2)^{.5}(\exp(-px^2) - \exp(-qx^2))\} \quad \text{...eq.16}$$

$$\text{and } k = 1/\{1-\exp(-g) + m(1-\exp(-m)) + (q/2)^{.5}(\exp(-p) - \exp(-q))\}$$

Such that p,q,g,>=0 with the proviso that if g=0 then m<>0, p=q=0 or if m=0 then g>0 which is the DEF case.

The saturation level is given at time=infinity by y=k(1+m) except when only p=0 when the upper bound is replaced by a line :-

$$y = k(1+m+(q/2)^{.5}x).$$

The second form replaces the values of g,p,q by

$$g = 1/2x_g^2,\ p = 1/2x_p^2 \text{ and } q = 1/2x_q^2$$

$$y = k\{1-\exp(-(x/x_g)^2/2) + m(1-\exp(-mx)) + x(\exp(-(x/x_p)^2/2)-\exp(-(x/x_q)^2/2))/x_q\} \quad \text{...eq.17}$$

A further aspect of interpretation is, that in both these models the value of m may be changed proportionately, as it has a linear effect on the saturation level. However only in the present model can the initial slope be changed in proportion to m^2, thereby giving an indication of the starting rate of growth. This property is utilised for setting the initial value of m in the algorithm.

It can be seen that this model may have one to four parameter solutions one of which has no saturation level. The wide range of growth curve possibilities combined with its interpretive value have led the author to select this model for the indepth performance tests. A comparison of the properties of the model selected, for exhaustive testing, with those of the traditional models are summarized in table 3.1.

Model	Number of parameters *	Equation number **	pros	cons
Berny	1 to 4	16	Easy reduction of parameters giving DEF, modified and other models. Forecast tests against later observations more than halve SPE given by the most accurate traditional model.	Relies on empirically assessed adjustment of confidence limit results for automatic selection of the best parameter combination.
Extended logistic	3 (2)	18	*** More flexible than DEF component of new model for peaks occurring in middle 50% of life cycle.	Cannot cater for high cost starts. Cannot viably be used at origin. Requires analysis of information matrix to aid in selecting most appropriate model. Saturation level limits use for exponentially growing curves.
Gompertz	3 (2)	19		
Cumulative lognormal	3 (2)	20		

* Bracketed number of parameters gives equivalence with the new model
** Equations 18 to 20 refer to section 4.2
*** Overcome in generalised DEF see Berny 1987c

TABLE 3.1 Comparison of the properties of the classical models and models based on the author's new growth curve conecpts

CHAPTER 4

NEW MODEL TESTING AND ITS USE

This chapter describes the techniques applied to the four parameter model, full analysis of which was described under 3.42b (equations 16 and 17). The same process has been applied to the other models for comparative testing, but some minor details differ; in particular the setting up of the initial search conditions.

4.0 Algorithm

In order to facilitate rapid curve fit solutions, a procedure has been evolved to find an efficient path through local minima leading to an acceptable SPE (Standard Percentage Error) minimum for each of the possible parameter combinations. To achieve this, the search for m and g parameters was made in linear step variations of m and x_g respectively. The limits on the steps being determined from practical experience using the data from Meade's paper and ten other projects. Prior to constructing the algorithm, an interactive routine was set up in order to study the nature of possible good local solutions. This procedure also threw light on the nature of the p and q parameters, which resulted in the use of wide search limits for x_p (p) and x_q (q), namely from .5/d to 1.5d.

The present version of the algorithm has been tested against a large number of highly varied data profiles. Whilst it is known that marginally better solutions can be found, the algorithm's construction was aimed at a broad spectrum of good results. The processing time for the four parameter solution is relatively slow. Experience has shown that the algorithmic analysis takes about 1.5 minutes per period on an IBM PC operating at 4.77 Mherz. However, it is worth noting that this time is more than halved on 80286 based micros. Potentially faster routines can be constructed, but tests of pilot schemes showed that the simpler routine devised was more reliable, effective and robust.

a) Starting values of the search

As the result of experimentation it was found sufficient to use three point smoothing. The starting value for m was determined from half the smoothed value of the first period and utilised the property mentioned earlier that at x=0, $dy/dx=km^2$.

The range of search for values of g was based on the maximum smoothed slope value and an estimate of g from a search minimising the SPE of the DEF using 90% discounting.

b) Search procedure

1. The search procedure adopted was to find solutions for each of the parameter combinations, giving up to 7 equations.

   Solution (1) m or g only

   Solution (2) g and m

   Solution (3) g,q,m=0 or m<>0, case with no saturation level

   Solution (4) g,q,p with m=0 or m<>0

2. The value of m was set equal to zero in situations where only marginal improvements were found to occur with the small, but non-zero values of m.

3. The four parameter search starts with the DEF value of g and searches for a good m solution using wide limits on p and q. This is repeated for a range of 30% about x_g for the starting value of g, using the value of m

found previously. If the solution for g appears to be closer to the maximum slope estimate of the g value, then a repeat search is advocated using the maximum slope value of g, and finally, the half way value between these estimates of 'g' is taken. The use of the two extra starting values for g is a form of the algorithm used only in rare cases.

4.1 Solution criteria

The above search procedure generates seven solutions for which two distinct methods of selecting the best parameter combination have been devised. The first of these assumes prediction will be based on the best curve fit, a method which is fairly widely viewed as inadequate for sound forecasting e.g. Meade (1984). The second is based on a 'fit/forecast' procedure. Whilst the latter is undoubtedly a more rigorous method, it is a more lengthy procedure, hence its use has been given the status of a special option in the commercial package VISIER.

4.10 Overview of selection procedures

A refined form of selection procedure for comparing different models was to inspect the information matrix or variance-covariance matrix. This was seen as impractical in commercial usage as judgement of the credibility of a solution using this matrix is a highly skilled technique. For this reason estimates of the confidence limits were used to replace the above requirement, as they are directly dependent on the information matrix and more readily understood. The confidence limit estimates allowed a fully automated selection procedure to be created in the commercial software.

These limits are directly dependent on the parameter standard errors and are used as the basis for adjudging whether a given model is statistically reliable. In order to simplify future discussion of these quantities three further abbreviations are defined.

Three abbreviations :-

<u>SE of y :-</u>

This is the standard error of a single predicted value of y at some future time t after the last measured value (F^*_d), see appendix 1 for the basis of the SE of y approximation.

<u>SE :-</u>

This is the curve fit standard error = sqr(SSE/n)=SPE*F_d, where normally n= d and F_d=last measured value see 3.1.

<u>SFS :-</u>

SFS is the abbreviation for a statistically feasible solution. This is a solution with parameters whose values exceed their standard errors as deduced from the information matrix see appendix 1.

A selection procedure for determining the best of the seven possible models had to be established. Two procedures were found to be practical, namely a weighting and a fit/forecast method. The latter is a standard procedure, described for instance by Meade (1984), and is traditionally combined with evaluation of the information matrix. The weighting method is more appropriate to models of the kind developed in this thesis as any one general model is the "parent" of the subset of models which it generates. An additional factor was considered, namely the saturation level which is viewed as an important criterion when comparing traditional models eg: Oliver (1980).

4.11 Weighting method

Where curve fitting only was considered, a simple weighting scheme was used for deciding which parameter combination gives the best solution. The information matrix was calculated for each parameter combination. The selected solution has the largest number of feasible parameters and the lowest SPE, with a weighting adjustment related to the number of degrees of freedom.

4.12 Trial fit/forecast method

The more reliable test procedure was to take a trial curve fit using the first 50% to 70% of the known data, after suppressing any obvious discontinuities. Using the above parameters the trial forecast SPE of the remainder of the data was found for each parameter combination. In addition, other statistical measures may be used as support criteria, in particular the mean error and mean absolute percentage error (MAPE).

4.13 Solution selection procedure

When selecting solutions from the trial curve fits, the statistical feasibility of each curve fit is not initially considered, only the trial forecast SPE's are compared. For instance due to the fact that the four parameter case requires generally more than 15 periods for SFSs, non-feasible trial fit/forecast results should not be ignored because they are likely to exclude valid forecasts using all the data.

This premise was based on the assumption that as the number of parameters increased then both the uncertainty in parameter estimates and the level of variation in the prediction increased. The latter implies that a fit/forecast is more likely to give unacceptable confidence limits with a small data subset and a high number of parameters.

This increases the possibility of rejecting a parameter combination only based on the trial fit/forecast data. The fact that a subset of data did not give a SFS did not seem sufficient evidence to reject a solution, because the prediction would only be made from the full data set. Hence it was seen as justifiable to accept forecasts from parameter combinations which gave SFS using the full data set, regardless of the statistical soundness of the fit/forecast.

The solution with the lowest forecasting SPE was then used for parameter estimation relative to the eventual forecast. Should the curve fit using the selected solution still be nonfeasible, then the next best solution was used, supported by the intuitive aspect of forecasting.

4.14 Evaluation of traditional solution criteria

This section considers stochastic solution criteria. These have been embodied in the analysis of the test results of the traditional and new multi-component models.

4.14a Saturation levels

A common guide to the value of a forecast is the saturation level and its SE (value at infinity). In the context of this work these measures can no longer be judged to give the same level of guidance.

For instance, the solution using p=0 has no saturation level, hence only the SE of y can be considered in this case. Secondly the SE of the asymptotes of the remaining new model solutions are a function of all the parameters. Therefore this does not create a sound basis of comparison with the results given by the traditional models which depend on the estimate of k alone.

4.14b Autocorrelation

Mar-Molinero (1980) raised the question of improving forecasts by considering autocorrelation. Based on the conducted tests, the need for autocorrelation was seen as potentially useful in a minority of cases. Experience has shown that in the majority of problems the potential improvement in results was at best marginal, whilst in some the results appeared to deteriorate.

Autocorrelation was found useful as an additional guide in the search for the best parameter combination in difficult cases. For this reason the present algorithm includes facilities to test autocorrelation solutions (see equation 22 appendix 2.13) and to evaluate the Durbin-Watson statistic together with an estimation of the autocorrelation coefficients (first order).

4.14c Variance-covariance or information matrix

As the information matrix and SSE both contribute to the SE of y no advantage was seen to be gained by including such statistical information in this report. This information is readily available as it is a prerequisite to the estimation the SE of y.

In addition it has been found useful to evaluate the SE of the period values of the growth curve, the evaluation technique is shown in appendix 1.

4.2 Case studies

As was indicated in the section describing the area of study, two groups of data were analysed. An in depth evaluation of the case studies quoted by Meade (1984) and two Construction Industry projects. These latter were used to illustrate bimodality and suppression of discontinuities in growth curves. Furthermore the use of the model to forecast given pre-knowledge of a project cost and duration was demonstrated.

In the case of the Meade data, only those traditional models which have given the best solutions were used for the purpose of comparison. These were the logistic, Gompertz and cumulative lognormal.

4.20 Summary of models used.

In order to create a fully consistent comparative basis the relevant models and their notation are defined.

Summary of symbols :-

Y^*_t = predicted cumulative value to time t

F^*_t = measured cumulative value to time t

and $Y_t = Y^*_t - F^*_0$ = predicted value - measured value at $t = 0$

$x = t/D$ and $y_t = Y_t/Y_D$

where Y_D = specified value at $t=D$

n = sample size

such that if the data is measured in equal intervals and starts from $t=1$ then $n=d$, where d = the last period of the trial fit or total sample duration.

Traditional models :-

In order to distinguish between the presented model and the traditional models the parameter symbols for the latter will be a,b,c and e.

Extended logistic (Oliver 1981) :-

$$Y^*_t = a/(1+b\exp(-ct)) + e \qquad \text{...eq.18}$$

Gompertz :-

$$Y^*_t = a \exp(-b(\exp(-ct))) \qquad \ldots eq.19$$

Cumulative lognormal :-

$$Y^*_t = a/(2 \cdot \pi\ c^2)^{.5} \int_0^t \exp(-((Ln(z)-b)/c)^2/2)/z\ dz \qquad \ldots eq.\ 20$$

An interactive routine was used for fitting the above models. The method utilised the procedure described in section 3.31 for the determination of the constant a in equations 18 to 20. In line with methodology generally used when evaluating parameters for the above models the first period is set at t=1. This has the effect that F^*_0 is set at zero.

The new model developed from the research embodied in this work which was to be judged against those defined above, is given by equation 16. This model has had a further generalisation to that given in equation 16 where n replaces the power 2 in the DEF function i.e. $\exp(-gx^n)$ see Berny (1987c). However the simpler form with n=2 was found adequate for the case studies. It is the four parameter formulation which is independent of a dominant peak but has interdependent accelerator and decelerator elements.

Meade (1984) analysed his data using the following equation :-

a) logistic see eq.1

b) Gompertz see eq.19

c) cumulative lognormal see eq.20

d) extended logistic (Meade based on Bass 1969) which failed to improve results in all the studies was :-

$$Y^*_t = (a-b_1 c \exp(-b_1-b_2 t))/(1+b_2 c \exp(-b_1-b_2 t)) \quad \ldots eq.21$$

For the purpose of brevity the statistical measures which are tabulated have been limited to the following list :-

a. The standard percentage error (SPE) and for comparison purposes the SE, $SE=F^*_d*SPE/100$.

b. The mean percentage error (MPE), $ME=F^*_d*MPE/100$.

c. The parameter values and their standard errors.

d. The 50% ahead prediction and its SE of y as a percentage of the former labelled SE* at the head of the tables.

For the purpose of consistency only results derived from the suggested algorithm are presented, although as has been pointed out better evaluations can be found in some instances.

Finally, where published statistical information was incomplete for the purpose of these case studies, the necessary analyses of the Meade data were repeated on the IBM PC which was used for this research. This exercise has led to some minor differences between results. The information matrix data has been fully checked out against the MACC published figures, this was adjudged to confirm the calculating formulation given in appendix 1.

NOTE :-

All the graphical exhibits in appendices 2 and 3 commence with period zero (t=0), hence the traditional models have a one period lag i.e. their period one is at t=0.

4.21 Analysis of the case studies

A detailed analysis of the four supportive traditional case studies is given in appendix 2.

These case studies were necessary to compliment the testing procedure in order to establish that the proposed model is sound in general forecasting terms. By this means the new concepts which form the basis of the proposed model will also be tested. This work also assisted in the general development of the methodology. In particular these cases represented the only information available, published to date, which permit full statistical validation of the selected model, see 3.42b.

A model which cannot be regarded as statistically feasible is one which will permit a very wide variety of cost profiles, all of which appear to have the same curve fit, this renders any attempt to produce a viable model meaningless. It is precisely the lack of this information about the earlier models, including that of the DHSS, that made them highly suspect.

Thus a sound basis was established by comparing performance using data which had been analysed earlier by experts in the forecasting field. The same strict regime was applied to two construction projects. These projects were selected for their highly diverse properties in order to establish an even wider spectrum of applicability than is currently used as a test bed by forecasting specialists.

4.21a Four classical growth curves

As stated at the outset of this thesis the first group of case studies were marketing problems taken from Meade (1984). Meade presents them as a comparison between adoption and consumption processes, which he describes as follows.

Consumption process :-

"A consumption process requires an initial decision to use a product and a sequence of subsequent decisions first to continue usage and secondly to determine the level of usage."

Adoption process :-

"An adoption process requires a single decision by a non-owner to become an owner."

Meade enlarges on the above, stating that in the case of the consumption process it "can fall theoretically to zero" and that this is not the case for adoption. Based on the definition of a growth curve, which describes how some quantity associated with a growth phenomenon increases with time, the consumption process represents the rate of change of the use of this quantity. This leads one to assume that the cumulative value of usage or total consumption to date would be the correct descriptor of the consumption process growth curve.

The above conclusion allows the original definitions to stand, but the analysis requires that the data given in the two adoption case studies may be treated as growth curve measurements. The data presented in the consumption process cases being, annual production levels, must therefore be cumulated to gain the best result. It is understandable how such anomaly has arisen. The S-shape of consumption process data i.e. the graph of annual production against years leads one to assume this is a growth curve. It is suspected that this type of fallacy may well exist in other areas of study.

Significantly the "re-analysis" of Meade's definitions proved to be valid for his case studies and thus rejects Meade's proposal that a consumption process is not a growth process.

4.21b Construction industry studies

Two projects were selected to demonstrate the use of growth curves in project management within the Construction Industry. The first project (A40 roadworks), which had a duration of 31 months, clearly demonstrates bimodality and thus was thought to be a good example of the use of the four parameter solution. The second project (Factory construction) was completed in 18 months and was more typical of medium sized construction jobs.

The analysis follows the same pattern as the previous case studies. In order to achieve this, the time span was split into three portions, the first two being a fit/forecast. The first two portions were then combined and used for a curve fit, from which the last portion of data was forecast to give an indication of the model's full forecasting behaviour.

A further analysis was made using the end cost and project duration for the unity point. This procedure carries one penalty, which is that a discrepancy arises between the forecast value at the point from which the prediction was made and the actual value at this time. For the sake of consistency the two are matched by a simple discounting calculation which has the effect of increasing the SPE value, but in practice the deterioration in SPE was found to be small.

The breakdown of the time divisions for fit/forecasting had to be made judiciously. Sufficient data must be available in order that smoothed pattern can be projected. Any known discontinuities should be removed, which in the case of bimodal data the algorithmic procedure incorporated an analysis to give a warning of new work phases, thus assisting in this value judgement. It was found that the new model was not likely to detect a peak until about 30% from it.

As the cumulative curve profile rises the curve fit procedure has to distinguish between an exponential climb or an 'S' curve climb. Given that the observations do not include a peak, it is nevertheless possible to detect one. This is achieved by analysing the slow down in the rate of climb which indicates if there is a peak. It has been found in practice that the exponential form of the model is predominantly selected, by the curve fit procedure, unless the observations indicate a peak to occur about 50% ahead in time.

In the case of the observations which include a peak the analysis in the last paragraph does not apply. However this property is an important feature of the new model as it explains one aspect of the improved forecasting facility. A further feature of this analysis is to detect a peak after a dip.

The case studies for the construction projects are given in appendix 3.

4.21c Summary of results

The full analysis of the case studies showed conclusively that the new models are significantly better than any current models. The curve fit results showed an average drop of 25% but more importantly the fit/forecasts showed a 60% improvement. These results are summarised in table 4.1

CASE STUDY	Traditional model	FULL - CURVE FIT		FIT - FORECAST		Notes
		SPE (%)	% drop (Berny vs trad.)	SPE (%)	% drop (Berny vs trad)	
PVC	(1) (2) Gompertz	0.53 1.90	72.1	N/R	N/R	Cumulative case see A 2.1
Titanium Dioxide	(1) (2) Cumulative lognormal	2.92 2.88	-1.4 *	5.4 7.1	23.9 *	Should be cumulated as in PVC case hence not comparable in average see A 2.2
Tractors in Spain	(1) (2) Logistic	.94 1.10	14.5	3.8 (1) 14.8 (2)	74.3	10 year forecast see A 2.3
Television sales	(1) (2) Gompertz	.96 1.17	17.9	11.1 (1) 21.4 (2)	48.1	28 year curve 16 year forecast
				.6 (1) 2.7 (2)	77.8	4 year forecast see A 2.4
A 40	(1) (2) Cumulative lognormal	2.30 2.68	14.2	2.1 4.3	51.2	See A 3.1
Factory	(1) (2) Logistic	1.39 1.98	29.8	1.3 2.3	43.5	See A 3.2
Average drop			29.7		59.0	

(1) Berny model
(2) Best of models tested by other authors
* Excluded from average

N/R = not reliable (statistically) hence no comparison

TABLE 4.1 *Summary of the results of case studies*

4.3 Application to project management

This section considers interpretation of project management information by the algorithm. The algorithmic testing procedure is described and general properties resulting from exhaustive testing are evaluated.

4.30 Interpretation

One of the practical uses of the author's new model, in the form given in 3.42b equation 17, is that the effect of changes in parameters can be fairly simply related to the shape of the growth curve. The constants refer to either the starting rate of the project or the points when project activity peaks or drops to a minimum.

An algorithm was constructed to evaluate the parameters of a given model providing knowledge exists of the unity point and that cumulative values during first 10% and half way through a project are known. For the case of a project which slows down during its cycle the 'bimodal' solution of the model is required. The information needed to form such a curve is time and value to date at the two peaks and dip.

The parameters generated by this algorithm allow a curve to be simulated reasonably representing an interpretation of the practical information relating to a project. In this manner this procedure may be used to allow project budgets to be developed from a minimum of information.

It is envisaged that by using a graphical interactive mode the effects of parameter changes can be more easily studied. For instance the normal growth process is likely to relate to the major activities in the project. A linear change in the position of the normal growth peak (x_g) implies a change in strategy. Thus if the peak activity time is moved and the cost profile alters in a favourable manner, this indicates the validity of such a decision.

The graphical interactive mode allows such a strategy to be tested and it may thus also indicate the opposite i.e. a change in peak position, which has little visual effect on the period cost profile, implies that it is unlikely to be a useful decision. A similar procedure may be used with the other parameters, however the parameter m changes are exaggerated as the starting rate is proportional to m^2 while x_p and x_q change linearly.

Foreknowledge of events affecting a growth curve can, not only be utilised through the discontinuity methodology, but also by the judicious use of the secondary process. In these ways the interpretive aspects of growth curves may have a wider application, as they permit an element of quantitative support for the intuitive side of prediction.

4.31 General results

A large number of projects were collected, however, only 80 out of 200 of the sets of data collected under SERC, see Berny & Howes (1985), could be analysed with varying levels of detail. The majority had too short a time span to be of any real analytic value. For this reason it was also found that the remaining projects were dispersed too widely on a construction category basis. Due to the fact that there were only an average of 10 projects per category, this did not permit any statistically viable analysis to be made of the parameters. This limited the planned exercise to establish a means of indicating typical parameters by category.

For the above reason it was deemed unsuitable to attempt to establish a historical precedent for parameter values by class of project. This aspect is not ruled out, but merely set aside as a potential source of further research. However it is suspected that such a historical basis can have a limited value. It was anticipated that only the two parameter form of the model (solution 2) was likely to generate widely applicable useful results. The following analysis supports this view.

As an example of an historical parameter data base, tests were conducted using results from factory projects. This category had more case studies which could be analysed, than were generally found among the collected information.

The historical data base consisted of interpretive parameters (see equation 17), and SPE values. The size of the percentage variation of the standard deviation from the parameter values was regarded as a means of judging the validity of including projects in one category.

Tests were conducted on eight projects in this factory group and the following results obtained.

Mean of parameter g, represented by x_g in percentage form, occurred at :-

x_g = 29.4% with a standard deviation 6.1%

Mean of m, the control of the initial growth curve slope

m = 1.31 with a standard deviation of .67

The variation of x_p and x_q was not assessable as only 5 projects gave viable four parameter solutions.

The above results show that the dispersion is too large to warrant any firm conclusions although it is interesting to note that the most reliable value was the peak activity time given by x_g. It shows that the peak occurs one third of the way through the project. The result falls in line with earlier findings by Berny and Howes (1983) which showed that the DHSS constants indicated a similar result.

Further analyses were made of the remaining 40 projects. This may be subdivided as follows :-

a) Algorithm testing.

b) Confidence limits.

c) General curve fitting.

d) Testing of the unity point philosophy.

4.32 Algorithm testing

In order to create viable software it was necessary to adjust the boundaries of the search procedures after the principle mechanism was relatively well established.

Eleven construction projects with highly diverse properties, plus five other projects were used to adjust the algorithm in order that all gave reasonable results. This was checked against an interactive system which allowed an in depth search for 'ideal' results.

It was found that the results of curve fitting with this data gave a range of SPE=.5% to SPE= 4%, the latter value being the highest SPE which should be regarded as giving adequate results.

4.33 Confidence limits

The same projects also permitted the confidence limit procedure to be tested and evaluated. Comparison with the traditional models established that the confidence limit results appeared to be more realistic.

The implication of the above is that one should be able to give a band-width within which future costs should lie. However, if one forecasts too early this information has little value for the completed project because the confidence limits diverge too rapidly indicating the model cannot be relied upon.

The two unity points , mentioned in the detailed case studies (see figure A 3.4), were used and showed a valid example of the use of alternative projections. The projection which passes through a future project end point has to have its confidence limits compared with the second projection. If the second curve, using the last measured value as a unity point, is not lagging or leading by a large amount the confidence limits are likely to overlap. If this is the case one can take this as a realistic basis of regarding the projection through a future project end as viable.

In contrast with this situation, a large lag or lead may be an indication of reality. If this is the case then once again the confidence limits should be used as reasonable estimates of the likely spread within which the future costs will be found.

The above descriptive uses of confidence limits are a direct result of the four classical case studies.

Summary of confidence limit uses.

Confidence limits give a reasonable indication of the range of future costs providing the following criteria hold.

(i) If the project cost and duration are reliable then using this information for the unity point gives a cost breakdown which is valid between the present confidence limits (e.g. 95%).

(ii) If the cost breakdown given by setting the unity point to occur at the last measured value has confidence limits overlapping with those given by (i) above then the lag or lead may be treated as viable.

Should this overlap not arise then either the information available is insufficient, as in the advanced factory case study, or the project is truly behaving in an unexpected manner.

4.34 General curve fitting

This section of the work was very extensive, its main purpose was to establish the extent to which the model can be relied upon. This was clearly established and it suggested that the upper limit of 4% for the SPE may be used as an indicator of reasonable simulation properties. This in turn permitted the testing of model interpretation. By selecting four points in the curve, parameters were generated which in turn gave curves for comparison with the original data. By this process it was found that it was more realistic to compare with smoothed data and generate parameters from such data. This approach was regarded as fairly reliable. The importance of the smoothed data is, however, critical - due to the fact that the model only produces a "best trend". It cannot simulate day to day vagaries. This generated a very important counterpart of this device by using the interactive graphics module, forecasting value judgements can be made, supplementing the "dry" statistical results of reliability.

Summary

A viable upper limit in good forecasting was established to have an SPE of 4%.

A means of generating cost profile trend projections using unity points was devised which relies on initial judgement of smoothed data or alternatively management estimates as described earlier.

A graphical routine was also devised which allows "reliability analysis" to be conducted.

4.35 Testing of the unity point

This section can only be regarded as a strong indicator of results. Only 15 projects were analysed in this manner and further analysis is seen as appropriate to refine results.

The general conclusions that were established are summarised below :-

(1) The weighting scheme developed earlier (see 4.11) to select a given solution from the individual curve fit results has proved to give an adequate means of establishing a good model solution, in particular when the unity point was set at the project end.

(2) The case where the difference in trial forecast errors was marginal, when comparing the two projections generated by the two unity points, the extrapolation from the last measured value is probably realistic of the actual ongoing project.

(3) If the drop in SPE of the trial forecast, when comparing two projections as above (2), was significant e.g. 90% to 6%, then in such an instance the poor curve extrapolation was the result of either :-

selecting a point from which to forecast was too early in the project or,

an inappropriate point of time selected by chance.

In such a case, unless other factors indicate its reliability, the curve generated by use of the unity point at the last measured value should be abandoned.

In general trials it was found that forecast errors up to about 10% may be treated as reliable. This indication has to be qualified by the percentage of time allowed for trial forecasts, typically over eight periods or more. However further analysis of such a criterion could enhance this aspect of analysis.

4.4 Further studies

Further studies in the following areas could be appropriate :-

(1) The unity point tests for 'what if?' projections (see 7.31) were only conducted at the half way point in the data i.e. 100% ahead forecasts were made. Further exploration of the limitations would be advantageous.

(2) Further analysis of appropriate data, by category, should be undertaken to establish a historical data base for the two parameter model.

(3) Confidence limit profiles have not been fully studied only 10%, 30% and 50% ahead forecasts confidence limits have been documented.

Summary of Practical Work with Project Data

The marketing and project data was used to support the production of reliable software, in order to establish the general reliability of the new model and show its improved performance in principle over all currently used models.

The statistical soundness of solutions and confidence limit methodology was established.

The unity point at the project end was explored adequately to test its validity and established a realistic method of projecting at least 100% ahead in time, see figure A 3.4 in appendix 3.

Interpretive devices were tested and established as viable tools of the "what if?" class.

4.5 Summary of model development

It was judged that in order to gain improvements in quantitative project management, a necessary first step was to discover and validate a better and more reliable model, capable of generating a much greater variety of known cost profiles. This prime aim was achieved.

This section of the research set out to give a sound basis for a model to represent the cost profile of Construction projects. It was found necessary to ensure that a mechanism was available to validate the statistical viability of models. This was achieved by introducing standard routines for determining the information matrix estimation procedures for non-linear equations. The previous models, such as the combined model, see Berny & Howes (1983), were difficult to check for statistical feasibility. When an equation is generated from a curve fit and is to be used for extrapolative purposes, it is essential that its statistical feasibility is upheld. If the latter is not the case, too many alternative extrapolations of equal validity could be generated as to make the results worthless.

Some uses of this model have been explored and established. However, limitations of time and appropriate data have not permitted its full potential to be adequately evaluated. In view of the versatility of the model, the new concepts from which it was developed and its allied properties it is considered a crucial step forward and justifies further exploration.

CHAPTER 5

RISK ANALYSIS

The second fundamental issue noted at the outset of this research (see 1.21a) was the need to find a probability distribution to measure the valuation of project duration which can have all its parameters generated from user information, which is not always the case e.g. in the beta distribution see Farnum and Stanton (1987).

The practical need for risk assessment was probably most forcefully brought to the author's attention by a government body indicating that there was an average 30% overrun in time and in cost on projects. The reasons for overruns have been well surveyed and recent example was given by Arditi et al (1985). As Yates (1986) points out the construction industry has been aware of shortfalls in this area since the 1970's.

5.0 Risk analysis methods

Probably the earliest industrial use of risk methods was PERT/RISK which originally referred to estimates of the probable time variations of an activity and assuming their dependence, was used to calculate the probable variations of a project duration. The most frequently stated project assessment methods use the beta distribution which is frequently supplemented by the rectangular, triangular and normal probability distributions. For instance Cooper et al (1985) use the above distributions but acknowledge the existence of interdependence which is lacking in PERT/RISK. The need to account for time interdependence, as exemplified by network analysis, can be overcome by simulation a good example is shown by Pugh and Soden (1986). Their methodology however, replaces the beta distribution by a combined normal distribution. This allows a simpler user input without losing the essential property of skewness.

Other techniques fall into two main areas namely sensitivity analysis, Perry (1986), who combines the results with decision tree methods and simulation. The latter approach was used primarily for construction projects. For a wider use, Neuburger (1986), advocates decision tree methods as a broad brush approach. For example he uses a 'risk-chance' technique for company acquisition. As a final example, Baker (1986), who was concerned with BP oil projects, specifies the need for rapid answers and easy to use computer methodology. Software, developed by BP, is stated to allow interactive modelling as it is supported by available data to improve the accuracy of the generated probability distribution.

The distribution to be described in this work and the forecast-based methodology with which it interfaces, has the necessary practical properties as described by Baker (1986) above and in addition considers the periodic variations which may be permitted to ensure that a project remains within the statistical limits the system sets. The author further believes that the methodology is even easier to use than the BP Oil technique.

5.1 User information requirements

The user requirements, and his/her ability to give valid information, was seen as central to the need to reconsider the underlying premises for generating a new probability distribution.

The information input for risk functions demands one to estimate the modal value and its extremes e.g. for activity time the most likely, shortest and longest possible duration. Alternative data sets may be used to generate these parameters e.g. the variance or mean value, however considerable difficulty arises in the actual estimation of many of these values. Discussion with project managers has shown the following to be more realistic.

a) The most likely (mode) is a natural part of any estimate.

b) The shortest time or lowest cost.

This may reasonably be estimated as it assumes good workmanship with ideal working conditions having minimal interference. If this is not available the lowest limit with an associated probability may be used.

c) An estimate of chance to exceed the mode.

The above implies no upper limit to the variate e.g. time. This is the most difficult to estimate but can be more easily assimilated if a limited choice is given namely high, medium or low probability.

The lower limit (b above) could be preset from other information e.g. by use of confidence limit information. An alternative approach to the probability evaluation (c above) can utilise estimates of percentage completion.

The choice of which distribution to use in conjunction with the above values is still a problem however. If the current selection requirements can be covered by approximately one function, this would add further simplification. Some of the advantages of the new risk function proposed in this work are that it does not depend on an upper limit and allows control of the lower limit. In addition the function allows simple parameter determination.

5.2 New risk probability distribution

In order to establish a basis for the new function, the Beta distribution will first be analysed. This will be shown to allow a generalisation to be developed, which leads to derivation of the new distribution.

5.20 Beta distribution

The probability density function (pdf) of the standardised beta normalized density function, Raiffa and Shlaifer (1961) may be stated as :-

$$B(x) = x^{p-1}(1-x)^{q-1}B(p,q)$$

such that $0 =< x <= 1$ and $p,q > 0$

where x is the variate and may for instance represent time with a relative span from 0 to 1.

The beta function may be defined as :-

$$B(p,q) = \int_0^1 z^{p-1}(1-z)^{q-1}dz = \Gamma(p)\Gamma(q)/\Gamma(p+q) \qquad \text{...eq.22}$$

In the case of PERT network analysis, p and q are replaced by p+1 and q+1 thus $p,q>-1$

The pdf may be interpretted in the following manner:-

Let $B(x) = dQ/dx$ where $Q(x)$ is the cumulative density function (cdf) from 0 to x.

The definition of Q(x) allows that dQ/dx is proportional to the time expended (x) and remaining time (1-x) i.e. upper and lower limits are assumed. In general one may consider the distribution functions currently used as based on two functions $f_1(x)$ and $f_2(1-x)$.

Statistical arguments of decision analysis type are then used to deduce the nature of these functions. This results in the case of the beta function that :-

$$f_1(x) = x^{p-1} \quad \text{and} \quad f_2(1-x) = (1-x)^{q-1}$$

5.21 A generalisation

Based on the previous argument, the central issue when developing a pdf is that an upper and lower limit exists for the variate. Alternative premises are used for other well known distributions such as the Poisson, however the basis for the beta and similar distributions was seen as worthy of deeper consideration.

A generalisation for the formulation for the pdf may be given as follows :-

$$dQ/dx = gf_1(x)f_2(1-x)f_3(Q(x))f_4(1-Q(x)) \qquad \text{...eq.23}$$

Q(x) is the cdf from 0 to x and g is a proportionality constant.

Let $P(x)$ be the probability to exceed time x,

then $P(x) = 1-Q(x)$ and $x = T - T_0$

where T = actual time or cost in the financial planning case

T_0 = lower limit of time or cost.

then a possible simplification is :-

$$dP/dx = -gf_1(x)f_4(P(x)) \qquad \text{...eq.24}$$

such that $x>=0$ and $0=<P<=1$

i.e. the pdf of the remaining probability (P) is dependent on a function of the variate x e.g. time expended and the remaining cdf value P.

The above analysis can be employed to easily generate the beta distribution by suppressing the last two functions (f_3 and f_4) in dQ/dx equation 23. This procedure could likewise be used for a large variety of other functions, its advantage being that basic principles may be more easily interpreted.

5.22 The new function

Following the example of the beta function $f_1(x) = x^{m-1}$ with $m>1$ and $f_4(P(x)) = P(x)$, given $d^2P/dx^2 = 0$ at the mode then it may be shown that :-

The cdf is $P(x) = P_M \exp\,[(1-1/m)(1-(x/x_M)^m)]$ ...eq.25

the pdf is $dP/dx = (P_M/x_M{}^m)(m-1)x^{m-1}\exp[(1-1/m)(1-(x/x_M)^m)]$ ...eq.26

where $m>1$, $x_M = T_M - T_O$ and T_M is the mode hence $x_M>0$ and P_M is the probability to exceed the mode.

The proposed risk function (equations 25 and 26) has the property, not present in the beta function, that $dP(x)/dx$ tends to zero as x tends to infinity and at x=0.

Furthermore when x=0 then $P(0) = P_M \exp\,(1-1/m)$ or

$m=1/(1-\ln(P(0)/P_M))$

such that $P(0)>P_M>P(0)/e>0$ as $m>1$. It also follows that :-

The mean $= H\ P(0)\ \Gamma(1+1/m)$,

where $H = x_M/(1-1/m)^{1/m}$

and

variance $= H^2\ P(0)[\Gamma(1+2/m)-P(0)\{\Gamma(1+1/m)\}^2]$

Based on the Weierstrass definition of the gamma function it can be shown that the variance is always greater than zero for m>1.

P(0) however is the probability to exceed the lowest value of the variate, which would normally have a value of 1.

Under these conditions $m = 1/(1+\ln P_M)$ and $P_M > 1/e$, hence the permitted probability to exceed the mode is approximately 37% or more.

The above function has one minor disadvantage that the expected value and variance have to be found iteratively, although they may easily be approximated using e.g. Stirling's equation for gamma functions. However the parameters can be determined by simple calculation given the formulation below.

a) The mode $X_M = T_M - T_O$

b) The probability to exceed the mode P_M at T_M

c) The probability to exceed the lowest estimate P(0) at T_O

The last (c) allows a user, who is uncertain about the lower limit, to associate a probability with his/her estimate of the optimistic value.

5.3 Implementation

The commonest set of distributions, used in present day computer packages, are the square, triangular, normal, mixed normal and beta. The new function exhibits properties which approximate the square, triangular and beta see figure 5.1.

NOTATION:

T = absolute time or cost.

P = probability to exceed T.

SUFFICES:-

O = most optimistic estimate.

M = modal (most likely) estimate.

NOTE

$\frac{dP}{dT} = \frac{dP}{dx}$
with $x = T - T_0$

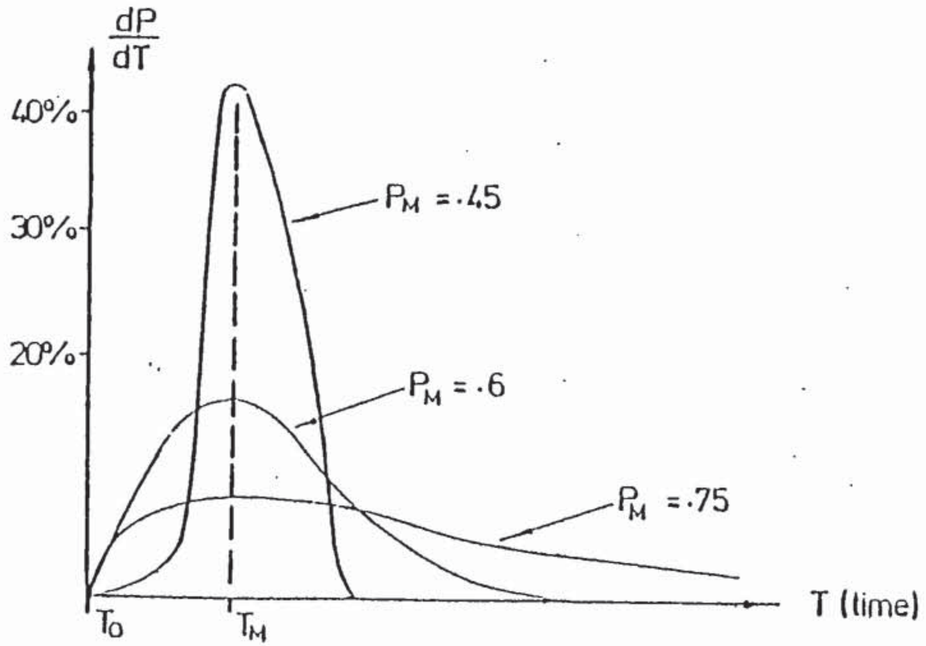


FIGURE 5.1 Berny probability distribution for 100% probability to exceed the optimistic estimate and constant mode

In order to test the proposed distribution, data was sought which would show typical overruns. However while well known organisations have analysed their projects and find a consistent pattern of overrun both in cost and time, they are not very keen to disclose this information. It was however possible to acquire data on underruns. Such data is much less common, however understandably publicly more presentable. For these reasons it was only possible to acquire one set of data, namely a sample of 30 cost underruns.

The model can be used for both underruns and overruns. By setting the value of T_0 as a minimum, one may analyse overruns or vice-versa for underruns. The latter form was used for the available test data on underruns.

The results are summed up in figure 5.2. The chi-squared test showed that the new function performed better than the beta distribution, however with no other case studies available this result may be fortuitous. Nevertheless it is an encouraging support of the new distribution.

It is essential that various risk functions should be tested on historic data. For instance repeats of an activity of identical kind could be used in this context. This would clarify the distinction between risk functions. Furthermore it would give a basis from which to simulate more precise estimates of project outcomes, whether they be under. or overruns of times, costs etc.

Practical use will, hopefully, be tested by implementation of software, called VISIER, the methodology is described in the risk assessor section of appendix 5, which incorporates these risk concepts. One aspect of this software is the use of

NOTES:-

(1) graph points at mid-range values of the distribution of costs from 30 projects.

(2) Chi-squared results:-
$\chi^2 (\nu = 7, \alpha = 0.05) = 14.07$
χ^2(observed Beta) = 5.35
χ^2(observed Berny) = 3.60

(3) Distribution parameters:-
Beta $p = 1.16, q = 1.48, l = 0.243$
Berny $R_M = 0.73, T_M = 25$

LEGEND:
—— Observations
– – Beta distribution
–·– Berny distribution

HORIZONTAL SCALE:-
Represents percentage change from planned costs.
(A) adjusted to scale 0 to 1 in line with Beta distribution with mode = 0.25 (25%)
(B) actual percentage change of planned costs in a range -75% to +25% with a mode = 0%

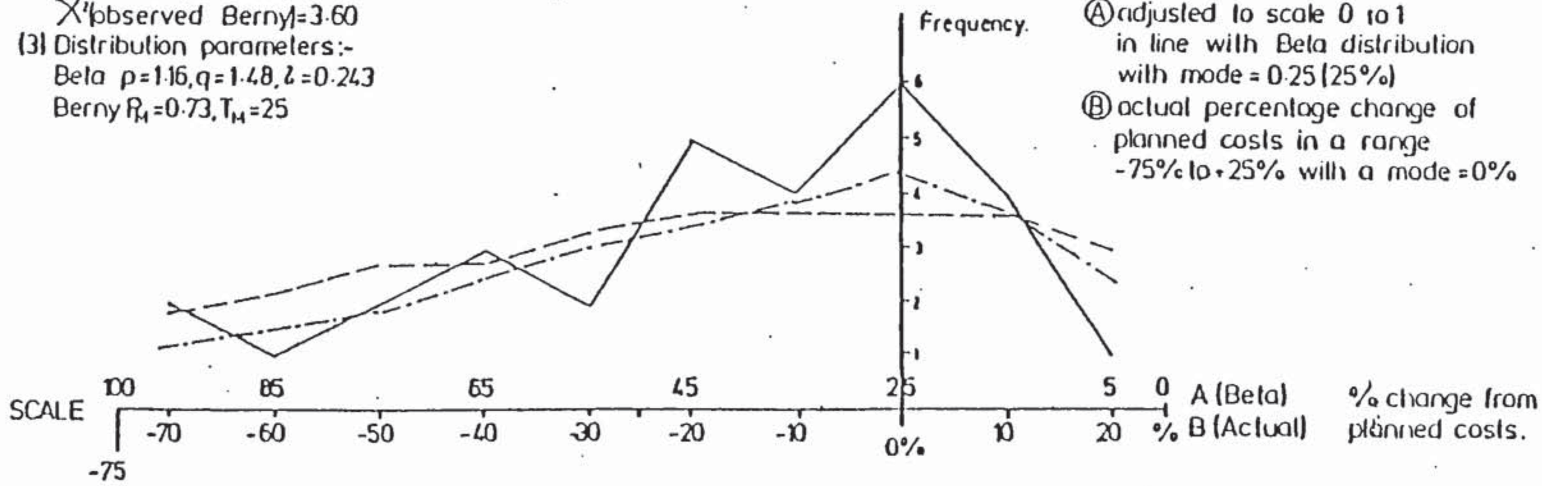


FIGURE 5.2 Test of Beta and Berny probability distributions against data, collected over three months, on job cost underruns

percentage completion as an alternative estimator of parameters. This is justified on the grounds that planned time is definable as the most likely i.e. the mode and the expected time is estimated while work progresses giving the mean value. The same argument naturally applies to cost risks. The minimum time is estimated by finding that point in time when the upper confidence limit of the best curve fit equals the modal (planned) value. This is made possible by virtue of the excellent fitting properties of a new growth curve model Berny (1987c).

CHAPTER 6

GENERAL APPLICATIONS WITH EMPHASIS ON SOFTWARE STRUCTURE

At the outset of this research project, the overall aim was to improve planning and control techniques for management. This was shown to require fundamental research into medium to long term forecasting and risk analysis. The utilisation of a proposed new techniques was viewed as the stepping stone towards the production of a practical tool for project management. This has been developed by writing appropriate software (see chapter 8), its span of applicability is the concern of this chapter.

6.0 Major areas of application

The application of the software generated as a result of this research, falls into two major categories, namely medium to long term forecasting alone, and secondly planning and control. Practical growth forecasting applies to industrial, commercial and scientific problems. In contrast planning and control require detailed project management mechanisms which should incorporate risk analysis as well as the facility for growth curve projections.

6.1 Forecasting applications

Forecasting alone may be applied to demographic issues for instance population and actuarial projections. An ecological example has been brought to the attention of the author by the Forestry Commission for the study of tree growth in problems relating to re-afforestation. Currently, control in growth of cells in Bio-engineering and related biological work, has been noted by Stone (1980). Economic and market forecasting uses are well established, Meade (1984) and Mar-Molinero (1980). A further area which may able to make use of the pure forecasting aspects, is material life-cycle studies relating, for instance, to construction industry technology.

An alternative view of the forecasting properties is its inverse, i.e. use in the decay situation for example in Nuclear Physics. The simplest isotope decay model is exponential decay. A more complex form is isotope daughter product decay, Berny

(1960). The new model concepts may have pertinence in practical aspects of Nuclear Physics which involve complex predictive problems such as in Nuclear Power Stations technology.

6.2 Management applications.

The general use of the software package, developed around the concepts described in the earlier sections of this thesis, has been discussed in depth with different commercial organisations. It has become evident as a result of those discussions that the potential applications also overlap with some of the pure forecasting uses.

The areas of application which have been ascertained by the author are briefly outlined under classified headings.

6.20 Marketing

This area has been mainly associated with pure forecasting, however extensive discussions with a marketing manager have confirmed that not only is there a valid use of risk assessment, but planning for production, monitoring and control, should play a vital role in a manufacturing companies corporate strategy, (see 8.1).

6.21 Computing

The commercial design of software is a multi-phase process, Putnam (1982). The life cycle from proposal to implementation involves planning, systems analysis, coding, testing and field trials. The managing of software design differs from construction design, in that procedures for the assessment of progress are more difficult to define.

6.22 Engineering.

Both the shipbuilding and aero-industries have shown interest in this overall area. A helicopter production cost-breakdown was presented to the author and proved to give a useful insight into the needs of 'what if?' routines.

6.23 Evaluation of projects and corporate strategy.

The facility to create cost-breakdowns of projects, whether at the research-and-development or construction end of the spectrum, (see 1.2) allows the analysis of corporate strategies, e.g. :- for banks, Dugdale (1978). The methodology to allow for multi-project analysis, has been incorporated in the software. It allows the importation of prepared projects and other financial breakdown into one global plan.

The analytic facilities which are available such as discounted cash flow, allow income evaluation from, for instance: property development or acquisition, Neuberger (1986).

In both multi-project and property management analyses the initial evaluation may subsequently be monitored and controlled, see chapter 7.

6.24 Architectural design.

The earliest data used in this research originates from the architects office. In this instance the project context is more multi-disciplinary and less site oriented, but clearly requires the same type of management as construction applications.

6.25 Construction.

The construction industry, which was the major source of data and feedback, is activity oriented. The gained knowledge was critical to the software developments discussed in detail in the next section.

6.3 Application software structure

This section discusses the means by which, the theoretical considerations developed in this thesis, were transformed for effective use in a practical environment, namely the construction industry.

The construction project data analysis and subsequent feedback has provided the additional advantage of a test bed for evaluating the general applicability of the system. From the outset it was intended to build a project management control system which would demonstrate the full potential of the theory previously described and would also provide a basis whereby the project as a whole, or in part, could be analysed.

Organisations concerned with project management will each have their own particular requirements and it was, therefore, anticipated that the methodology will only fully meet the needs of individual firms with software tailoring facilities. However, the basic concepts of management budgeting, monitoring, control and predictions built into the methodology were found to be applicable to the majority of projects.

The project management process maybe divided into three stages, namely planning, monitoring and control, and forecasting with risk analysis. The overlap between these processes emphasisied the need for understanding the inter-relationships and hence the importance of a reliable feedback system. The principles have been well established in industry. This is exemplified by the Property Services Agencies (PSA) methodology (1986), who use these stages in their project work as described in the aforementioned document. The current work was discussed with this important construction group. PSA have acknowledged the relevance and importance of the new facilities which have been developed. Their awareness of unsolved management problems is highlighted in their methodology document.

6.30 Software development

The software development was guided by the needs of the three stages. This breakdown was achieved by modularisation of the package.

6.30a Stage 1 (planning)

A Budget Forecasting and Data Preparation module was developed which is easy to use, as well as being quick and reliable. The budgeting system generates a database which is common to the monitoring and control modules. Because of the volume of data and preparation involved, it was considered essential that the system proposals should be integrated into one software suite interfacing the forecast and risk programmes. The principal task of the forecasting system was to enhance the reliability of a project plan. The nature of the expenditure curves modelled can be identified in terms of parameter constants, or an interpretive measure based on percentage elapsed project time. This enables knowledge to be accumulated on project expenditure which can then be used to determine other project forecasts at some future date.

6.30b Stage 2 (monitoring and control)

A Monitoring and Control system was developed to give the manager a basis on which to make decisions regarding the project. It was intended that adverse trends in expenditure would be quickly identified, thereby, giving early warning of likely losses if corrective action was not taken.

The system would also assess actual progress relative to the budget proposals previously determined, hence gains or shortfalls in performance would be periodically brought to the manager's attention. The resultant information would give management clear indications relating to cost, valuation, progress, profitability and past trend patterns.

6.30c Stage 3 (forecasting and risk analysis)

A projection or forecasting system was designed to enable estimation of future project expenditure and duration. In addition, revised project cost and duration was examined. Both could be analysed in terms of confidence cost envelopes and time risk assessment.

The software system, forms the basis for adaption to individual commercial applications, such as the current incorporation with network packages, and will additionally, it is hoped, provide scope for future experimentation and research.

6.31 Broad system requirements

The systems developed have been based within an IBM PC environment and software has been developed utilising GW-Basic and will run on MS DOS version 2.1, or higher versions of Disk Operating Systems (DOS). The major commercial tailoring facilities have been resolved by confirming that the system could operate in conjunction with database and spreadsheets e.g. Framework and Lotus 123. The important practical factor that a common database is used, required the creation of a special job operations module which ensures: easily used data files and effective data transmission between the forecasting, monitoring and other sub-systems. The common database allowed the computer programmes to be designed on a modular basis.

A hard disk computer is the normal requirement for running the system because of the extensive nature of the software and its data generation. The software space is more than half megabyte in the uncompiled form and the maximum of 112 data files/project could occupy a further 150 kilobytes.

As a matter of general development policy it was vital that the systems proposed were easy to use, requiring little initial computer knowledge. The size of the system database, made it essential that a job operations module allowed fully automated file handling; and the straight forward editing procedures facilitate the amendment, updating and processing of entered data.

6.32 The forecasting system

The object of this system was to facilitate the marriage of the new model philosophy to a practical means of creating project forecasts in a quick and efficient manner and thus permitted an interactive solution. Two knowledge based programs were developed, which enable the user to easily define cost profiles in response to his/her own knowledge of the specific property of an activity or project.

The essential data required by the forecasting system for the purpose of generating budgets needed to be precisely formulated in the following terms :-

1. Total estimate of project cost.
2. Envisaged project duration.
3. Cost to peak expenditure and the time at which it occurs.
4. The initial project cost. (first 10% of job duration).
5. In the case of perceived bi-modality, one has to additionally give the time and cost-to-date at which dip occurs replacing 3 above.

Alternatively these values may be interpretated in the form of percentages of elapsed time and cost.

The total data input processes 1 to 5 have the potential to allow the user to study alternative strategies at the plan and subsequent ongoing stages of management control.

An important aspect of the forecasting system's design was the potential to produce more detailed and reliable project budgets quickly and easily. This justified the decision to develop the knowledge based (expert system style) conversational mode subsystems.

6.33 Development of the plan

Previous work in this field, particularly that carried out by the DHSS, relied entirely on establishing data relating to past project performance which then formed the yardstick for assessing future projects. The problem with this approach, if used exclusively, is that it raises the question of how relevant are past results in present circumstances and how adaptable to change are the constants thereby produced. The advantage with the proposed model and associated system was that the user could start to use it for forecasting immediately by adopting the following procedures.

6.33a Project without activity breakdown.

In the absence of past experience and data available, the user can, with the aid of an expert system style interrogative process, select a curve shape which he feels will represent the project. The software developed enables the curve to be represented in terms of parameter values. These are retrieved from, either a stored parameter database or a user defined generator. This approach has the advantage that the nature of the curve is quantified as the cost profile performance may be evaluated by comparison with the actual data of a completed project. Subsequently the results can be stored for future reference when preparing budgets.

6.33b Activity based project.

The forecast could be modelled on a cost centre (activity) basis utilising a project time scale. The cost centre period costs could be determined by periodic allocation or by adapting the curve selection as described in (6.33a).

This approach is extremely fast and facilitates easy preparation. For example, the initial project budget could be produced rapidly and then fine tuning of the budget could be made by the use of the programme editing facilities. These are analogous to the spreadsheet 'what if?' principles.

The traditional method of allocating detailed bill of quantities costs over a contract programme is both outdated and relatively unnecessarily slow.

The number of cost centres utilised can be selected according to the requirements of the user. For normal construction projects it was not envisaged that these could exceed thirty in number, although the system does make provision for up to 55. Major activities may be built up from minor ones and imported into the overall project.

With increasing familiarity the user would be able to formulate the project cost centre expenditure in terms of parameter values. Consequently, by utilising past experience in the form of records kept and his/her own intuition, greater flexibility in the selection of an appropriate curve to forecast the expenditure is envisaged.

6.34 The philosophy of construction industry cost centres

Over the past few years, it has been a matter of accepted principle to break a project down, into a number of financial packages, in order to more readily identify costs and to exert greater financial control. The number of cost centres selected for a project will vary according to size, complexity, value of work and the requirements of the Client.

An important factor in evaluating the number and size of cost centres, relates to: the appropriate level of control necessary compared with the cost of implementation and the likely benefits to be accrued. The actual breakdown into cost centres must of course be consistent with natural breaks in the work, representing clearly identifiable boundaries which effectively isolate one package of cost from another.

The potential accuracy of forecasting expenditure budgets is increased with the level of breakdown B.Fine (1972) and hence the cost centre approach can be particularly useful where little previous known data exists regarding project costs. One way to produce a project growth curve is to cumulate straight line cost centre expenditures with each set to a timescale. This stance has the advantage that the reliance on derived data is considerably reduced and the emphasis is moved to the nature of the project in hand. Hence, the observed physical requirements are taken directly into account and have a direct bearing on the project time span.

Another factor which tends to question the entire reliance on a statistically derived growth curve for a project (without cost centre analysis) is the host of individual clients' requirements which can considerably affect the total project. For example, the construction time required and special phasing arrangements necessary to satisfy specific project needs. It will no doubt be fully appreciated that a hospital project, of a certain value, to be completed in five years may have a completely different shaped expenditure pattern to one to be completed in two and a half years, due to the less complex facilities the shorter project is likely to possess.

By incorporating cost centres set to a timescale, a direct relationship can be achieved between the contract programme and the expenditure envisaged.

Before adopting a range of cost centres it will be necessary to prepare a contract programme which clearly and accurately represents the construction operations. The overall construction programme can either be scheduled as a bar chart or as a network programme. The construction activities shown in the programme will need to be grouped together under cost centre headings and the expenditure associated with each cost centre must be determined from the estimated cost of the work. The amount of work and difficulty experienced in accurately allocating estimated costs, will largely depend on the method by which the quantities have been prepared. A bill of quantities produced in accordance with the Standard Method of Measurement of Building Work does present some difficulties in this respect, since not all items contained can be directly related to the project activities. In general, the more operationally based bill of quantity (as generated by quantity surveyors) enables the aforesaid relationship to be directly established in an easier and more accurate manner.

It is worth noting at this stage that some of the most advanced computer based network programming packages, particularly those evolved in the USA (PCS- Digital Corp., PMS- IBM), have comprehensive facilities for establishing cost centres described by their cost and timespan as alternatives to procedure logic. A possible development to interface the proposed forecasting and prediction sub-systems with network (CPA) programmes has been commenced. It is anticipated that this will increase the versatility of software in the context of complex planning and control.

Once the cost centre data has been quantified financially, its anticipated cash flow relative to time is determined. The nature of the expenditure profiles may take many different forms as depicted in figure 6.1 which shows examples derived from the new model.

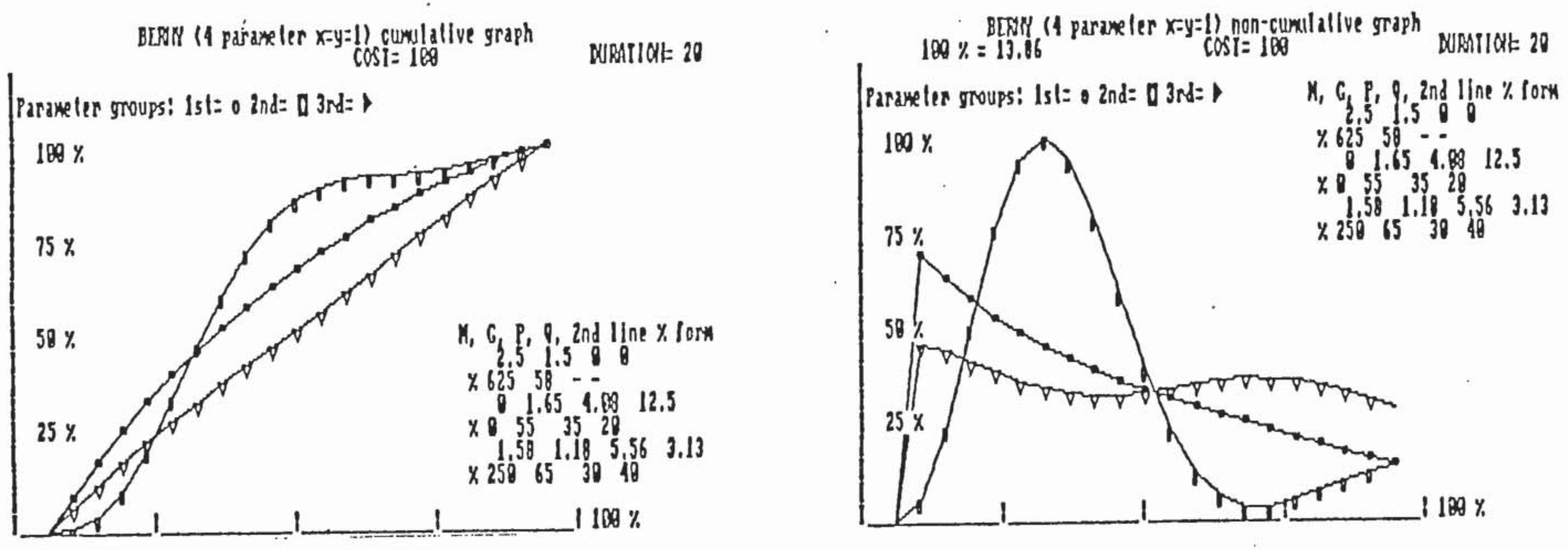


FIGURE 6.1 Typical new model growth curves

It is generally considered that the adoption of cost centres offers the best possible means of modelling project expenditure. This process can be carried out quickly and efficiently in accordance with individual project requirements. Furthermore it enables management to have additional control facilities beyond those which rely on projection and risk analysis feedback.

6.4 Integration of software

The previous section considered the software structure required by management which would assist them both in basic project planning, monitoring and control. The interface with forecasting and risk analysis techniques was developed within the software.

The underlying pattern of the structure was that these devices are not used in isolation, but may be integrated. This created a more dynamic device, with the inherent benefits of more constructive feedback to the decision maker.

In tandem with the integration described, it was further seen that the well accepted advantages of 'what if?' techniques could not only be incorporated, but the new technology could further enhance them. The 'what if?' is a major concern of the next chapter.

CHAPTER 7

PROJECT MANAGEMENT APPLICATIONS

Existing project control systems have been observed, as a result of discussions at conferences, to suffer from inadequate feedback and a shortfall of accurate enough information to give adequate corrective measures. For this reason, management aids have been introduced, most commonly software packages. However a need exists, for more reliable and flexible aids which go beyond currently available project and financial modelling devices, outlined in a survey summary Berny (1988a). In particular these packages do not reflect the need for reliable forecasting and contingency planning of the overall financial and monitoring facets of a company.

7.0 Introduction

The basic objective of this chapter is to describe a viable and flexible means of work planning, monitoring and forecasting both projects (in the widest terms) and corporate strategies. In addition the usage of the newly proposed method of risk analysis enhanced by the new forecasting techniques, will be placed into the context of assessing probable variations in the financial and temporal outcome of work.

Many predictive aspects of project management have been researched, however the term predictive is open to wide interpretation. The specific concern in this thesis is with the application of quantitative predictive methods related to forecasting. This should not suggest that qualitative prediction, for instance the impact of project organisational structure, is excluded since such considerations usually precede and interface critically with the quantitative aspects of planning, see 'Project Start Up' (1984).

The second objective was the need to determine a viable means of planning and monitoring. This has been addressed in the form of a bar chart analysis. Perhaps more importantly a 'what if?' methodology has been extended to both the planning and the monitoring phases.

The effects of changing the quantitative nature of the project may be tested via the 'what if?' procedures. The results of such changes may be inspected graphically to improve the assimilation of the extensive information planning generates.

The issues which had to be addressed, see Berny (1987a), in approaching the above goals successfully were :-

a) Methods of generating budgets which are appropriate to a practical method of project planning.

b) For the purpose of monitoring, a means by which current activities and future activities may be simply "re-programmed".

c) Overall project future trends to be soundly established.

The major factor linking all the above was the need to find a robust model around which the facilities required of a quantitative management aid could be built.

7.1 Planning

The current methods of quantitatively analysing a project have generally relied on network planning. However the barchart approach has received much favour in industry.

The view taken by the author is that network procedures can best be used to analyse specific complex activities which involve a strict procedure and have a complex structure. The resulting activity time breakdown, expanded to account for cost-breakdown, would then be incorporated in a barchart. The term 'cost' is used loosely in the sense that cost could be used to measure, for instance, man-hours or material quantities, as well as monetary outlays or returns.

Information, from specific activities, may be deduced by network analysis and the description of extra activities to be specified within the structure of a barchart. Utilising this data, it should be possible to build a profile of the project. This requires two major quantitative inputs. Specifically the nature of the 'cost' and time breakdown and secondly the ability to alter the activity time slot and cost to give the project a dynamic structure as reality befits the situation.

Each activity 'cost' profile is established as a result of the newly proposed modelling concepts. The profile is generated by the model and enhanced through editing and the flexible structure created by the 'what if?' facility, contained in the software developments. The importance and principle of the need to assess alternative plans, was stressed by Woodward (1975), who also shows the need to pay attention to the risk aspects at the earliest stages of quantification of plans and their progress. The simplest means of establishing 'cost' profiles, is to choose from a series of 7 curves, given in figure 7.1, these may be selected by an 'expert - system' style procedure, Berny (1988b).

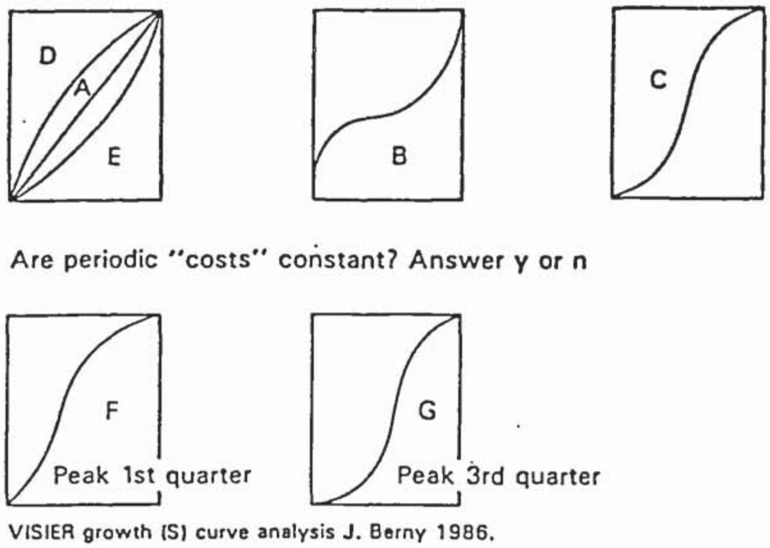


Figure 7.1 Expert system style cumulative cost profile selector

The above process was extended to allow user defined cost-breakdowns. The basis being the result of extensive testing of projects. This showed remarkable uniformity in the interpretive results for hotels, see Table 7.1, this is a result of further data collection subsequent to that described in 4.31. The use of this technique for activities was justified by the 'parent-child' analogy.

Interpretive Measure	Mean (%)	(Standard deviation/ mean) as %
Percentage cost 10% through project	5	45
Percentage cost to peak period	47	15
Peak period as percentage of job length	43	21
Curve fit (SPE) percentage (noise)	3.4	30
Costs of projects £24,000 to £2,000,000 (8300% spread)		

Table 7.1 Curve fit results of 30 hotel projects.

The interpretive parameters are :-

a) peak period;

b) 'cost' to the peak period;

c) 'cost' when 10% of the project duration has expired.

These parameters can then be translated into the model parameters. Table 7.1 states the results of studying a specific category (hotels) of project data, which had an 8300% spread in costs. The analysis showed an average of + or - 27% variation in the interpretive parameters values. This indicates that the projects in this category have broadly similar properties and hence common planning features.

An alternative means of planning is that previous projects for similar work may be reused with minor alterations. This is easily achieved in the software which has this design feature incorporated in the editor module. The ideal format is that an organisation produces a group of typical project shells which can be upgraded and expanded and/or contracted. New activity cost breakdowns may be generated by the 'expert system style' cost profile selection feature described earlier.

In addition to the above, planning may be assisted by considering alternative plans i.e. by 'what if?', and assessing potential variations and risks which may be encountered.

The 'what if?' mode permits one to move, expand, contract and mask activities. At the corporate level, currently active and new projects may be combined and studied in the same 'what if?' manner.

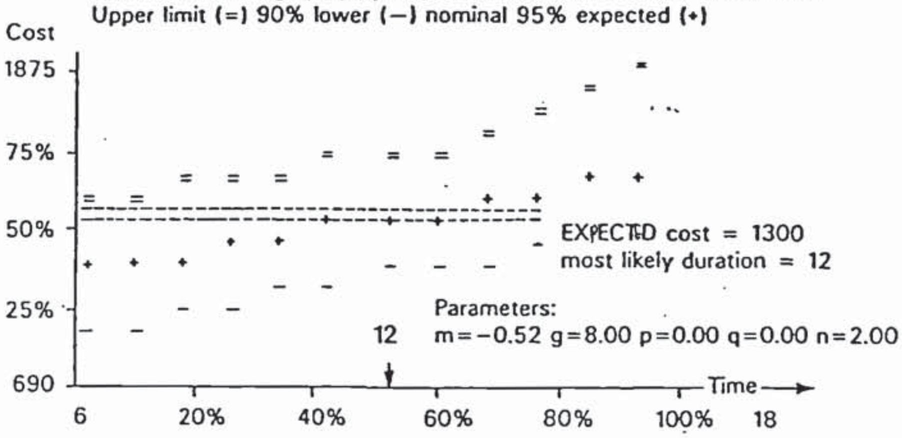
The information may be subjected to the contingency/risk analysis. This generates confidence limits and probability histograms of time against 'cost' (see figure 7.2).

7.2 Current work forecasts

An active project requires careful monitoring, Woodward (1975), and thus collection and assimilation of cost and time data is of utmost importance. The updating is dependent on monitoring methods. The assimilation of this information was seen to be best aided by forecasting.

Risk Analysis

Confidence band graph project plan/monitor (cib/a-cib87) 12-02-1986
Upper limit (=) 90% lower (—) nominal 95% expected (+)



Model statistically sound, average monthly cost variations = 12% (monthly)
Upper 90% lower 95% confidence limits span from 6 to 18
span of expected cost from 1102·156 to 1531·935

Title project plan/monitor (cib/a-cib87) 12-02-1986

Case A to exceed max time 20 prob = 9% min time 6 reset at 7
Case B to exceed most likely time 12 prob = 60%
Expected time=14.0 'worth' = 1,384.6

prd.	Cum. % prob.	Prob. % / prd.	Expected cost
7	100	5	1,090.319
9	95	15	1,195.434
11	80	20	1,266.166
13	60	19	1,334.758
15	41	17	1,408.571
17	24	11	1,489.014
19	13	7	1,576.762
21	6	3	1,762.48
23	3	2	1,776.893
25	1	1	1,890.79

Graph of prob % /period

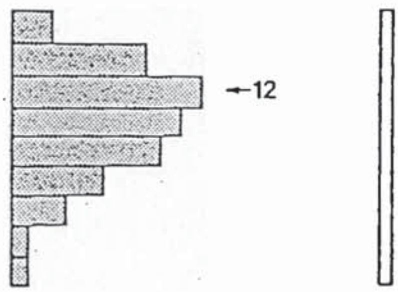


FIGURE 7.2 Confidence limit forecast curve (TOP)
Time risk probability histogram (BOTTOM)

7.20 Monitoring

Monitoring is not only concerned with the methods of establishing performance, which may lead to replanning; but also the decision nature associated with the process of management in the light of determined targets or objectives. It is equally important for the client and management team to estimate the actual outcome of a large project i.e. estimate total costs, duration, profits and other 'cost' implications of the current status of work. In order to obtain answers to these 'predictive' questions, various projections of the model trend are made, including risk assessments or contingency analyses.

7.20a Interactive model adjustment

The ability of the new forecasting model to allow parameter adjustment was utilised to assist the monitoring process, which thus aided real time forecasting. This was considered a major innovation of this section and was achieved by interactively updating the model parameters from the plan. This emphasized the robust nature of this model, as it enabled the software to redistribute moneys or resources in a realistic manner (i.e. non-linear) for each active phase of work.

The result of this facility is to give a 'forecast' break down of the remainder of a current activity in line with the initial 'cost' profile, demonstrated in figure 7.3, which shows the extent to which a cost profile is adjusted to current

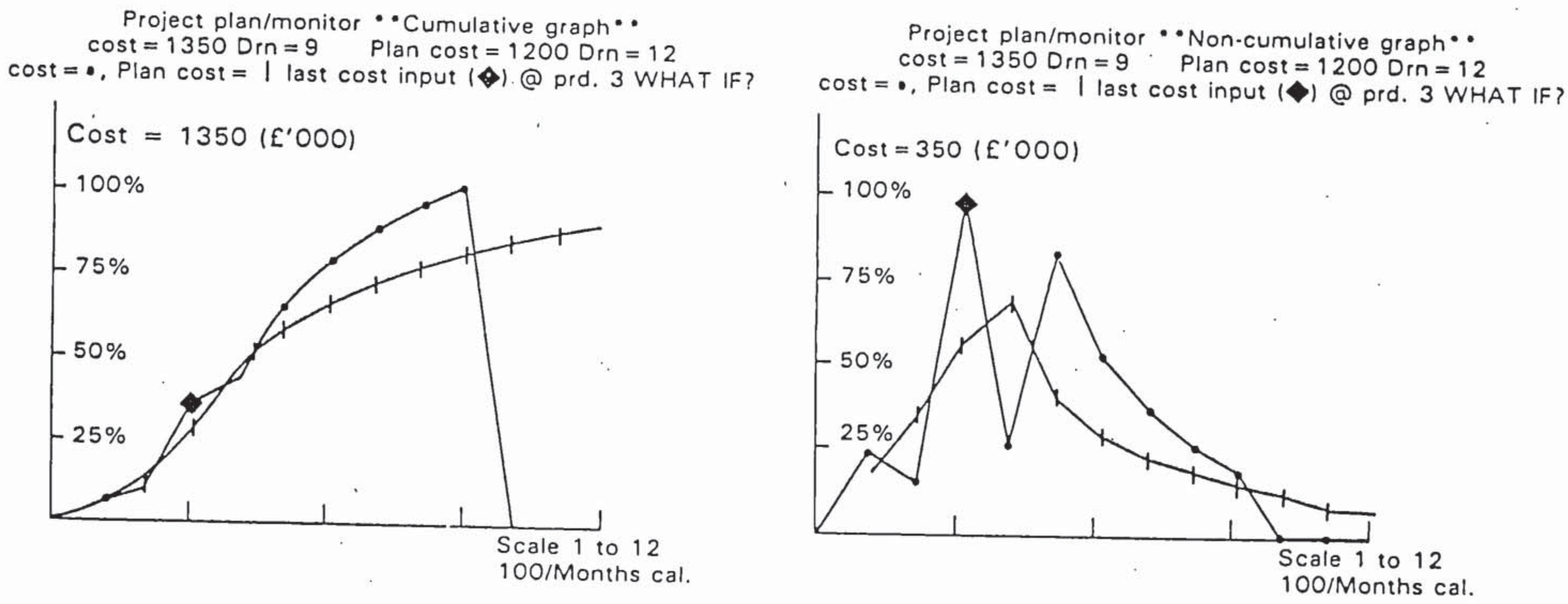


FIGURE 7.3 *Interactive model adjustment effect on cost profiles of a planned project and 3 month update with "What if ?"*

update information. For large activities the interpreted parameters may be altered if required.

Furthermore, activities may be masked to show underlying trends and individual activities inspected. All the features are supported graphically. One example showing the usefulness of this device, is to allow one to inspect the effect which a long delay in an activity may have on the overall project outcome.

7.20b <u>Reliability of model</u>

A critical factor related to the above procedures was the need to assess the performance of the model. The statistical testing procedure of the model's validity allowed for the extension to subsequent testing of contingency/risk aspects of the current plan. The procedures were tested on over fifty construction projects, 98% of which showed the model fitted the project data with a mean error of less than or equal to 3.4%.

There was only one single project for which the curve fit parameters had associated with them, errors greater than the parameter values themselves. In such an instance it would be dangerous to use the model, since it gives an unreliable description of the project data. The reason for this anomaly is that parameter values would be realistically replaced by a large number of alternatives, generating many apparently valid but widely differing curves to represent the actual data. This is clearly unacceptable and hence the model is rejected and is said to give a statistically invalid trend. The author has observed that no other commercially available systems include this automated test of the model validity and for the above reasons would appear to cast a shadow of doubt on commonly used models. Further, the results generated, other than on this one project, strongly support the validity of the use of the proposed model.

These tests are included in all analytic procedures to preclude misuse of the model. A further success of the above analysis, was to show that projects in similar categories can be described by a limited number of interpretive parameters.

7.20c <u>Historical database</u>

As a result of the above results, and the devised technology, a historical file of past records of project models may be created.

The historical file, by category, may be inspected and statistical interpretive parameter information generated, (see section 7.1). This enlarges the basis of generating curve profiles or alternatively historic model parameters values may be used. The process of generating this information is extremely simple, once the relevant data has been collected and input.

7.3 Control

An active project normally requires continual intervention by the management team who control progress and are ultimately responsible to the client or his representatives. Rarely does a large project run exactly to plan, thus changes in future actions have to be made. These arise in many forms but as a generalisation may be divided into two major categories, namely :-

(i) those due to client changes.

(ii) problems which hamper planned progress.

7.30 Introduction

The above classification of reasons for changes in project management is important in terms of the control techniques. Client changes e.g. adding an extra lift, may involve considerable reorganisation, hence only 'what if?' techniques interfaced with risk assessment are of practical value for management decision support. Programme of work interference, due to for instance site problems, can be detected by a variety of projection and control analysis procedures as well as 'what if?' and risk assessment techniques.

7.31 Activity and project forecast "What if?"

The following section is based on feedback from industry, eg. PSA (1986), and reflects the research response to their needs. It is a result of several iterations of the field trial feedback loop.

The 'what if?' term derives from spreadsheet origins Lewis (1987). This author's interpretation of 'what if?' is that it is the process of inspecting the effect (in this instance) on a financial model, created by changing one or more of its 'parameter' values.

This principle has been extended to allow :-

(i) activities to be changed in 'cost'

(ii) activity cost-profiles altered

(iii) activity start and end times changed

(iv) activities removed

(v) new activities added

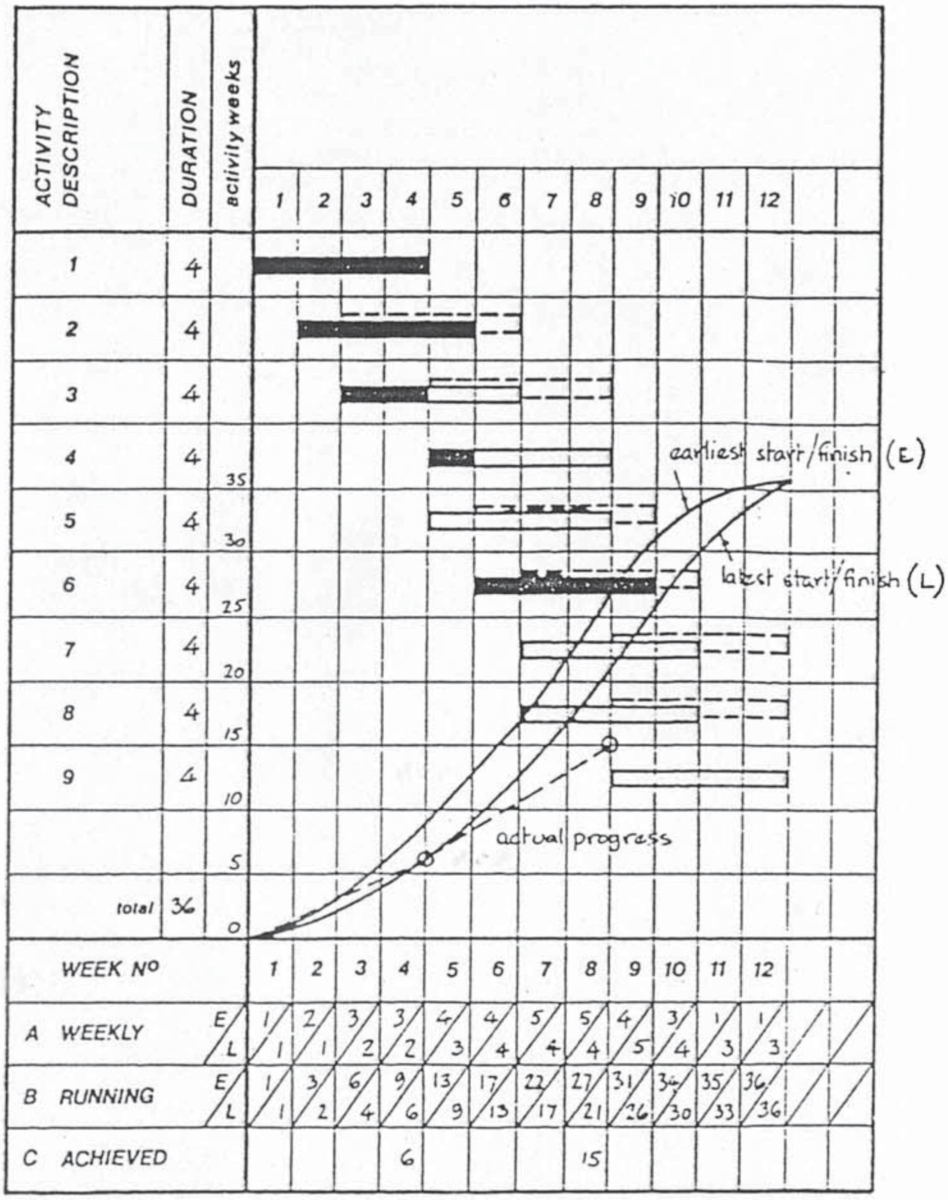
7.31a Activity "What if"

The selection mechanism for activities which may require investigation was based on a search for high cost regions in activities and the whole project. This procedure, which also acts as an early warning system, has been called 'control analysis'. Current procedures are fairly limited, (see figure 7.4 courtesy of PSA (1986)).

7.31b Forecast "What if"

Forecasting may also be used in a 'what if?' manner by testing the effect of changing job costs and durations and projecting from current updates, and the model is then evaluated for the statistical reliability of such changes. This procedure has the advantage that forecast results may be studied with greater flexibility. This facility is seen as more satisfactory by management than to rely on one projection only, enabling them to investigate alternative job outcomes by using "forecast 'what if?'".

By using the above scenario of 'what if?' and comparing with plan information or current management update information, the decisions used for progress control, can be supported in a less ad hoc manner.



MONITORING PROGRESS ON A BAR CHART WITH FLOAT

FIGURE 7.4 Illustration of common project monitoring procedures
Courtesy of PSA (Property Services Agency 1986)

7.31c Percentage completion

A further advance was made in relation to the current updating information, due to the nature of the new model which permits the parameters to be interactively adjusted, based on the current updates. This feature gives a more accurate basis for comparisons.

Measures such as percentage completion estimates of the proportion of the work on an activity, normally adjudged on allocation of time and/or cost, are used as further indications of progress.

7.31d Analytic support to "What ifs ?"

It has been found that the "what if?" techniques required further support in some situations. These are listed below :-

a) Discounted cashflow and internal rate of return

b) Inflation upgrade of plan cost breakdowns

c) Comparison with past work of a similar kind

d) Introduction of lag when comparing, for instance, cost and income.

The next section elaborates on the various forecasting procedures which add to the control support system.

7.32 Forecasting

Assessment of future outcomes is an essential part of control. Given the 'what if?' alternatives which may be tested under forecasting conditions, the nature of control decisions may be usefully improved.

The predictive viability is highly dependent on the good trend following properties of the new model. A graphical 'confidence generator' is that the model should follow the smoothed pattern of actual data. It was found that the most severe test was to observe the monthly trends given in figure 7.5. However the visual analysis is also supported by a statistical one which rejects the use of trends which do not follow the known data realistically. In order to achieve a reasonable degree of confidence in future projections, three modes of forecasting may be used.

a) The last six data values are forecast (if sufficient information is available).

b) The trend may be projected through an anticipated end cost and job duration ('what if?' mode).

c) The best trend may be extrapolated to any point in the future.

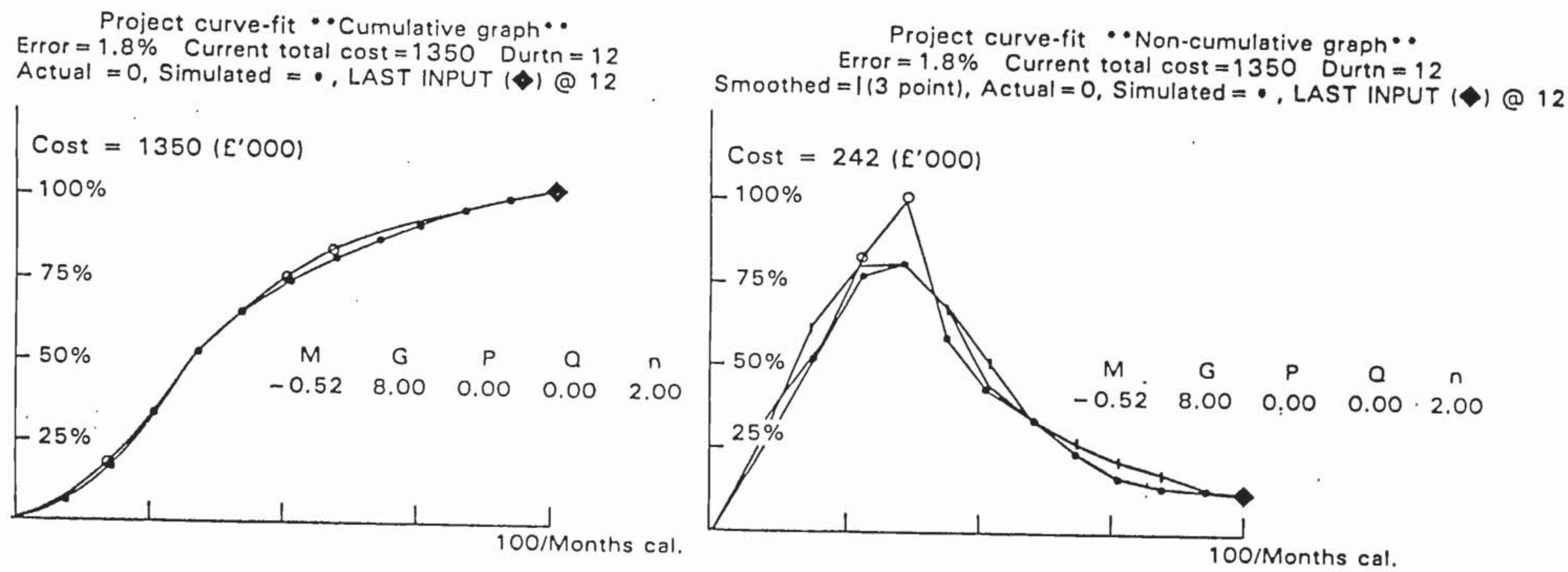


FIGURE 7.5 *Curve fit of planned cost and smoothed data 'test' of reliability*

It was viewed as important to extend these facilities to contingency/risk plans. Data generated in either of the last two modes may be processed through the contingency/risk mode to further enhance the decision making process and give a cost envelope.

An additional feature which has been embodied in the software is that either some data may be suppressed or comparisons may be made with project data which has had specific activities/projects masked, i.e. temporarily given zero costs in order to see the cost profile without the cost presence of the masked data. For instance, holidays or high cost equipment installation may usefully be excluded from forecasting but reinput at a later stage for overall project forecasts.

The last section is concerned with risk aspects, which relate to forecasting.

7.33 Contingency / Risk analysis implications

The statistical work combined with the evidence that the new model does give a high degree of replicability of project 'cost' profiles, permitted the development and simplification in contingency/risk planning methodology. The nature of information given by this analysis is given in figure 7.2.

Previous work has required detailed investigations of each activity e.g. Pugh and Seldon (1986) and Bennett (1986). However, if a model can represent the activity or project accurately, this, it is argued, would permit using the model as basis of such probabilistic investigation and considerably reduce the need for time consuming indepth studies which the present systems require, Berny (1987b).

7.4 Software

A large suite of computer programmes has been written to accommodate these facilities.

This software is called :- VISIER (Visual Interpretation and Evaluation of Risks) and a summary given by Berny (1987a).

Some of the output generated by this software, is illustrated in the graphs and diagrams of this chapter. A flowchart showing typical use of VISIER is given in appendix 4.

A specific feature of VISIER is that all processed information is represented graphically. Thus in the 'what if?' stage of budget planning, this assists assimilation of large amounts of information and one hopes would enhance the confidence of decision makers.

Evidence of the viability of the software is not at present commercially extensive but is strongly supported. However the robustness has been demonstrated as indicated by the cost breakdown span from £100,000 to that of large industrial project worth 100's of millions (£) viably constructed by these techniques.

7.41 Software related developments

The current status of the work has generated a potentially viable management aid. Nevertheless the effect of generating a new aid is to open avenues for further development. At present, this is envisaged on three fronts. The first is a continual appraisal, improvement and expansion of the software, supported by industrial feedback. Secondly, the work has raised many new areas of research which should be explored, for instance, the testing of alternative models and expanding the use of interpretive parameters. However the most important area is to extend the developments in risk / contingency analysis but this requires good and extensive collaboration with industry. The last two are expected to have an important impact on the software in the future.

CHAPTER 8

SOFTWARE IMPLEMENTATION

The pragmatic issues, see 1.2, contain the proposals for the implementation of the management usage and resolution of the fundamental problems. Due to the complex nature of the management planning and control framework the only possible means of implementation was seen to be via the aegis of modern computers. The extensive suite of programs has and is undergoing continual evaluation by several industrial organisations, principally in the construction industry.

8.0 Introduction

The software development, extensively described in chapters 6 and 7 originated from the requirements to improve forecasting and risk analysis technologies. It was further necessary to develop approximation routines to enable interactive adjustments to model parameters to be made, in order that they reflect updated information. The risk analysis required the development of an algorithm to find expected values and to handle low probability limitations.

8.1 Software usage development

The formal setting of the software, which surrounded the above algorithmic routines, was based on a logical structure which would reflect the pertinent system requirements for practical management use. A summary flow chart and also detailed flow charts of the software modules are shown in appendix 4.

The commonest procedural sequence was that after initial planning 'what if?' routines would be used and access to risk assessment would be required. The monitoring and control procedures were also interfaced in a like manner. In order to achieve adequate analysis a separate inspection programme was written which allowed comparisons to be made, for instance the plan and its 'what if?'. This section was supported by analytic devices, namely discounted cash flow, control analysis, inflation and lag routines. An overview of the software given in appendix 5 contains a full printout of the demonstration tutorial disk screens. It will be observed that user friendliness has been strictly observed, its importance is stressed by Lewis (1985) and supported by software house acceptance.

To further improve the communication with a user, specific terms relating to a given field of work e.g. marketing as opposed to construction terminology, was made user definable.

File 'jargon' was totally removed, because computer tree structures are complex, amongst other file devices, thus a computer 'naivé' user can be shielded from the DOS system. A user is asked to set up a category of work, e.g. housing, and state a job code. This necessity is enhanced as a project could generate within one job code over 112 files. Thereafter the user retrieves, for instance, the plan or current work 'what if?' information. Help facilities are not always to the liking of a user, thus a slot was created in each programme which may carry the user's own notes.

In view of the mass of data which is generated, a graphic display was always given to ease the assimilation of all activity or project cost breakdowns. The same facility has been extended to the information processed for comparative purposes, forecasting and risk.

The whole of this aspect of work has been supported by continual industrial feedback to ensure that a potential user had a satisfactory product, as for instance outlined by Davies (1986). The industrial feedback showed the need for two other facilities namely, an indication within the programmes of the next stage of work and the idea of creating project shells, (typical job proto-types).

The above facilities have been well received by industry and were further seen as a means of improving communication between the client and management. In addition the time constraint imposed by business requirements showed the need to add a 'get out' clause which stored the current work and allowed the user to return another day and continue with the same job.

The problem of job-coding was further alleviated by storing both a library of essential information about ongoing, completed work and of all past work i.e. copies, deletions and new jobs created. It is the latter facility combined with 'what if?' editing facilities that allowed project shells to be used.

8.2 Field trials results

The reaction of the industry to the software was critical to its development, The major number of trial responses was from the project management sources. The second critical source of feedback was from the marketing field. A major electrical group's switch-gear marketing was reassessed successfully using VISIER and showed the potential in this field of work. In addition commercial usage was explored.

The two fields of operation, marketing and project management were seen as opposite ends of the spectrum of application and hence show the broad base viability of pragmatism of this thesis.

The section is based on both interview and questionnaire reports (see appendix 6)

The final form of the software largely resulted from the feedback from these industries and the questionnaires were left until version 1 of VISIER was agreed to.

In total software trials were conducted over 18 months and showed that the software, called VISIER, is genuinely seen by practitioners, in commerce, project management and marketing, as a new aid to management. Nine groups carried out the trials, in varying degrees of depth, three in each of the above fields of operation. The managers concluded that this package is unique and cannot be compared with any other known software.

This latter aspect was well researched in the field trials. The only available "roughly" comparable software which has some functions which aim at similar uses are summarised in table 8.1.

The field trial results are outlined below :-

a) <u>User friendliness</u>

"Eminently usable" but some technology would be unfamiliar to the newcomer.

b) <u>Modular usefulness</u>

The Planner 'What if?' was found exceptionally useful. Global use of VISIER showed all modules were either very useful, although some less so depending on the discipline (marketing, commerce or project management).

COMPARISONS WITH VISIER

PACKAGE	COST	BAR CHART	NETWORK ANALYSIS	CORPORATE MULTI-PROJECTS	FORECAST STATS	RISK DECISION AID	WHAT IF?	COST BREAK DOWN	OTHER FEATURES
ARTEMIS	£5,000	TIME	YES	?	NONE	S = 1000s	TO PU	LINEAR	D SYSTEM
HORNET	£3,500	TIME	YES	?	NONE	NONE	TO PU	LINEAR	D LINK
OPEN PLAN & OPERA	£4,680	TIME	YES	NO	NONE	S = 1000s	TO PU	LINEAR	
KORKUS & PERK	£9,000	TIME	YES	NO	NONE	SPECIAL	TO PU	LINEAR	D LINK
NICHOLS	£5,000	NO	NO	YES	P DHSS	S = 100s	NO	DHSS	
VISIER	£3,500	COST & TIME	NO	YES	P & R BERNY	FORECAST BASED REAL TIME	C & T, PU & FR	BERNY	*

LEGEND

D = Database　P = Protection　R = Reliability tests

TO = Time Only　C = Cost　T = Time

PU = Plan and Update　FR = Forecast and Risk　S = Simulation Steps

DHSS implies false conclusions

BERNY automates 14 models

Please note that the table is based on the Author's understanding of these packages.

*

VISIER also includes

Expert System style portfolio and Marketing Strategy, Discount and IRR

TABLE 8.1　*Comparison of VISIER with commercial software with features of a "similar" class*

c) Learning

This was seen by all concerned to be a slow process. It was suggested that five, one-day sessions over six months were required for full familiarization, a year to fully utilise VISIER. (One day allows you to start using VISIER.)

d) Scope

This was found to be continually expanding - new uses have been found by all parties including the author. The overall scope is not estimable at present. The newest sphere of use has been for town planning work.

e) Future expansion / shortfalls

A possible expansion of graphics has been suggested. But in line with 'scope' above, not until VISIER is fully explored can this be judged further. No shortfalls were observed.

f) Practical advantages

The consensus view was that VISIER should give at least a 10% saving in marketing and contracting, with respect to both costs and time. Greater competitiveness as a result of using VISIER was anticipated by project managers.

VISIER was seen by all commercial and industrial management who had been approached, at different feedback levels, as a commercially viable tool and as filling an empty niche in the software market. Its only drawback was its avant garde nature, thus the requirement for carefully planned marketing. It was further anticipated that VISIER would have slow initial acceptance.

CHAPTER 9

CONCLUSIONS

This research project consisted of two major fields of study namely the development of new theories in forecasting and risk analysis and secondly their implementation, in particular, in management.

The research successfully developed new theoretical concepts for deterministic growth curve forecasting and a new basis for generating stochastic risk distributions. This was supported by extensive tests.

These developments guided the exploration of the means by which to practically implement these theoretical results in the guise of relevant software.

The conclusions are broken down into three sections, the first two being concerned with the theoretical developments and the last considers the pragmatic issues.

9.0 Introduction

This research set out to assist with control and budgeting of projects in the Construction Industry. The experience gained by studying a large portfolio of projects has been invaluable and has contributed directly to the underlying research and its applicability. Since cumulative costs of projects always exhibit a growth curve pattern, a study was made to formulate a better model to represent project cost-breakdowns. The research showed up many weak areas in current growth curve modelling. There is no doubt that many of the sophisticated techniques used to improve the use of traditional deterministic models developed in 1825, have their place. However, the rather different data characteristics produced by Construction Industry projects, forced the author to evaluate the foundations of deterministic model building upon which current models are based. Stochastic innovations based on the earlier models had not shown a high degree of success, hence intuitive judgement indicated that the model basis, rather than the technology, was the key to progress, which successfully led to the primary theoretical developments in this thesis.

Risk analysis, which is a very broad subject, was only considered in the context of the major practical drawback of the beta distribution and related interpretative problems. The current methodology is not considered adequate for "naive management use" hence there was seen a need to achieve more "user friendly" risk assessment procedures.

The pragmatic issues considered in this research were to generate a practical means of integrating theory with the needs of management and thus to develop new aspects of information feedback, which could not have been achieved without the new theoretical techniques.

9.1 Summary of findings in forecasting.

The fundamental objective of this section of research was in the first instance to prescribe more fully the definition of the term growth curve. This led to the introduction of the concept of the secondary process and produced a means of controlling the initial rate of growth and catering for bimodality. The six case studies analysed, which are highly diverse in nature, upheld the validity of the secondary process concept. Furthermore, these studies showed that a two parameter solution derived from the new model was generally a better predictor than the equivalent traditional models, which is seen to be a direct result of the concept of the secondary process and possibly due to the introduction of the, hitherto apparently unknown, differential of the error function (DEF) equation.

The forecasting results proved the robustness of the model on two premises. A small number (c. a minimum of 15) observations were required for a feasible curve fit in the case of the 4 parameter solution, while many traditional models failed in this situation. Secondly, one case study (PVC) showed that all traditional models failed, however, the new model did give a statistically valid solution.

The comparative tests showed a notable reduction in trial forecast errors, namely an average 59% lowering of SPE errors in the seven trial forecast comparisons with traditional models. The split was 67% drop in the classical (marketing) studies and 57% in the project case studies.

The research did not reject the use of traditional models, but added to their potential use by both improving their analysis and simplifying their applicability by specifying a point through which a growth curve was constrained to pass (unity point). The latter is a direct result of the needs of the unity point used in project control. Perhaps it was a consequence of analysing growth curves from a radically different standpoint, which led to the discovery of the flaw in the analysis of traditional consumption process. This emphasised the need to ensure that data has the characteristics of growth curve models, i.e. :- to be applicable, the observations should be monotonically increasing.

The varied nature of the case studies investigated, brought to light the need to improve comparative statistical measures. The standard percentage error (SPE) usefully replaces the standard error (SE) and can be said to indicate good results if the SPE value is below 3%. In a like manner the mean error was replaced by the mean percentage error (MPE). The nature of the solutions had also invalidated the strict use of the saturation level (value at infinity) and its SE. These measures have been replaced by the 50% ahead forecast and its related standard error approximation, giving more realistic measures of future performance and reducing the need to inspect the information matrix. No evidence has been found to encourage the use of autocorrelation, however the question of the use of the Kalman filter technique is left unanswered.

Finally it was demonstrated that the new model is open to a variety of practical interpretive uses. For instance the sensitivity of projects to external influences may be studied by building non-cumulative curve profiles from either :-

the elements of the secondary process

or by setting peak and trough points.

Project control is directly assisted from an early stage using this approach, certainly before half the project is complete and subject to estimates of the project cost and duration.

9.10 Future investigations

It was impossible to consider all aspects of this study, in particular alternative models which have been found. Future work is envisaged in the area of further developing and evaluating alternative models, such as those already cited. Another area of research could be to find new secondary process elements and possibly test the inclusion of more elements in models, in particular from the interpretive viewpoint.

Statistical and technical issues exist that are also worthy of further investigations. For instance specific findings indicated that the question of statistical feasibility requires clarification. It might also be useful to attempt to minimise the mean absolute percentage error as a method of finding best solutions, however such a study would have to consider the impact of information matrix theory.

Equal importance should be attached to furthering the practical use of these models, in particular the practical points raised in 9.1 have only so far been rudimentarily tested. The fields of study considered presently by research as opposed to the practising public, require expansion. These areas should include problems of decay as well as growth. Potential areas of investigation would be - biology, demography, materials science, corporate planning and a large vista of industries who are concerned with project control. Finally a database could be constructed which related similar problems and could be used as a basis for correlating results and also act as an historical record from which to build projections of the possible outcome of a new growth process.

It is felt that the case studies presented cover a wide enough spectrum to justify the statement that; the scene has been set for fresh work to be conducted in growth curve forecasting.

9.2 Risk analysis

An alternative approach to deriving risk functions was found and used to develop a new distribution function for risk analysis. It is believed this methodology should have a wider application. Particular importance was attached to the requirements dictated by the needs of industry. In this respect, an example has been set by responding to feedback from the project management users of risk analysis.

The nature of the new risk distribution overcomes the need to specify its variance, which is a requirement of most continuous probability distributions used in risk analysis. It is particularly relevant in the case of the use of the beta distribution in PERT/RISK. This uses an approximation of the equation derived for the mean when the standard deviation is 1/6, i.e.:- a single value selected from a range 0 to 1/2, thus notably reducing the potential use of PERT. This result is graphically depicted in figure 9.1

The new function developed in this work requires no approximations of parameter estimates, as is the case for the commonest method of using the beta distribution. It further has the advantage of approximately encompassing the simple risk functions commonly in use. The evaluation of industrial needs has led to a practical means of establishing risk functions, thus supporting the principles initially set out to establish this new function. It is hoped that the current impetus of industry to pay attention to analysis of risks will generate greater practical feedback, which will foster further analysis as exemplified in the above proposals.

9.3 Pragmatic developments

The new theories have been transformed into robust and dynamic computer techniques, which form the basic structure of the developed integrated software. The package amalgamates the use of quantitative management information and generates relevant feedback for the decision maker.

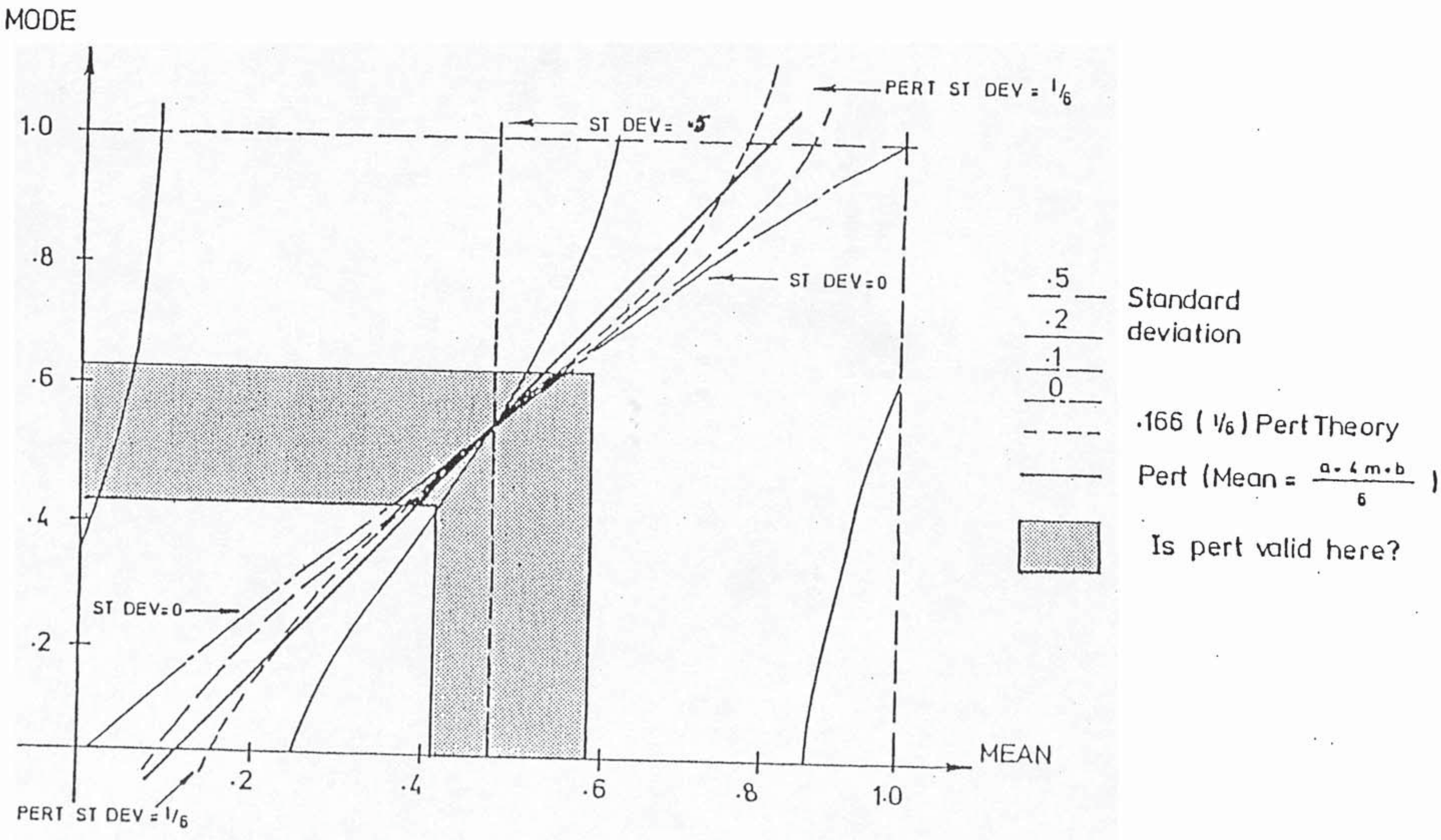


FIGURE 9.1 *Limitations of the use of the Beta distribution in PERT/RISK analysis*

The field trials have confirmed that this software fills many of the management aid shortfalls, highlighted in a recent indepth survey. Both these investigations show areas of future research which relate to this package.

It is envisaged that future theoretical investigations, for instance cyclical growth curve models, would expand the versatility of the software. Some developments have already been made which expand the software facilities by interfacing with available packages. This is regarded as an ongoing area of work with useful feedback on all the interface components.

However, the primary area of study of pragmatic value, directly relating to the current work, is in expert systems. It has been proposed to utilize this software, which has minor expert system style devices, as a test bed of knowledge based system research.

9.4 General conclusion

The major result of this research was the establishment of a set of new concepts for generating deterministic growth curve models . This has led to progress in the area on several fronts and tests have shown that a significant drop in forecasting errors has been achieved. The research has been expanded to show that the new range of models will allow for seasonal adjustments.

Testing the new set of models with relevant data has shown that this methodology is also applicable in the marketing field. In addition, features which are common to specific categories of construction projects, have been found to encompass a very broad range of both cost and duration.

The above work has also led to the development of a new continuous risk distribution function. This function's primary advantage is that it overcomes the major drawback of the beta distribution, used in PERT types of analyses which assume a constant variance.

Non-linear regression methods were adopted and further developed for: testing the validity of curve fits and generating approximate confidence limits.

These techniques have been amalgamated into a software package; which permits a major simplification as well as a more rapid and precise means of planning, control and risk assessment for both projects and marketing strategies. The structural basis assumes that bar chart information is known.

This software resulting from the theoretical breakthroughs has currently successfully undergone field trials and has been well received in commerce generally and in particular the construction and software industries.

The aims of this research have been achieved, culminating in new theories and a totally new aid for management. In turn these successes have opened up many new avenues of future research.

A P P E N D I C E S

APPENDIX 1

STATISTICAL THEORY

This appendix describes the theoretical basis for the statistical analysis used to support the curve fitting procedures. It was found necessary to adapt current methods in order to cater for multi-component models. Furthermore, approximations were deduced for confidence limit computation and thus avoid the time consuming simulation procedures.

A 1.0 Information matrix

The main purpose of this section is to describe the technique adopted, for the purpose of ensuring the reliability of all curve fits, to generate an informaton or variance-covariance matrix.

Given the SSE $=\Sigma(Y_t - F_t)^2)$

and $F_t = F^*_t - F^*_o$, $F_t = Y_t$ + residual

where F^*_o = measured value at t = 0

The standard error (SE) is :-

SE = sqr(SSE/n) for the traditional models

for the proportionate case this takes the form :-

se = sqr(SSE/n)/F^*_D

where D = time of last measurement or fit end
and n is the number of observations.

It may then be shown based on Goldfeld and Quandt (1972) that if :-

$b_{jk} = \Sigma_t \, \partial y_t/\partial a_j \partial y_t/\partial a_k$,

where a_i are the model parameters
and j,k have values 1 to m,
such that m = number of parameters

Then :-

(1) y_t may be replaced by Y^*_t and se by SE

(2) if y_t=f(t)/f(D)

then it is advisable to use the relationship :-

$\partial y_t/\partial a_i = (\partial f(t)/\partial a_i - y_t * \partial f(D)/\partial a_i)/f(D)$

The information matrix is given by :-

V_{jk} = se $[\, b_{jk} \,]^{-1}$,

where $[\, b_{jk} \,]$ is the matrix of the b_{jk} values above
V_{jk} = covariance of parameters a_j and a_k,

but if j=k then V_{jj} is the variance of a_j

The adjustment for the number of degrees of freedom for the SE of the parameters has not been made, based on a point raised by Box and Jenkins 1976 that :-

"arguments can be advanced for using the divisor n-m rather than n, but for moderate sample sizes the correction makes little difference".

However when testing the validity of the procedures above against published results the correction was incorporated. The resulting evaluations then conformed with the MACC results quoted by Meade for parameter SE values.

Subsequently this adjustment was not used in order to give consistent information and due to the Box and Jenkins point cited above. Hence the SE values of parameters and curve fit refer to the sample size (n). For the purpose of comparison with the SE values of parameters in Meade (1984), a first approximation is to multiply the appropriate results given in appendix 2 by the sqare root of (n/(n-m)).

A 1.1 Approximate confidence limits

This calculation is based on Gregg et al. (1964) and Williams (1959). The procedure is a first approximation based on the Taylor's series expansion and is an extension of the method used for estimating the SE of a single value in linear regression.

The standard error of a single value is :-

SE of y =

$$SE \text{ sqr}\{ 1 + 1/n + \Sigma_j \Sigma_k (\partial y_t/\partial a_j \partial y_t/\partial a_k * V_{jk})\}$$

The tolerance limits of y = y + or - t*(SE of y), t being the t-distribution value. The SE of a single periodic value is given by the above formulation, but for t>0 the value y_t in the differential is replaced by $y_t - y_{t-1}$ the periodic value at time t.

APPENDIX 2

FOUR CLASSICAL GROWTH CURVE STUDIES

The major case studies considered in published literature were reviewed by Meade (1984). Those selected for detailed analysis by Meade, outlined in section 4.21, have been used as a test-bed for this thesis and are analysed in this appendix.

A 2.0 Introduction

For the purpose of clarification of one of the major marketing issues considered by Meade, the introductory section of 4.21 which primarily consists of quotations, is restated below.

Meade presents market growth as a comparison between adoption and consumption processes, which he describes as follows.

Consumption process :-

"A consumption process requires an initial decision to use a product and a sequence of subsequent decisions first to continue usage and secondly to determine the level of usage."

Adoption process :-

"An adoption process requires a single decision by a non-owner to become an owner."

Meade further states that :-

"A consumption process can fall theoretically to zero" and that "this is not the case for adoption."

Based on the definition of a growth curve, that it describes how some quantity associated with a growth phenomenon inreases with time, the consumption process represents the rate of change of the use of a product. This indicates a reasonable conclusion would be that :-

'The cumulative value of usage or total consumption to date would be the correct descriptor of the consumption process growth curve.'

CONSUMPTION PROCESSES

The following case study, on the consumption of PVC per person annually, demonstrates the above proposition: namely that a cumulated consumption process is a growth curve.

A 2.1 U.K. consumption of PVC

Hutchesson (1967) studied the annual usage of PVC, he used data from 1950 to 1965. Meade (1984) reanalysed this data using the models quoted in section 4.20.

The result of testing the above models showed the full curve fit gave nearly identical SE values for the two logistics and Gompertz in the range .22 to .24. The Gompertz was the only model which had parameter values with smaller corresponding SE values, hence it was selected for the comparative tests.

As all these models were non-feasible this meant a large number of variations on the generated constants would serve equally well to create a diversity of curves all giving good curve fits.

Meade concludes his analysis stating :-

"That none of these (four) growth curves can offer any useful information to the forecaster".

In view of the suggested alternate solutions to this case study the non-cumulative or annual per capita use of PVC i.e. the original data will be reviewed first.

a) Non-cumulative PVC (original data)

The author's model gave a SFS with a best curve fit SE=.29, which was higher than that of the Gompertz with a SE=.24. This result was given by the three parameter form of the fourth solution (m=0) illustrated in figure a 2.1 .

The study was continued and a fit/forecast was made using the first 11 years (1950-1960) to determine the parameters for the trial forecast. This resulted in a SFS only for the one parameter case, which had a full curve fit SE=.34. The trial forecasts for all four solutions had SPE values which only differed marginally from 17% to 20%. Thus it was found preferable to base the final forecasting decision on the full curve fit results. For this reason the three paramater form of solution four was selected for forecasting from 1965.

This gave for the 50% ahead forecast in 1973 a SE of y of 7.3% of the predicted value of 16.7 or a 95% confidence range of 14.2 to 19.4. This result is compatible with the actual 1973 value of 14lb/head of PVC, see table a 2.1. The corresponding result for the Gompertz gave a SE of y=15.3% of 20.8lb/head of PVC.

In order to illustrate the spread of the confidence limits with predictions further in the future the last value quoted by Meade was selected. This gave a production level of 17lb/head for an 80% ahead forecast of 13 years (1978). The author's model gave 95% confidence limits of 16.4 to 28.2 about 23.4lb/head of PVC. The 1978 estimate of the Gompertz gave 33lb/head ranging from 14 to 52.

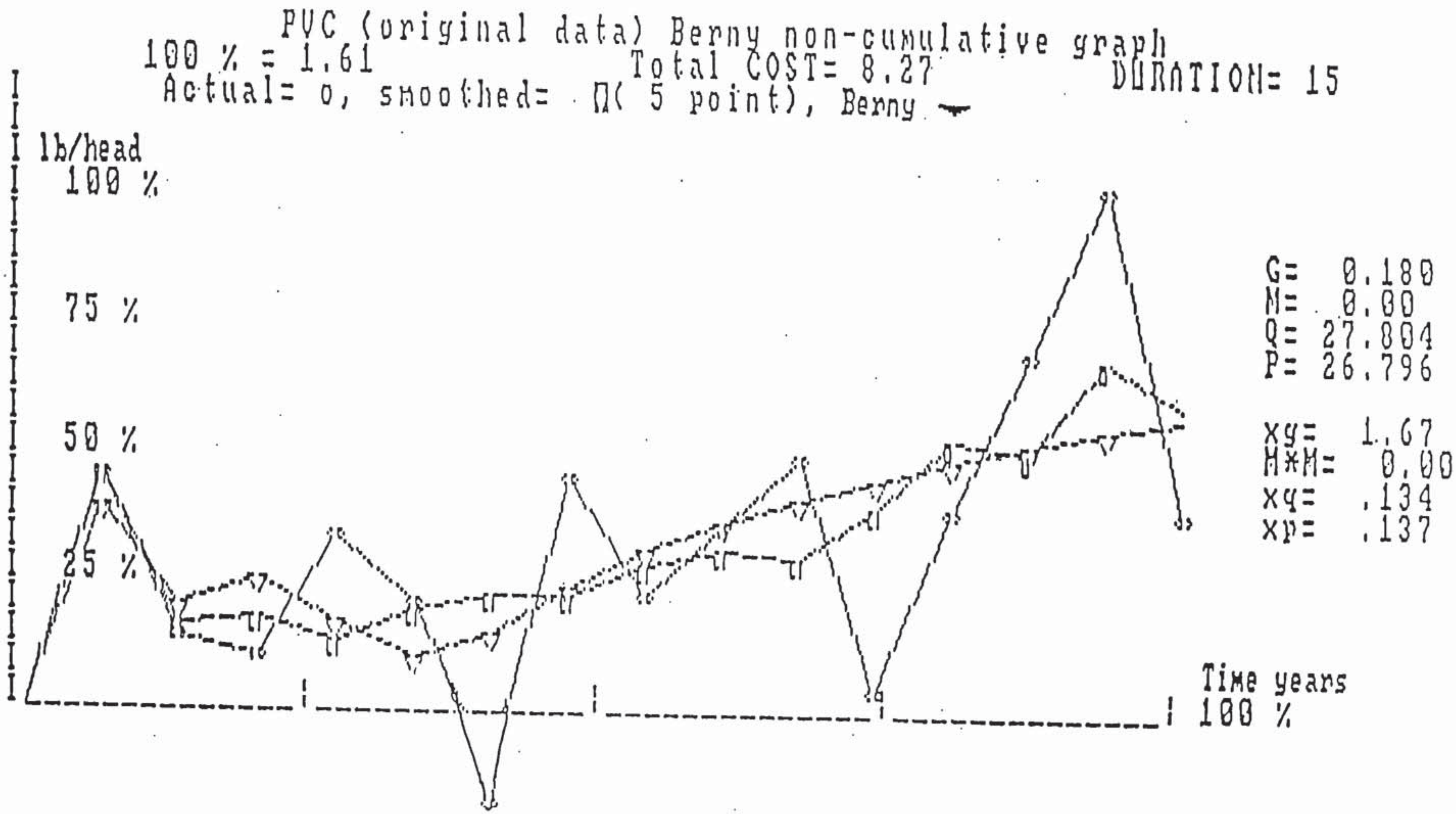


FIGURE A 2.1 *Non-cumulative 'original' PVC data compared with Berny model and smoothed data*

Model	TEST span	Parameter values (SE of parameters)				Statistical measures SPE	(SE)	MPE	50% ahead forecast value	(SE*)	period	Comment
Berny		g	m	q	p							
	0 to 10	.32	.27	16.1	15.2	4.65%	(.18)	-0.56%	10.2	(26.8%)	1965	forecast fit
		(.81)	(.12)	(11.4)	(10.8)							
	11 to 15	"	"	"	"	18.2%	(.77)	6.3%				trial forecast
	0 to 15	.18	0	27.8	26.8	3.47%	(.29)	-1.20%	16.7	(7.3%)	1973	full curve fit
		(.12)	-	(10.8)	(10.2)							
Gompertz		a	b	c	-							
	1 to 16	367	6.65	.035	-	2.85%	(.24)	1.06%	20.8	(15.3%)	1973	full curve fit
		(760)	(1.9)	(.017)	-							

TABLE A 2.1 Results from the analysis of the Non-cumulative case PVC (1950-1965) (original data)

Model	TEST span	Parameter values (SE of parameters)				Statistical measures SPE	(SE)	MPE	50% ahead forecast value	(SE*)	period	Comment
Berny		g	m	q	p							
	0 to 15	.277	0	1.18	1.61	0.53%	(.28)	-0.08%	12.5	(21%)	1973	full curve fit
		(.14)	-	(.07)	(.30)							
Gompertz		a	b	c	-							
	1 to 16	244	6.70	.092	-	1.90%	(1.01)	.85%	8.0	(29%)	1973	full curve fit
		(86.9)	(0.2)	(.01)	-							

TABLE A 2.2 Results from the analysis of the Cumulative case PVC (1950-1965)

Note :- In ALL the subesequent tables the SE* is used to represent the SE of y values given as a % of Y*)

b) Cumulative PVC data

In order to test the proposal that a consumption process is in fact only a true growth curve if the data is cumulated, the author's model and the Gompertz were fitted to the data giving total production to date.

One aspect of this analysis was that it exemplified the alternative starting value for the g parameter. The range between the maximum smoothed slope and DEF solutions for x_a gave periods 14 to 29. The best solutions used x_a=21, giving the unusually low SPE%=.53 (SE=.28).

The results of this test showed not only that both models gave SFS but more realistic estimates, see table A 2.2 and the graphical comparison shown in figuresA 2.2/3.The SE of y values have been calculated for the periodic case (see appendix A 1.1) in order to evaluate the annual production levels. For instance in 1971 (30% ahead) the production was 12lb/head. The Gompertz underestimates this value giving 8.2 with 95% confidence limits 3.9 to 12.5, whilst previously it overestimated giving limits 12.8 to 20.9 about 16.9. The author predicts 12.5 with a range of 8.5 to 16.5, the earlier range was 12.6 to 16.3 about 14.5. The 50% ahead forecasts gave 95% confidence limits of 7.0 to 18.1 about 12.5 and in the Gompertz case these limits were 3.1 to 12.9 about 8.0lb/head compared to the 1973 value of 14lb/head.

The author's values decline in quality but nevertheless show an improvement on the earlier forecasts. Both models were statistically feasible in the second case adding credibility to the proposal that this data should have beeen cumulated. While the pragmatic view may be that growth curves cannot indicate annual fluctuations about trends, hence should not be used in such cases, it is not unrealistic to use them if they can be relied upon.

A 2.2 U.K. consumption of titanium dioxide

Harrison and Pearce (1972) studied the annual production of titanium dioxide using data from 1925 to 1968. Meade enlarged upon this.. analysis and showed using a fit/forecast of 34 years (1925-1958) with a trial forecast of 10 years (1959-1968), that the cumulative lognormal was the most likely model to give good forecasts. The author used the same time division and obtained a SPE=5.4% (SE=.11) for the forecast period, which is 26% lower than the SPE=7.1% (SE=.15) given by the cumulative lognormal. However the best curve fits of the complete data set gave nearly identical results shown in table A 2.3.

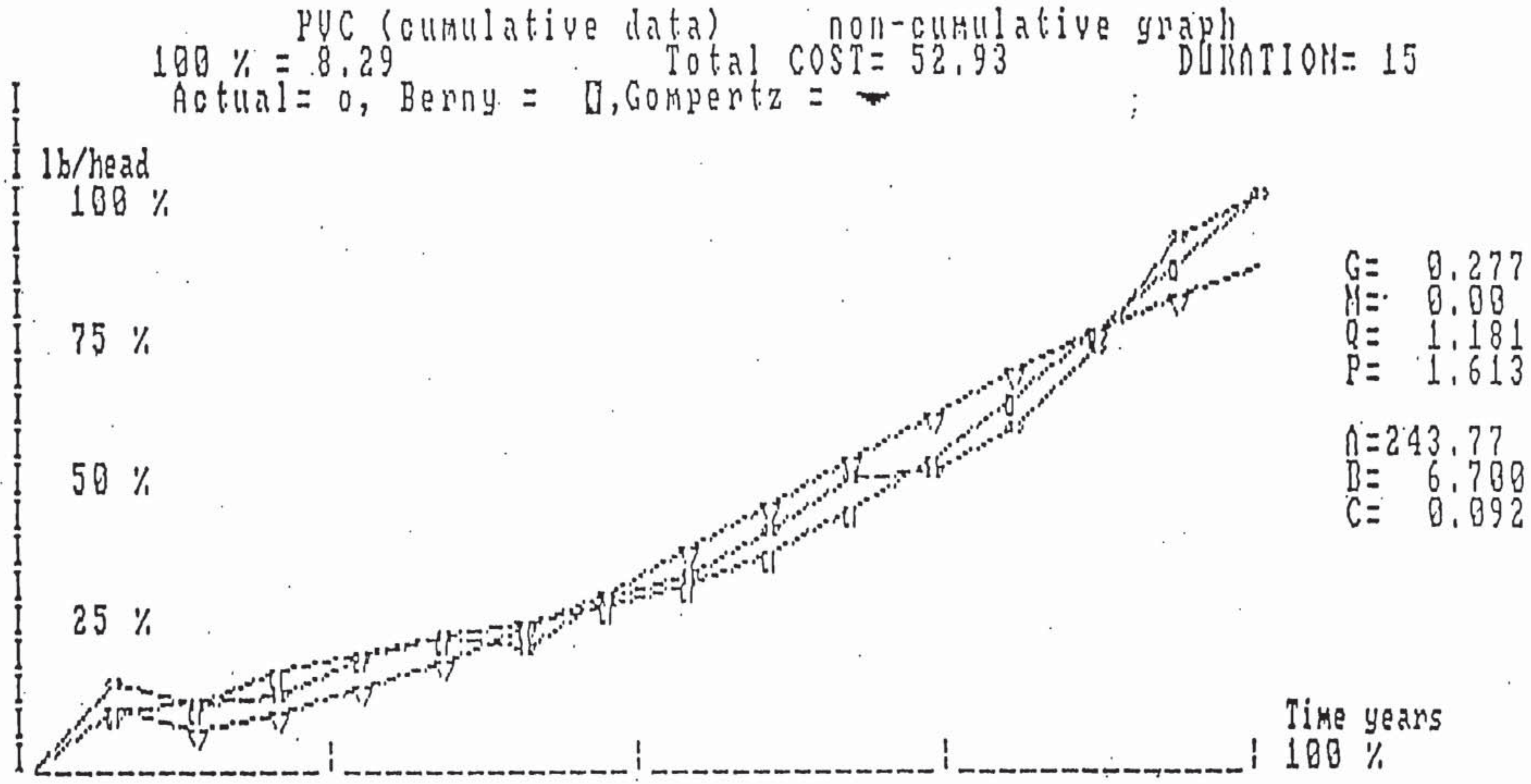


FIGURE A 2.2 *Non-cumulative 'cumulated' PVC data compared with Berny model and Gompertz model*

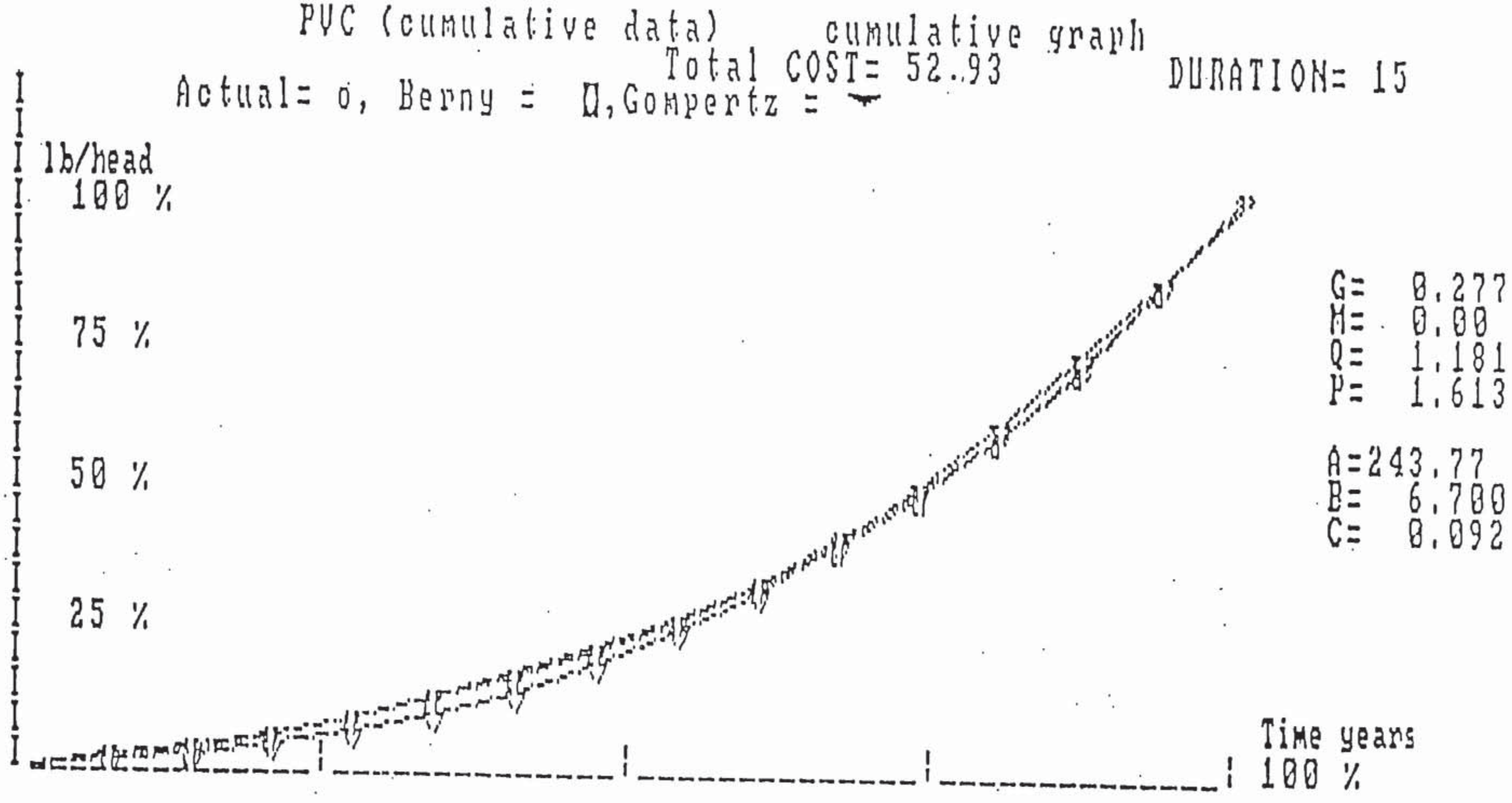


FIGURE A 2.3 *Cumulative 'cumulated' PVC data compared with Berny model and Gompertz model*

In order to judge the forecasting performance of the selected models a prediction was made for 1979 (11 years ahead), which had a production level of 3.43kg/head. The resulting production estimates for 1979 were 3.66kg/head for the new model and 3.61kg/head for the cumulative lognormal.

The change in annual production shows extremely irregular features. This is illustrated in figure A 2.4, which has only 42 periods. The loss of 2 periods was caused by the fact that production stopped in 1926 and 1927. As the initial value (F^*_o) was equal to the 1925 production level the next two years were suppressed. A further consequence of these fluctuations required the unusual step in the model solution to set the specified values one year before the last data values. Thus D=30 (1957) and Y^*_D=2.1 for the trial fit and D=40 (1967) with Y^*_D=2.85 for the full data set. The irregular nature of the data, as well as the fact that it represents a marketing consumption process, indicates the need to cumulate the data as was found valid in the PVC case.

The present solution does have one extra feature which results from the fit/forecast choice of the p=m=0 parameter case. It describes a trend showing a constant annual increase would occur in the foreseeable future. Inspection of figure A 2.4 indicates that this may be a plausible result.

ADOPTION PROCESSES

The last half of this appendix considers Adoption processes. The best known example used in current growth forecasting studies is the following case study on "Spanish Tractor Industry" introduced by Mar-Molinero (1980).

A 2.3 Tractors in Spain

Mar-Molinero 1980 (MM) made a special study of the tractor economy in Spain for which he selected the logistic after an extensive study of the tractor industry in Europe. In particular he was concerned with the interaction between the production levels of tractors and the economic climate in Spain which is a largely agrarian nation. He fitted data spanning the years 1951 to 1976. In an endeavour to improve results he introduced an autocorrelation correction by generating a lagged error term.

Model	TEST span	Parameter values (SE of parameters)				Statistical measures SPE	(SE)	MPE	50% ahead forecast value	(SE*)	period	Comment
Berny		g	m	q	p							
	0 to 33	1.10	0	5.41	0	4.33%	(.091)	0.09%	3.24	(13.4%)	1975	forecast fit
		(1.32)	-	(2.19)	-							
	34 to 43	"	"	"	"	5.40%	(.11)	3.86%				trial forecast
	0 to 43	1.38	0	8.66	0	2.92%	(.088)	0.85%	4.30	(4.0%)	1988	full curve fit
		(0.45)	-	(1.28)	-							
Cum logn		a	b	c	-							
	1 to 34	4.13	3.49	.686	-	4.33%	(.091)	0.21%	3.06	(.1%)	1975	forecast fit
		(1.20)	(.24)	(.11)	-							
	35 to 44	"	"	"	"	7.10%	(.149)	5.73%				trial forecast
	1 to 44	6.12	3.82	.819	-	2.88%	(.087)	1.98%	4.13	(5.2%)	1988	full curve fit
		(1.22)	(.19)	(.085)	-							

TABLE A 2.3 Results for the analysis of Titanium Dioxide production (1925-1968)

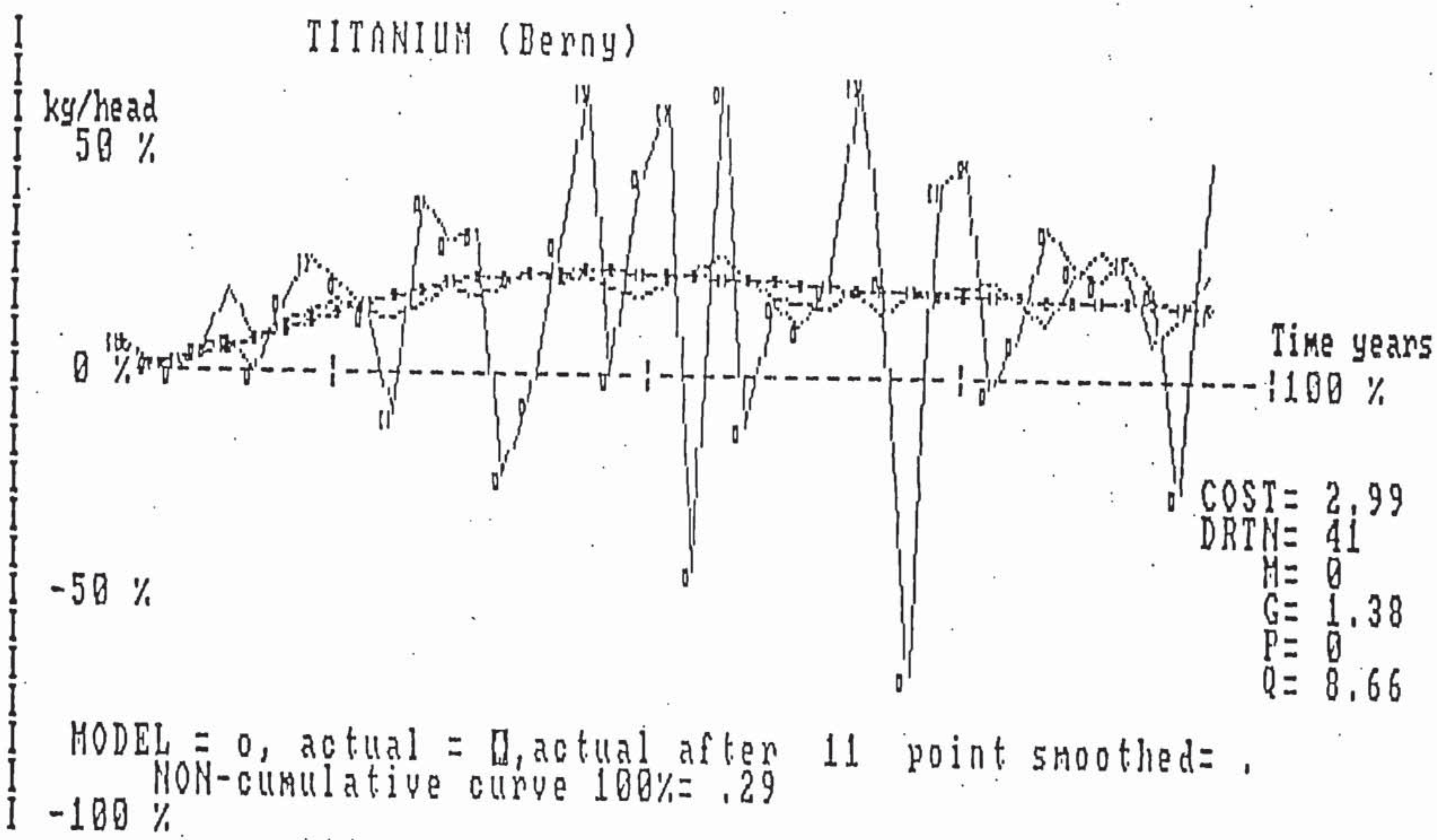


FIGURE A 2.4 *Non-cumulative Titanium Dioxide data compared with Berny model and smoothed data*

The method used for correcting first order autocorrelation may be applied to any of the presented functions in the following manner.

Let y_t represent the model function then:

$Y^*_t = y_t + \delta\ (F^*_{t-1} - y_{t-1})$ ...eq. 27

where δ is the autocorrelation constant

and $Y^*_0 = F^*_0$, otherwise $t>=1$

Furthermore if there is no more data then a trend pattern may be generated by replacing F^*_{t-1} by Y^*_{t-1}.

Meade (1984) continued with the above study and made a fit/forecast using 16 years (1951-1966) for the fit and 10 years (1967-1976) for the trial forecast. Meade's investigation validated MM's choice to use the logistic. Hence this model was compared with the author's. An alternative approach was taken by Oliver (1981) who showed that marginal improvements could be made using his extended logistic see equation 18. The effectiveness of Oliver's model improved with shorter time spans, however this point has not been investigated further.

The use of the logistic has been fully explored in this case study. Tests were made using the extended logistic and the two parameter form shown below, but in the latter case only a full curve fit was made. The author's model was also subjected to autocorrelation and had the same order of success as in MM's case. While the curve fit improved with autocorrelation it made no significant difference to the forecast using the trend extension of equation 27.

Three parameter (Oliver) logistic :-

$Y^*_t = e+(Y^*_D-e)(1+b\ \exp(-cD))/(1+b\ \exp(-ct))$

where D=26 and Y^*_D was taken from the last measured value in 1976

The fourth solution, with m=0, was selected for forecasting purposes. This solution gave a trial forecast SPE=3.79% (SE=.64) which was 64% lower than the best logistic result and 76% lower than the result quoted by Meade. It was not statistically feasible, but markedly better than the next best solution (no. 3 with m=0) which had a SPE=5.15% (SE=.87) and was a SFS. The best logistic result, given by using a statistically derived saturation level, had a trial forecast SPE=10.5% (SE=1.77). This was in itself 34% lower than the result quoted by Meade, who obtained a SE=2.63 or SPE=15.6%. Table A 2.4 summarises these results but only makes comparisons with Meade's results.

Model	TEST	Parameter values				Statistical measures			50% ahead forecast			Comment
	span	(SE of parameters)				SPE	(SE)	MPE	value	(SE*)	period	
Berny		g	m	q	p							
	0 to 13	.193	0	.596	0	2.96%	(.385)	-0.04%	29.1	(15.9%)	1972	forecast fit
		(.094)	-	(.699)	-							
	14 to 25	"	"	"	"	7.12%	(.927)	-6.61%				trial forecast
Logstc.		a		b	c							
	1 to 14	265740		290000	.19	2.27%	(.927)	0.04%	50.0	(3.4% ?)	1972	forecast fit
		$(3.9*10^7)$		$(4.3*10^7)$	(.01)							
	15 to 26	"	"	"	"	310%	(40.3)	-225%				trial forecast

Note :- ? This result may be caused by a matrix instability due to the size of the parameters

Model	TEST	Parameter values				Statistical measures			50% ahead forecast			Comment
	span	(SE of parameters)				SPE	(SE)	MPE	value	(SE*)	period	
Berny		g	m	q	p							
	0 to 15	.227	0	3.29	3.54	1.77%	(.296)	0.85%	34.7	(20.4%)	1974	forecast fit
		(.467)	-	(1.11)	(.65)							
	16 to 25	"	"	"	"	3.79%	(.642)	-14.50%				trial forecast
	0 to 25	.372	0	12.25	12.76	1.02%	(.411)	0.63%	73.2	(2.1%)	1989	full curve fit
		(.039)	-	(1.87)	(1.91)							
Logstc.		a	b	c	e							
	1 to 16	54.0	61.8	.21	-	2.12%	(.358)	0.17%	30.6	(17.1%)	1974	forecast fit
		(18.5)	(16.0)	(.02)								(author's results)
	15 to 26	"	"	"	"	14.8%	(2.52)	-14.5%				trial forecast
	1 to 26	52.1	50.2	.195	-	1,10%	(.442)	0.07%	50.8	(2.5%)	1988	full curve fit
		(1.40)	(2.36)	(.005)	-							

TABLE A 2.4 Results from the analysis of Spanish tractor production (1951-1976)

A recent strategy for forecasting from models, using a Kalman filter, has been developed by Meade (1985). In this work Meade used the logistic as the base model for this case study. However, his fit/forecast used 14 years for the fitting procedure, hence his trial forecast spanned the years 1965 to 1976. The least squares estimate for the logistic grossly overestimated forecasts (MAPE=89%), but the new model gave sound results (MAPE=3.05%). Both 12 years ahead fit/forecasts using least squares are included in table A 2.4. The Kalman filter method gave a mean absolute percentage error (MAPE) value of 5.05%. MAPE was used as it was adjudged by Meade that it was a better indicator of the performance of his technique.

The Kalman filter improves the performance of the logistic and presumably this effect would be more pronounced using the new models. This argument fosters the proposal that better models are as important as techniques for advances in growth curve forecasts.

All the forecasts from the full curve fit indicated far higher levels of production than those generally indicated by the logistic. For this reason the Spanish Embassy was approached, who supplied data upto 1982. The 1982 production was 56.26 (in 10000's), this confirmed the author's estimate namely 55.67 (1% lower), see table A 2.5.

The logistic results gave a range of marginally differing values for 1982 the best being Oliver's logistic in the author's two parameter form giving 49.3 (12.4% lower than the 1982 value) showing some validity in Oliver's extension and the parametric simplification. The logistic appeared unable to give a higher saturation level than the three parameter Oliver model giving a value of 55.8, thus this type of model gave a maximum which was reached between 1981 and 1982, the 1981 production level was 54.81.

The success of the new model with this case study is attributed to the fact that it generated a trend which followed the smoothed data pattern. This profile showed continual growth as opposed to an early peak indicated by the logistic. These points are clearly illustrated in figure A 2.5 and A 2.6. The first compares the new model with the logistic and the next gives a comparison with the smoothed data trace.

The autocorrelation correction showed little impact on the solution as indicated in the list below. This case study illustrates the value of the reduced parameter versions of the logistic suggested earlier and the statistically derived saturation levels as well as the robustness of models built on the concept of a secondary process.

1982 predictions Logistic	Y*	SE	SPE%
1) Meade parameters	47.46	.442	1.10
with autocorrelation	48.04	.320	.80
2) Saturation 'a' derived statistically	47.56	.438	1.09
3) Two parameter case	47.55	.438	1.09
4) Oliver	48.30	.433	1.08
with autocorrelation	48.42	.319	.79
5) Three parameter Oliver	49.25	.434	1.08
with autocorrelation	49.29	.319	.79

1982 predictions Berny	Y*	SE	SPE%
1) Third solution (m=0)	52.98	.364	.94
with autocorrelation	52.92	.291	.75
2) Fourth solution (m=0)	55.67	.411	1.02
with autocorrelation	55.49	.294	.73
ACTUAL 1982 production	56.26		

TABLE A 2.5 Results using all variants of the logistic and Berny models for Spanish tractor production

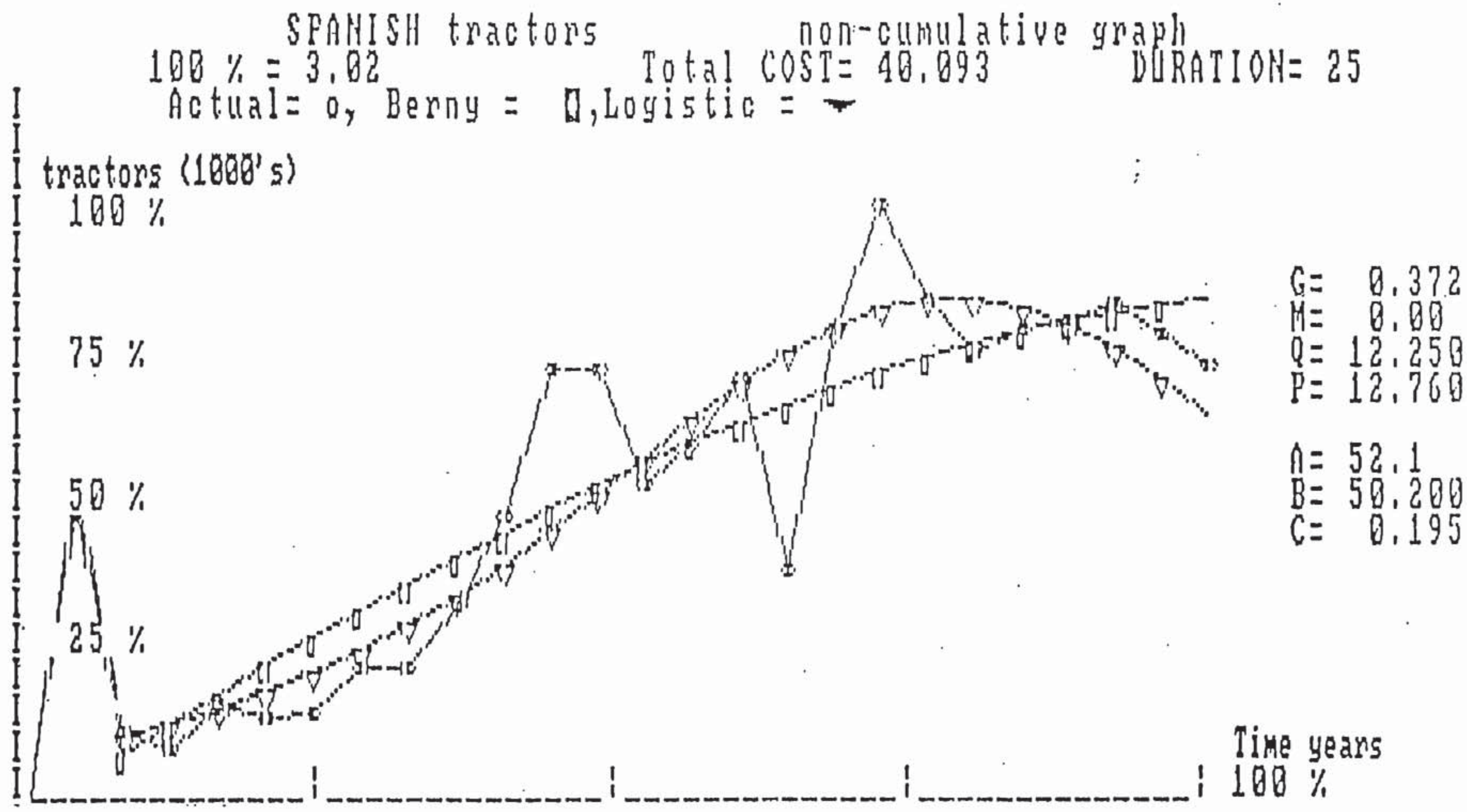


FIGURE A 2.5 *Non-cumulative Spanish Tractors data compared with Berny model and logistic model*

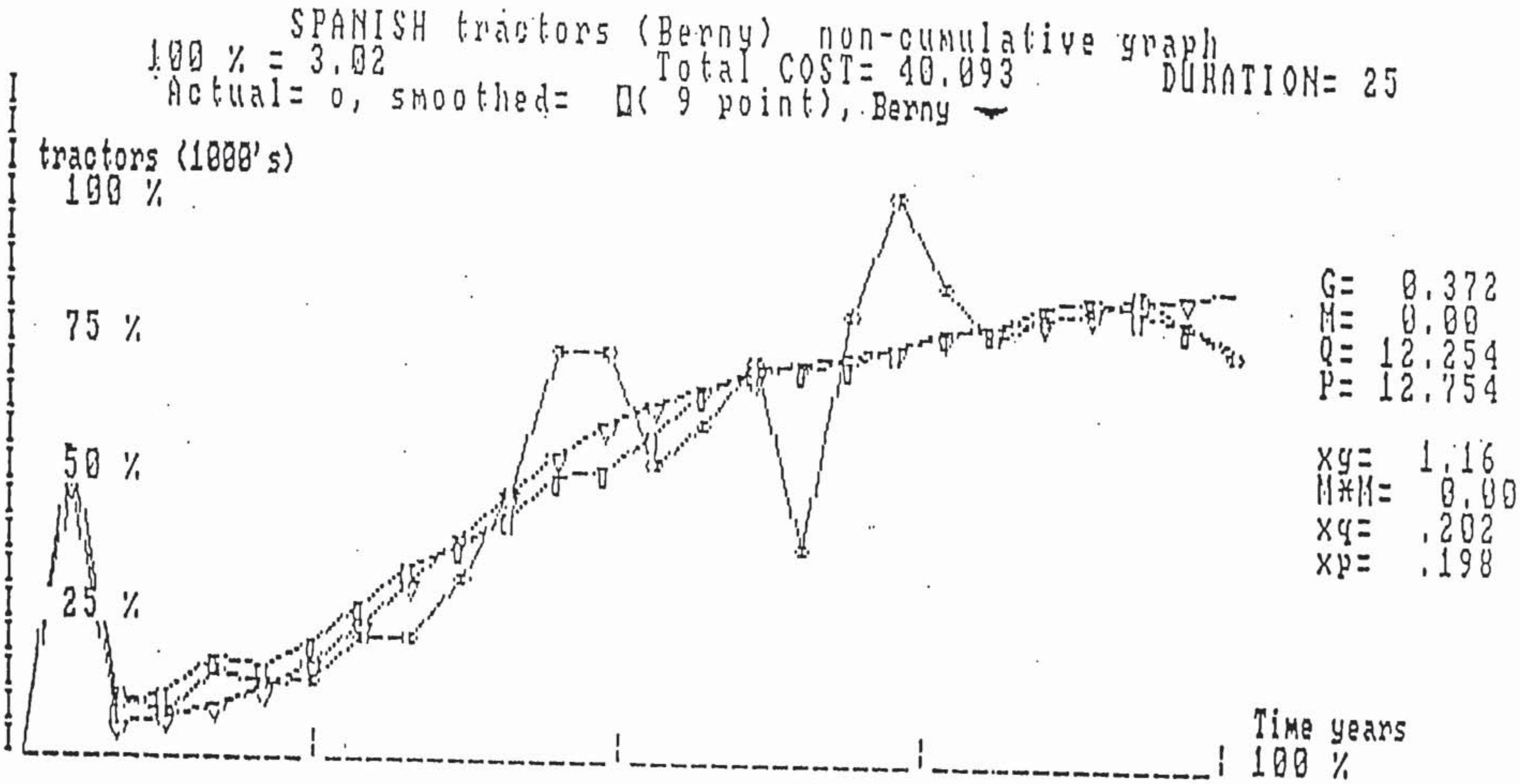


FIGURE A 2.6 *Non-cumulative Spanish Tractors data compared with Berny model and smoothed data*

A 2.4 Colour Television

This case study was concerned with the new rentings and sales of colour televisons in the U.K. between 1967 an 1978. The data was taken from Bagot (1979) and categorised by Meade as an adoption process. The quarterly measured value of the growth curve was the proportion of households owning or renting at least one colour television set.

It was observed by Meade (1984) that there was a marked drop in sales at the beginning of 1975. For this reason he chose to make two fit/forecasts, namely 1967 to 1974 (28 periods) and 1967 to 1977 (40 periods). Meade tested the four models listed in section 4.2 and found that the three which gave SFSs had unsatisfactory saturation levels. Meade selected the Gompertz and cumulative lognormal as a base for forecasting analysis. The logistic was marginally better for forecasting between 1974-1978, but its saturation level (53.2) was lower than the 1978 value (68.11), hence this model was discarded.

The comparison with traditional models is shown in table A 2.6. The trial forecast SPE values are 48% (1974-1978) and 77% (1977-1978) of the lowest respective SPEs given by the two selected traditional models. The full curve fit was also compared with the cumulative lognormal and Gompertz. The last model gave a much better fit than the lognormal hence it was used to compare with the new model see figures A 2.7 and A 2.8. This shows that a better trend pattern exists near the end of the data, which is attributed to the initial slope element.

Meade 1985 also used this data to test his development of the Kalman filter technique and utilised the Gompertz as the underlying model. A similar drop in trial forecast errors, to those found above also, occurred when testing a 12 quarters ahead prediction. The least squares results are included in table A 2.6, the MAPE=.99% for the new model and 1.99% for the Gompertz based Kalman filter trend.

Meade (1984) proposed a solution to the problem of the drop in sales after 1974, by making the saturation level of the remaining models (Gompertz and cumulative lognormal) dependent on the Index of real personal disposable income (rpdi).

To represent the rpdi he introduced :-

parameters a* and b* to replace a,

thus $a=a^* \exp(b^* z_i)$
and z_i is the ith. period rpdi index.

Model	TEST span	Parameter values (SE of parameters)				Statistical measures SPE	(SE)	MPE	50% ahead forecast value	(SE*)	period	Comment
Berny		g	m	q	p							
	0 to 27	.659	0	3.63	5.21	0.96%	(.397)	0.4%	71.3	(6.8%)	1977 Q2	forecast fit
		(.207)	-	(.488)	(.181)							
	28 to 43	"	"	"	"	11.1%	(4.96)	-10.1%				trial forecast
Gompertz		a	b	c	-							
	1 to 28	101	9.31	.085	-	1.20%	(.497)	0.3%	77.3	(4.4%)	1977 Q2	forecast fit
		(8.55)	(.44)	(.005)	-							(author's results)
	29 to 44	"	"	"	"	21.4%	(8.89)	-19.5%				trial forecast
Berny		g	m	q	p							
	0 to 31	.918	0	5.63	7.97	1.61%	(.791)	0.87%	72.5	(6.0%)	1978 Q4	forecast fit
		(.184)	-	(.783)	(.428)							
	32 to 43	"	"	"	"	1.23%	(.606)	1.19%				trial forecast
Gompertz		a	b	c	-							
	1 to 32	74.6	11.6	.106	-	1.30%	(.680)	0.45%	69.5	(3.0%)	1978 Q4	forecast fit
		(2.95)	(.75)	(.005)	-							(author's results)
	33 to 44	"	"	"	"	2.60%	(1.28)	-1.60%				trial forecast
Berny		g	m	q	p							
	0 to 39	1.453	.299	9.41	14.47	1.16%	(.725)	0.23%	79.4	(2.0%)	1981 Q4	forecast fit
		(.074)	(.166)	(.53)	(.99)							
	40 to 43	"	"	"	"	0.60%	(.372)	0.54%				trial forecast
Cum logn		a	b	c	-							
	1 to 40	75.5	3.29	.435	-	1.31%	(.820)	0.42%	73.1	(2.0%)	1981 Q4	forecast fit
		(2.04)	(.02)	(.014)	-							(author's results)
	41 to 44	"	"	"	"	2.67%	(1.07)	2.54%				trial forecast
Berny		g	m	q	p							
	0 to 43	1.767	.312	11.1	17.0	0.96%	(.652)	0.01%	82.1	(1.4%)	1983 Q2	full curve fit
		(.047)	(.103)	(.50)	(.97)							
Gompertz		a	b	c	-							
	1 to 44	75.1	10.7	.102	-	1.17%	(.797)	0.14%	74.1	(1.5%)	1983 Q2	full curve fit
		(.886)	(.451)	(.002)	-							

TABLE A 2.6 Results from the analysis of Colour television sales (Q.1 1967 - Q.4 1977)

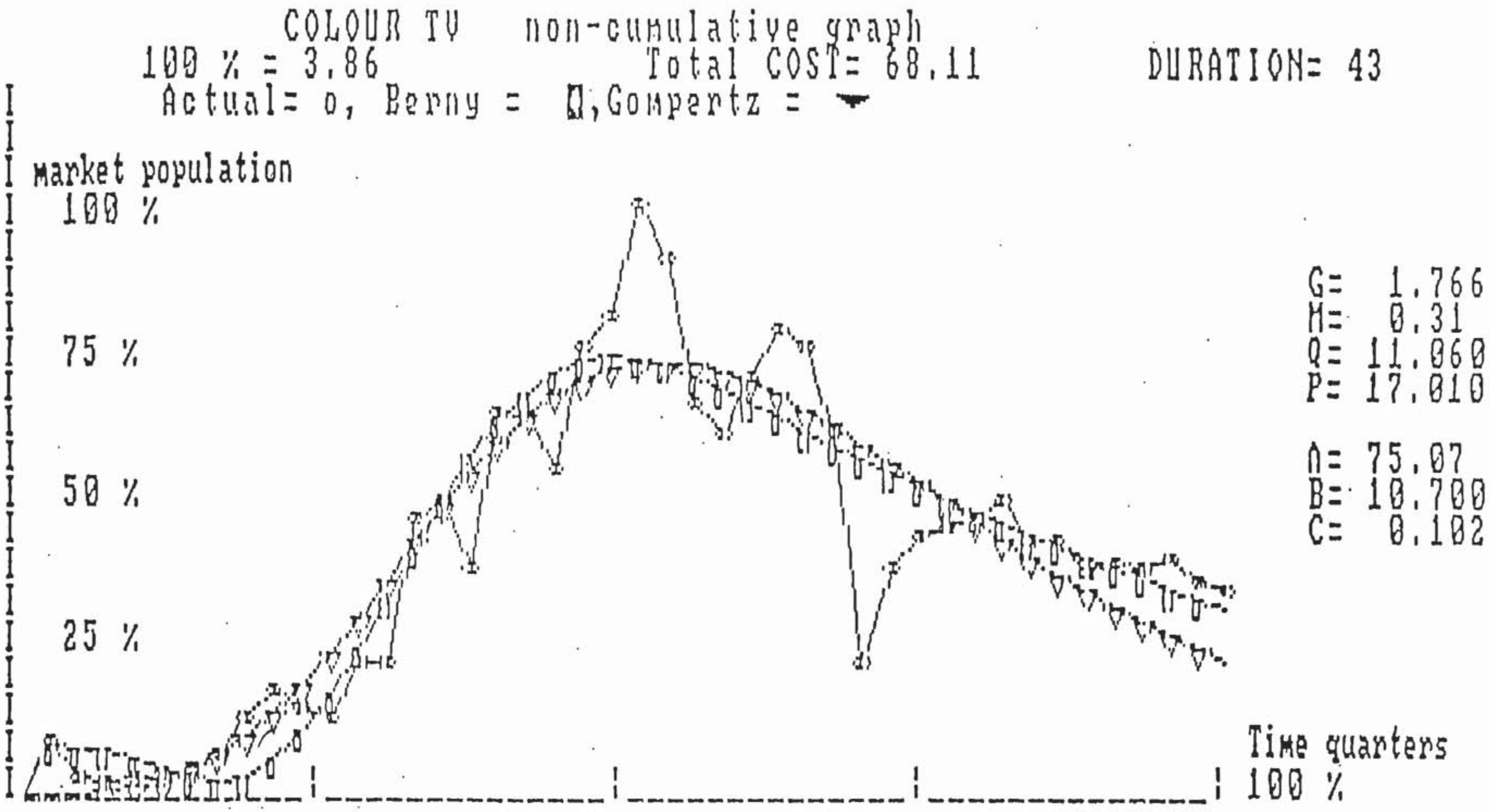


FIGURE A 2.7 *Non-cumulative Colour Television data compared ; with Berny model and Gompertz model*

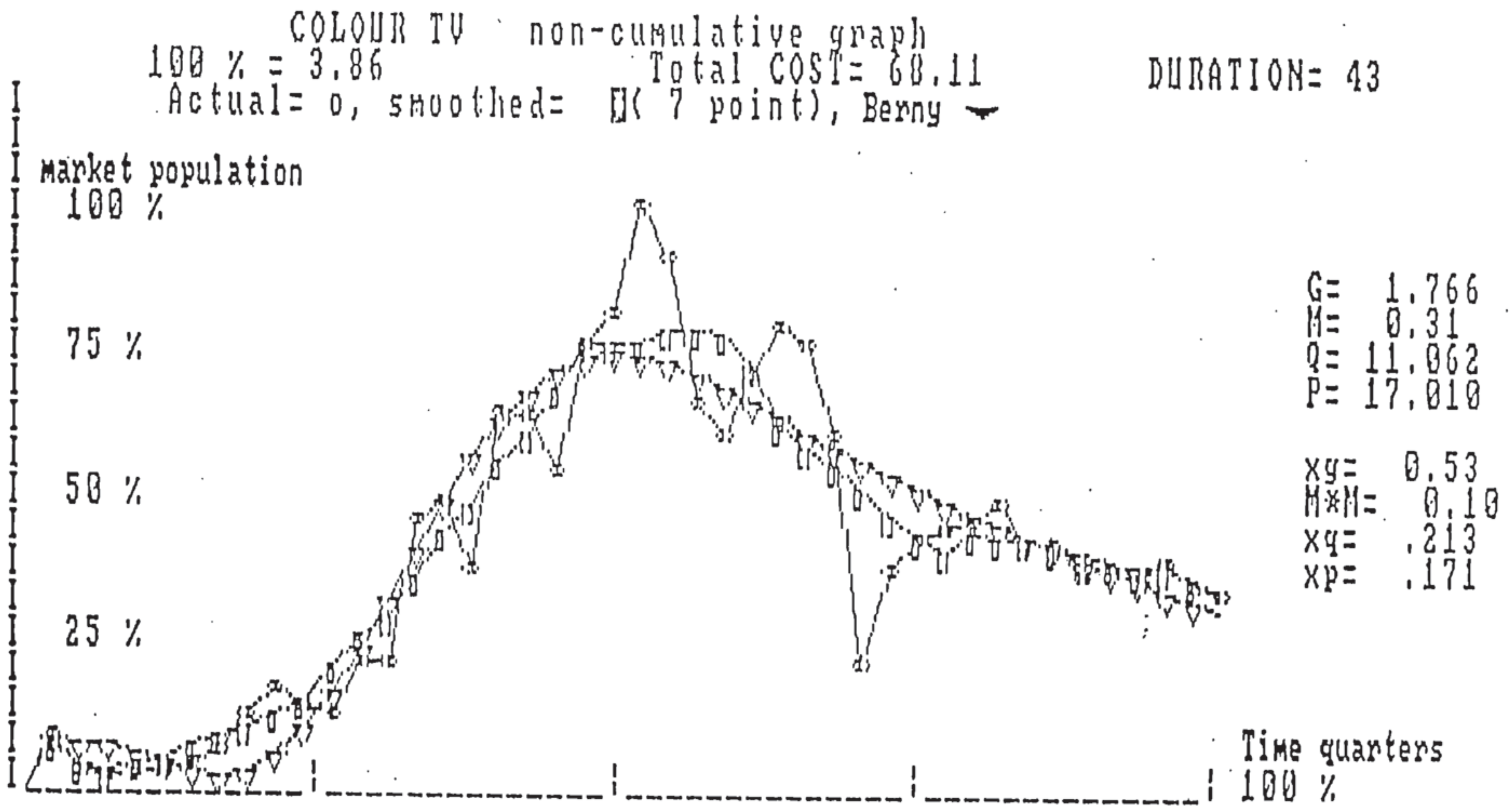


FIGURE A 2.8 Non-cumulative Colour Television data compared with Berny model and smoothed data

The trial forecast results for the four quarters (1977-1978) gave a SE=.35 using Meade's models. The new model, which used no external information, gave a marginally higher SE=.37.

If we regard the rpdi index as a parameter, then his new models have in effect five parameters. As has been shown such models can be reduced to four parameters by use of a specified point and thus his formulations are compatible with the four parameter solution. This assumes the rpdi index is linearly forecast. Meade adopted this condition to make preictions from 1978. His results gave a sales value for period 52 (Q.4 1980) of 75.8 with 95% confidence limits of 72.2 to 78.6. The new model gives comparable results, the expected sales being 75.6 with 95% confidence limits estimated to be 74.1 to 77.1.

The need to use additional data to assist forecasts is viewed as a weakness, to be avoided if possible, and probably created the wider confidence limits shown above for the Meade model.

APPENDIX 3

CONSTRUCTION INDUSTRY CASE STUDIES

The need to enlarge the scope of growth curve studies was highlighted at the outset of this thesis. The observation, that there is a dearth of analytic work in the rich field of project management, is being corrected in this appendix by the inclusion of two Construction Industry projects. These were subjected to the same rigourous considerations, as the marketing cases in the previous appendix.

A 3.0 Introduction

The projects considered for inclusion were selected on two premises, namely type of work and growth curve characteristics not found in the earlier marketing studies. The first was the cost breakdown of civil engineering work on the A40 road. The monthly costs exhibited bimodality, which was stated to be caused by a dramatic drop in the work load during severe winter weather. The second analysis, the building of an advanced factory, was selected because of the nature of the work and that the growth showed a 'discontinuity' in the cumulative cost profile. The reason for this was an exceptionally high cost of one months duration, possibly attributed to machinery instalment, typical of new industrial estates.

A 3.1 A40 roadworks

The major reason for the inclusion of this project is to demonstrate the effect on the prediction of paying attention to bimodality in the growth curve model. It is also being used to show the use of the unity point when it is placed at the anticipated project end. In order to achieve a comparison with the listed traditional models the two parameter solution is included.

The data was subdivided for the fit/forecast commencing at period 0 (with initial cost zero) fitting to period 18 and the forecast was made to period 24 in order to establish a viable solution. This resulted in solution four being selected with a forecast SPE=9.3%.

The second stage in the testing was to forecast from period 24 to period 31. This showed that the forecast given by solution two was marginally better than that of solution four in terms of the SPE (1.9% versus 2.0%), but the MPE was 40% lower for the four parameter forecast. The latter indicates that the predicted values passed more smoothly through the data in solution four. Nevertheless the purpose of this exercise was to establish the superiority of solution four for bimodality, this is illustrated in figure A 3.1 and is justified as its full curve fit gave a SPE=2.3% using the algorithm.

The comparison between the traditional models and solution two is included in the summary given in table A 3.1. The initial testing for fit/forecasts showed that the cumulative lognormal gave the best prediction from period 18 to 24 with a SPE=18.4%, which is nevertheless higher than a SPE=12.6% for solution two, thus solution two has its SPE 30% lower than the lognormal but the trial forecast for solution four

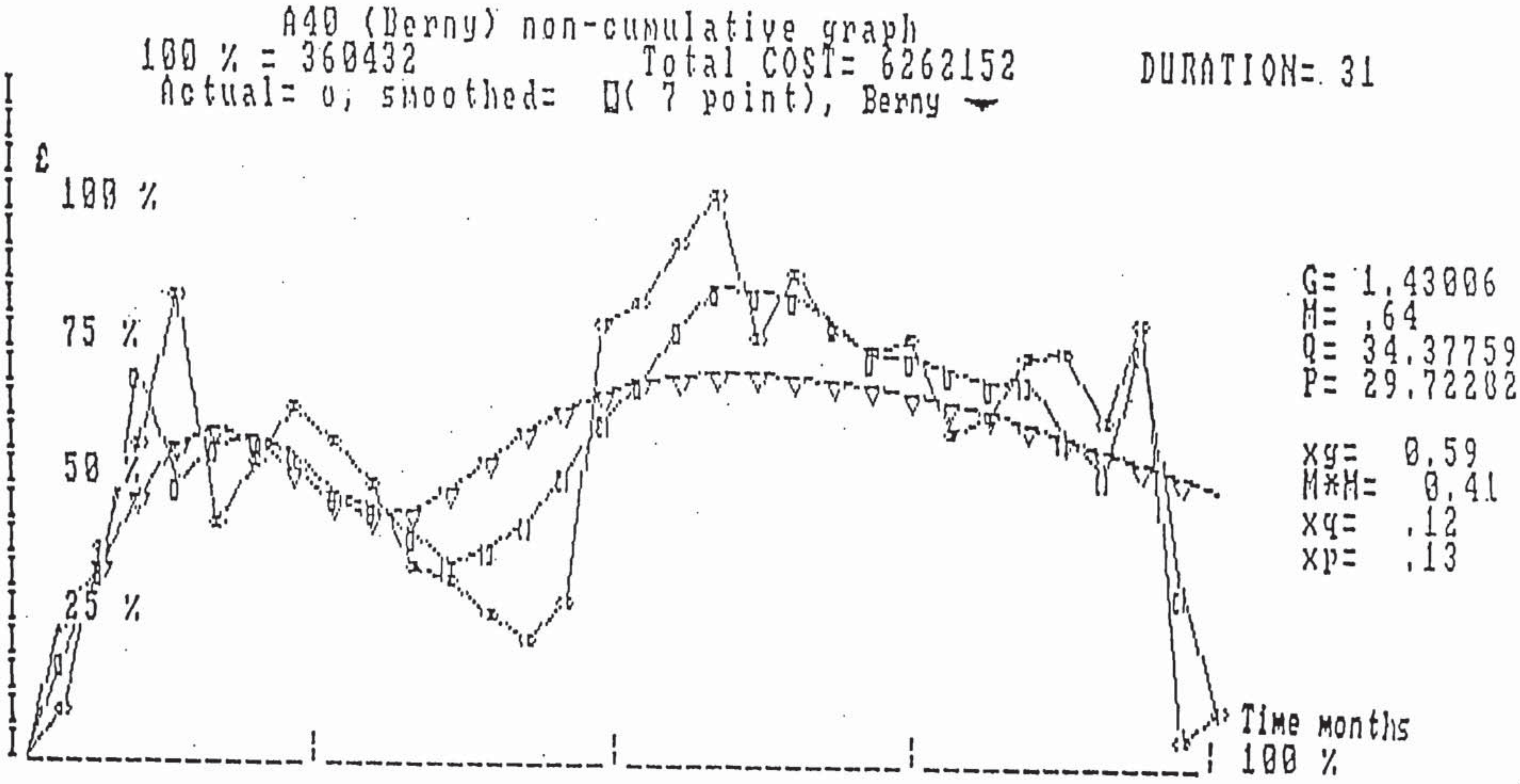


FIGURE A 3.1 *Non-cumulative A 40 Roadworks data compared with Berny model and smoothed data*

Model	TEST span	Parameter values (SE of parameters)				Statistical measures SPE	$(SE*10^6)$	MPE	50% ahead forecast value$*10^6$	(SE%)	period	Comment
Berny		g	m	q	p							
	0 to 18	.794	.533	11.4	9.1	4.55%	(.153)	-2.5%	4.934	(6.0%)	mth 27	forecast fit
		(.705)	(.27)	(5.6)	(3.7)							
	19 to 24	"	"	"	"	9.30%	(.314)	8.3%				trial forecast
Cum logn		$a*10^{17}$	b	c	-							
	1 to 18	5.39	47	6.5	-	3.83%	(.129)	-1.1%	4.769	(5.9%)	mth 27	forecast fit
		(.306)	(2.4)	(.11)	-							
	19 to 24	"	"	"	"	18.06%	(.608)	16.4%				trial forecast
Berny		g	m	q	p							
	0 to 24	.751	.559	24.0	20.9	2.83%	(.141)	-1.30%	7.33	(8.4%)	mth 36	forecast fit
		(.387)	(.117)	(7.3)	(16.8)							
	25 to 31	"	"	"	"	2.07%	(.103)	-0.23%				trial forecast
Cum logn		$a*10^{17}$	b	c	-							
	1 to 24	136.7	44.8	5.8	-	3.31%	(.165)	0.13%	8.15	(4.8%)	mth 36	forecast fit
		(9.66)	(3.2)	(.09)	-							
	25 to 31	"	"	"	"	4.27%	(.212)	-1.97%				trial forecast
Berny		g	m	q	p							
	0 to 31	1.425	.637	34.4	29.7	2.30%	(.144)	-0.20%	8.02	(3.7%)	mth 47	full curve fit
		(.225)	(.15)	(10.6)	(9.7)							
	0 to 31	1.126	.756	0	0	2.87%	(.179)	0.84%	8.38	(4.0%)	mth 47	full curve fit
		(.206)	(.058)	-	-							
Cum logn		$a*10^{17}$	b	c	-							
	1 to 31	186.7	46.5	6.0	-	2.68%	(.168)	0.04%	10.96	(16.5%)	mth 47	full curve fit
		(67.1)	(15.6)	(.31)	-							

TABLE A 3.1 Analysis of results for A40 Roadworks was 49% lower. Continuing with these tests the forecast from period 24 to 31 gave a SPE=4.3% for the cumulative lognormal whilst solution two had a SPE=1.9%, which is 56% lower. Finally the full curve fit gave a SPE=2.7% for the traditional model and solution two had a SPE=2.9%. The last SPE values give a clear indication of the need of the four parameter solution which gave a best SPE=1.9%. This analysis underlines the advantage gained by the concept of the secondary process.

The unity point may be used also as an aid in project control. Two cost profiles may be generated. One using the last measured value for the unity point and the second using the anticipated project end. These projections into the future will give an indication of lag or lead in the actual work.

The results showed that forecasting from period 16, which is too early to account for the bimodality and as anticipated the forecast was poor. It gave a SPE=90% for the trial prediction when the unity point was set at period 16. Some improvements could have been made to this result by subjecting the data to deeper analyisis in the manner described for the earlier case studies.

The intention of this exercise was to use the algorithm in the most straightforward manner. This required that selection was automated using the weighting scheme and that no use was made of alternative starting values of 'g '. The setting of the unity point at the project end dropped the SPE to 6%, thus confirming the overshoot and illustrating the fact that projections can be succesfully made at much earlier points in a project than would normally be considered by an analyst.

An additional facility in the suit of programmes checks for new project phases which are the major cuase of bimodality, this was confirmed in this case study. This indicator not only forwarns of potential hazards in prediction but also alerts the analyst to the possibilty of commencing forecasts at a later stage in a project by discarding the earlier data.

A 3.2 Advance factory construction

This case study was primarily concerned with demonstrating the value of suppression to overcome the problem of known discontiniuities. It also served as a demonstration of both the limitations governing detection of bimodality by the model and as a further example of the unity point concepts.

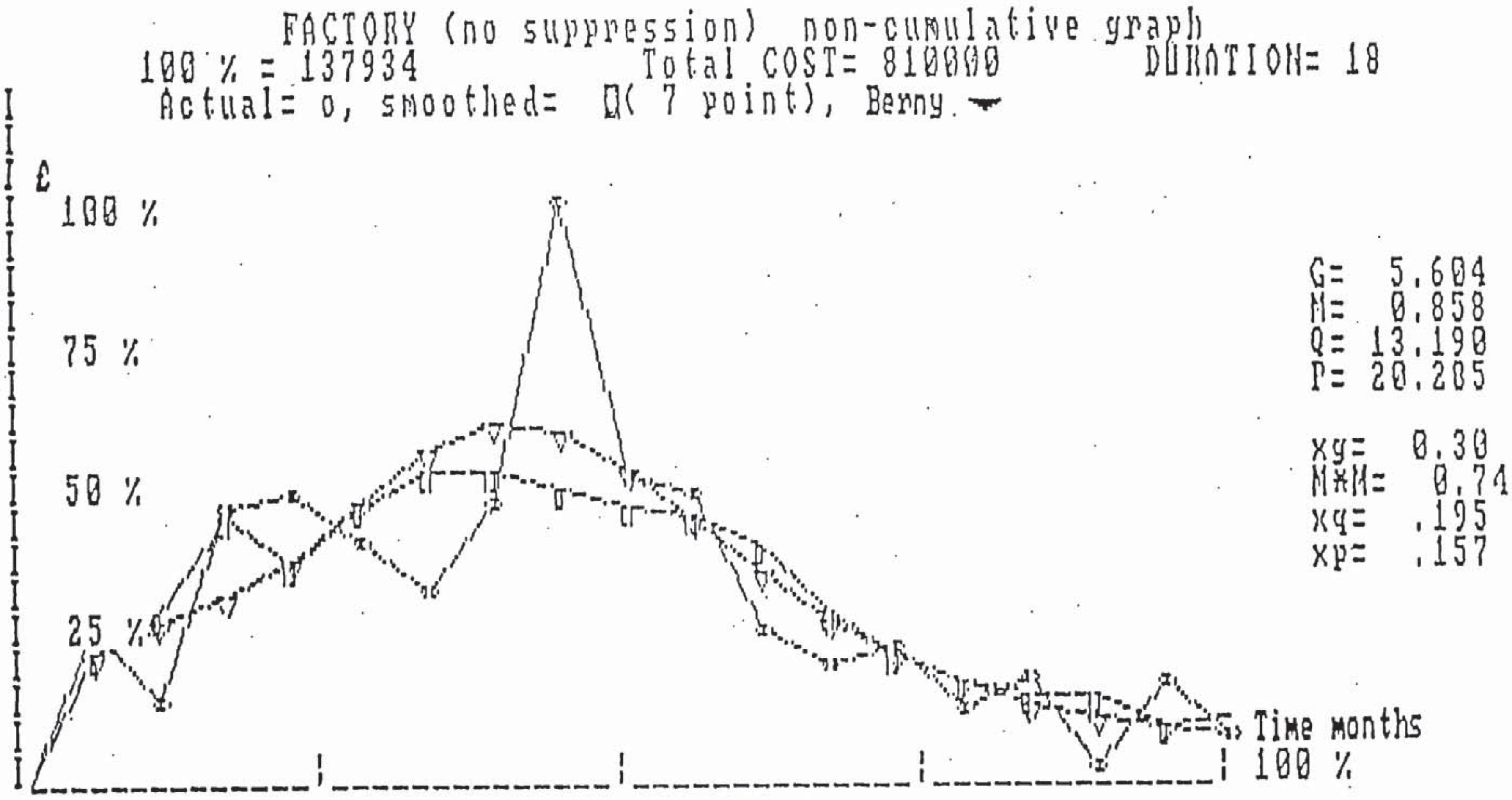


FIGURE A 3.2 *Non-cumulative Advanced Factory data compared with Berny model and smoothed data*

In order to adhere to the proposed method of analysis (see 4.21b) three time divisions were created, namely periods 0 to 7, 8 to 12 and 13 to 18. These were based on inspection ofthe period cost profile, see figure A 3.2, and constrained by the small sample size. However the cost pattern indicated that the first time division was unlikely to be viable. This proved to be the case as the fit/forecast tests could not clearly indicate which solution should be used for further tests.

On the basis of the fit/forecast from period 11 to 18 solution 2 would have been selected with or without suppression hence this solution will be compared with the traditional models. The corresponding forecast SPE values for solution 2 were 1.3% when period 8 was suppressed and 4.8% without suppression. This large difference was due to the error caused by the high cost of period 8, this is clearly indicated in figure A 3.2.

In view of the deterioration of results cauesd by the high cost period (8), the comparative tests were only conducted on data with suppresion. The prediction from period 11 to 17 indicated that the logistic was the best traditional model. The trial forecast using the logistic gave a SPE=2.28%, whilst solution 2 gave a SPE=1.34% (41% lower). The final curve fit gave a similar order of results, see table A 3.2. The SPE=1.98% for the logistic and 1.39% for the new model.

Another effect of suppression on the solutions was to show the limitatins of the model to detect bimodality. Thus before suppression solution four, illustrated in figure A 3.2, gave a trial forecast SPE=1.8% and it was statistically feasible. After suppressing period 8 solution four was not statistically feasible and gave a SPE=1.5%.

The nonfeasibility of solution four after suppression, which reduced its duration to 17 periods, can be seen from figure A 3.3 to have noncumulative cost peaks at periods 4 and 8. These are too close together and the trough is not low enough to allow the model to fully cater for this bimodality. It is an indication of the fact that the high period cost created a reasonable degree of bimodality and thus solution four which had no suppression gave a SFS and also the best solution under these circumstances. It has been found that the resolution between peaks needs to be about 6 periods or more for detection by the model. It is for this reason that projects with highly irregular period cost profiles are unlikely to utilise solution four.

Model	TEST span	Parameter values (SE of parameters)				Statistical measures SPE	($SE*10^6$)	MPE	50% ahead forecast value$*10^6$	(SE%)	period	Comment
Berny		g	m	q	p							
	0 to 11	1.905	.730	0	0	1.77%	(9943)	0.14%	676490	(2.4%)	mnth 17	forecast fit
		(.184)	(.10)	-	-							
	12 to 17	"	"	"	"	1.34%	(7528)	-10.3%				trial forecast
	0 to 17	4.755	.830	0	0	1.39%	(9311)	-0.40%	848322	(1.2%)	mnth 26	full curve fit
		(.099)	(.08)	-	-							
Logistic		$a*10^6$	b	c	-							
	1 to 11	8.014	23	.44	-	2.68%	(17970)	-0.09%	793821	(5.8%)	mnth 17	forecast fit
		(.047)	(3.2)	(.035)	-							
	12 to 17	"	"	"	"	2.28%	(15304)	-1.8%				trial forecast
	1 to 17	.7908	22	.44	-	1.98%	(15843)	-0.14%	790624	(2.4%)	mnth 26	full curve fit
		(.0095)	(2.4)	(.018)	-							

TABLE A 3.2 Analysis of results for Advanced Factory Construction with one month suppressed

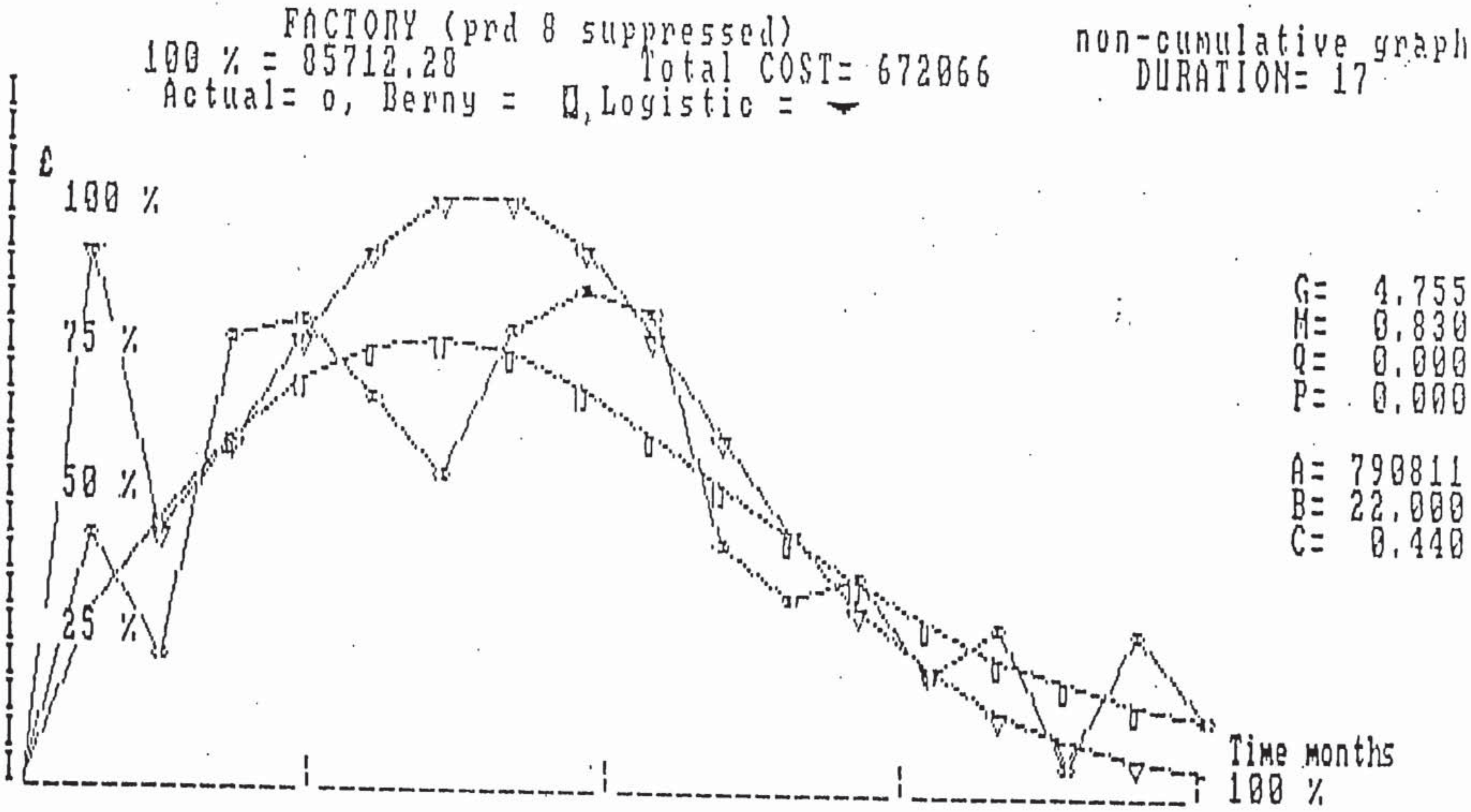


FIGURE A 3.3 *Non-cumulative Advanced Factory data, with period 8 suppressed, compared with Berny model and logistic model*

This project also typifies the use of the unity point at a project end as shown in figure A 3.4. It demonstrates the effectiveness of this technique for 'what if?' forecasting; as described in the previous case study, but under even more severe conditions. Period 9 was selected for this critical test, with suppression this is period 8. This choice was based on the earlier result that the data was not amenable to a fit/forecast between periods 7 and 11. The results showed a drop of trial forecast fram a SPE=46% to 3% when the unity point was set at the project end and no suppression was made.

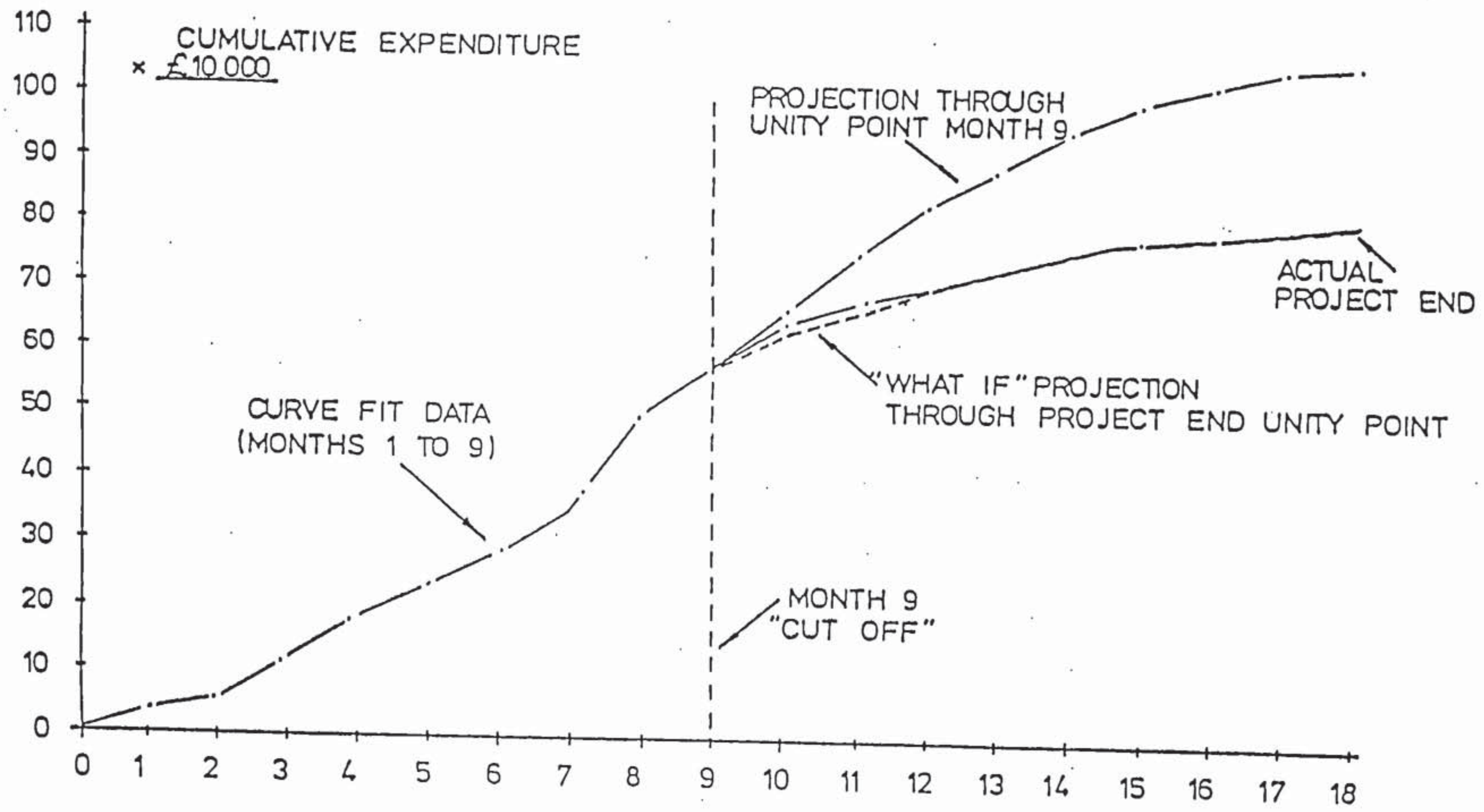


FIGURE A 3.4 *Cumulative Advanced Factory data demonstrating the "What if ?" forecast analysis*

A P P E N D I X 4

VISIER DEMONSTRATION/TUTORIAL SCREENS

The screens reproduced in this appendix are those currently used to demonstrate and support initial instruction in the use of this software. They are the result of an intensive exercise with members of industry who helped to ensure this device was upto, I quote "Lotus 123 standards". This has been subsequently upheld as Computer Associates, the developers of Supercalc, have accepted the package as an expansion of the latter and Super Project Expert (CPA program).

V I S I E R DEMO START UP
=========================

This is a special section used to alter the timing of AUTOMATED SCREEN answers

Typically faster computers e.g. IBM AT may require increment 4

Current delay increment is 0

Do you wish to alter the time delays ans. 'y or n'

The delay may be increased incrementally

Your choice is 0 to 9 increments.

0 gives no change 9 effectively multiplies by 10

Typical alternatives: Amstrad 2, Olivetti M24 3, '286 or 386' machines 4

INPUT your increment (0 is for 1983 OLD IBM PC) you selected 0

press space bar

" V I S I E R "

demonstration

FINANCIAL PLANNER
RISK ANALYSER &
REAL TIME FORECASTER

Corporate, Project
Market & Growth plans

VIS ual
I nterpretation
& E valuation
of R isks

Press space bar to continue

V I S I E R contact:- J. Berny B Sc. M Inst P, MBIM, C Phys, FOR
39 Buckingham Av., E. Molesey, Surrey KT8 9SY. tel. 01-979 9219

V I S I E R introduction

========================

The demo/tutorial follows the MENU along the route shown below
There are 40 screens, min. 'demo' time 12 mins. but most likely 30 mins.

Plan expenditure on a project
:
↓
Update with actual costs and forecast new cost breakdown
:
↓
Demonstrate in depth cash flow inspection facilities
:
↓
Explain 'What If ?' forecasting facilities
↓
Perform one Risk Assessment cycle

Press space bar to continue

V I S I E R MENU

==================

This is an automated & cut down version of the full V I S I E R package
VISIER TEXT usually gold on dark blue is OVER-LAYED with commentaries

a Planner
b What if ? planner
c Monitor current updates
d What if ? current work
e Inspection <to activity level>, control, DCF & inflation
f Forecasters & WHAT IF? projections tested for viability
g Contingency/risk analysis
h Past work analyser and summary
i H E L P
j EDITOR & PRINTER >what if?<
k JOB OPERATIONS

graphics in modules a to g

***** * ONLY PRESS the SPACE BAR ALL answers are automated * *****

Press SPACE BAR

Expanded menu

a Planner
Used at the start of planning and also allows import of earlier VISIER work
* e.g. for corporate plans. Building of project shells.

b What if ? planner
to allow alternatives & comparisons to be made with a plan

c Monitor current updates
forecast adjustments based on update information, time & 'cost' % completion

d What if ? current work
as for 'plan what if?' ‹b to d include comparative graphics›

press SPACE BAR to continue

e Inspection ‹to activity level›, control, DCF & inflation
lag, retention, discounted cash flow comparisons,
* control by detection of high cost regions of work

f Forecasters & WHAT IF? projections tested for viability
a variety of 'cost' forecasts or projections is available

g Contingency/risk analysis
forecaster based high, expected and low cost graph, also time histograms

h Past work analyser and summary
from activity level upwards,
* statistical analysis of 'cost', time and parameters

press SPACE BAR to continue

i H E L P
These screens

j EDITOR & PRINTER ›what if?‹
with extra WHAT IF? facilities, allows new plan creation by
* adding, changing or deleting from plan shells

k JOB OPERATIONS
copy to temporary files or plan shells ‹reusable plans›, delete and create

l
NOTE :- EDITING allows ANY COMBINATION of activities to be FULLY INVESTIGATED

The DEMONSTRATION starts with module (a) ==> PRESS SPACE BAR

V I S I E R - demonstration

P L A N N E R
module a

JOB TITLE :-HI RISE OFFICE (sample job)

Client :- VISIER

COST or RESOURCE - TIME BAR CHARTS

a Whole project or 1 activity
b Multi activity project, use separate activities for high & low cost eg hols
c Multi-project (corporate plan)

CORPORATE PLAN (uses c)

PROJECT PLAN (use b or c) — WHOLE PROJECT (use a) — INVESTMENTS etc. (use a or b)

'centres' of activity <== minor activities — 'parcels' of work

The above flow-chart shows the breadth of user options (ANY no. of elements)
Using GROWTH 'S' curves allows straightforward COST-TIME BAR CHART building.

press SPACE BAR to continue

V I S I E R - demonstration

P L A N N E R
module a

This SCREEN prepares the 'job' outline
The answers are in pink blocks
Is SMALLEST time UNIT:-
1 Mnths (cal) 2 Weeks 3 Qutrs 4 F'tnghts
5 Mnths (4wk) 6 Years 7 other :- No. ?

Anticipated Project Cost and Duration input
these may be changed while planning

Durtn = 18 Mnths cal , PLAN expenditure (£) 3200000

Your UNIT of MEASURE has been restricted :-
to £ '000s and your PLAN expenditure = 3200 £ '000

expenditure outlay prior to PLAN start ==> 20 e.g. for bidding costs

---

The 20 £ '000 will ONLY be added to the final plan category new
F3 & ↵ redo screen if required Job code a-smith
F1 & ↵ menu press SPACE BAR to continue

V I S I E R - demonstration

P L A N N E R
module a

This SCREEN prepares method of cost input
The answers are in pink blocks

Both the previous screen and this one outline the project size
and also the manner of cost or resource generating.

INPUT NUMBER OF activities

UP TO 55 then ↵ =========> 2

Each activity may have a CODE

The monthly costs may be simply generated --> by use of GROWTH 'S' CURVES

---

category new
F3 & ↵ redo screen if required Job code a-smith
F1 & ↵ menu press SPACE BAR to continue

V I S I E R - demonstration
This SCREEN cost-time BAR CHART BUILDING
The answers are in pink blocks

P L A N N E R module a

The remaining screens ask for activity outlines
Projects or major activities, previously prepared in VISIER, may be imoorted

If a multi-activity plan is started one may STOP exit to menu & ED

Enter NAME of activity 1 or TO STOP use F1 & ↵
upto end of 'line' ===> STRUCTURE

ENTER expenditure (£ '000) of :- activity (1) ===> 1800
% profit (on expenditure) 10 % (this creates an income plan)

Cumulated expenditure= 1800 £ '000 remainder of plan expenditure 1400 £ '000
Start at 1 End at period 12 Mnths cal

---

press SPACE BAR to continue category new

F1 & ↵ STOP (Edit if needed)

V I S I E R - demonstration

P L A N N E R module a

Planned expenditure= 3,200.00 £ '000
duration = 18 Mnths cal

Resource or cost generating

The next SCREEN shows the basic GROWTH and 'S-shaped' curves.

Selection of an S-shaped curve is made by question & answer

The monthly costs may be simply generated. N.B. DHSS type is curve C

---

category new
Job code a-smith
press SPACE BAR to continue

```
┌─────────────────────┐  Plan expenditure =     3,200.00 £ '000
│    M O N I T O R    │   duration = 18  Months cal
│  current  updates   │  JOB TITLE :- HI RISE OFFICE (sample job)
│   module  c         │
└─────────────────────┘  VISIER     PROJECT PLAN
```

This section allows the input one or more monthly updates and has the option to allow changes in current & future activities to be made.

It UPDATES the last PROJECTION based on current information

```
          Project changes & current update

     How many periods to input  then ↵  ==) 1
     Are there any alterations to the plan ?  Ans. -> y
     Is the START DATE :- 9 Jun 1987 still valid ?  Ans. -> y
```

```
                                                      category new
     loading data                                 Job code a-smith

             press SPACE  BAR to continue
```

```
                              V I S I E R  demonstration
┌─────────────────────┐  Plan expenditure =     3,200.00 £ '000
│    M O N I T O R    │   duration = 18  Months cal
│  current  updates   │
│   module  c         │
└─────────────────────┘

          Activity  1    Structure

Activity expenditure BUDGET = 2000  Starts at Months cal 1  Ends at 13

   Corresponding planned Activity @ period  3  was  261.59
          Activity period  3  expenditure  then ↵  ===) 450


Prd. 3 expenditure         450.00 cum.= 950  leaving=      1,050.00

                           press SPACE BAR
```

```
                                                      category new
                                                  Job code a-smith
```

V I S I E R demonstration

M O N I T O R current updates module c

Plan expenditure = 3,200.00 £ '000
duration = 18 Months cal

Activity 1 Structure

Activity expenditure BUDGET = 2000 Starts at Months cal 1 Ends at 13

Can you give % expenditure completion of this activity ? Ans. -> y

INPUT % completion then ↵ ==> 35 %

INPUT % completion then ↵ ==> 20 %
New start @ 1 end @ 15 with expenditure = 2714.286 £ '000

---

category new
Job code a-smith

press SPACE BAR to continue

V I S I E R demonstration

M O N I T O R current updates module c

Plan expenditure = 3,200.00 £ '000
duration = 18 Months cal

Summary of updates

Since last update (2) expected expenditure= 4114.286 duration= 18

This is 21.01 % over 3400

with 0 % change of duration= 18

Are these changes valid ? Ans. -> y

---

press SPACE BAR to continue

category new
Job code a-smith

V I S I E R demonstration
Plan expenditure = 3,200.00 £ '000
duration = 18 Months cal

M O N I T O R current updates module c

a Return to menu when data stored
b Graphs or for What if? to cashflow (module e) when data stored
c PRINT, GRAPHS & SCREEN review when data stored
d PRINT (screen display automated) only when data stored
e GRAPHS & SCREEN review only when data stored

Storing data Job code a-smith

HI RISE OFFICE (sample job) expenditure Updated to 3 Months cal
The following SCREENS show the cost-time bar chart of the current activity

Activity expenditure :- * timescale: Months cal *

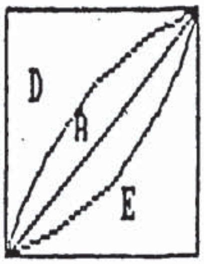
	1	2	3	4	5	6
		>> Structure <<				
Actl/fcst:--> *	200.0	300.0	450.0	307.1	297.5	265.4
	(-actual-)	(-actual-)	(-actual-)	-forecast-	-forecast-	-forecast-
Planned:--> #	144.0	234.0	270.0	288.0	234.0	198.0

TOT NON-CUM actl	200.0	300.0	450.0	---	---	---
TOTAL CUM actual	200.0	500.0	950.0	---	---	---
TOTAL CUM plan	144.0	378.0	648.0	936.0	1,170.0	1,368.0
CUM % over planned	39 %	32 %	47 %	---	---	---

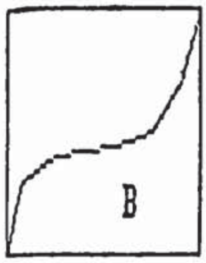
press SPACE BAR to continue

V I S I E R Growth (S) curve analysis (C) J Berny 1986

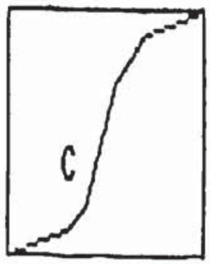
Investment or Overhead



Mining



DHSS or Banking

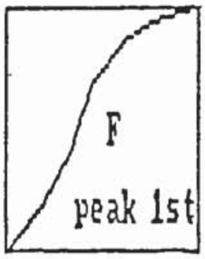


Curve type ? (H for question & answer selection)

Curve C has a work peak in the middle of the activity & a medium start up

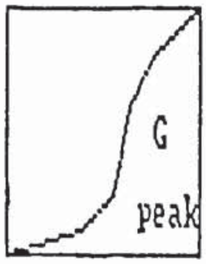
Curve :: c :: selected.

Factory



peak 1st qtr

Hotel



peak 3rd qtr

-*-*-* Press SPACE BAR to continue -*-*-*

V I S I E R - demonstration

P L A N N E R module a	Planned expenditure= 3,200.00 £ '000 duration = 18 Mnths cal

The planning session is now complete

This is normally followed by :-

a Cost - time bar chart, Summary and Graphics

TRANSFERRING to CURRENT UPDATE module (c)

category new

press SPACE BAR to continue Job code a-smith

HI RISE OFFICE (sample job) expenditure Updated to 3 Months cal
The following SCREENS show the cost-time bar chart of the current activity

Activity expenditure :- * timescale: Months cal *

	7	8	9	10	11	12
	>> Structure <<					
Actl/fcst:--> *	221.7	176.0	134.8	101.4	76.7	59.4
	-forecast-	-forecast-	-forecast-	-forecast-	-forecast-	-forecast-
Planned:--> #	144.0	90.0	72.0	54.0	36.0	36.0
TOT NON-CUM actl	---	---	---	---	---	---
TOTAL CUM actual	---	---	---	---	---	---
TOTAL CUM plan	1,512.0	1,602.0	1,730.0	1,840.0	1,946.0	2,080.0
CUM % over planned	---	---	---	---	---	---

press SPACE BAR to continue

HI RISE OFFICE (sample job) expenditure Updated to 3 Months cal
The following SCREENS show the cost-time bar chart of the current activity

Activity expenditure :- * timescale: Months cal *

	13	14	15	16	17	18
	>> Structure <<					
Actl/fcst:--> *	48.0	40.6	35.8	0.0	0.0	0.0
	-forecast-	-forecast-	-forecast-			
Planned:--> #	0.0	0.0	0.0	0.0	0.0	0.0
TOT NON-CUM actl	---	---	---	---	---	---
TOTAL CUM actual	---	---	---	---	---	---
TOTAL CUM plan	2,220.0	2,402.0	2,612.0	2,836.0	3,032.0	3,200.0
CUM % over planned	---	---	---	---	---	---

press SPACE BAR to continue

V I S I E R demonstration
Plan expenditure = 3,200.00 £ '000
duration = 18 Months cal

M O N I T O R
current updates
module c

SUMMARY of HI RISE OFFICE (sample job)

Original expenditure (ex-start) = 3200 (£ '000)
Current expenditure (ex-start) = 4114.286 (£ '000) exceeded by 29 %

Last update was at period 2 and had expenditure= 3400

---

press SPACE BAR to continue   category new
Job code a-smith

V I S I E R demonstration
Plan expenditure = 3,200.00 £ '000
duration = 18 Months cal

M O N I T O R
current updates
module c

Summary details

Activity (1) Structure

Original data:- expenditure budget= 2000 (£ '000) exceeded by 36 %

with timespan from 1 to 13 Months cal with duration 13 exceeded by 15 %

Current data :-
expenditure budget = 2714.286 (£ '000) with timespan from 1 to 15 Months cal
% completion cost based= 35 % time based= 20 %

---

press SPACE BAR to continue   category new
Job code a-smith

V I S I E R contact:

GRAPHS

J. Berny 01-979-9219

Use Esc to smooth monitored data
the smoothed curve gives an indication of the model's reliability
& allows tests of other parameters

Key F10 & then press space BAR to exit from graphics

Key F2 to print graphs and if required graph data

PRESS space bar to CONTINUE

V I S I E R demonstration

Plan expenditure= 3,200.00 £ '000 over 18 Mnth

JOB:-HI RISE OFFICE (sample job)

INSPECTION
control, inflation
module e

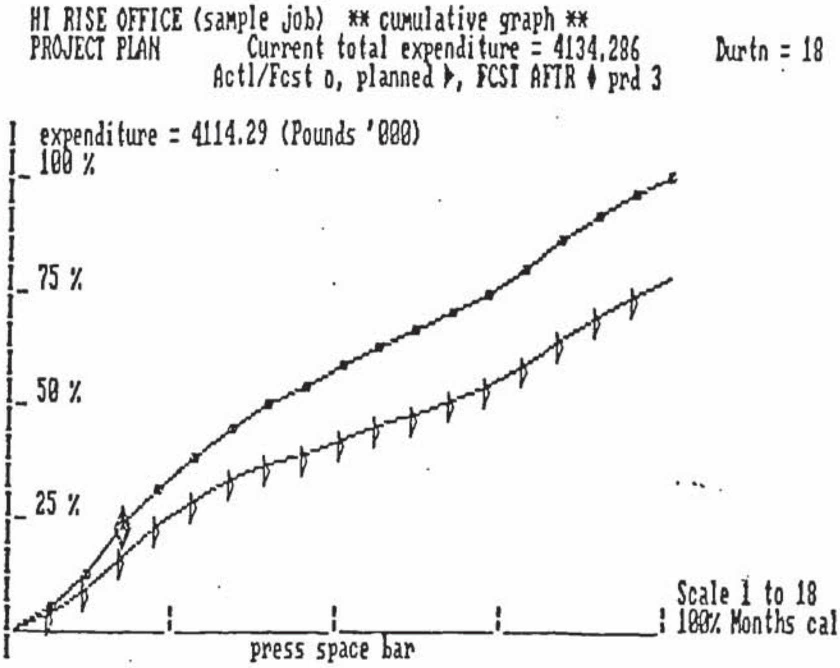
This section allows inspection of :-

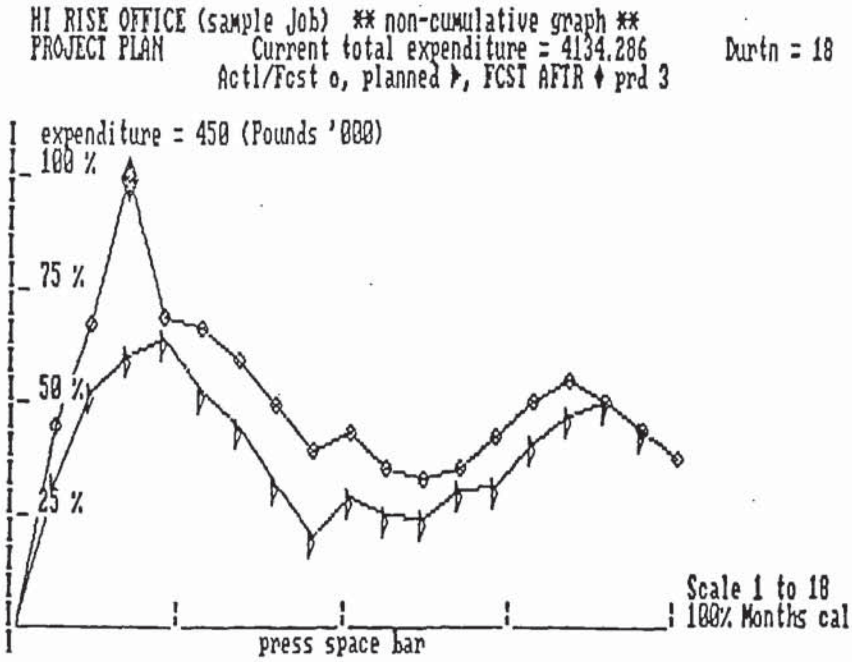
a Two cashflows with or without discounting (-------
b Single cashflow only
c Single cashflow with discounting

prefix 1 = a expenditure
prefix 2 = b income
prefix 3 = e rent returns
Select prefix (may include lag...) No.1
prefix is a expenditure

i.e. expenditure will be inspected

press SPACE BAR


HI RISE OFFICE (sample job) ** cumulative graph **
PROJECT PLAN Current total expenditure = 4134.286 Durtn = 18
Actl/Fcst o, planned ▸, FCST AFTR ♦ prd 3
expenditure = 4114.29 (Pounds '000)
100 %
75 %
50 %
25 %
Scale 1 to 18
100% Months cal
press space bar


HI RISE OFFICE (sample Job) ** non-cumulative graph **
PROJECT PLAN Current total expenditure = 4134.286 Durtn = 18
Actl/Fcst o, planned ▸, FCST AFTR ♦ prd 3
expenditure = 450 (Pounds '000)
100 %
75 %
50 %
25 %
Scale 1 to 18
100% Months cal
press space bar

V I S I E R demonstration

INSPECTION control, inflation module e

Plan expenditure= 3,200.00 £ '000 over 18 Mnth
JOB:-HI RISE OFFICE (sample job)

Having selected two cashflows we need to pick the required pair

Two cashflows with or without discounting MEASURING:-

a plan data vs. plan what if? expenditure
b current updates vs. current what if?
c plan data vs. current updates <------

Plan will be compared with Current update (1st. two periods only)

---

press SPACE BAR

V I S I E R demonstration

INSPECTION

Plan expenditure= 3,400.00 £ '000 over 18 Mnth
JOB:-HI RISE OFFICE (sample job)

Monthly cashflow		Whole plan cashflow plan data vs. current updates		
Mnth cal	Actl. Expndtr.	Plan	Difference	% diff.
1	200.0	144.0	56.0	38.90
2	300.0	234.0	66.0	28.20
expenditure forecasts				
3	261.6	270.0	-8.4	-3.10
4	268.4	288.0	-19.6	-6.80
5	243.7	234.0	9.7	4.10
6	201.4	198.0	3.4	1.70
7	155.0	144.0	11.0	7.60
8	113.7	90.0	23.7	26.40
9	137.9	128.0	9.9	7.70
10	115.8	110.0	5.8	5.30

---

press SPACE BAR

Job code a-smith

INSPECTION

V I S I E R demonstration

Plan expenditure= 3,400.00 £ '000 over 18 Mnth

JOB:-HI RISE OFFICE (sample job)

Monthly cashflow — Whole plan cashflow plan data vs. current updates

Mnth cal	FCST. Expndtr.	Plan	Difference	% diff.
11	115.7	106.0	9.7	9.10
12	135.0	134.0	1.0	0.80
13	171.8	140.0	31.8	22.70
14	182.0	182.0	0.0	0.00
15	210.0	210.0	0.0	0.00
16	224.0	224.0	0.0	0.00
17	196.0	196.0	0.0	0.00
18	168.0	168.0	0.0	0.00

press SPACE BAR

Job code a-smith

INSPECTION

V I S I E R demonstration

Plan expenditure= 3,400.00 £ '000 over 18 Mnth

JOB:-HI RISE OFFICE (sample job)

Cumulative cashflow — Whole plan cashflow plan data vs. current updates

Mnth cal	Actl. Expndtr.	Plan	Difference	% diff.
1	200.0	144.0	56.0	38.90
2	500.0	378.0	122.0	32.30
expenditure forecasts				
3	761.6	648.0	113.6	17.50
4	1,030.0	936.0	94.0	10.00
5	1,273.7	1,170.0	103.7	8.90
6	1,475.1	1,368.0	107.1	7.80
7	1,630.1	1,512.0	118.1	7.80
8	1,743.8	1,602.0	141.8	8.90
9	1,881.7	1,730.0	151.7	8.80
10	1,997.5	1,840.0	157.5	8.60

press SPACE BAR

Job code a-smith

V I S I E R demonstration

INSPECTION

Plan expenditure= .3,400.00 £ '000 over 18 Mnth
JOB:-HI RISE OFFICE (sample job)

Cumulative cashflow		Whole plan cashflow plan data vs. current updates		
Mnth cal	FCST. Expndtr.	Plan	Difference	% diff.
11	2,113.2	1,946.0	167.2	8.60
12	2,248 2	2,080.0	168.2	8.10
13	2,420.0	2,220.0	200.0	9.00
14	2,602.0	2,402.0	200.0	8.30
15	2,812.0	2,612.0	200.0	7.70
16	3,036.0	2,836.0	200.0	7.10
17	3,232.0	3,032.0	200.0	6.60
18	3,400.0	3,200.0	200.0	6.30

press SPACE BAR

Job code a-smith

V I S I E R demonstration

INSPECTION

Plan expenditure= 3,400.00 £ '000 over 18 Mnth
JOB:-HI RISE OFFICE (sample job)

Cumulative cashflow		Whole plan cashflow plan data vs. current updates		
Mnth cal	FCST. Expndtr.	Plan	Difference	% diff.
11	2,113.2	1,946.0	167.2	8.60
12	2,248.2	2,080.0	168.2	8.10
13	2,420.0	2,220.0	200.0	9.00
14	2,602.0	2,402.0	200.0	8.30
15	2,812.0	2,612.0	200.0	7.70
16	3,036.0	2,836.0	200.0	7.10
17	3,232.0	3,032.0	200.0	6.60
18	3,400.0	3,200.0	200.0	6.30

The final part demonstrates V I S I E R 's extra ability in Project Management Integration by considering the JOB RISK aspects

press SPACE BAR

Job code a-smith

V I S I E R demonstration

F O R E C A S T E R module f

JOB TITLE :-HI RISE OFFICE (sample job)
Client :- VISIER

a plan
b plan with 'what if ?'
c update
d update with 'what if ?'

Risk assessment requires 1 good curve fit, it usually takes c.15 minutes replacing good simulations taking many hours for the same task.

Additionally an historic data base of curve fit results is maintained

The other types of projection are given on the next SCREEN

PLAN data (a) was selected
Press SPACE BAR to continue
Job code a-smith

F1 & ↵ menu

V I S I E R demonstration

F O R E C A S T E R module f

OPTIONS
=======

1 Past data (3 parameter) curve fit analysis. For historical records

2 Best curve fit and projection from last known value

3 WHAT IF ? Projection & reliability (by changing JOB duration & 'cost')

4 Fit/forecast (last 6 periods of given data are predicted & tested)

Press SPACE BAR to continue
Job code a-smith

F1 & then ↵ menu

V I S I E R demonstration

R I S K ASSESSOR module g

TITLE HI Rise Block

Cost= 3000 £'000 start cost= 20 £'000 duration= 18 Mnths cal

Curve fit error (cumulative variation) = 3.8 %
approximate periodic variation= 5.32 %

The following calculations are based on a model with parameters :-

g= 3.4 , m= 1.6 , p= 0 , q= 0 , n= 5

Start cost (20 £'000) will be excluded

---

category new
press SPACE BAR Job code a-smith
F1 & then ↵ Menu

V I S I E R demonstration

R I S K ASSESSOR module g	Project expenditure= 3,000.00 £ '000 duration= 18 Mnths cal

Uses of risk analysis currently available in VISIER
Each complete risk evaluation consists of the following elements :-

1. Statistical evaluation of viable cost variations
   These are likely cost variations within which a project or activity
   may be relied upon, to remain inside the limits given in 2 & 3 below

2. Generation of a cost envelope
   This sets the likely upper and lower cost of the project or activity

3. Cost and time analysis
   Determines both the expected cost and time, and their deviations
   A histogram shows the likelihood of project or activity completion

---

category new
press SPACE BAR Job code a-smith
F1 & ↵ Menu

V I S I E R demonstration

R I S K ASSESSOR module g	Project expenditure= 3,000.00 £ '000 duration= 18 Mnths cal

1. Statistical evaluation of viable cost variations

Average cumulative 'cost' variation calculations show that the model is statistically good upto 9.6% usable to 23.8% then UNRELIABLE
NOTE: Risks increase as the variation or curve fit error increases.

INPUT mean cumulative 'cost' variation % ? 8 %
Periodic variation is approximately = 11 %

The confidence curve limits are based on the above, but also require the chance to exceed the upper costs limit to be set.

---

category new

press SPACE BAR Job code a-smith

F1 & ↵ Menu

V I S I E R demonstration

R I S K ASSESSOR module g	Project expenditure= 3,000.00 £ '000 duration= 18 Mnths cal

2. Generation of the cost envelope

The confidence curve statistical limits set the cost envelope shape which gives the likely upper and lower costs at any given time

The lower confidence curve limit has been preset to give a 1 in 10 chance for the costs to be lower than the quoted costs at any given time.

Chance to exceed upper cost value :- a (1:7) b (1:10) c (1:20) d (1:100)
INPUT a letter d was selected

---

category new

press SPACE BAR Job code a-smith

F1 & ↵ Menu

```
             *****   R I S K    A N A L Y S I S  *****
          CONFIDENCE BAND GRAPH HI Rise Block  (new\a-smith) 01-01-1980
          Upper limit (=) (1:100) lower (-) nominal (1:10) expected (*)
 expenditure
4687     :
         :                                  =  =  =  =  =  =
         :                        =  = =  =
 75% --  :                     =                           *  *
         :               =  =  .     *  *  *  *  * *  *
3000     :::::::::::=:::::::*::*::::-::-::-::-:-::-::-::-::
___ Expected expenditure = 3000
         :     =        *  -  -
         := =         * -                         Most likely duration = 18
         :     *  *  -                         Parameters:-
 25% --  :  *       -                     m= 1.60 g= 3.40 p= 0.00 q= 0.00 n= 5.00
         :* -  -
         :-
984  ____:_________:_________:_________:_________:_________:  Time --->
         : 9      20 %      40 %      60 %      80 %      100 %  27
Model statistcly. Sound, cost variation= 8 % ( limits 10 % & 24 % )
        Upper(1:100) lower (1:10) confidence limits span from 15  to 31
        span of expenditure from 2571.613  to 3340.556
                 press SPACE  BAR
```

V I S I E R demonstration

R I S K ASSESSOR module g	Project expenditure= 3,000.00 £ '000 duration= 18 Mnths cal

3. Cost and time analysis

The TIME- PROBABILITY histogram bars display the % probability of completion at a given time (over 2 periods).

In addition listings are given of the expected cost and the % probability of completion by a given period and that period or later.

The chance to exceed the most likely duration is normally 50 %

hence only a practical range of probabilities of 40 % to 80 % is allowed.

---

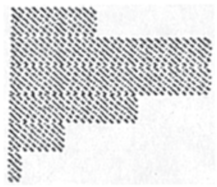
category new

press SPACE BAR — Job code a-smith

F1 & ↵ Menu

TITLE HI Rise Block (new\a-smith) 01-01-198 Most likely drn. 18
'Cost'= 3000 LO (1:10) 2672.1 HI (1:100) 3623.8 vartn./prd 11.2 %

Case A to exceed max time 30 prob = 0 % min time 15
Case B to exceed most likely time 18 with MEDIUM probability = 60 %
Expected time=19.5 & 'worth'= 3,086.5 high value= 3,710.3



prd.	CumX prob	prob % /prd	GRAPH of prob % /period	Expected cost
15	100	12		2571.613
17	88	28		2900.568
19	60	28		3065.83
21	32	19		3144.87
23	13	9		3199.602
25	4	3		3244.722
27	1	1		3282.483

press SPACE BAR

TITLE HI Rise Block (new\a-smith) 01-01-198 Most likely drn. 18
'Cost'= 3000 LO (1:10) 2672.1 HI (1:100) 3623.8 vartn./prd 11.2 %

Case A to exceed max time 30 prob = 17 % min time 15
Case B to exceed most likely time 18 with HIGH probability = 75 %
Expected time=23.1 & 'worth'= 3,201.8 high value= 3,825.6

prd.	Cum% prob	prob % /prd	GRAPH of prob % /period	Expected cost
15	100	10		2571.613
17	90	15		2900.568
19	75	14		3065.83
21	61	13		3144.87
23	48	12		3199.602
25	36	9		3244.722
27	27	7		3282.483
29	20	6		3314.094
31	14	4		3340.556
33	10	3		3362.709
35	7	2		3381.253

Old Probl to exceed 30 = 0 %, 18 = 60 % Exp time 19.5 wrth 3086.50

Press space bar

A P P E N D I X 5

VISIER FLOW CHARTS

The first flow chart in this section depicts a broad brush route through the software, covering the introdoctory training in its use. The next five give a more detailed view of the pathways through the software, which an experienced user would be likely to take. Each of the flow charts F1 to F5 are concerned with a specific module.

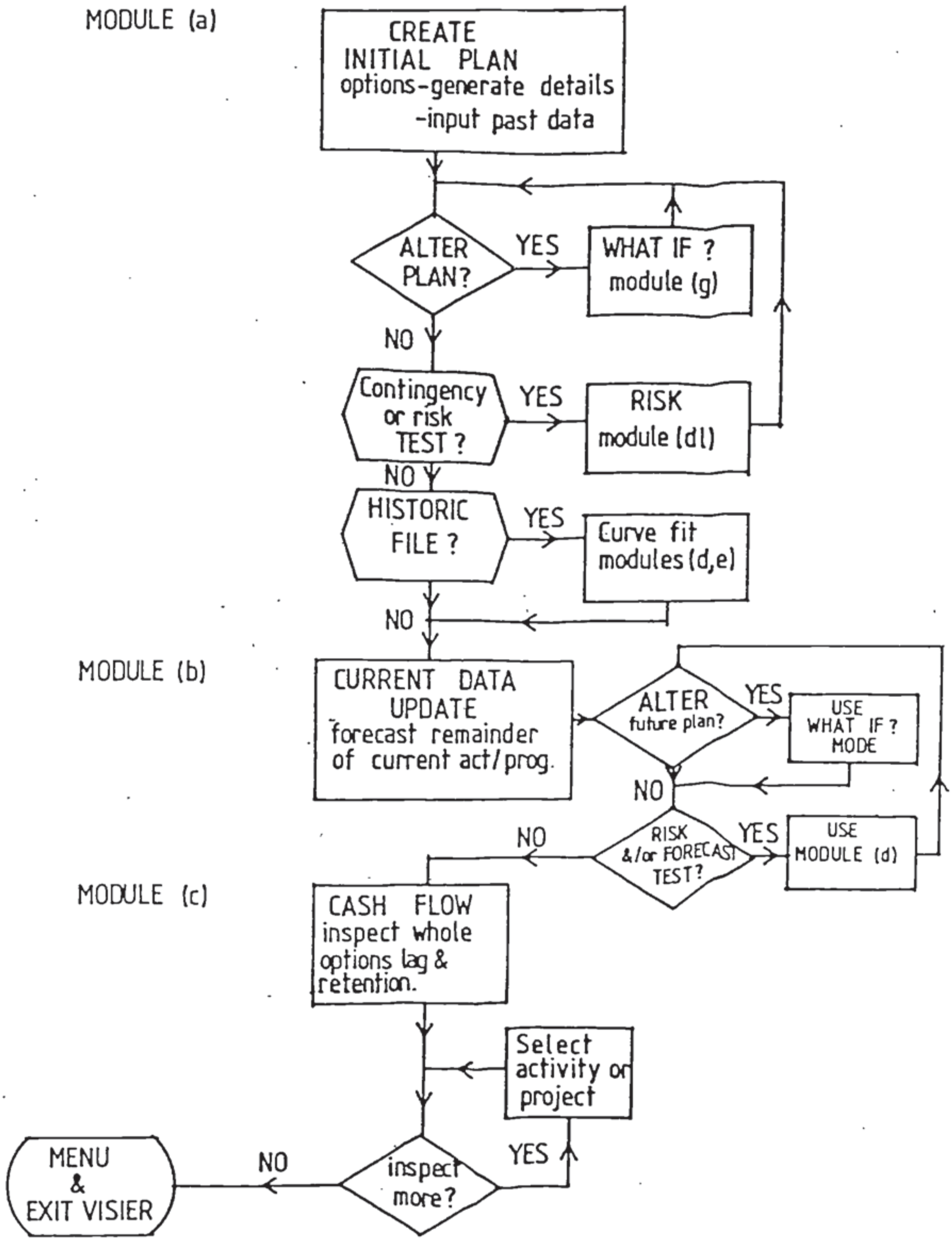
List of flow charts :-

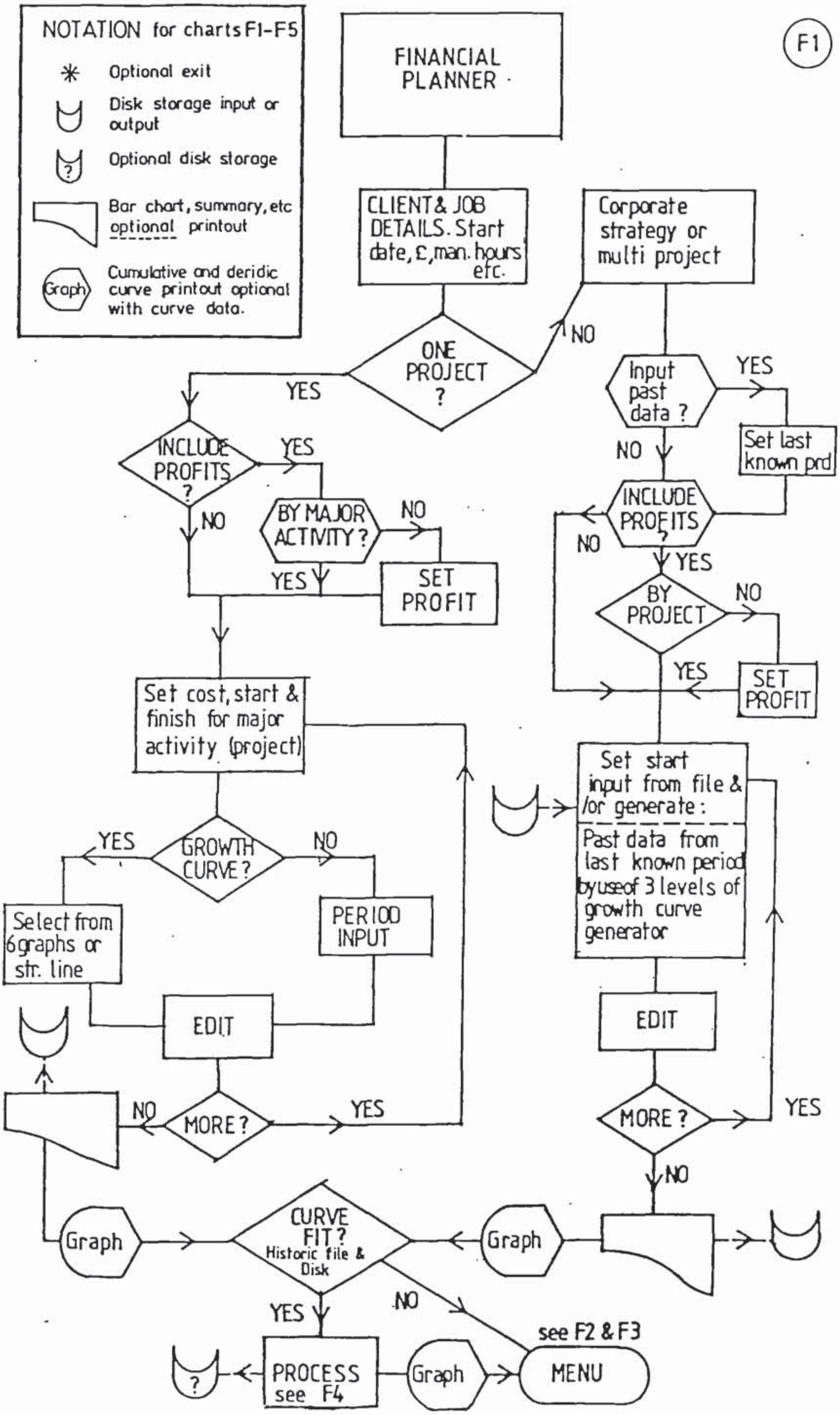
	Overview
F1	Planner
F2	What if planner
F3	Work in progress and its what if
F4	Forecaster
F5	Risk analyser

OUTLINE ROUTE THROUGH VISIER

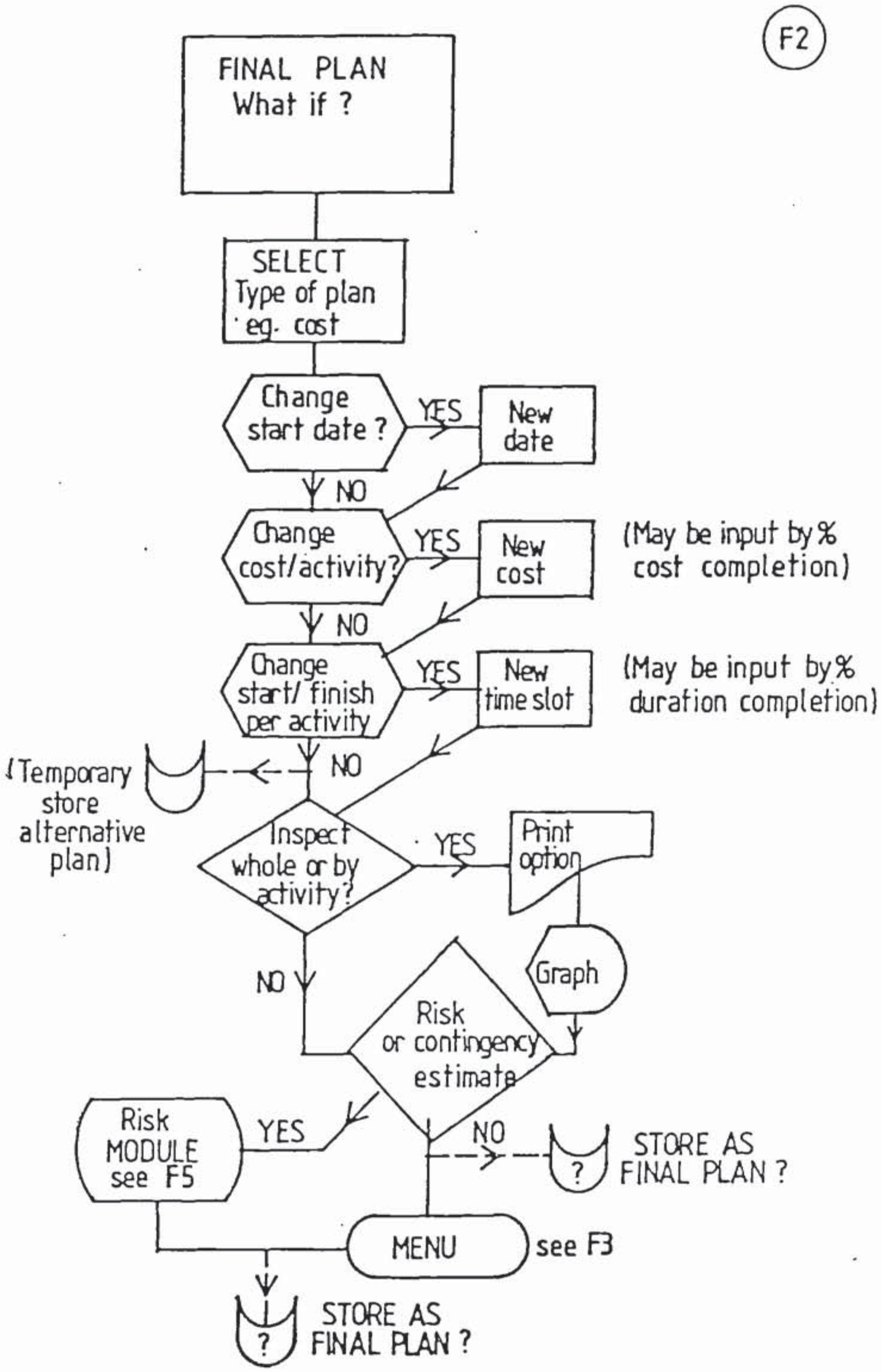
For projects using major activities or projects for for corporate strategy

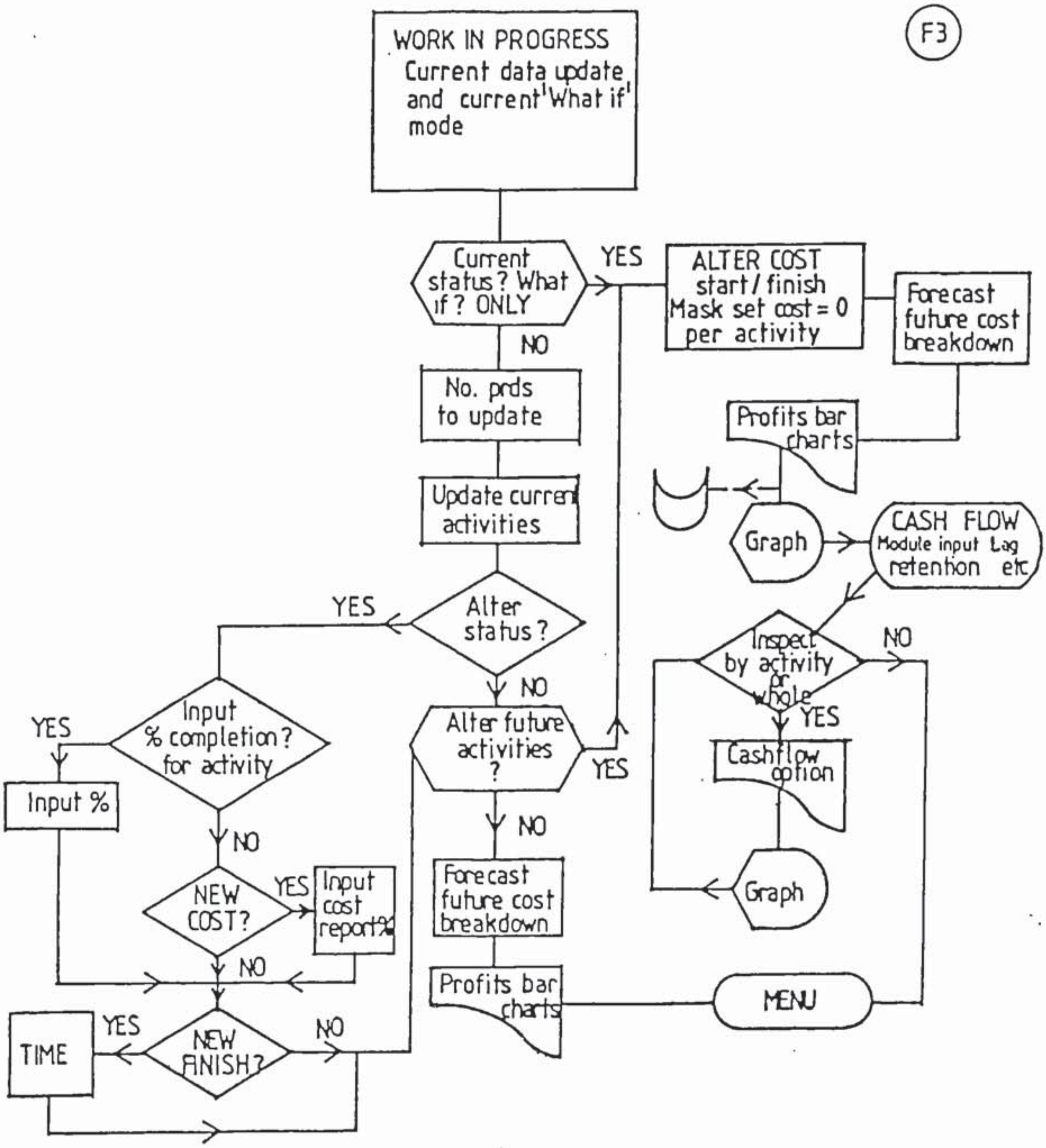
(FINANCE; RESOURCES, MARKETING)



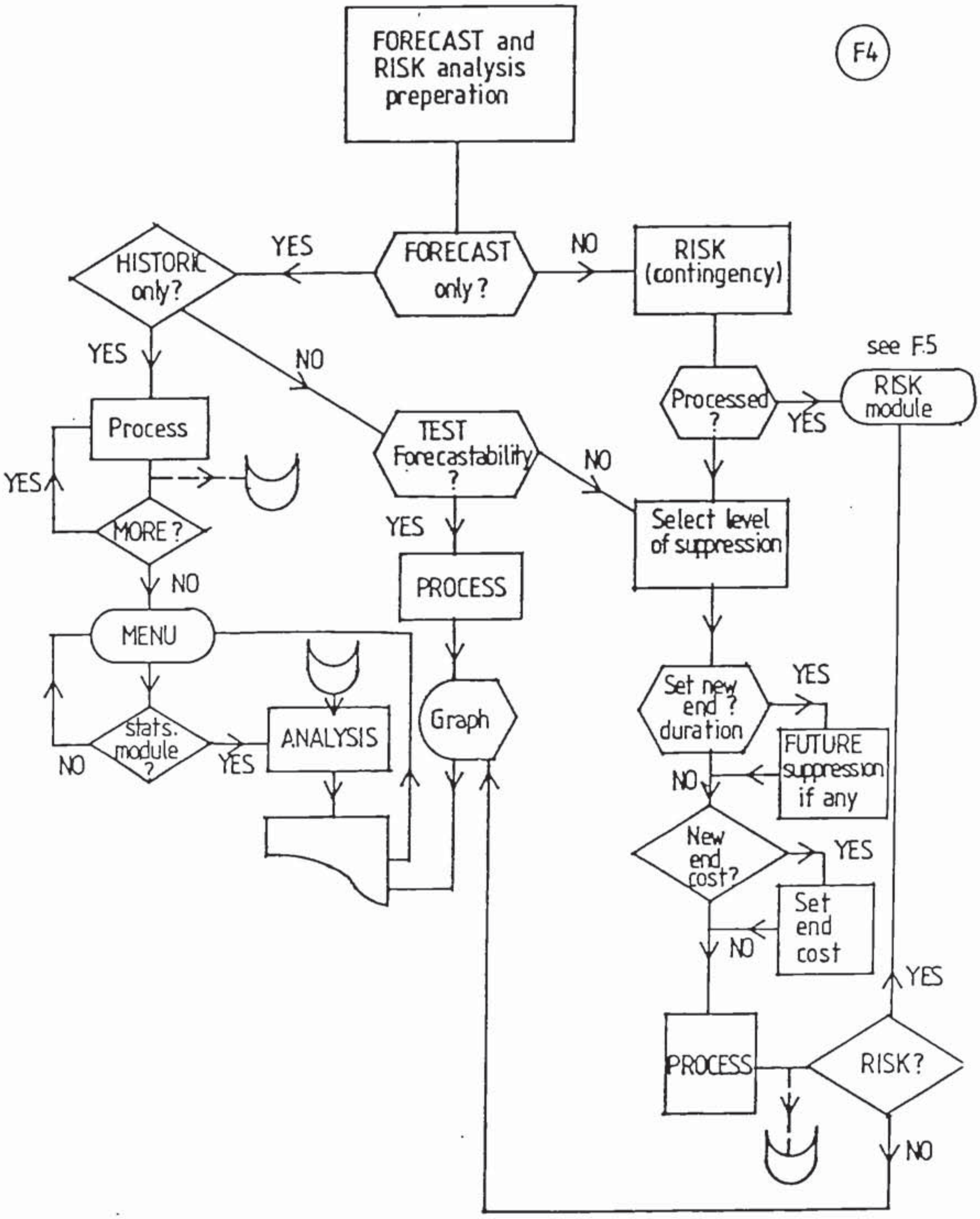


NOTATION for charts F1-F5
Optional exit
Disk storage input or output
Optional disk storage
Bar chart, summary, etc optional printout
Cumulative and deridic curve printout optional with curve data.
Graph
F1
FINANCIAL PLANNER
CLIENT & JOB DETAILS. Start date, £, man. hours etc.
ONE PROJECT ?
YES
NO
Corporate strategy or multi project
Input past data ?
Set last known prd
INCLUDE PROFITS ?
BY MAJOR ACTIVITY ?
SET PROFIT
BY PROJECT
Set cost, start & finish for major activity (project)
GROWTH CURVE ?
Select from 6 graphs or str. line
PERIOD INPUT
EDIT
MORE ?
Set start input from file & /or generate :
Past data from last known period by use of 3 levels of growth curve generator
CURVE FIT ? Historic file & Disk
PROCESS see F4
MENU
see F2 & F3

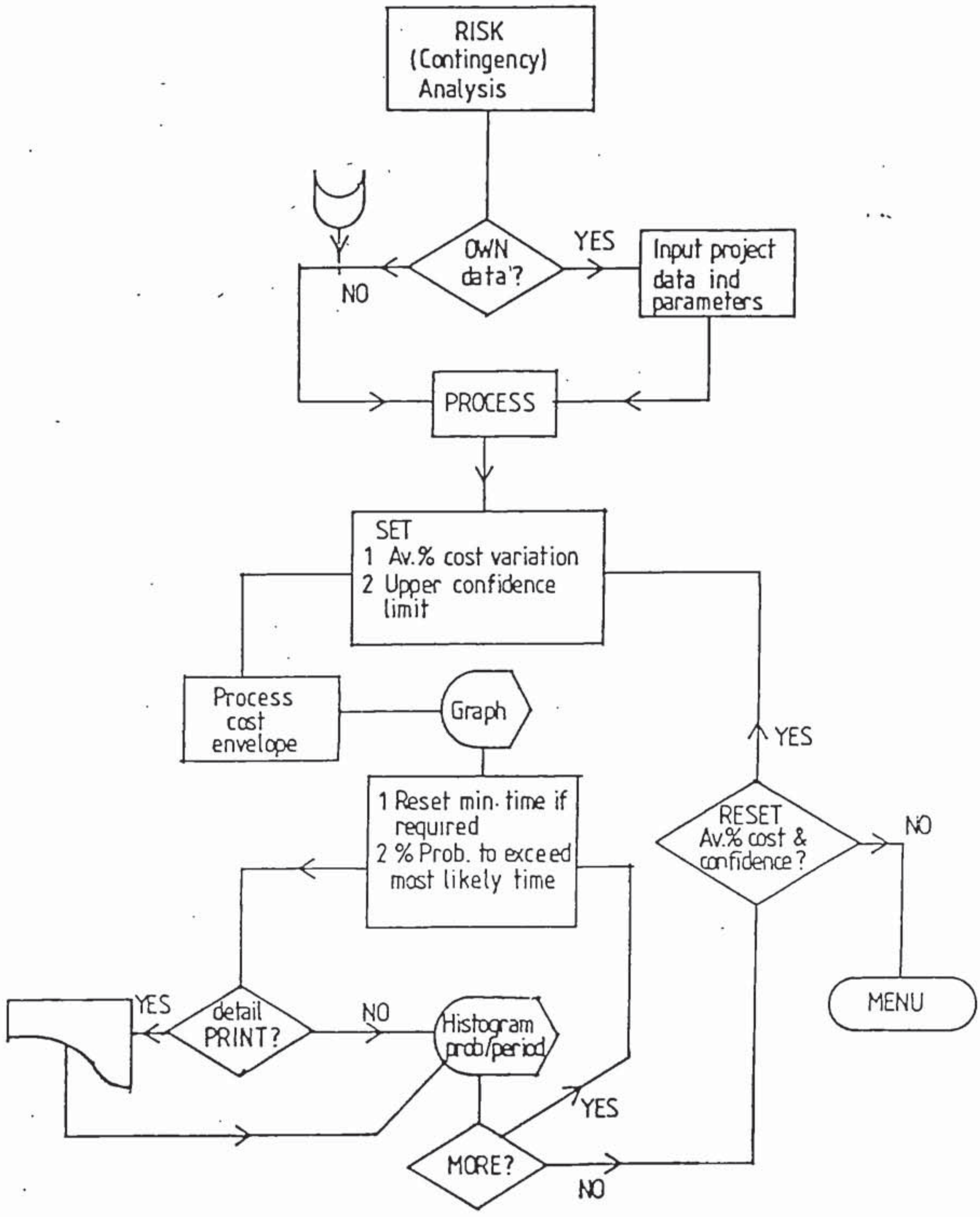

F2
FINAL PLAN
What if ?
SELECT
Type of plan
eg. cost
Change
start date ?
YES
New
date
NO
Change
cost/activity?
YES
New
cost
(May be input by %
cost completion)
NO
Change
start/ finish
per activity
YES
New
time slot
(May be input by %
duration completion)
(Temporary
store
alternative
plan)
NO
Inspect
whole or by
activity?
YES
Print
option
Graph
NO
Risk
or contingency
estimate
Risk
MODULE
see F5
YES
NO
STORE AS
FINAL PLAN ?
MENU
see F3
STORE AS
FINAL PLAN ?



F3
WORK IN PROGRESS
Current data update and current 'What if' mode
Current status? What if? ONLY
YES
NO
ALTER COST start / finish Mask set cost = 0 per activity
Forecast future cost breakdown
Profits bar charts
Graph
CASH FLOW Module input Lag retention etc
No. prds to update
Update current activities
Alter status?
YES
NO
Inspect by activity or whole
NO
YES
Cashflow option
Graph
Input % completion? for activity
YES
Input %
NO
Alter future activities ?
YES
NO
Forecast future cost breakdown
Profits bar charts
MENU
NEW COST?
YES
Input cost report%
NO
NEW FINISH?
YES
NO
TIME


FORECAST and
RISK analysis
preperation
F4
FORECAST
only ?
YES
NO
HISTORIC
only?
RISK
(contingency)
YES
NO
see F.5
Processed
?
RISK
module
YES
Process
TEST
Forecastability
?
NO
YES
MORE ?
YES
Select level
of suppression
NO
PROCESS
MENU
Set new
end ?
duration
YES
stats.
module
?
ANALYSIS
Graph
FUTURE
suppression
if any
NO
YES
NO
New
end
cost?
YES
Set
end
cost
NO
YES
PROCESS
RISK?
NO

F5


RISK
(Contingency)
Analysis
OWN data'?
YES
NO
Input project data ind parameters
PROCESS
SET
1 Av.% cost variation
2 Upper confidence limit
Process cost envelope
Graph
1 Reset min. time if required
2 % Prob. to exceed most likely time
YES
RESET Av.% cost & confidence?
NO
MENU
YES
detail PRINT?
NO
Histogram prob/period
YES
MORE?
NO

APPENDIX 6

VISIER FIELD TRIAL QUESTIONNAIRE

The questionnaire was designed specifically for managers and other potential users of the software. They contributed most constructively and in great depth. The insight, generated by their feedback, enabled the author to ensure the software development would be acceptable in both industry and commerce.

The response to the questionnaire was only the last stage in the feedback loop. For this reason it was not designed on more usual marketing survey principles; but as a means of confirming, some of the most pertinent issues resulting from this extensive field exercise. All persons involved replied fully. Their answers were very instructive and are summed up in chapter 8.

VISIER

USER FEED BACK

RATING

Scale - referred to as RATING

1 = least 'use' 5 = maximum 'usefulness'

SECTION A

QUANTITIVE

1. Specific.

Please give RATING, critisism and/or comment on each module.

(a)	Planner	[RATE]
(b)	What if? Planner	[RATE]
(c)	Current update	[RATE]
(d)	Update what if?	[RATE]
(e)	Inspection	[RATE]
(f)	Forecaster	[RATE]
(g)	Risk assessor	[RATE]
(h)	Historical feedback and stats	[RATE]
(i)	Help	[RATE]
(j)	Editor and Project Shells	[RATE]
(k)	Job operations	[RATE]

EXTRA comments on specific modules :- if none say NIL

2. When will VISIER be used, for what purpose and by whom (type and level of personnel).

E.g. Before plan ... mainly what if?
During work ...
After completion ...
Corporate strategy etc.

3. General impressions:

PROS CONS

Aspect	Rating	Comment
User friendliness		
Ease of use		
Flexibility or adaptability to:		
Job Type		
Job Size		
Top management need		
Middle management need		
'Line' management need		
Use in corporate strategy		
Use in Marketing		
Use in Project		
Other:- SPECIFY		

4. Envisaged learning time and depth with respect to 'whom'.

Management Levels	Top	Middle	Line	Other:- SPECIFY
Days (no) (training)				
Depth rating 5 = in depth by level of management				
Interval Between days of training				
Total Assimilation (months)				
Rating				

SECTION B

QUALITATIVE

1. How could VISIER be envisaged as fitting into your Company/ Organisation.

2. Which software (known to you as relevant) does VISIER:-

   (1) Interface with

   (a) VISIER use before the software

   (b) VISIER while using other software

   (c) VISIER use after software

   (ii) Supercede

   (iii) Stands alone from the software.

3. Give comparative 'valuation' with software, known to you, relating to part or whole of VISIER.

   Please add 'description of part' and RATING in both 'parts' and VISIER.

4. (a) Current version of VISIER 'suggestions box':-

   (i) Extra features

   (ii) Surplus features.

   (b) FUTURE version:-

   (i) How many months before emergence of new version after launch say end of October 1987.

   (ii) Extra features

   (iii) Upgraded features.

5. Potential market:-

   (a) in your sphere

   (b) views on horizontal market

   i.e. other sphere of work.

6. Potential value of VISIER, assume your area of 'work'.

   i.e. ....

Management Levels	Top	Middle	Line
% saving of man-year			
% saving man-cost/year			

Gross Profit Range	Less than £5 mill.	5 to 30 mill.	30 mill. and over
% profit gain by company p.a.			
% overall cost saving by company p.a.			

7. Please give a one sentence (if possible) description of VISIER as you see it.

8. Any other comments.

SECTION C

PROGNOSIS OF QUESTIONS

PLEASE STATE any question in which you felt:-

(a) should not have been asked

(b) is too 'tight'

(c) should not have been asked for a RATING

(d) Glaring omissions

REFERENCES

Arditi D., Guzin T.A., and Gurdamar S., 'Cost overruns in public projects', International Journal of Project Management, 3 (1985), 218 - 224.

Bagot, P.A., 'The growth of colour television ownership in the UK - a sales forecasting study', M.Sc. report, Imperial College, London (1979).

Baker, R.W., 'Handling uncertainty', International Journal of Project Management, 4 (1986), 205 - 210.

Battersby, A., 'Network analysis for planning and scheduling', MacMillan, 3rd Edition, Appendix 1, (1970).

Bennett, J., 'Simulating the management task on construction projects;, Association of Researchers in Construction Management, Conference: Towards the Nineteen Nineties, (1986).

Berny, J., 'Half Life measurement of Ir 194', Physics Degree Dissertation, Birbeck College, University of London (1960).

Berny, J., Jarmain, M., Thomson, K., Thorley, G., 'York county exhibition centre', Portfolio used by Knight, Frank and Rutley, (1978).

Berny, J., 'S-curves and their application', Special lecture to the Construction Study Unit, Brunel University, (1984).

Berny, J., 'New concepts in deterministic growth curve forecasting', submitted to Journal of Forecasting (1985).

....

Recommended 1987 to submit to Journal of Applied Statistics: accepted 1988 with anticipated publication year 1989.

Berny, J., 'Uncertainty to advantage', presentation of draft paper on VISIER: for CIB W-65 conference London 1987, OR newsletter, 17 no.4 (1987a), 9.

Berny, J., 'A new distribution for risk analysis', under revision, submitted to Journal of the Operational Research Society, (1987b).

Berny, J., 'Forecasting and risk analysis for management: a summary of theory', Doctoral Working Paper 121, University of Aston, (1987c).

Berny, J., 'Summary and critique of S-L. Tan's M.Sc. dissertation: project managers and computers', Report for Knowledge Based Systems Group of the Association of Project Managers, April (1988a).

Berny, J., 'Project planning expert system style', to be published by INTERNET Conference, Glasgow, (1988b).

Berny, J., and Howes, R., 'Project management and control using real time budgeting and forecasting models', Chartered Institute of Building: Construction papers, 2 (1983), 19 - 40.

Berny, J., and Howes, R., 'Project management and control using growth curve models applied to budgeting and forecasting within the Construction Industry', SERC Report No. GR/C/42187, (1985).

Box, G.E.P., and Jenkins, G.M., 'Time series analysis: forecasting and Control', Holden-Day (1976), 226 - 228.

Bromilow, F.J., 'Measurement and scheduling of construction time and cost performance in the building industry', Melbourne, Australia CSIRO, Division of Building Research, (1973).

Cooper, D.F., MacDonald, D.H., and Chapman, C.B., 'Risk analysis of a construction cost estimate', International Journal of Project Management, 3 (1985), 141 - 149.

Davies, H., 'Managing with computers', Paper presented on behalf of Laing Manageement Contracting at the Construction Computing Conference, Barbican, London, (1986).
<H.Davies is currently with BURO4>.

Dugdale, I., 'Corporate planning and control systems in William and Glyn's Bank', Long Range Planning, 11 (1978), 41 - 46.

Eilon, S., 'Aspects of management', Pergamon International Press, Chapter 23, (1979).

European Space Agency, Data retrieval system (search 1985).

Farnum, N.R., and Stanton, L.W., 'Some results conerning the estimation of beta distribution parameters in PERT', Journal of the Operational Research Society, 38 (1987), 287 - 290.

Fildes, R., Jailand, M., and Wood, D., 'Forecasting in conditions of uncertainty', Long Rang Planning, 11 (1978), 29 - 38.

Fine, B., 'Effect of project cost centre breakdown on forecasting budgets', Publications by Fine, Curtis and Gross (1972).

Goldfeld, S.M., and Quandt, R.E., 'Nonlinear methods in Econometrics', North-Holland (1972), 49 - 62.

Gregg, J.V., Hassell, C.H. and Richardson, J.T., 'Mathematical Trend Curves: an aid to forecasting', ICI Monograph No. 1, Oliver and Boyd (1964).

Harrison, P.J., and Pearce, S.F., 'The use of trend curves: an aid to forecasting', Industrial Marketing Management, 2 (1972), 149 - 170.

Hudson, K.W,. 'DHSS expenditure forecasting method', Chartered Quantity Surveyer, 5 (1978), 42 - 46.

Hutchesson, B.N.P., 'Market research and forecasting for the chemical industry: the state of the art', IMRA Journal, 3 (1967), 242 - 260.

Lewis, C.D., 'Industrial and Business Forecasting Methods: a practical guide to exponential smoothing and curve fitting', Butterworth (1982), 112 - 115.

Lewis, C.D., 'Forecaster - an exercise in user friendliness', Journal of the Operational Research Society, 36 (1985), 621 - 623.

Lewis, C.D., 'Management with microcomputers', Basil Blackwell, 3rd Edition, (1987).

(MACC):- Maddison Academic Computing Centre, UWHAUS program, University of Wisconsin, 1210, West Drayton Street, Madison, Wisconsin, 53706, (1984).

Mar-Molinero, C., 'Tractors in Spain: a logistic analysis', Journal of the Operational Research Society, 31 (1980), 141 - 152.

Meade, N., 'The use of growth curves in forecasting market development: a review and appraisal', Journal of Forecasting, 3 (1984), 429 - 451.

Meade, N., 'Forecasting using growth curves: an adaptve approach', Journal of the Operational Research Society, 36 (1985), 1103 - 1115.

Miller, R.W., 'How to plan and control with PERT', Harvard Business Review, March - April, (1962).

Neuburger, K., 'Accessing the risks in acquisitions - risk chance analysis', Long Rang Planning, 11 (1986), 41 - 88.

Norden, P.V., 'On the anatomy of development projects', IRE Transactions on Engineering Management, EM7(1) (1960), 34 - 42.

Oliver, F.R., 'Tractors in Spain: a further logistic analysis', Journal of the Operational Research Society, 32 (1981), 499 - 502.

Oliver, M.O., 'A Bayesian model to predict saturation and logistic growth', Journal of the Operational Research Society, 38 (1987), 49 - 56.

Perry, J.G., 'Risk management: an approach for project managers', International Journal of Project Management, 4 (1986), 211 - 216.

Pilcher, R., 'Appraisal and control of projects', McGrawhill, (1973).

Project Start-up, Proceedings of the 9th International Expert Seminar: INTERNET, (1984).

(PSA): Public Service Ageency, 'Planned progress monitoring', Internal document of PSA, (1986).

Pugh, L.A. and Seldon, R.G., 'Use of risk analysis techniques in assessing the confidence of project cost estimates and schedules', International Journal of Project Management, 4 (1986), 158 - 162.

Putman, L., 'Software cost estimating', DPMA Conferences: 5959 W Century Blvd, Suite 1016, PO Box 91295, Los Angeles, California 90009, USA, (1982).

Putnam, L.H., and Fitzsimmons, A., 'Estimating software costs', Datamation, 25 (1979), 171 - 178.

Raiffa, H., and Schlaifer, R., 'Applied statistical decision theory', Division of Research, Harvard Business School, (1961), 216 - 220.

Rektorys, K., 'Survey of Applicable Mathematics', Illiffe Books (1969), 200 - 204.

Sparkes, J.R., and McHugh, A.K., 'Awareness and use of forecasting techniques in British industry', Journal of the Operational Research Society, 3 (1984), 37 - 42.

Spanish Embassy (records office), 'Fuenie: Direccion General de la Produccion Agraria', Medios de Produccion: tractores, cosechadoras y motores de riego (1984), 595.

Stone, R., 'Sigmoids', Bulletin in Applied Statistics, 7 (1980), 59 - 119.

Tan Sor-Lan, 'Project Management and Computers', M.Sc. dissertation is Property Development, South Bank Polytechnic (1987).

Trimble, G., 'Financial control: the forward look', Construction News, (1974).

Tucker, S.N., 'Formulating construction cash flow curves using a reliability theory analogy', Construction Management and Economics, 4 (1986), 179 - 188.

Williams, E.J., 'Regression analysis', Wiley (1959), 98 - 106.

Woodward, J.F., 'Construction management and design', Macmillan, (1975), 31 - 43 and 149 - 180.

Yates, A., 'Risk analysis - an example of its application in the forecasting of construction cost', Quantity Surveyors research and development forum, 13th March, (1986).