A COMPUTER BASED DECISION AID FOR FORECASTING MONTHLY TIME SERIES

ROBERT HUGH EDMUNDSON

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

1987

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or any other institute of higher learning, except where due acknowledgement is made in the text of the thesis.

R.H. Edmundson

TO

Mary, Helen, and Peter.

ACKNOWLEDGEMENTS

Without the encouragement and support afforded me by many people this work would have often been in jeopardy. My family and friends have shown unflagging interest, and this has helped me along my way.

First among those who deserve special thanks stand my wife, Mary, and children, Helen and Peter. They have forgiven me my periodic remoteness, when they said I was in "Ph.D mode" with all interrupts disabled. Mary has also contributed greatly with the grind of proof reading.

I am grateful to all my colleagues who undertook extra burdens in order to lighten my load. Certain of those colleagues must be mentioned for their special help. My supervisor, Cyril Brookes, has done a remarkable job of guidance, and I thank him. Michael Lawrence gave unstintingly of his special knowledge and interest, and Marcus O'Connor was ever available to bring encouragement and a critical mind to my aid.

Mary Ackerman contributed her technical skills, and this alone would have earned my gratitude. However, her day to day support and concern for the project warrant special recognition. I often reflect on my good fortune to work with such a person, who is capable of making sense of my tortuous logic and willing to do so.

ABSTRACT

An interactive graphical decision aid for forecasting monthly time series of economic data, called GRAFFECT, was designed, implemented, and tested. The testing was carried out using 68 time series of real data. The series were forecast by two groups of subjects, the first comprising three experienced time series analysts and the second, 35 postgraduate students inexperienced in time series analysis.

The testing of GRAFFECT revealed, for the first time, judgmental extrapolations to be significantly more accurate than deseasonalised single exponential smoothing for a six month forecast horizon. The error rate was lower than any previously reported for the time series by a single forecasting method. This showed a significant improvement in judgmental forecasting accuracy for both novice and experienced forecasters.

The design of GRAFFECT was based, in part, on a study of the characteristics of time series and three methods of extrapolation. That study showed that judges appeared to have advantages over deseasonalised single exponential smoothing for series with subtle seasonal, and high noise characteristics.

Following the accuracy studies, the sub-tasks in the extrapolation were examined. Seasonal pattern

identification was examined in an ANOVA study conducted on nine time series, forecast using GRAFFECT and a hard copy plot of the series by ten postgraduate subjects. The use of GRAFFECT significantly reduced the difference in seasonal pattern from that determined using a ratio to centred moving average method. This provided a reduction in the advantage that deseasonalised single exponential smoothing had been observed to have in high seasonality series with low noise characteristics. However, that gain was accompanied by the loss of advantage for low seasonality series with high noise.

The data from the accuracy studies was examined to determine whether automatic processes for seasonal identification, trend identification, or extrapolation from the noise residual would improve the forecast. In no case did the automatic process give rise to an improvement. The analysis of the trend identification function showed that for series in which the judges identified trend there was a decided advantage to be gained over the use of deseasonalised single exponential smoothing.

CONTENTS

1. INTRODUCTION

1.1 GENERAL OBJECTIVES

The purpose of this thesis is to report on the development and evaluation of a computer based, interactive, graphical aid for forecasters making judgmental extrapolations of monthly economic time series (GRAFFECT). The phases of the work carried out were:

a) the characteristics of judgmental extrapolation relative to statistical methods were examined to determine:

   i) Situations in which statistical techniques should replace judgmental extrapolation, and

   ii) the opportunities to improve the judgmental process.

b) GRAFFECT was designed and constructed on the basis of the results from ii), and the implications of the human information processing literature.

c) The forecasting aid was then tested and evaluated to determine:

   i) the accuracy of forecast achieved in its use, relative to statistical forecasting methods and to unaided judgmental extrapolation,

   ii) how well the subjects performed the sub-tasks in the extrapolation. This was intended to throw light on certain human information processing issues such as the ability to identify seasonality signals in a noisy time series.

1.2 JUSTIFICATION FOR THE PROJECT

There are two important aspects in the justification of the work carried out. The first refers to the need for forecasting. The second addresses the way in which the forecast is to be made, either judgmentally or using a statistical algorithm.

The role of forecasting in business and government administration is well established. For instance, it has been demonstrated that forecasting sales turnover is a critical business function (Dalrymple, 1987), and that forecasting provides the basis for budget development (Welsch, 1971). The importance of forecasting as an economic activity has been reflected in the volume of academic literature concerned with the methods and accuracy of forecasting. For example, Armstrong (1985) provides a review of the major issues in forecasting and has a bibliography section extending to more than 130 pages, containing over one thousand references.

Although practitioners and academics agree on the importance of forecasting as an economic activity, there is an implied conflict between the two groups concerning the role of judgment in generating forecasts. As will be shown below, a majority of commercial practitioners use subjective forecasting procedures (Dalrymple 1987), whereas academics have concentrated heavily on "objective" methods. Most of the academic literature referred to in the previous paragraph concerns non-subjective extrapolation. Scot Armstrong is a recognised member of the "forecasting" community with a reputation for thoroughness, especially in reviewing the literature. That thoroughness provides confidence in using the bibliography in Armstrong (1985) as an indicator of the nature of academic interest in forecasting. A search of the 130 pages of bibliography reveals approximately 10 references that are primarily concerned with subjective forecasting. There are three broad classes into which the judgmental or subjective forecasting literature falls. Those classes contain papers that conclude:

a) That judgmental forecasting is not as accurate as statistical processes, eg.:

Adam & Ebert (1976), used simulated data and concluded that judges did not handle random components well.

Carbone & Gorr (1985), used 10 real time series, and limited the support provided to calculators and drawing instruments. The authors concluded their paper with a sweeping and dramatic statement that appears to generalise their results far beyond the 10 series examined:

> "Beware of eyeball judgmental extrapolation from time series graphics-too much reliance on judgment is not advisable" Carbone & Gorr (1985, p160).

b) That judgmental extrapolation is as accurate as statistical processes:

Lawrence (1983), a small scale study with 13 subjects showing relatively poorer judgmental performance over short horizons, and better over longer horizons.

Lawrence, Edmundson and O'Connor (1985), used 111 real time series with both experienced and inexperienced forecasters. On average judgmental extrapolation was as good as the better statistical extrapolations.

c) That, in the main, practitioners adopt judgmental processes:

Lawrence (1983), a field study that showed that none of ten large, and highly computerised, Australian companies used statistical forecasting techniques. Four had tried, but had discontinued their use.

Dalrymple (1987), a large survey showed that 70% of U.S. company forecasts are made subjectively.

In addition to the rather limited support for judgmental forecasting outlined above, there is evidence from other fields of study that humans can perform well for tasks in which the judge has some expertise. Edwards (1983) comments:

> "...even without tools, experts can in fact do a remarkably good job" Edwards (1983, p 511).

This conclusion was based on the results of studies concerning probability assessment by weather forecasters[1], and to a lesser extent by physicians[2]. In an other area, Brown and Rozeff (1978) show that analysts perform well in predicting corporate earnings. Phillips (1983) lends some support to the view of Edwards (1983), and in discussing probability assessment[3] points out that while there is a strong suggestion from the literature that judgment under

---

[1] Murphy and Winkler (1977).

[2] Lusted et al (1982).

[3] Much of the original work on which the popular view of human capabilities is based was concerned with probability assessment, see Edwards (1983) at page 508.

uncertainty is flawed, the generalisation of the results of the studies is a value judgment. He continues:

> "The major conclusion of this paper is that we do not yet know how good people are at judging uncertainty, and that people may well be capable of making precise, reliable and accurate assessments of probability" Phillips (1983, p 526).

Despite the forgoing, the work of Lawrence (1983), and Lawrence, Edmundson and O'Connor (1985), plus the reported preference for judgmental methods shown by practitioners (Dalrymple 1987), the academic forecasting community has been slow to endorse judgmental forecasting. Edwards (1983) might provide a partial explanation for this:

> "I hear the message that man is a "cognitive cripple" from a wide variety of nonspychologists (sic) these days. I encounter it in refusals to accept manuscripts submitted for publication showing men performing such tasks well;..." Edwards (1983, p 509)

The outcome of this is that little attention has been paid to means of supporting human judgment in forecasting economic time series. There has, however, been considerable energy devoted to the development of extrapolative techniques designed to replace the judge. Those techniques vary greatly in sophistication, but not much in accuracy (see for example Makridakis et al 1982[4]).

---

[4] The issue of relative accuracy of the time series extrapolation methods is discussed in sections 2.2 and 2.3. There it is shown that there is no "best" method for all time series, and that judgmental extrapolation is as viable a candidate method as the better of the statistical methods on the basis of accuracy of extrapolation.

Even if the extrapolation were performed statistically, judgment would be required in the consideration of non-time series data. In a field study in one consumer based industry Edmundson, Lawrence, & O'Connor (1987) showed that the accuracy of a sales forecast was primarily determined by knowledge lying outside the time series values themselves. Judgmental revision of an extrapolation requires detailed knowledge of the prior process. This would be difficult, especially for the more sophisticated statistical methods.

This analysis demonstrates that, on pragmatic and academic grounds, further research is warranted into the support of the judgmental forecasting task, and other related issues including the evaluation of the impact of interactive graphics, and decomposition on the time series forecasting task. The potential scope of the task is huge, encompassing the following:

* supporting the extrapolation decision,
* creating an environment in which one or more judges may work on a forecast, including the provision for simulation of effects of the forecast on, say, budgets,
* means to support the use of codified and soft data external to the time series, in either a separate part of the decision process or holistically with the extrapolation.

Therefore a narrower scope is required. This thesis is particularly concerned with attempts to improve the judgmental extrapolation process by the provision of a decision aid that could form the core of subsequent decision support systems. Therefore, the decision aid was designed bearing in mind the need to subsequently expand the investigation to cover the latter two points mentioned above.

1.3 RESEARCH METHODOLOGY ISSUES

1.3.1 THE FORECASTING ACTIVITIES SELECTED FOR STUDY

The forecasting task, as it is carried out in business, may require the consideration of:

* data from the time series history,
* information about internal and external events that affected the time series values,
* known business plans,
* anticipated external pressures that could affect the outcome for the forecast series, such as competitive reaction and broad economic events.

As described in the previous section, the scope of the project is limited to the consideration of time series data only. Statistical time series extrapolation methods can only, directly, take account of that data. The cues presented to the subjects generating judgmental forecasts were limited to the time series data, and did not include any information on the nature of the series or the years over which the data was collected. This places the methods under consideration on equivalent footing, in terms of the data sets from which the extrapolations are developed[5]. This abstraction permits the examination of the extrapolation process in the absence of confounding effects from non-time series data. It also limits the applicability of the results to this particular issue.

1.3.2 CHOICE OF ERROR MEASURE

There are a large number of available error measures, but none is recognised as generally applicable[6]. In the studies reported, the error measure selected is the Mean Absolute Percentage Error "MAPE". Selection of this measure, which is more fully discussed in chapters 3 and 6, reflects its common acceptance and use by practitioners (Carbone and Armstrong 1982). The MAPE also provides a conservative basis for the comparison of judgmental and statistical accuracy because it does not weight extreme errors as heavily as squared error measures. Lawrence, Edmundson and O'Connor (1985) showed that judgmental forecasts tended to avoid the occasional very large errors that statistical processes produce.

---

[5] This was the basis of the Lawrence, Edmundson and O'Connor (1985) study of the comparative accuracy of relatively unsupported judges and statistical methods. That study forms an important foundation for this dissertation, and is fully discussed in section 2.3

[6] Makridakis, Wheelwright and McGee (1983).

1.3.3 THE TIME SERIES USED

There are two broad strategies to be found in the literature for obtaining experimental time series. The series for some studies are generated by statistical processes[7]. This approach has some useful qualities, the underlying process of the series is known, therefore it is possible to determine the "true" seasonal or trend characteristics of the series as opposed to the characteristics apparent from the particular values used in the study. This approach suffers from the disadvantage that the external validity of the results is questionable, in that there is little assurance that the series used reflect the nature of real commercial time series. There is also the difficulty that it is no easier to determine the nature of the series as portrayed in the values used than for a real series. For instance, random events can cause the series to display apparent trend or seasonal characteristics in the short term.

The second approach to acquisition of experimental series is to use real series[8]. The true nature of those series may not be entirely discernible because of the complexity of the driving processes. This may be more of a disadvantage than that for generated series, because in the latter there is at least an a priori expectation concerning the nature of the series. The external

---

[7] For instance Adam and Ebert (1976).

[8] For instance Carbone and Gorr (1985), Lawrence, Edmundson and O'Connor (1985).

validity of the results is less threatened, at least there are commercial instances of the series studied. The extent to which the results might be generally applicable depends on the selection of series.

In this dissertation the series used are real series kindly made available to researchers by the Makridakis et al (1982). The series are 68 monthly series that were randomly selected from a database of 1001 real series for use in the forecasting competition ("M-Competition") reported by those authors. This does not guarantee, however, that the particular series, or the mix of the series reflect the general situation.

1.3.4 EXPERIMENTAL SETTING AND SUBJECTS

The studies carried out were designed as laboratory experiments, and in the main used postgraduate students for subjects. This approach is open to the criticism[9] that the external validity is threatened:

> "There is considerable evidence to suggest that the external validity of decision making research that relies on laboratory simulations of real-world decision problems is low" Ebbesen and Konecni (1980, p 42)

While those criticisms are valid, and are discussed in chapters 6 and 7, they are mitigated by the need to decompose the overall problem into manageable tasks. More particularly, in this instance there is some evidence that the use of postgraduate students in a laboratory setting

---

[9] See Winkler and Murphy (1973).

does not differ from the use of commercial forecasters in a similar setting (Edmundson, Lawrence, and O'Connor 1986).

1.4 STRUCTURE OF THE DISSERTATION

Chapter 2:

presents a review of the forecasting literature. The literature reviewed mainly supports the studies carried out in chapter 3. It considers:

* the role of forecasting in business.
* statistical forecasting comparative accuracy literature.
* judgmental forecasting literature.
* literature on choosing forecasting methods.

Chapter 3:

explores the nature of time series and some of the processes used for extrapolation. It presents:

* a replication of the Lawrence, Edmundson and O'Connor (1985) study to determine that judgment is sufficiently reproducible to warrant consideration.
* a discriminant analysis study which reports the development and testing of candidate metrics to be used in establishing a means to choose between forecasting methods. The results of the study generated implications for the design of GRAFFECT

Chapter 4:

presents a review of the human information processing literature. This literature, along with the implications from the studies in chapter 3 support the design of GRAFFECT reported in chapter 5. It covers:

* general HIP literature

* forecasting HIP literature

* strategies to improve judgmental extrapolation, ie decision decomposition and form and presentation of cue data.

Chapter 5:

describes the design and operation of GRAFFECT.

Chapter 6:

presents studies examining the overall accuracy of forecasts generated using GRAFFECT, and time taken to make the forecast. Two studies are reported using experienced and naive subjects respectively. In addition, the discriminant analysis study reported in ch 3 is replicated with the GRAFFECT method as a candidate.

Chapter 7:

presents studies that examine seasonal pattern identification in the extrapolation. This and the following two chapters are concerned with determining whether there are sub-tasks in the extrapolation process that might be better performed automatically.

Chapter 8:

presents studies concerned with trend identification in the extrapolation process.

Chapter 9:

presents studies concerned with the extrapolation from the noise residual that results from the decomposition of the extrapolation task.

Chapter 10:

provides the summary and conclusions, discussion of the limitations of the studies, and the suggestions for future research.

1.5 REFERENCES

Adam E.E., and Ebert E.R., "A comparison of human and statistical forecasting", *AIIE Transactions*, 8,1, (1976), 120-127.

Armstrong J.S., *Long-range forecasting from crystal ball to computer*, Second Edition, John Wiley and Sons, New York, (1985).

Brown L.D., and Rozeff M.S., "The superiority of analyst forecasts as measures of expectations: evidence from earnings", *Journal of Finance*, 33, (1978), 1-16.

Carbone R., and Gorr W.L. "The accuracy of judgmental forecasting of time series.", *Decision Sciences*, 16, (1985), 153-160.

Carbone R., and Armstrong J.S., "Evaluation of extrapolative forecasting methods: results of a survey of academicians and practitioners", *Journal of Forecasting*, 1, (1982), 215-217

Dalrymple D.J., "Sales forecasting practices: results from a United States survey" *International Journal of Forecasting*, forthcoming (1987).

Ebbesen E.B., and Konecni V.J., "On the external validity of decision making research: what do we know about decisions in the real world?", in *Cognitive Processes in Choice and Decision Behaviour*, ed. Wallsten T.S., LEA, (1980)

Edmundson R.H., Lawrence M.J.,, & O'Connor M.J., "The use of non time series information in sales forecasting : a case study." *Information Systems Research Reports*, University of New South Wales. (1987).

Edwards W., "Human cognitive capabilities, representativeness, and ground rules for research", in *Analysing and Aiding Decision Processes*, ed Humphreys P., Svenson O., and Vari A., Akademiai Kiado, Budapest, (1983), p 507-513.

Lawrence M.J. "An Exploration of Some Practical Issues in the Use of Quantitative Forecasting Models", *Journal of Forecasting*, 2 (1983), 169-179.

Lawrence M.J., Edmundson R.H., and O'Connor M.J. "An Examination of the Accuracy of Judgment Extrapolation of Time Series" *International Journal of Forecasting* 1 (1985), 25-35

Lusted L.B., Roberts H.V., Wallace D.L., Lahiff M., Edwards W., Loop J.W., Bell R.S., Thornbury J.R., Seale D.L., Steele J.P., and Fryback D.B., "Efficacy of diagnostic radiological procedures" in *Practical Evaluation: Case Studies in Simplifying Complex Decision Problems*, ed Snapper K., Information Resources Press, Washington, (1982)

Makridakis S., Andersen,A., Carbone,R., Fildes,R., Hibon,M., Lewandowski,R., Newton,J., Parzen, E., and Winkler,R. "The accuracy of extrapolative (time series) methods : results of a forecasting competition", *Journal of Forecasting*, Vol 1, no 2, (1982), 111-153.

Makridakis S., Wheelwright, S.S., and McGee, V.E., *Forecasting: methods and applications*, 2nd ed., Wiley, New York, (1983).

Murphy A.H., and Winkler R., "Reliability of subjective probability forecasts of precipitation and temperature", *Journal of the Royal Statistical Society*, Series C, 26, (1977), 41-47

Phillips L.D., "A theoretical perspective on heuristics and biases in probabilistic thinking" in *Analysing and Aiding Decision Processes*, ed Humphreys P., Svenson O., and Vari A., Akademiai Kiado, Budapest, (1983), p 525-543.

Welsch G.A., *Budgeting: Profit Planning and Control*, 3rd edn, Prentice Hall, Englewood Cliffs N.J., (1971).

Winkler R., and Murphy A.H., "Experiments in the laboratory and real world", *Organisational Behaviour and Human Performance*, 10, (1973), 252-270.

2. REVIEW OF THE FORECASTING LITERATURE

2.1 INTRODUCTION

2.1.1 CHAPTER OVERVIEW

This chapter contains a review of the academic literature concerned with forecasting. The major sections of the chapter address the following issues:

2.2 Statistical forecasting methods

This section introduces the comparative accuracy research from the forecasting literature. It establishes that there is no "best" forecasting method for all series types. It also identifies the statistical methods that form the benchmarks used in this dissertation for determination of comparative accuracy, and the source of the database of time series used for testing.

2.3 Judgmental forecasting

In this section, the role of judgment in extrapolation is discussed, in terms of its use in business and its acceptance by the forecasting academia. It is shown that judgment is viable in extrapolation despite fairly entrenched views to the contrary found in the literature.

2.4 Choosing a forecasting method

There have been several calls for effective means to choose a forecasting method to suit a given series. The review of the literature shows, however, that no such means have been developed, and that the published attempts to find such a rule have not considered judgment as a candidate method.

2.1.2 OVERVIEW OF THE LITERATURE

The volume of academic forecasting literature since the 1960s has reflected the importance of forecasting as an economic activity. There has been no waning of interest in the area either as an academic or a practical issue. Jenkins, in reflecting on 25 years of forecasting experience commented:

> "It is no accident that forecasting as a management activity has assumed increasing importance in recent years. The world that we are passing on to our children and grandchildren will be a good deal more uncertain than the relatively stable environment that prevailed during the period 1945-1975." Jenkins (1982, p3)

Given the more turbulent economic environment to which Jenkins referred, it is understandable that some of the assumptions and attitudes that were held in the forecasting community should have been re-evaluated. Makridakis (1981) illustrated this well. The thrust of his paper centered on the effect of the likelihood of a failure of the assumption of constancy on statistical forecasting methods. Drawing upon the human information processing literature[1] for support Makridakis stated a view that had currency at that time, that human judgment would fare poorly in comparison with statistical methods. He continued:

> "Complaining, for instance, about the pure predictive ability of statistical forecasting makes little sense when the alternative, i.e. human judgment, can be even worse and is

---

[1] The literature is reviewed in chapter 4

> definitely more costly." Makridakis (1981, p309)

Both supposed attributes of judgmental extrapolation are addressed in this dissertation. The evaluation and improvement of the accuracy of judgment is the major aim. Chapter 3 considers, in part, the reliability of the Lawrence, Edmundson and O'Connor (1985) finding that judges are no worse than statistical extrapolative models. Then, in chapter 6, the accuracy of judgment aided by an interactive graphical tool is evaluated, and the cost of making judgmental extrapolations is also considered.

Despite his poor opinion of judgment, Makridakis (1981) did point out that there was no choice but to have human intervention if the established patterns and relationships broke down. The role of the human judge was not specified precisely, though its use in adjustment of the statistical forecast or of the forecasting model was implied. It is postulated that in addition the roles include generation of the initial extrapolation in appropriate cases.

Not all the forecasting literature is applicable to this study. Forecasting economic time series may be accomplished by a wide variety of means. Broadly, those means may be classified as either:

1) Extrapolative, or

2) Causative.

Of those two classes, only the extrapolative methods are considered in this research. This does not imply that the causative or econometric methods lack accuracy, interest, or use[2]. The extrapolative methods are attractive objects of research, however, by virtue of their lack of reliance on data outside the time series coupled with their common use in business. Armstrong, in reviewing this branch of the forecasting literature, commented:

> "I am defining extrapolation as methods that rely solely on historical data from the series to be forecast. No other information is used. This class of methods is widely used in forecasting, especially for inventory control, process control, and in situations where other relevant data are not available." Armstrong (1984, p52)

Armstrong (1984) traced the development of extrapolative forecasting from 1960. He illustrated the state of the art in 1960 by an anecdotal look at a forecasting application that relied upon human judgment supported by the provision of graphs "for the more important items". The development of statistical forecasting methods of varying sophistication since that time has been reflected in the academic literature. In his review of the literature comparing the accuracy of simple and sophisticated methods Armstrong (1984) discovered 39 studies, dated from 1960 to 1983. Those studies all considered the comparative accuracy of <u>statistical</u> techniques only.

---

[2] See Klein, L.R., & Burmeister, E, (1976).

In the last five years there has been evidence of some interest in the accuracy of forecasts generated by human judges. The recent results of Lawrence (1983), and Lawrence, Edmundson, and O'Connor (1985) which showed that judgment can be as accurate as the statistical methods prompted Armstrong to say:

> "These studies suggest that further research is needed on when and how to use judgment for extrapolation." Armstrong (1984, p57)

It is not clear what was meant by "when and how" to use judgment. But the following are the possibilities, (for the issues addressed in this dissertation, the major chapter numbers are included):

* For what time series (when?), (ch 3, ch 6)
* For what commercial circumstances (when?),
* For which tasks (how?):
    * Extrapolation, for direct use or for inclusion in an averaging process, (ch 3, ch 6)
    * For part of the extrapolation process, (ch 3, ch 7, ch 8, ch 9)
    * Amendment of statistical forecasts, in the light of external circumstances or otherwise,
    * Evaluation and adjustment of statistical models,
* The process of judgmental extrapolation, and its improvement (how?), (ch 3 to ch 9).

Evaluation of the role of judgment in the light of factors external to the time series is beyond the scope of this dissertation. As discussed in the conclusions in chapter 10, long term applied research into those issues is required, based on the results of the investigations reported here. Issues of adjustment of statistical forecasts, and averaging of forecasts are the subject of other current investigations.

2.2 STATISTICAL FORECASTING METHODS

2.2.1 OVERVIEW

In November 1984 Armstrong (1984) published a survey of studies from the extrapolative forecasting literature that compared forecasting methods. He found, and analysed 39 studies that compared the accuracy of statistical techniques. The objective of the analysis was to evaluate the effectiveness of the increasing sophistication and complexity of the techniques available that has been seen since the early 1960s. Armstrong found that, according to his classification of sophistication, there was little support for the added complexity of the sophisticated models.

Table 2.1 summarises the review of the studies in Armstrong's Table 1 (Armstrong 1984, p 54). The table shows the number of studies, in half-decade time periods, that revealed an accuracy advantage to either sophisticated or simple methods and those that were indeterminate.

	PERIOD					
	60-64	65-69	70-74	75-79	80-83	Total
Number of Studies finding:						
sophisticated models best	1	2	4	4	-	11
no advantage for sophist/simple	-	3	6	9	3	21
simple models best	-	-	-	2	5	7

Table 2.1 Twenty three years of comparative accuracy studies on statistical extrapolation

Two observations of interest may be made from the above table:

1) There has been a growing level of interest in comparative accuracy of forecasting methods from the mid 1970s. It is possible to point to the instability in world economics since the oil crises of 1974/75 and to infer that interest in forecasting increases at such times[3].

2) In more recent experiments, and perhaps as the time series examined have become more tainted with the instability since the mid 1970s, there has been a swing away from reporting success of sophisticated techniques towards either indeterminate results or success of simple methods.

Those casual observations must not be given too much weight, there are several alternative explanations for the particular pattern of incidences of reporting study

---

[3] Indeed Makridakis (1981, p 308) did draw such implications

outcomes. The explanations, for instance, might include any or all of the following:

* "fashion" effects in which one study stimulated a batch of "me too" studies.

* a selection process of authors, editors, and publishing houses. After a point there is no interest in publishing yet another study that confirms prior results but a contradictory study arouses new interest.

None the less, the range of results does indicate the indeterminacy of the overall superiority between candidate statistical extrapolative methods:

> "More important, Table 1 provides little evidence to suggest that sophistication beyond methods available in the 1960s has had any payoff." Armstrong (1984, p55)

As an alternative to sophistication Armstrong (1984) proposed that selection techniques be developed[4]. In part, chapter 3 is directed at finding a method for selection between forecasting methods.

A closer look at the "comparative accuracy" of statistical forecasting is provided in the immediately following section.

2.2.2 THE COMPARATIVE ACCURACY OF STATISTICAL METHODS

The forecasting literature has predominantly considered statistical methods, thus the review

---

[4] It does not follow, however, that the development of selection methods is an alternative to the development of extrapolation methods.

establishes the general framework within which this dissertation is placed. The time series and statistical forecast data used in the dissertation was drawn from the major study reviewed in section 2.2.2.1 below.

As indicated above, there have been 39 studies of the comparative accuracy of statistical forecasting techniques between 1960 and 1983. The growth in interest has been dramatic since the mid 1970's, and this is reflected in the scale and rigour of the studies reported since that time.

Newbold and Granger (1974) are recognised as providing one of the earliest major studies of comparative accuracy. They analysed 88 time series and compared the accuracy of exponential smoothing methods, stepwise autoregression and ARIMA (Box Jenkins) methods. They reported that the ARIMA method achieved better results 65% of the time, confirming a prior finding of Reid (1969).

Makridakis and Hibon (1979), conducted a study using 111 time series and some 22 forecasting methods. Their results were at variance with those of Newbold and Granger (1974), in that it was shown that exponential smoothing models performed better than moving average models, especially for the shorter forecast horizons up to 6 months ahead. This result was qualified by the observation that for series with very low randomness the ARMA did do better.

Following the presentation of the Makridakis and Hibon (1979) study, there was an ongoing discussion that lead to the most recent major study. This is the so called "M-Competition" reported by Makridakis et al (1982). There follows a detailed review of that study, which forms the platform for this dissertation. The results from that study, and the data used were kindly made available in machine readable form by the authors.

2.2.2.1 The M-Competition, Makridakis et al (1982)

The M-Competition was a major study of forecasting accuracy. It comprised seven experts forecasting up to 1001 time series with up to 24 techniques. Certain techniques, such as Box Jenkins, required such effort[5] that only 111 time series were forecast. The 111 series were selected by taking every ninth series from the 1001 starting from a randomly selected point. The authors excluded from their reports the results obtained for certain time series which had an MAPE greater than 1000%.

The stated intentions of the authors was not to find a best forecasting method:

> "What is important, therefore, is not to look for 'winners' or 'losers', but rather to understand how various forecasting approaches and methods differ from each other an how information can be provided so that forecasting users can be able to make rational choices for their situation." Makridakis et al (1982, p111).

---

[5] In the case of Box Jenkins it was reported that each series required about one hour of human effort Makridakis et al (1982, p112).

The study was a 'post sample' study, involving the retention of data for the evaluation of forecast accuracy, rather than a model fitting study. Accuracy was evaluated in terms of five measures:

Mean Average Percentage Error

Mean Square Error

Average Ranking

Medians of Absolute Percentage Error

Percentage Better

With such a number of forecasting methods, and the range of error measures used, the report was very detailed, and the results rather complex. The authors drew many of their conclusions from the sub-set of data containing 111 time series. They observed that performance varied with:

Accuracy measure used,

Type of series, and

Forecast horizon.

The results of the M-Competition were not clear cut, and the authors did not provide an extensive discussion or interpretation of the results. They did, however, draw a number of inferences from the data which are summarised as follows:

* Deseasonalised single exponential smoothing performed well for <u>monthly series</u>, but not for annual series.

* Simple techniques performed better for micro series, while sophisticated techniques were better for macro series.

* Deseasonalised single exponential smoothing, and other relatively simple methods performed well for seasonal series.

* High noise series were better forecast by simple techniques.

* Simple methods such as deseasonalised single exponential smoothing performed well at short time horizons.

* Sophisticated methods such as Box Jenkins may have had problems with high trend series.

Interestingly, there was no evidence of differences in accuracy in forecasting time series from the pre and post 1975 period. The authors give no indication of the number of series that were compared in coming to this conclusion.

From the point of view of developing alternative forecasting techniques there was an important conclusion that in determining the seasonal characteristics of the time series the authors could not improve on a simple ratio-to-moving average (centred) calculation. They tried two processes based on the sophisticated CENSUS II method, but neither improved on the simple method. In a re-analysis of the data reported in the competition, McLaughlin (1983) indicated the importance of the estimation of the seasonal component by showing that only six of 15 methods improved on a naive-2 forecast[6], two of the six methods comprised using the composite of a number of individual forecasting methods.

In choosing the methods for inclusion in the competition, the authors of the M-Competition identified the following means of obtaining forecasts:

> "(a) purely judgmental approaches;(b) causal or explanatory (e.g. economic or regression methods; (c) extrapolative (time series) methods; and (d) any combination of the above" Makridakis et al (1982, p111).

The study was then limited to the extrapolative methods. The techniques evaluated implied that the authors' class (c) only contained methods that were substantially objective[7]. Thus they did not evaluate a judgmental extrapolation method. Lawrence, Edmundson, and O'Connor (1985) used the same time series data, and the statistical forecasts from the M-Competition to extend the comparisons to include extrapolations made judgmentally. This is more fully discussed in section 2.3 following.

Although the M-Competition did not reveal a testable set of rules for forecasting method selection it did indicate some general principles that might be used. More importantly the study gave impetus for further studies to address the issue of developing

---

[6] Naive-2 forecasting adopts as a forecast the last observation in the time series after taking account of seasonality.

[7] Though consider the inclusion of Box-Jenkins with a reported human input of one hour!

decision rules for the choice of forecasting method[8] and improving understanding of forecasting:

> "Specific hypotheses should be formulated and tested to determine which methods will be most effective in given situations. In other words, the goals now should be to find specific guidelines to help make better forecasts in a given situation." Armstrong and Lusk (1983, p 261).

2.2.3 CONCLUSIONS FROM THE STATISTICAL FORECASTING LITERATURE

The results of the "M-Competition" encapsulate the general view of the recent literature. There is no single method of extrapolation that performs best for all time series and circumstances. This raises two avenues for improvement in forecasting in general:

* Improve current methods so that any choice may be made from better candidates, and perhaps a dominant candidate might emerge,

* Find a means to select between candidate methods.

Both those issues are addressed in this dissertation, the latter in chapters 3 and 6. The former, in this case the improvement in judgmental extrapolation, is the major thrust of the dissertation.

The conclusions drawn by the authors of the "M-Competition" that are listed in the previous section form the basis for the selection of one of the benchmarks used throughout the dissertation. Deseasonalised single

---

[8] This is examined later in chapters 3 and 6.

exponential smoothing was shown to be a preferred method in the following circumstances:

* For monthly series
* For seasonal series
* For high noise series
* For short forecast horizons

The scope of the dissertation is limited to consideration of monthly series. As will be discussed in chapter 4, there is some interest generated from the human information processing literature in the performance of judgment in the presence of high noise, and in particular in the detection of signal such as seasonality in the presence of high noise. The forecast periods selected for examination, months 1-6 and 7-12, are the short and mid range forecast horizons considered in the "M-Competition". Thus, the single most stringent test for judgment would appear to be deseasonalised single exponential smoothing. This method has a further, practical advantage as a benchmark, its simplicity makes it a very attractive alternative for managers currently making judgmental forecasts who have not got the technical expertise or resources to adopt a sophisticated method.

The other benchmark used in the dissertation is Box-Jenkins. This is a sophisticated forecasting technique. It did not perform outstandingly in the "M-Competition", but it is selected because of its mystique and reputation, particularly among academic forecasters. It seems that no comparison of accuracy can be quite complete unless Box-Jenkins is included as a candidate.

2.3 JUDGMENTAL FORECASTING

2.3.1 IS INTEREST IN JUDGMENTAL FORECASTING JUSTIFIED?

Despite the "academic" attitudes towards judgmental forecasting implied in the quotation from Makridakis (1981) discussed in section 2.1 above (that judgment is worse and more costly), there is evidence that practitioners are prepared to espouse judgment. Dalrymple (1987) reviews the not insubstantial literature[7], and concluded that about 70% of commercial forecasts are made subjectively. Armstrong (1985), in commenting on the literature said:

> "Most forecasts are made with subjective methods....It also seems that the more important the forecast, the more likely it is that subjective methods will be used. Yet in many of these situations, objective methods would be more appropriate" Armstrong (1985, p73)

Armstrong is clearly a most respected and influential member of the international academic forecasting community. But his comment above raises an extraordinary enigma. It has been observed that over the last twenty years there has been no diminution in the use of subjective forecasting methods by the practitioners. Those practitioners make forecasts of immediate, and in

---

[7] See for example Cerullo and Avila (1975), Rothe (1978), Sparkes and McHugh (1984), Mentzer and Cox (1984), and Dalrymple (1985).

many cases vital importance. Business forecasts form the basis of decisions concerning matters such as cash budgeting which affect the very existence of the organisation. In the same twenty years there has been considerable publication of academic opinion that objective methods of forecasting are superior to subjective methods. There are few explanations of this dichotomy, including:

1) That practitioners have not kept abreast of the academic literature, and do not know of the possibility of improvement in forecasting accuracy.

2) That the standards by which academics evaluate forecasting methods are not those used by practitioners.

3) That the standards are the same, but that there is some fault in the evaluation and reporting processes.

As far as the first of these is concerned, it must be doubted that the whole body of forecasters have remained ignorant of the possibilities of gaining real benefit. Material is published in both practitioners and academic journals, and the services of forecasting experts and the availability of forecasting packages are well evident in the market. Lawrence (1983) reported a small field study involving ten large, and highly computerised, companies in Australia. None of the companies studied used a statistical forecasting method at the time of the study. Four had used statistical methods but had discontinued their use. Lawrence (1983) commented:

> "...the lack of use probably does not reflect lack of exposure to the techniques, since a number of organizations have experimented with them, but rather lack of *perceived* success with their implementation." Lawrence (1983. p 170).

The second and third explanations are not as easily dismissed. There is evidence, from the forecasting literature cited previously in this chapter, that there has always been a role for judgment. The statistical methods are not able to take external information into account in arriving at a forecast. Even in this role there is little or no direct evidence of the effect of the subjective inclusion of market information in the forecast. Carbone et al (1983) found that subjective adjustment of statistical forecasts did not give rise to improved accuracy, but the subjects did not make the revision in the light of any extra information.

In very recent times there has been some limited evidence that human judgment is not less accurate than statistical methods, even using the standards applied by the academic forecasting community. Lawrence, Edmundson and O'Connor (1985) established that human judgment was as accurate, on average, in forecasting economic time series as the better of the statistical techniques examined by Makridakis et al (1982) in the so called "M Competition". Lawrence, Edmundson, and O'Connor (1985) reported that there were, however, great relative differences in accuracy for individual time series, but that the standard deviation of the error of the forecast was lower for judgment than for statistical methods.

There is no direct evidence on which to draw in making a conclusion as to why practitioners use judgment in the face of the arguments from forecasting academia. Perhaps practitioners use judgment in making many of their forecasts because it has proved to them to be the "best" in those circumstances, where "best" is determined with respect to accuracy, cost, ease of use, and the comfort of the decision maker. The observations made by Lawrence, Edmundson and O'Connor (1985) would fit such a view:

* Judgment is as accurate overall, but not for every time series, as statistical methods.

  This would imply that the experienced forecaster would be found applying either judgment or statistics according to circumstances.

* The standard deviation of the error of the forecast is lower for judgmental than for statistical forecasts.

  For what Armstrong (1985) called "more important" forecasts it would not be unexpected, therefore, that the method with the lower chance of extreme error might be adopted.

From the foregoing discussion it is considered that pursuit of improvement in the formulation of forecasts by consideration of the use, and the improvement of judgment is justified. From a pragmatic viewpoint, the improvement of judgmental forecasting methods would find wide application with practitioner forecasters who have adopted judgmental methods to date. From a more academic viewpoint, the choice of whether or not to use judgment in particular instances is of direct interest, addressing, as it does, the possibility that practitioners are not necessarily making the best choices at present. This is considered further in the next section.

2.3.2 IS JUDGMENT A VALID CANDIDATE FOR CHOICE OF FORECASTING METHOD?

There is a theme in the forecasting literature that implies that human judgment is not as accurate as statistical methods in forecasting time series. Makridakis and Hibon (1979)[10] concluded that there was sufficient evidence to support the view that clinical judgment would outperform statistical methods only in very rare cases.

Forecasting competitions in recent times have failed to reveal a statistically based forecasting methodology that was significantly better than its competitors in all circumstances. The work of Lawrence, Edmundson and O'Connor (1985) extended those competitions to include various unsophisticated judgmental forecasting methods. Again, the broad outcome was that, on average, there is no significant difference between methods although Carbone and Gorr (1985) using a sample of time series from one competition did conclude that a difference in

---

[10] Recently re-published in Makridakis et al (1984)

accuracy existed. The Lawrence, Edmundson, and O'Connor (1985) results showed that the variation in relative accuracy between statistical methods over the range of time series found by Makridakis et al (1982) also held when judgmental methods were examined.

The standard deviation of the forecast error for judgmental methods has been shown by Lawrence, Edmundson, and O'Connor (1985) to be lower than that of the statistical methods considered. Although this factor is not of direct relevance to the dissertation, it does highlight the role of judgment at least in circumstances where the avoidance of very large errors is desirable. There are many commercial situations in which marginal "mean" accuracy would be sacrificed in order to ensure that very large forecast errors did not occur.

In the commentary on Armstrong (1984), Ayres (1984) conjectured that the variation in performance between forecasting methods arose from the fact that the methods responded to different cues in the data:

> "Each of the statistical methods tested only recovers part of the coded information. In fact, the surprising conclusions cited by Armstrong can perhaps be explained on the hypothesis that even the 'best' statistical method recovers only a fairly small part of the coded information in the time series. The fact that subjective 'eyeball' extrapolations are as good as computer extrapolations, at least in some situations, would seem to substantiate this hypothesis." Ayres commentary on Armstrong (1984, p61).

The matters mentioned above, and the fact that commercial forecasters continue to use judgmental forecasting processes, indicate that there are good reasons to consider judgment as a candidate when making a choice of forecasting method. Indeed, the pragmatics of the issue would suggest that evaluation of the choice between statistical and judgmental methods is vital. This is certainly the view of Armstrong (1985):

> "In my opinion, the choice between subjective and objective methods is the most important decision to be made in the methodology tree" Armstrong (1985, p73).

The choice made by the practitioner is made in a complex situation. It is necessary to consider the following factors, and any trade-off that exists between them:

1) The relative accuracy of the extrapolation by the candidate methods,

2) The chances of a wild error, and the costs of such an error,

3) The need to consider external information in making the forecast,

4) The efficiency of the forecasting process.

The consideration of some of those issues, and certainly the interaction between them, calls for the use of fine judgment. If it were possible to provide the decision maker with better information about the relative accuracy of the extrapolation for the instance at hand rather than overall, then this could give rise to a better decision. In the following section, the issue of choosing between methods merely on the basis of extrapolative accuracy is considered.

2.4 CHOOSING A FORECASTING METHOD

The fact that certain methods appear to perform well for particular time series stimulates the desire to derive an a priori means to classify time series, and to identify the characteristics of the time series and processes that give rise to those differences[11]. In the conclusions to their report on the M Competition Makridakis et al (1982) expressed this opinion:

> "If the forecasting user can discriminate in his choice of methods depending upon the type of data...then he or she could do considerably better than using a single method across all situations....Overall, there are considerable gains to be made in forecasting accuracy by being selective". (Makridakis et al 1982, p145)

Fildes (1985) attempted to quantify the gains to be made by correctly selecting the appropriate forecasting method:

> "there is typically between 10 and 20% to be gained by selecting the model with best aggregate performance measured in MAPE." Fildes (1985, p8)

Fildes (1985) continues to show that the commercial effect of such a gain realized in an inventory forecasting environment could be worth 2% of the total inventory commitment.

Although Makridakis et al (1982), and Fildes (1985) were addressing the question of choice between statistical

---

[11] See the commentary by Reid on Makridakis and Hibon (1979)

forecasting methods the Lawrence, Edmundson, and O'Connor (1985) study established that unsophisticated judgmental forecasting methods were viable alternatives.

The lack of an established rule for the selection of a forecasting method may be inferred from the above quotation from Makridakis et al (1982). This confirmed the opinion of Fildes and Howell (1979):

> "The difficulty that confronts the forecaster is that there is as yet no theoretical basis on which to choose a forecasting model appropriate to his situation" Fildes and Howell (1979, p297).

Fildes (1985) lists eight methods for forecasting method selection, of those that are truly selective[12] the procedure is based on the comparison of the model errors either on the single series or an aggregate of series. Such an approach may be useful but may be expensive to adopt if there is a risk of the time series changing character, in which case the full extrapolation evaluation with each model would be required rather frequently. Finally, the issue is put succinctly by Schnaars (1984) who comments:

> "Still rare, however, are comparable studies that attempt to discover those situations in which one type of forecasting model might be expected to outperform another" Schnaars (1984, p290)

For the initial approach to the task of selecting a forecasting technique it has been determined to consider the supposed position of the practitioner. Lawrence (1983)

---

[12] Some methods listed by Fildes (1985) are not selective in the true sense, for instance he lists a rule that selects the first of two models, another to select the second, and yet another to average the two models under consideration.

has shown that at least some practitioners do use judgmental methods, and Dalrymple (1987) goes further to show that judgment is commonly used in practice. The task facing such a forecaster is thus to decide when to use a statistical process in preference to judgment. Chapter 3 considers the development of a reliable decision rule, based on objectively determinable metrics, to permit a method to be established to indicate the advisability or not of using a judgmental method. The characteristics of the decision rules also, incidentally, indicate matters of concern in devising a method to overcome shortcomings in judgmental forecasting.

There is little comment in the literature concerning the identification of metrics for forecast method selection. In considering selection between competing statistical forecasting techniques Fildes and Howell (1979) stated that the ex ante performance of the methods must be considered because:

> "Ample evidence exists to show that good ex post fit does not lead to good (ex ante) forecasts, and, more seriously, that ex ante forecasting performance does not even correlate closely with ex post fit." Fildes and Howell (1979, p 304).

Some general indications of characteristics of forecasting methods may be gleaned from the literature, however. For instance, Chatfield and Prothero (1973) commented that the Box Jenkins method was "generally not so good for ...seasonal series with a large random component." Chatfield and Prothero (1973, p296).

Makridakis and Hibon (1979) regressed MAPE of model fit, and MAPE of forecast against a number of metrics derived from time series decomposition. The authors did not come to a firm conclusion, but felt that a high noise component in the decomposition might lead to over fitting with sophisticated forecasting techniques.

Reid (1971) considered a number of adaptive forecasting methods and developed a decision tree for choosing between them based mainly on subjectively determined characteristics. the characteristics he used were the length of the series, seasonality, randomness, peakedness, and time horizon. Schnaars (1984) considered the effect of amounts of data, aggregation level, product type, and stability[13] on the accuracy of a sample of forecasting methods. Schnaars (1984) examined annual series only, and concluded that stability in the history of the time series was an important factor in the selection of an extrapolation method. The "stability" was determined subjectively, and the competing methods considered by Schnaars (1984) varied from random walk to triple exponential smoothing. Schnaars concluded:

> "In sum, market forecasters can expect to do no better than a simple random-walk model when forecasting unstable data series. These series can be identified from an inspection of scatter diagrams." Schnaars (1984, p 296).

Finally, in the "M-Competition" the classifications of

---

[13] Stability was determined both judgmentally and using autoregression.

the time series used in analysis were:

1) seasonal/nonseasonal series,

2) macroeconomic/microeconomic series,

3) demographic/industry series,

4) monthly/quarterly/annual series.

That study concluded that deseasonalised single exponential smoothing performed relatively well for monthly time series in general, Holt's method[14] did well if there was trend present, simple methods did better for micro series and sophisticated methods did better for macro data. It is not certain what constituted a 'sophisticated' method, and there was no evidence that the 'more sophisticated' of two competing methods was likely to be the better for forecasting a macro series.

The classification categories were also problematic because they were broad and overlapped. Furthermore some important characteristics of a time series may well not have been captured by these metrics. For example, the seasonal/nonseasonal classification does not capture the stability aspects of the seasonal component.

It is possible that the fuzziness of the categories used in the M competition has contributed substantially to the lack of success in identifying rules that point to the clear superiority of any one method in any given conditions.

---

[14] Holt's method has a trend factor which DSE lacks, see Makridakis, Wheelwright, and Mcgee (1983)

None of the studies mentioned addressed the question of discriminating between judgmental and statistical forecasting techniques. Neither did they express their results in a form to enable a decision to be made in the light of objective metrics. The discriminant analysis studies reported in chapters 3 and 6 address both those issues. In the latter case, the judgmental extrapolation considered was supported by a computerised data presentation tool.

2.5 CONCLUSIONS

The literature reviewed above indicates that there is a valid role for judgment in time series extrapolation. Forecasting in the commercial environment would benefit from any improvement in the accuracy of commonly used extrapolative methods. The substantial use, in business, of judgmental forecasting gives impetus to the need to develop methods of improving the accuracy of that process. There would also be significant benefit if an objective rule could be developed to choose between extrapolative methods.

Chapter 3 reports a study aimed at identifying such a rule, based on the characteristics of the time series. The results of that study have implications both for the selection of extrapolation method, and for the development of improved methods. For instance, the finding that judgment was disadvantaged in the presence of high seasonal and low noise characteristics in the series, lead to the development of data display strategies designed to overcome the difficulty.

In chapters 4 onwards a computerised aid for judgmental extrapolation, based on the results from chapter 3 and the human information processing literature, is developed and evaluated.

2.6 REFERENCES

Armstrong J.S., "Forecasting by extrapolation: conclusions from 25 years of research", *Interfaces* 14:6, (1984), 52-66.

Armstrong J.S., *Long-range forecasting from crystal ball to computer*, Second Edition, John Wiley and Sons, New York, (1985).

Armstrong J.S., & Lusk E.J., Commentary on the Makridakis Time Series Competition (M-Competition), *Journal of Forecasting*, 2, (1983), 259-311.

Armstrong J.S., Denniston W.B. Jr., and Gordon M.M., The use of the decomposition principle in making judgments" *Organizational Behavior and Human Performance*, 13, (1975). 257-263.

Ayres R.U., Commentary on Armstrong J.S., "Forecasting by extrapolation: conclusions from 25 years of research", *Interfaces* 14:6, (1984), 61-62

Carbone R., Andersen A., Corriveau Y., and Corson P.P., "Comparing for different time series methods the value of technical expertise, individual analysis, and judgmental adjustment" *Management Science* 29, (1983), 559-566.

Carbone R., and Gorr W.L. "The accuracy of judgmental forecasting of time series.", *Decision Sciences*, 16, (1985), 153-160.

Cerullo M.J., and Avila A., "Sales forecasting practices: a survey" *Managerial Planning*, 24, (1975), 33-39.

Chatfield C., and Prothero D.L., "Box-Jenkins seasonal forecasting: problems in a case study", *Journal of the Royal Statistical Society, Series A*, 136 part 3, (1973), 295-336.

Dalrymple D.J., "Sales forecasting practices in business: results from a 1983 U.S. survey." Working paper, Graduate School of Business, Indiana University.

Dalrymple D.J., "Sales forecasting practices: results from a United States survey" *International Journal of Forecasting*, forthcoming (1987).

Dickhaut J.W., and Eggleton I.R.C., "An examination of the processes underlying comparative judgments of numerical stimuli." *Journal of Accounting Research*, Spring (1975), 38-72.

Edmundson R.H. "Metrics for a priori selection of forecasting methods: a preliminary investigation." *Fourth International Symposium on Forecasting*, London (1984).

Fildes R., "Gains through univariate forecasting model selection" *Fifth International Symposium on Forecasting*, Montreal, (1985).

Fildes R., and Howell S., "On selecting a forecasting model", *TIMS Studies in the Management Sciences* 12, (1979), 297-312.

Hogarth R.M., *Judgment and choice*, Wiley, Chichester, (1980).

Hogarth R.M., and Makridakis S., "Forecasting and planning: an evaluation", Unpublished manuscript, INSEAD, Fontainebleau, France, (1979).

Hogarth R.M., and Makridakis S.. "Forecasting and planning: an evaluation", *Management Science*, 27,2 (1981), 115-138.

Jenkins G.M., "Some practical aspects of forecasting in organizations", *Journal of Forecasting* 1, (1982), 3-21.

Klein L.R., & Burmeister, E,. *Econometric model performance*, University of Pennsylvania Press, Philadelphia, (1976).

Lawrence M.J. "An Exploration of Some Practical Issues in the Use of Quantitative Forecasting Models", *Journal of Forecasting*, 2 (1983), 169-179.

Lawrence M.J., Edmundson R.H, and O'Connor M.J., "The accuracy of combining judgmental and statistical forecasts", *Management Science*, 32,12, (1986), 1521-1532.

Lawrence M.J., Edmundson R.H., and O'Connor M.J. "An Examination of the Accuracy of Judgment Extrapolation of Time Series" *International Journal of Forecasting* 1 (1985), 25-35

Makridakis S., "Forecasting accuracy and the assumption of constancy", *Omega* 9:3, (1981), 307-311

Makridakis S., and Hibon, M., "Accuracy of forecasting: an empirical investigation", *Journal of the Royal Statistical Society, Series A*, 142, (1979), 97-145

Makridakis S., and Wheelwright S.C., *Forecasting Methods and Applications*. Wiley. Santa Barbera. (1978).

Makridakis S., Andersen,A., Carbone,R., Fildes,R., Hibon,M., Lewandowski,R., Newton,J., Parzen, E., and Winkler,R. "The accuracy of extrapolative (time series) methods : results of a forecasting competition", *Journal of Forecasting*, Vol 1, no 2, (1982), 111-153.

Makridakis S., Andersen,A., Carbone,R., Fildes,R., Hibon,M., Lewandowski,R., Newton,J., Parzen, E., and Winkler,R. *The forecasting accuracy of major time series methods*, Wiley, Chichester, (1984).

Makridakis S., Wheelwright, S.S., and McGee, V.E., *Forecasting: methods and applications*, 2nd ed., Wiley, New York, (1983).

Mentzer J.T., and Cox J.E., "Familiarity, application, and performance of sales forecasting techniques" *Journal of Forecasting* 3, (1984), 27-36.

Mclaughlin R.L., "Forecasting models: sophisticated or naive?" Commentary on the Makridakis Time Series Competition (M-Competition), *Journal of Forecasting*, 2, (1983), 259-311.

Moriarty M.M., and Adams A.J., "Management judgment forecasts, composite models, and conditional efficiency", *Journal of Marketing Research*, vol xxi, (1984), 239-250.

Newbold P., and Granger C.W.J., "Experience with forecasting univariate time series and the combination of forecasts", *Journal of the Royal Statistical Society, Series A*, 137 part 2, (1974), 131-164.

Reid D.J., "A comparative study of time series prediction techniques on economic data" Ph.D. Thesis, University of Nottingham (1969).

Reid D.J., "Forecasting in action: a comparison of forecasting techniques in economic time series", *Joint Conference of O.R. Society's Group on Long Range Planning and Forecasting*. (1971).

Rothe J.T., "Effectiveness of sales forecasting methods" *Industrial Marketing Management*, 7, (1978), 114-118.

Schnaars, S.P., "Situational factors affecting forecast accuracy", *Journal of Marketing Research*, vol xxi, (1984), 290-297.

Sparkes J.R, and McHugh A.K., "Awareness and use of forecasting techniques in British industry" *Journal of Forecasting*, 3, (1984), 37-42.

Winkler R.L., and Makridakis, S., "The combination of forecasts", *Journal of the Royal Statistical Society*, 146 Part 2, (1983), 150-157.

3. INVESTIGATION OF TIME SERIES CHARACTERISTICS

3.1 INTRODUCTION

This chapter reports on two studies:

a) a replication of part of the Lawrence, Edmundson, and O'Connor (1985) study.

b) a discriminant analysis study to identify time series characteristics that would be useful in choosing between judgmental and statistical forecasting methods on the basis of expected accuracy, and as an indication of the means to improve the judgmental process.

3.1.1 OBJECTIVES OF THE INVESTIGATION

As discussed in chapter 2, it was shown by Lawrence (1983) and Dalrymple (1987) that it was common in business for forecasts to be made judgmentally. Armstrong (1985) commented that the use of judgmental forecasting would often not be the optimum. In order to achieve an improvement in business forecasting it would therefore seem necessary to improve the judgmental process and to provide the means for managers to be aided in their choice of methods. The major objective of this dissertation is to determine the practicality of aiding forecasters by developing and evaluating an experimental decision aid as reported in chapter 5. The major work reported in this chapter, the discriminant analysis study, was intended to form the basis of design decisions taken in the development of the decision aid. The objectives of the discriminant analysis study were:

a) to examine the possibilities of generating simple rules, based on objectively determinable metrics, for choosing between a judgmental forecasting method and one of two representative statistical forecasting techniques, and

b) to provide some insight into the characteristics of time series that are associated with relative inaccuracy in judgmental extrapolation.

The first objective concerns the possibility of deriving a rule to indicate to a manager that either deseasonalised single exponential smoothing or Box-Jenkins[1] would provide a more accurate pure extrapolation than a judgmental process supported by hard copy graphs.

The second objective concerns identification of the characteristics of a time series that are related to relatively poor judgmental forecasting. This information would then be used during the design of decision aids with the intention of alleviating the problem. In some cases it might be possible to modify the decision process to account for the characteristic, or to signal the advisability of substituting a statistical process.

However, before the major study was undertaken it was necessary to show that the judgmental method to be considered was a reliable and reproducible process. Thus two pre-requisite objectives were required which were

---

[1] The choice of methods for comparison is discussed in section 3.3.1.2.

addressed in the replication study:

c) to establish the reproducibility of a forecast for methods involving judgment [2], and

d) to establish the stability of the ranking of forecasting methods involving judgment with respect to statistical methods.

These are very similar objectives. The first is concerned with determining whether judgment is consistently accurate enough on average to warrant examination. The second concerns a series by series check on accuracy relative to statistical methods. Failure to achieve consistency at this level would indicate that a rule could not be established to aid in the selection between judgment and the statistical methods.

3.1.2 JUSTIFICATION OF THE STUDIES

As described in sections 2.2.1 and 2.2.2 there is little difference in average accuracy of forecasting a large number of time series between the better of the available methods. There is, however, great variation in the accuracy with which methods forecast individual time series [3]. The authors of the "M-Competition" (Makridakis et al 1982) were quoted in chapter 2 as concluding that

---

[2] That is, where there is a subjective input to the forecast it is necessary to show that the output is sufficiently stable over replications to justify the use of the method.

[3] For instance, in comparing deseasonalised single exponential smoothing and judgment from Lawrence, Edmundson and O'Connor (1985), in 40% of the series forecast one or other method had an MAPE 1.5 times the other

effective selection of method would give rise to great benefit, and Fildes (1985) quantified this by estimating a 10 to 20% gain in MAPE. Examination of the data used in this dissertation shows that the estimate is correct, on average, but that the potential gains on some series is far greater.

Although Makridakis et al (1982), and Fildes (1985) were addressing the question of choice between statistical forecasting methods the Lawrence, Edmundson, and O'Connor (1985) study established that unsophisticated judgmental forecasting methods were viable alternatives.

None of the prior literature, reviewed in section 2.2.3, addresses the question of discriminating between judgmental and statistical forecasting methods, although it was shown in section 2.2.2 that judgmental forecasting remains much used in business. Further, those studies are generally deficient in the identification of objectively determinable metrics to be used in the comparison of methods, and the proposal of objective decision rules.

The benefits to flow from the successful identification of time series characteristics useful for selecting between a judgmental forecasting method and a statistical method would be:

a) the provision, within a forecasting decision aid, of objective functions to flag time series that might be extrapolated better using statistical processes.

b) the possibility of structuring a forecasting aid to overcome, or alleviate, the shortcomings of the judgmental method.

3.2 THE REPLICATION STUDY

3.2.1 INTRODUCTION

As described in section 3.1.1 above, the replication study addresses two issues related to the classification of time series for discriminant analysis on the basis of the ranking of the MAPE[4] accuracy.

The second of these issues, stability of ranking, has importance even if it is shown that judgmental forecasts are reproducible. There may still be problems in obtaining consistent rankings of forecasting methods. The size of the variance of the forecasting process for judgmental methods relative to the difference in means could give rise to instability in the rankings obtained. Classifications on the basis of point estimates within an area of substantial overlap of the population distributions could produce results that would confound the discrimination process. In considering misclassification because of overlaps among the groups Eisenbeis comments:

> "..in the equal dispersion, two group case ... when each of the group assignment errors are random and equal, there is no effect on the classification errors....The assumption of random assignment errors is not particularly realistic"(Eisenbeis 1977, pp 889).

---

[4] MAPE is the mean absolute percentage error of forecast, it is discussed in section 2.2.2.

3.2.2 DESCRIPTION OF THE REPLICATION STUDY

The broad objectives c) and d) in section 3.1.1 were addressed in two testable hypotheses:

H3.1 There would be a difference in forecast accuracy, in terms of MAPE, between two judgmental forecasts.

H3.2 There would be a difference in the ranking of the MAPE errors of replications of a judgmental forecast with respect to a deseasonalised single exponential smoothing (DSE) forecast.

To test those hypotheses the part of the work of Lawrence, Edmundson and O'Connor (1985) in which the subjects were supported by a hard copy plot was replicated as follows:

* The replication was restricted to the forecasting of the 68 monthly series in the data base, omitting the quarterly and annual series. The maximum forecast horizon of 12 months in the replication compared with 18 months in the original.

* The subjects used in the replication were those previously used in the Lawrence, Edmundson and O'Connor (1985) study.

* The forecasting method used was the judgmental extrapolation of a hard copy plot[5] (GRAPH).

---

[5] Lawrence, Edmundson, and O'Connor called this the "GR" method, however in this document it is referred to as the "GRAPH" or judgmental hard copy method to distinguish it from screen based graphical methods discussed in later chapters.

The replication took place some 14 months after the original. In that time, the validation data, and indeed the details of series type and timing, had not been seen by the subjects. The random assignment of series to the subjects ensured that there was limited chance that the results would be affected by any learning effect.

Average MAPE errors were calculated for the months 1-6, 7-12, and the overall 1-12 forecast horizons. The forecast horizons 1-6 and 7-12 were to be considered in the discriminant analysis study. The analysis of the results was as follows:

a) The MAPE's for the replication were compared with those of the original experiment using a paired t-test. This addresses the issue of reproducibility raised in section 3.2.1.

b) The MAPE's were also used to determine whether the ranking of GRAPH against DSE remained stable in the replication [6]. The two sets of rankings were compared using Kendall's Tau Coefficient [7].

3.2.3 RESULTS OF THE REPLICATION STUDY

The results of the t-test analysis is exhibited in table 3.1 below. It was not possible to confirm the hypothesis H3.1. Over the full 12 months the two error rates were found to be drawn from the same population

[6] In this case, the ranks were determined subject to there being a difference between the two MAPE's of at least 10% of the lower MAPE.

[7] The Tau statistic was preferred over the Spearman Rank Correlation Coefficient because the Tau gives each inversion of a pair the same weight, see Hays (1981).

with a high level of probability. The worst case was the month 1-6 comparison which revealed a 68% probability that the two were drawn from different populations. It was determined that, subject to the analysis of the rankings, this result did not provide evidence to remove GRAPH from the discriminant analysis study.

	original		replication		2-tail
	mape	std dev	mape	std dev	prob
time horizon					
1-6	11.6	9.4	12.4	10.4	0.32
7-12	16.5	14.2	15.8	15.8	0.63
1-12	14.1	10.8	14.1	11.8	0.98

Table 3.1. Comparison of MAPE's for Replicated Judgmental Forecasts

The results of the analysis of the ranking of GRAPH and DSE in the original study and the replication are exhibited in Table 3.2. It was not possible to confirm hypothesis H3.2. Of the series that achieved a definite ranking in both the original and the replication[8] 36 out of 45 were the same for months 1-6, and 45 out of 51 were the same for months 7-12. The analysis showed that the rankings obtained were acceptably stable over the two sets of observations. This indicates that subjects of similar skills and motivation are capable of reproducing a GRAPH forecast performance relative to DSE, and that GRAPH is a viable method for inclusion in the discriminant analysis study.

---

[8] That is, disregarding the series that were unclassed because the difference in MAPE between GRAPH and DSE was less that 10% of the lower MAPE.

PERIOD	ORIGINAL BEST METHOD	REPLICATION BEST METHOD GRAPH	DSE	UNCL*
1-6	GRAPH	**14**	4	4
	DSE	5	**22**	6
	Uncl	3	6	4
7-12	GRAPH	**15**	3	4
	DSE	3	**30**	7
	Uncl	2	1	3

* Uncl indicates those series for which the difference in MAPEs was less than 10% of the lower of the two.

Table 3.2 Comparison Original and Replication Rankings

Using Kendall's Tau test, it was shown that the resulting classifications were not independent with a significance of 0.0048 (1-6), 0.0004 (7-12).

3.2.4 DISCUSSION OF THE REPLICATION STUDY

The results of the replication study provide justification for including the GRAPH method in the discriminant analysis study. The method, despite inherent subjectivity, is sufficiently reproducible to warrant the development of a decision rule. In particular, GRAPH produced a stable ranking with respect to the DSE error.

A question remains as to whether forecasting techniques that have a small judgmental component, such as Box Jenkins, are also sufficiently reproducible. There is a judgmental input to Box Jenkins forecasts in the model fitting phase, but the judgmental tasks are well defined, and the results of decisions may be reviewed in the light of objective statistics [9]. The effect of minor errors in model identification would be somewhat mitigated by the statistical processes of parameter estimation. It is therefore assumed that Box Jenkins forecasts would be no less reproducible by persons of similar training than the purely judgmental forecasts, and that inclusion of the method in the discriminant analysis study is warranted.

3.3 THE DISCRIMINANT ANALYSIS STUDY

3.3.1 INTRODUCTION

The discriminant analysis study addresses the objectives a) and b) established in section 3.1.1 concerning the identification of time series characteristics that distinguish series that are forecast better by either statistical or judgmental methods. There follows a description of the data used in the study, and the forecasting methods and forecast horizons analysed.

3.3.1.1 DESCRIPTION OF THE DATA

The authors of the "M-Competition", Makridakis et al (1982), kindly made available to independent investigators the time series data (including the 'validation' data) and the forecasts used in the competition. As described in chapter two, Lawrence, Edmundson, and O'Connor (1985) extended the scope of

---

[9] For example, the effect of identifying a particular seasonal characteristic may be reviewed in the light of the resulting autocorrelation coefficients, and the partial autocorrelation coefficients.

the "M-Competition" to include unsophisticated judgmental forecasting. As a result, the data available for the discriminant analysis study extended to forecasts of 111 time series by 26 methods.

It was determined to limit the investigation to the 68 monthly time series in the data base. The reasons for this decision were:

a) there were more monthly series than series aggregated either quarterly or yearly,

b) higher levels of aggregation can give rise to effects such as smoothing fluctuations, and masking seasonal characteristics,

c) it was conjectured that in business forecasting, time series would be encountered that were sampled at least monthly.

Forecasts were available for the 18 months of data set aside as validation data. Two forecast horizons were selected from the data, months 1-6 and 7-12. As discussed in section 2.2.2 Lawrence, Edmundson, and O'Connor (1985) found that there appeared to be a change in the relative accuracy of judgmental methods at about months 6 and 7. The months 1-6 forecast horizon covers the immediate, short term in business forecasts. The 12 month cut off is arbitrary, but it was selected as a reasonable trade off between the desire to have a long enough horizon to make some business sense, and the need to have as short an horizon as possible to reduce the risk of a failure in the assumption of constancy.

3.3.1.2 FORECASTING METHODS CONSIDERED

The judgmental method examined here is the extrapolation of a hard copy plot of the time series, and is referred to below as GRAPH. This 'eyeballing' method was evaluated by Lawrence, Edmundson, and O'Connor (1985) and has been fully described in section 2.2.2. It was shown to have an accuracy not significantly different to DSE (deseasonalised single exponential smoothing) for the two forecast horizons studied here. Despite the similarities in average error, the correlations between GRAPH and DSE were not high (0.6 for the 1-6 case and 0.5 for the 7-12 case[10]). As described in section 2.2.2 this gives rise to the possibility of gaining advantage if the better characteristics of each method can be captured. A final reason for including GRAPH is that it is a simple method to which managers might resort. It assumes that the seasonal characteristics of the data would receive attention, and that the extrapolation would be performed using a simple plot.

In the report of the "M-Competition" the authors refer to simple methods performing well for micro series and sophisticated methods performing well for macro series. This pointed to the inclusion of a representative method from each category.

---

[10] By comparison, the correlations between DSE and BJ were 0.7 for the 1-6 case and 0.9 for the 7-12 case.

Armstrong (1984) indicated that there was little to choose between forecasting methods, and it was not possible to choose two representative statistical methods on the basis of significant difference in average MAPE over the 68 monthly time series. Table 3.3 reports an extract from the results of the "M-Competition" on the basis of the 22 single (as opposed to averaging) methods, with the inclusion of the GRAPH results form Lawrence, Edmundson, and O'Connor (1985). The table indicates the very similar accuracy achieved by the better methods.

method	MAPE	rank/23
BAYES F	10.7	1
D ARR EXP	10.8	2
DSE	11.0	3
BJ	11.3	4
PARZEN	11.4	5
GRAPH	11.6	6
mean (of 23)	14.9	

Table 3.3 Performance of Selected Methods for months 1-6.

Since there was no statistically significant difference on which to base the choice between the methods it was decided to consider both the level of accuracy achieved in the "M-Competition" and the standing and acceptance of the methods.

The "simple" statistical process chosen was Deseasonalised Single Exponential Smoothing (DSE). DSE is a simple technique that utilises elements of decomposition and smoothing, it is cheap and easy to implement, and requires no specialised knowledge. The method performed well in the "M-Competition", ranking third in accuracy over the months 1-6 forecast horizon. DSE was bettered by a Bayesian forecasting technique (MAPE of 10.7) and deseasonalised adaptive response rate exponential smoothing (MAPE of 10.8). In selecting a "simple" method, the Bayesian technique would not be a candidate. The deseasonalised adaptive response rate exponential smoothing might qualify as a simple method, but the improvement over DSE was slight and would not warrant the added complexity. Finally, Dalrymple (1987) has shown that about 23% of organisations use exponential smoothing at some point in their forecasting procedures.

The choice of the sophisticated technique was not as clear cut. The Bayesian F method scored the lowest MAPE for the forecast horizon considered. However, it is not a widely accepted method, and Fildes and Lusk (1984) showed that only four percent of responding forecasting professionals in the U.S. felt that it was the preferred method for short term forecasts based on the time series values alone. On the other hand, Box Jenkins (BJ) [11] is one of the best recognised sophisticated forecasting techniques. It is an 'auto-regressive/ integrated/ moving average' (ARIMA)

---

[11] For a description of these and other forecasting methods see Makridakis, Wheelwright, and McGee (1983)

method, and as such is based on a general model of time series[12]. BJ requires substantial computing and knowledge resources:

> "In the case of one methodology, Box-Jenkins, many organizations that had tried the method no longer used it simply because it was too complex" Makridakis, Wheelwright, and Mcgee (1983, p786)

Despite this, the results of Fildes and Lusk (1984) show that 34% of U.S. professional forecasters rank this the best method for short term forecasting. The strong theoretical appeal of the BJ method has resulted in its inclusion in many studies of accuracy of forecasting methods. For that reason it was felt that it should be selected as the "sophisticated" method in this study.

3.3.2 DESCRIPTION OF THE DISCRIMINANT ANALYSIS STUDY

The methodology used comprised three steps:

1) POST-HOC CLASSIFICATION: a method was determined for post hoc classification of the time series according to the forecast method which was more accurate in pairwise comparison. The pairwise comparisons examined were GRAPH vs. DSE, and GRAPH vs. BJ.

2) METRICS DEVELOPMENT: metrics to describe time series characteristics were evolved. The metrics were based on the classical decomposition of the time series into trend, seasonal, and random sequences.

---

[12] See Makridakis, Wheelwright, and Mcgee (1983)

3) METRICS EVALUATION: the candidate metrics were evaluated to determine their ability to predict the classification determined in the first step.

There follows a more detailed description of each phase.

3.3.2.1 POST HOC CLASSIFICATION

As previously mentioned, the discrimination process could be confounded by classification on the basis of point estimates within an area of overlap of the population distributions. That is, if the expected errors of using two techniques for a particular time series fell in the same range, then the chance observation that one technique out performed the other in a single trial would not reflect any underlying process. In that case any attempt to distinguish between the methods would be fruitless. Inclusion of a number of such observations in a discriminant analysis would similarly cause difficulty. Ideally the classification would be carried out in the light of the population distributions, however, the data to hand did not permit the determination of the standard error of each forecasting processes for each time series. It was therefore necessary to adopt an approach that assumed that such standard errors were a fixed ratio of the means of the processes. Time series that failed to meet a "hurdle" percentage difference between MAPEs were excluded from the classification process.

For the purposes of this study a hurdle rate of 50% of the lower error rate was applied. Several hurdle rates between 10% and 90% were tried and figure 3.1 displays the percentage of cases correctly identified, and the significance of the discriminant function [13], for each hurdle rate. The highest percentage of cases correctly classified, and the most significant function, occurs at the 50% hurdle. Although there was a very good performance at the 80% and 90% hurdle rates, the number of cases in the analysis was lower than at the 50% hurdle.

[13] The significance was calculated by a transform of the Wilks' lambda to a variable with an approximate chi-square distribution, see Nie et al (1975).

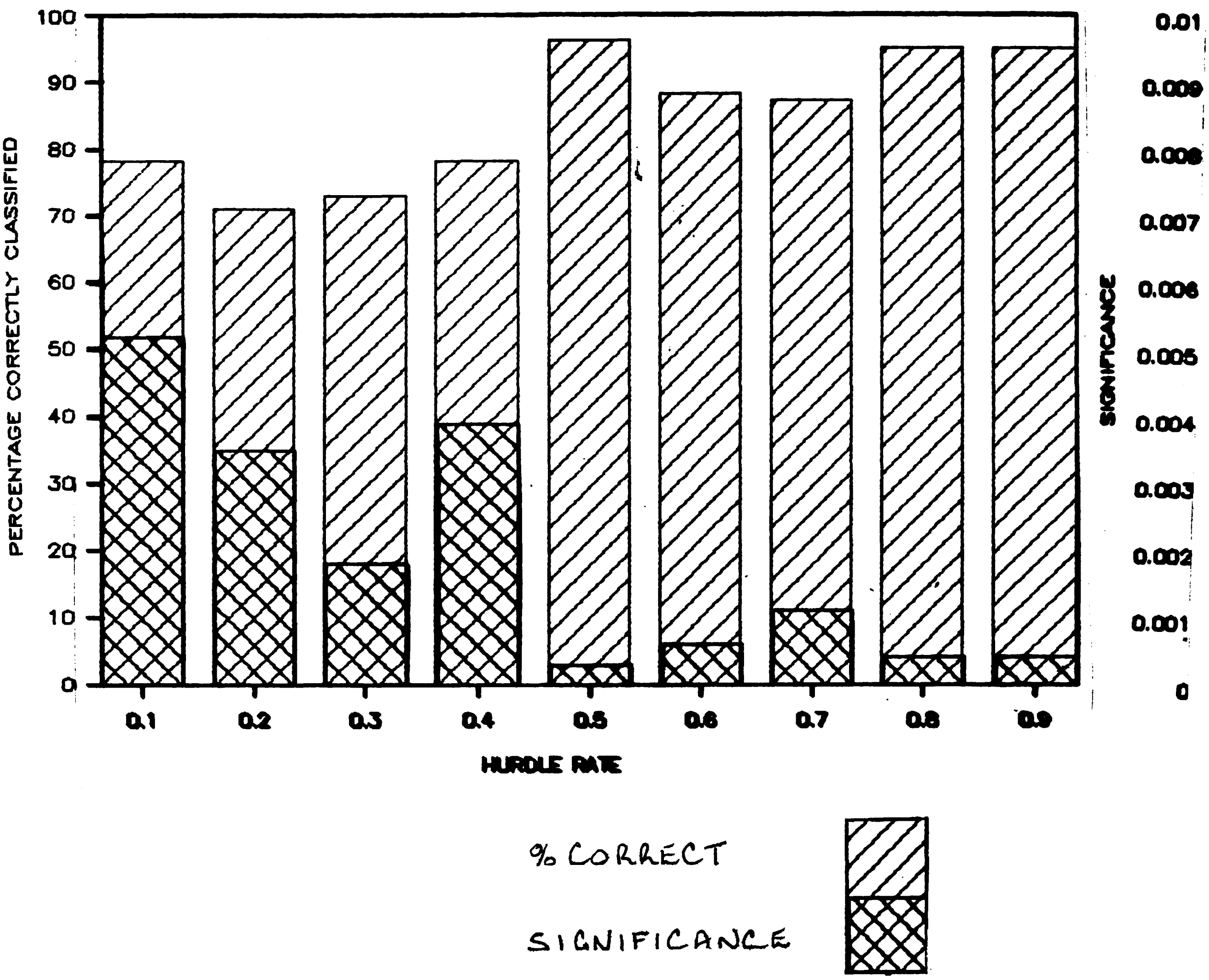


Figure 3.1 Cases Correctly Identified at Various Hurdle Rates

For the two pairwise classifications considered, a time series for which the difference in MAPEs failed to meet the hurdle rate was categorised as unclassed. The remaining time series were categorised according to which method was the most accurate, that is as GRAPH best (CL-J) or either of DSE best (CL-D) and BJ best (CL-B). Table 3.4 shows the class memberships for the cases considered when a hurdle of 50% is applied.

	HORIZON 1 to 6	7 to 12
CL-J	10	14
CL-D	17	27
UNCL	41	27
CL-J	7	10
CL-B	15	15
UNCL	46	43

Table 3.4 Class Membership with a 50% Hurdle.

3.3.2.2 DEVELOPMENT OF METRICS

The approach to the development of metrics was based upon classical decomposition theory [14] which considers the time series to be a function of cycle, seasonal, trend and noise factors. The detection of cyclic factors from the time series values is difficult, and impossible in the short term where it is confounded with trend and seasonality. Thus, aspects of seasonality, trend, and noise in the time series were examined. Approximately 20 candidate metrics were developed, many of them in more than one form [15]. Of the metrics developed, nine remained after preliminary evaluation [16] that considered:

a) the theoretical validity of the metric,

---

[14] See Makridakis, Wheelwright, and Mcgee (1983)

[15] For instance, metrics involving the use of absolute values were considered also with squared values.

[16] The evaluation was largely subjective, for instance, there was a high correlation between the Trend and Noise metrics described later, but it was considered reasonable to retain both since they potentially reflected different aspects of the time series.

b) the potential as a discriminator, based on pilot runs, and

c) the duplication of the metric, considering correlations, and the theoretical basis of the metrics.

In addition, the analysis considered "compound" metrics derived from the nine main metrics. This is discussed more fully in Appendix 3A where a detailed description of each metric may be found. Each metric is identified by a mnemonic, and they are clustered according to their characteristics into:

a) seasonality metrics,

b) trend metrics,

c) noise metrics, and

d) compound metrics.

The four metrics that were found to give significant results are illustrated below.

"SEAS"

The "seas" metric was computed from "moving average" seasonal factors. In Figure 3.2 below, the seasonal factors are plotted about the mean of "1". Seas was computed as the absolute sum of the twelve differences DSi.

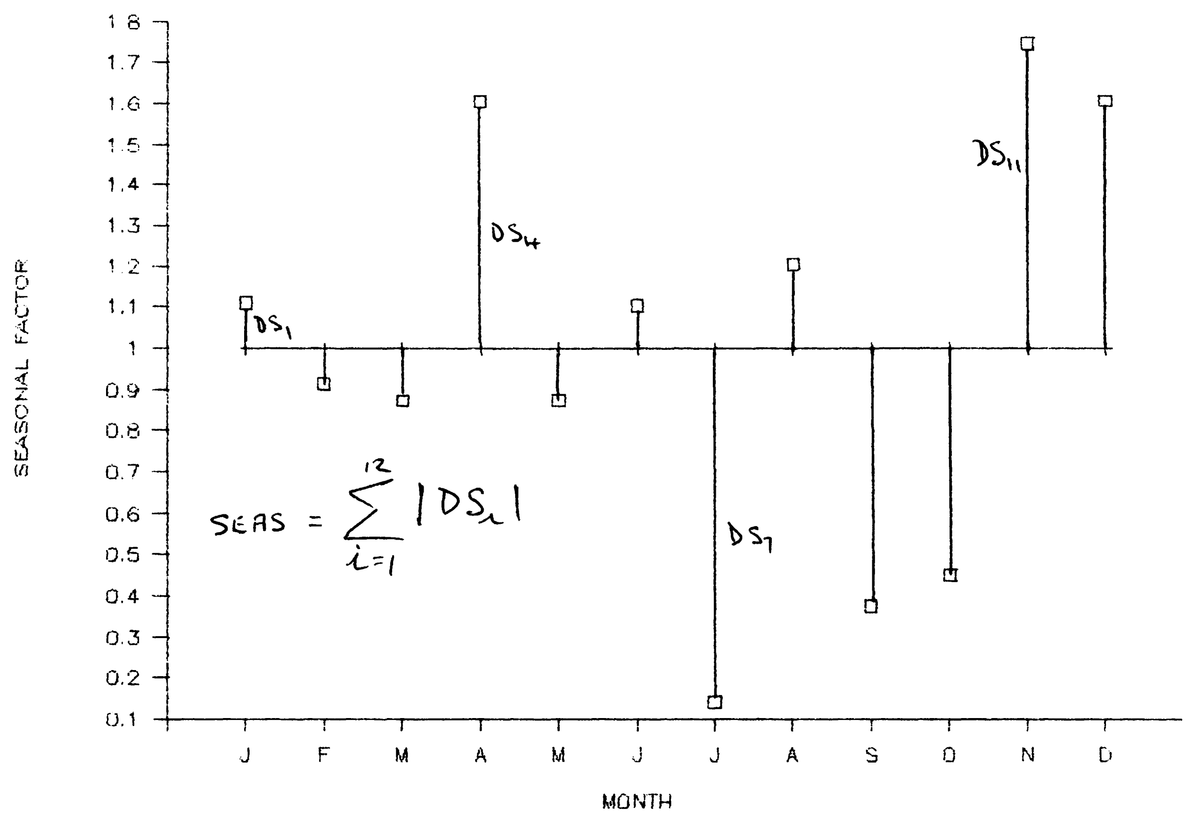


Figure 3.2 Derivation of the SEAS metric

"SYN"

This metric reflects the stability of the seasonal signal. In figure 3.3 below the raw time series data is plotted for the months of January through March for adjacent years. The figure shows that the January to February movement in the first year plotted is of the opposite sign to that of the following year.

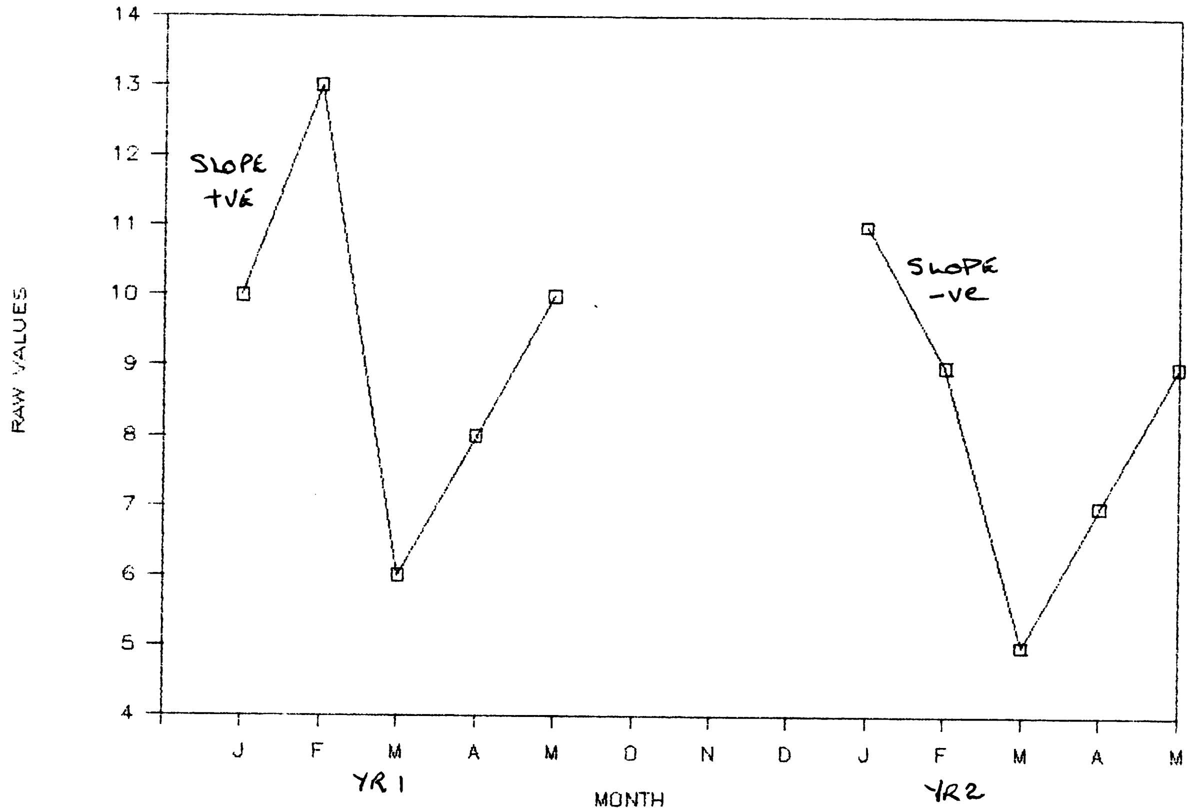


Figure 3.3 Derivation of the SYN metric

If all the January to February movements were in the same direction then the seasonal pattern would be very stable. Since, in the illustration, they are not, the metric would be incremented by an amount equal to 1/N, where N is the total number of pairwise comparisons possible. The February to March movements have the same sign, and therefore the metric would not be incremented for that pair of months, for that pair of years. The syn metric considers all the pairs of months, for all pairs of adjacent years

"NOISE"

The noise metric represents the dispersion of the data about the "signal" in the final 24 observations of the time series. The data was deseasonalised, using moving average seasonal factors, to reduce the impact of the seasonal signal on the metric. Then the absolute deviations (shown as DN<u>I</u> in figure 3.4 below) of the transformed series from a regression line were summed. To remove the effect of scale, the sum generated was divided by the value of period 12 of the regression function, and the metric expressed as an average monthly figure.

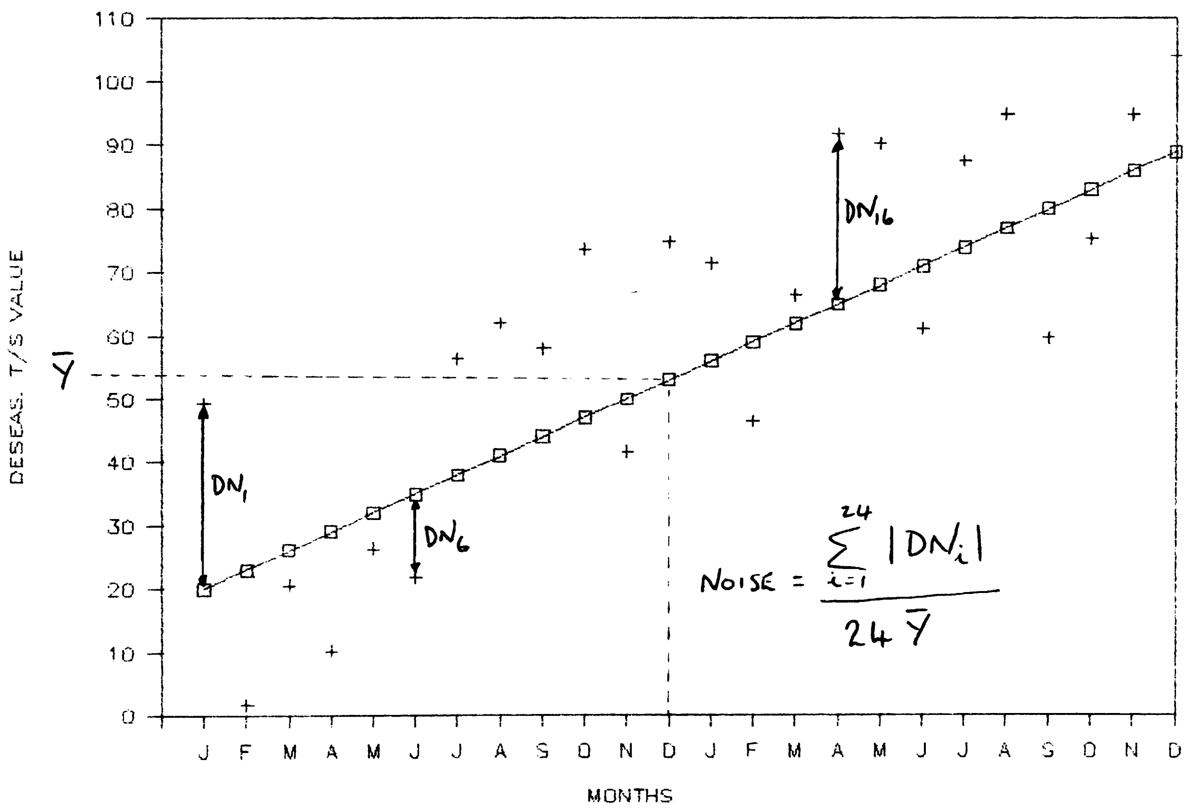


Figure 3.4 Derivation of the NOISE metric

"AUTORSQ"

The stability of the series was addressed by this metric. It was computed as the coefficient of determination of the lag 1 autoregression function fitted to the deseasonalised final 2 years data.

3.3.2.3 EVALUATION OF THE METRICS

Linear discriminant analysis (Nie et al 1975) was used to evaluate the ability of the metrics to discriminate between the classifications of the time series. As previously described and shown in table 3.5 below, the data was analysed for two forecast horizons for each of two pairwise comparisons.

	Forecast Horizons	
	Months 1-6	Months 7-12
Methods Compared		
GRAPH and DSE	GR/DSE_{1-6} (10/17)	GR/DSE_{7-12} (14/19)
GRAPH and BJ	GR/BJ_{1-6} (7/15)	GR/BJ_{7-12} (10/15)

Table 3.5 Cases Examined, with the Number of Series in Each

For each of the cases examined here stepwise linear discriminant analysis was used initially to determine which of the candidate metrics were likely to be useful. This was followed by direct linear discriminant analysis using those metrics. It was determined to evaluate the validity and stability of the discriminant function by performing repeated analysis with each time series being held out in turn, the so called "Jackknife" method. This permitted the evaluation of the function using Lachenbruch's U-method (Lachenbruch 1975) for the rate of mis-classification and the Jackknife statistic for the significance of the discriminant function coefficients (Mosteller & Tukey 1977). This gave an indication of the stability of the discriminant function.

3.3.3 RESULTS OF THE DISCRIMINANT ANALYSIS STUDY

Each case described above was analysed in the same manner. The analysis of the GR/DSE1-6 case is fully described below, followed by summary details of the overall results.

3.3.3.1 GR/DSE1-6

The stepwise analysis of the GR/DSE1-6 case with a 50% hurdle produced five metrics that appeared to act as discriminators. Table 3.6 shows the discriminators in order of entry into the discriminant function with their standardised and non-standardised coefficients [17]. The discriminant function evaluated at the GRAPH group centroid is 1.61, and at the DSE group centroid is -0.95.

[17] The size of the standardised coefficient is indicative of the relative contribution the metric made to the inter-group difference. The non-standardised coefficient is needed to construct an operational rule for discrimination.

discriminating variable	coefficient stand'ised	non-stand'ised
autrsq	1.736	6.516
seas/sesdev	1.736	0.008
noise	1.569	0.855
seas/syn	-1.583	-2.180
seas/noise	-1.259	-1.566
constant		-4.479

Table 3.6 GR/DSE$_{1-6}$ Discriminating Metric Coefficients

The Jackknife statistic provides a relatively unbiased test for the significance of the coefficients of the discriminating variables. Table 3.7 shows the jackknifed discriminant coefficients obtained by jackknifing with a "hold-out" of one case at a time. Included in the table is the significance of the coefficient, that is, the likelihood that the coefficient does equal zero.

	autorsq	seas/ sesdev	noise	seas/ syn	seas/ noise	const
coef	5.055	0.006	0.423	-1.738	0.197	-4.296
std error	0.984	0.004	0.252	0.933	1.786	1.150
signif'ce	0.00	n/s	0.1	0.08	n/s	0.00

Table 3.7 GR/DSE$_{1-6}$ Jackknifed coefficients

The Lachenbruch U method is a relatively less biased estimator of the population effect than the percentage of cases correctly classified in the whole sample. The method is based on the classification of the cases held out. Table 3.8 reports the accuracy of classification using that method.

	predicted group GRAPH	DSE
actual group		
GRAPH	8 80%	2 20%
DSE	5 29%	12 71%

Table 3.8 GR/DSE_{1-6} Classification Results

Overall, 74% of cases were correctly classified by the discriminant function. Interpretation of that result must include the consideration of the distribution of the classes, especially when the class sizes are not equal [18].

3.3.3.1.1 DECISION RULE FOR GR/DSE1-6

The classification of a time series may be determined from the results of applying the values of the discriminating variables to the Linear Discriminant Function for each case. This classification function is derived from the pooled within groups covariance matrix. The functions derived were:

GRAPH1-6 class:

12.69 * autorsq + 0.06 * seas/syn + 0.82 * noise - 7.17

DSE1-6 class:

6.14 * autorsq + 1.20 * seas/syn + 0.51 * noise - 2.03

---

[18] The prior probability of a GRAPH case in the sample is approximately 37%

The time series is assigned to the class that has the higher score from the Fisher functions given above.

3.3.3.2 SUMMARY OF THE RESULTS FROM THE FOUR CASES

The stepwise linear discriminant analysis used to select the candidate metrics for jackknifing provided an indication of the characteristics affecting accuracy for each method. Although the outcome of the jackknifing exercise provides more specific information on the likely "population" effects, the initial results hold interest for the further development of metrics. The fact that a metric that was useful at the stepwise, whole sample level was excluded by jackknifing shows that its effect was very "patchy". Future analysis might resolve this problem, leading to improvement in the achievable discrimination. The standardised coefficients for the metrics entering the stepwise discriminant analysis are shown in table 3.9 below with those subsequently eliminated by jackknifing in parentheses.

discriminating variable	standardised coefficients			
	GR/DSE_{1-6}	GR/DSE_{7-12}	GR/BJ_{1-6}	GR/BJ_{7-12}
autrsq	1.736	-	-	1.000
noise	1.569	-	1.153	-
mag	-	(1.009)	-	-
seas/sesdev	(1.736)	-	-	-
seas/syn	-1.583	-	(-0.914)	-
seas/noise	(-1.259)	-	(0.766)	-
noise*trend	-	(-0.766)	-	-

Table 3.9 Discriminating Metrics

In compiling the above table, metrics were considered as "excluded" if they failed to achieve a significance of 0.1 or better. This is a low level for acceptance by the standards used in social science hypothesis testing. In this instance it is justified in terms of the objectives of the study. A 90% chance of a metric providing a reliable means of choosing a forecasting method would be an acceptable base for a commercial decision. The actual jackknifed coefficients, and their levels of significance are provided in table 3.10. The equivalent Fisher functions for each case are given in appendix 3B.

discriminating variable	jackknifed coefficients			
	GR/DSE_{1-6}	GR/DSE_{7-12}	GR/BJ_{1-6}	GR/BJ_{7-12}
autrsq	5.055***	-	-	3.063***
noise	0.423*	-	0.789**	-
seas/syn	-1.738*	-	-	-
constant	-4.296***	-	-1.562**	-1.506***

(*** signifies sig at .01, ** at .05, * at .1)

Table 3.10 Discriminating Metrics

The table shows that, of the metrics developed, only three had any effect. Two, "autrsq" and "noise", were significant in two cases, and their operation was the same in each case.

The Lachenbruch tests for the effectiveness of the derived discriminant functions are summarised in table 3.11. The results are given in terms of the percentage of the cases used in the discriminant analysis (see table 3.5 above for the numbers in each case) that were correctly classified in the "hold out" run.

	PERCENTAGE CORRECTLY CLASSIFIED			
ACTUAL GROUP	GR/DSE_{1-6}	GR/DSE_{7-12}	GR/BJ_{1-6}	GR/BJ_{7-12}
GRAPH	80	14	57	70
DSE	71	85	-	-
BJ	-	-	100	73
OVERALL	74	61	86	72

Table 3.11 Lachenbruch Test of Cases Correctly Classified

From the table it can be seen that the classification in the GR/DSE1-6 case was excellent, especially considering the prior probability of a GRAPH series was only 0.37. A similar result holds in the GR/BJ1-6 case, though the identification of GRAPH series was not nearly so good (the prior probability was 0.32). As expected from the discussion above, the GR/DSE7-12 case did not reveal good results.

3.3.3.3 EFFECT OF THE DECISION RULES ON FORECAST ERROR

Of the four cases examined in this study, one failed to reveal a reliable or statistically significant discrimination function. That case, GR/DSE7-12, has not been included in the evaluation of the effect of the discriminant functions upon forecast error.

The remaining three cases, GR/DSE1-6, GR/BJ1-6, and GR/BJ7-12, have been evaluated against the full database of 68 monthly time series from the "M-Competition". Thus, this evaluation includes the large number of cases that were excluded from the discriminant analysis because of the small differences in MAPE between the forecasting methods.

In terms of correctly identifying series that should be judgmentally forecast the decision rules were not highly successful. Where the prediction was "GRAPH" it was correct 59% of the time in the GR/DSE1-6 case, 67% of the time in the GR/BJ1-6 case, and 64% of the time in the GR/BJ7-12 case. The type II error, failing to identify a GRAPH series, is somewhat worse.

In the GR/DSE1-6 case 32% of GRAPH series were identified, 39% were identified in the GR/BJ1-6 case, and 30% in the GR/BJ7-12 case.

Because of the great variation in relative accuracy between the methods, there is not uniform benefit to be gained across all series by correctly identifying the most accurate forecasting technique. For instance, there are series for which the difference in MAPE between the forecasting methods are small, and others for which the difference is large. Given a choice between two rules, each of which correctly identified half the series, the preferred rule would be that which correctly identified the greater number of the "large difference" series.

The commercial impact of the decision rule is more properly evaluated in terms of the expected improvement in the MAPE. This evaluation is carried out using a paired one tailed [17] t-test and is reported in table 3.12. The results of the comparison with the statistical methods must be conservative because GRAPH series are in the minority, therefore the majority of the "pairs" in the test have the MAPE of the statistical method in both samples. For a significant difference to be obtained the effect of

---

[17] the one tailed test is justified on the basis that application of the rule derived from the discriminant function is not likely to cause the MAPE to increase

the GRAPH series on the means and standard errors of the MAPE's would have to be large.

CASE	MAPE FOR				IMPROVEMENT	
	GRAPH	DSE	BJ	"RULE"	%	SIGNIF
GR/DSE_{1-6}	11.6	-	-	9.8	15.5	0.01
	-	11.0	-	9.8	10.9	0.08
GR/BJ_{1-6}	11.6	-	-	10.6	8.6	0.02
	-	-	11.3	10.6	6.2	0.04
GR/BJ_{7-12}	16.5	-	-	13.9	15.8	0.00
	-	-	16.3	13.9	14.7	0.12

Table 3.12 Accuracy Over All 68 Series

Table 3.13 reports an evaluation of the relative accuracies of the methods for the series that the decision rule indicates that the judgmental forecasting method be used. Again a one tailed paired t-test is used for the evaluation.

method	forecast horizon	mean MAPE	std dev	diff from stat method	signf of diff
DSE	1-6	13.5	17.9	-	-
GRAPH	1-6	8.7	7.5	4.8	0.00
(n=17)					
BJ	1-6	19.3	10.5	-	-
GRAPH	1-6	16.4	8.8	2.9	0.00
(n=18)					
BJ	7-12	21.0	31.6	-	-
GRAPH	7-12	13.5	14.2	7.5	0.12
(n=22)					

Table 3.13 Accuracy for "GRAPH" Series

3.3.4 DISCUSSION OF THE RESULTS

3.3.4.1 THE DECISION RULES

For the short time horizon, months 1-6, the results are encouraging. By means of the jackknife statistics it was possible to develop stable discriminant functions that permit an a priori classification between two methods with between 74% and 86% accuracy. This success was not repeated for the longer horizon, months 7-12, for although the jackknife method revealed a significant metric in the GR/BJ7-12 case this did not result in a decrease in MAPE. No significant metric was found in the GR/DSE7-12 case.

Examination of the time series data revealed that there was low correlation between the errors of the two horizons; 0.64 (p=0.000) for GRAPH, 0.57 (p=0.000) for DSE, and 0.45 (p=0.000) for BJ. The low correlation between the two time horizons for DSE in particular indicates that there may be a failure of the assumption of constancy in a number of the time series. The implication arises because the forecasting method provides a single smoothed forecast for all future months, which is adjusted for the seasonal pattern in the modelling data. It would therefore be expected that, if the basic structure of the time series remained constant, the error rates over time would be highly correlated.

For months 1-6, use of the simple decision rules developed for choosing when to use an unsophisticated judgmental forecasting method resulted in an improvement in accuracy. For the GR/DSE1-6 case the discriminant function performed well, both in terms of the number of cases correctly classified, and particularly in terms of the reduction in error of forecast. Although the decision rule did not capture all the information related to relative accuracy, it provided a clear, and statistically significant, advantage. The average MAPE over the horizon (9.8) was lower than any method reported by Makridakis et al (1982) for these time series, and is significantly better (p=0.08) than the MAPE for DSE (11.0).

In the GR/BJ1-6 case the discriminant function was effective, despite the failure to correctly identify a large number of GRAPH series. The error rate achieved was an MAPE of 10.6. This error is higher than the result obtained in the GR/DSE1-6 case, but is better than any single method reported by Makridakis et al (1982). The standard deviation of the forecast error was lower than that of either of the component methods.

Apart from the improvement in accuracy, the adoption of the decision rule in both the GR/DSE1-6 and the GR/BJ1-6 cases resulted in a standard deviation of forecast error that was lower than that of the candidate methods. This is an important factor in commercial decision making, where the avoidance of the consequences of very large forecast errors might be a major consideration[20].

3.3.4.2 FORECASTING AID DESIGN CONSIDERATIONS

The discriminant functions identified metrics related to the relative performance of judgment in extrapolation. The interpretation of the results, however, is somewhat speculative. <u>There is no justification to assume causation in the relationships found</u>. However, it is necessary to speculate on the results if progress is to be made towards understanding the processes involved. Three potentially useful metrics were revealed by the jackknifed discriminant functions, and these are discussed below.

---

[20] Consider a scenario in which a sales forecast is to be made for a product with a very short shelf life, and with high consumer loyalty.

"Autorsq" showed as a significant metric for the GR/DSE1-6 and the GR/BJ7-12 cases. The metric is the coefficient of determination of the "lag one" autoregression function fitted to the prior 24 months of deseasonalised data. A high value for this metric is an indication of a relatively stable series, and in both cases was an indication that GRAPH should be adopted. The following reasons might be advanced for this result:

a) GRAPH may be unable to cope with relatively unstable series.

and, as far as the GR/DSE1-6 case is concerned, [21]

b) GRAPH accommodates an autoregressive process in the data which DSE does not address.

c) There might be some trend in the data which influences autrsq. DSE does not have a trend adjustment and GRAPH does.

Only the first of these has implications for the design of a forecasting decision aid. The others would indicate the need to add capabilities to the DSE method.

The ratio "seas/syn" was a significantly useful metric in the GR/DSE1-6 case. It is a "signal to noise" measure, high values of which indicated the use of DSE. This implies that the GRAPH method has

[21] Neither of the following explanations apply to BJ which has the capability to model autoregressive series, and to handle trend.

difficulty in handling a strong seasonal signal in a low noise series but is better able to discern a subtle seasonal signal in the presence of noise. This result runs somewhat contrary to expectations [22], and is further examined in chapter 5.

The final significant metric was "noise", which was based on the absolute deviations of the series from the regression line. This metric was significant in both the GR/DSE1-6 and the GR/BJ1-6 cases, a high value for this metric indicated the use of GRAPH in each case. The fact that the effect is consistent over the two cases reinforces the implication that judgmental methods cope well with instability. This may be reconcilable with the results of Makridakis and Hibon (1979) who felt that the presence of high noise in a series could lead to over fitting with sophisticated methods such as BJ.

The conclusion that judgment handles noisy series well is contrary to one of the conclusions with respect to "autrsq". It is also contrary to the literature reviewed in chapter 4 concerning the effect of randomness on human judgment.

Further analysis is required to determine the actual effect of those characteristics, and whether a decision aid might improve the forecast. The factors

[22] See for instance an implication in Eggleton (1976) that the human judge would have difficulty in identifying an alternating pattern in the presence of noise.

identified have been considered in the development of the experimental decision aid reported in chapter 5.

3.4 CONCLUSIONS

The results obtained in this study may be partially reconciled with the findings of Makridakis et al (1982) discussed in section 2.2.1:

* Their comment that DSE was suited to monthly series cannot be examined here, since only monthly series were considered.

* Their rule of thumb that simple methods are suited to micro series and sophisticated methods perform better for macro series is not directly confirmed. However, macro series would, in general, exhibit lower noise than micro series and, in the GR/BJ1-6 case, be better suited to BJ. As far as the selection between DSE and GRAPH is concerned, it is not possible to apply the rule of thumb. It is not clear what is a sophisticated or a simple method, and it is equally unclear whether Makridakis et al (1982) were considering absolute classes or relative characteristics.

The implications of the results obtained for the design of a forecasting decision aid are discussed below, where each of the three discriminating metrics is considered. It was determined that the decision aid should address the issue of closing the difference between the judgmental process and the deseasonalised single exponential smoothing method. Those two processes were seen as similarly simple, and probably more direct alternatives than Box Jenkins and judgment.

3.4.1 AUTRSQ

The advantage to the judgmental process of high autocorrelation of lag one might have reflected either an unsuspected autoregressive aspect of judgmental extrapolation, or a"trend" related aspect. Although trend can give rise to autoregressive characteristics it does not fully explain the result achieved. There is a significant correlation (p=0.001) between the Trend metric and the Autrsq metric, but the value of the coefficient is low, at less than 0.4.

In the design of the decision aid it was felt that any advantage currently held by the judgmental method should not be jeopardised. Thus, the design of the display in the trend identification module would not significantly differ from that of the original hard copy plot. This matter is discussed more fully in chapter 4, where the human information processing literature is reviewed.

The other aspect of the decision aid design affected by this particular characteristic is the extrapolation, as opposed to the identification of any trend and seasonal models. In the case of the decision aid that task was designed to be carried out on the residual series generated by the removal of any trend and seasonal models. The effect of the removal of those models on the autoregressive nature of the series is indeterminate, though as will be discussed in chapter 8 the de-trending in particular could induce autocorellations. The effect of this on the design of the displays for the residual series extrapolation functions was to encourage the use of an ordinary plot of the residual series, again similar to that of the hard copy plot. That is, for those processes the judge should have a standard plot of all the decision values. As a first pass it was decided that summary cues would be omitted in case these masked the processing of the autocorrellation cues.

3.4.2 SEAS/SYN

The results here were rather enigmatic. The implication was that judgment performed well in response to a subtle signal to noise cue, and not so well in response to a stronger cue. It was felt that possibly the precision of the automatic process was an advantage in the case of the more stable and obvious signal. The automatic process in question was the ratio to centred moving average. That such a process could have an advantage in conditions of high seasonality gave rise to the implication that judges either:

1) fail to process all the cues adequately, or,

2) make computational assumptions or errors that lead to a result differing from the mean, or,

3) fail to express the result accurately.

Figure 3.5 below illustrates a series with high seasonality presented in the form used in the hard copy plot.

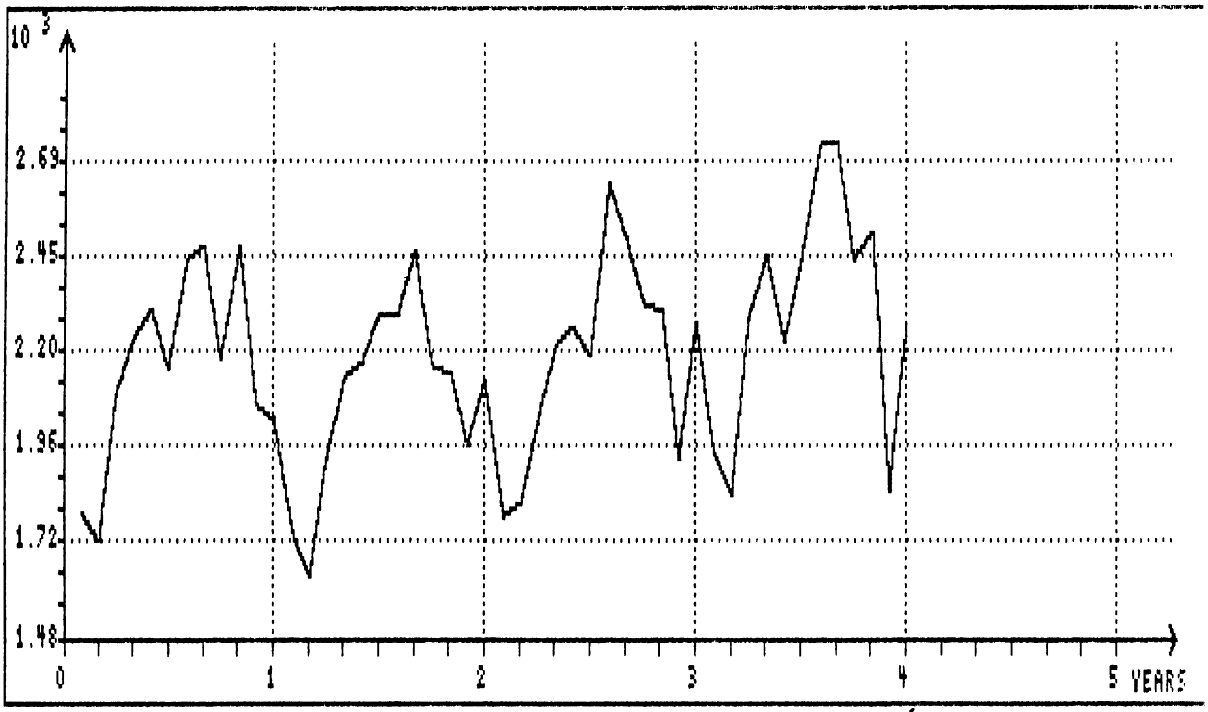


Figure 3.5 Hard Copy Plot of a High Seasonality Series

From such a display the seasonal pattern in the first of the four years, say, may appear remote to the judge, who may anchor on the most recent year. Scale errors and transcription errors could also arise because the drawing of the pattern is removed from the cues.

Finally, the estimation of the mean (or weighted average) seasonal factor from the four cues might be faulty, even if the judge set out to make such an estimate (category 2)).

It was determined to present the cue data with the years plotted in "parallel" rather than serially. Figure 3.6 shows such a display, the cue data is displayed in the lower half of the screen. The upper half of the screen contains a display of the modelled pattern, as it is input.

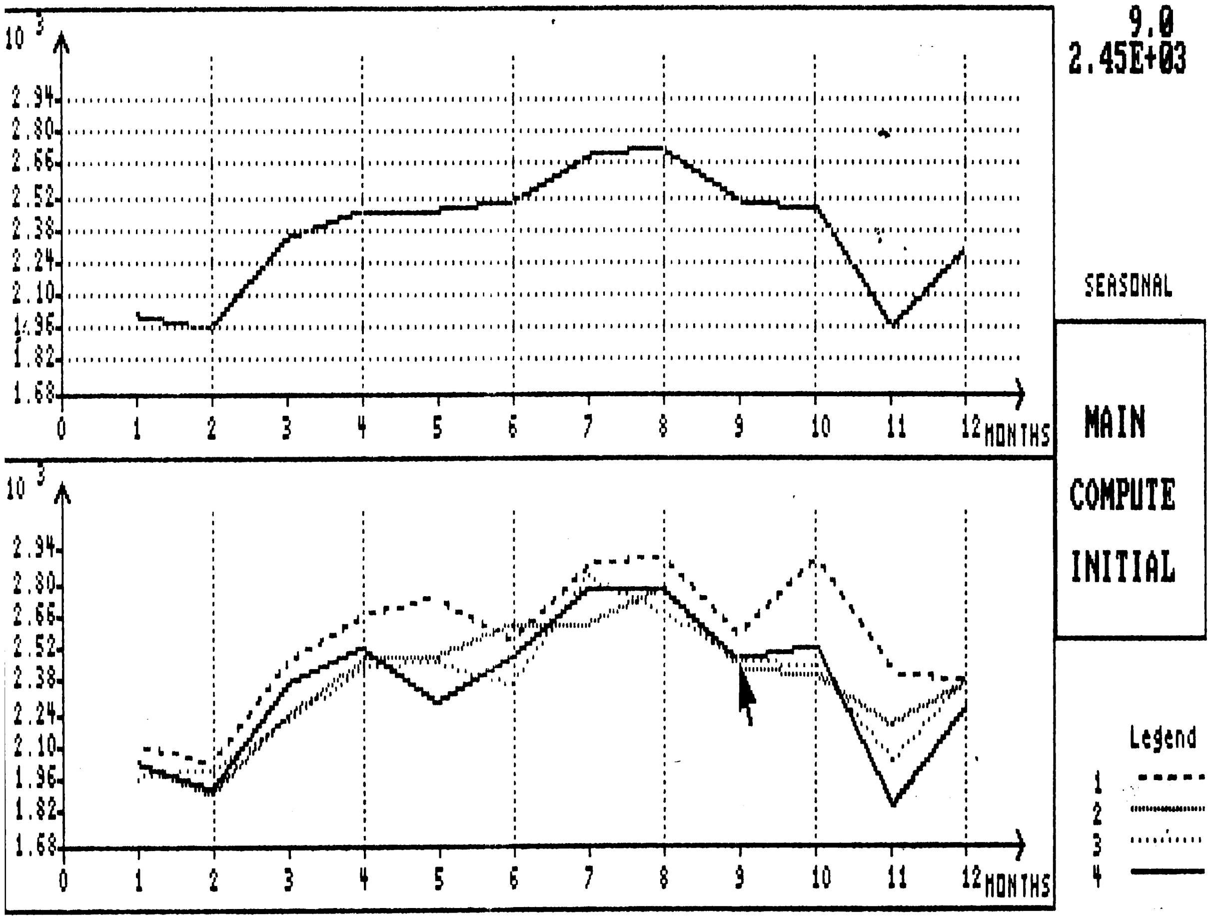


Figure 3.6 Seasonal Display Screen

This display was designed to assist in the detection of subtle seasonal signals, and more importantly, with appropriate data entry design to alleviate the possible difficulties described above. The four cues for each month are in a single column and coded for identification, reducing the likelihood that one or more be given too little (or too much) attention. Estimation of the mean position is more directly assisted, and with a design that allows the position to be input in the same column on the screen the chance of a transcription error occurring is diminished.

The task of seasonal pattern identification was anticipated to be one of the more challenging sub-tasks in the extrapolation, and a major determinant of overall accuracy. Therefore the data capture method was designed to permit:

* Simple iteration between identification and review processes.

* Model building by input of parameters either on the cue display or on the model display.

3.4.3 NOISE

The advantage of high noise cues to judgmental processes was unexpected. Unfortunately it was not clear that the advantage arose only from the noise characteristics. The noise metric was correlated at 0.7 with the trend metric. It was therefore possible that at least part of the advantage stemmed from the effect of trend characteristics. This would explain why the trend metric did not enter the analysis. It was expected that judgment would have an advantage in the presence of trend, because deseasonalised single exponential smoothing has no capability to model trend. This matter will be discussed later in chapter 8.

Initially it was considered that whatever was the cause of the advantage detected, it arose from the use of an ordinary serial plot of the data. Therefore, as a first pass at the design of the decision aid it was determined to permit the forecaster to extrapolate the residual series without the provision of additional cues that might mask the implied advantage.

3.5 LIMITATIONS

The results obtained here suffer from two possible shortcomings. The first concerns the representativeness of the time series examined, and the second concerns the possibility that significant characteristics of time series have been omitted from the analysis.

The first of those problems is a threat to the external validity, or the capacity to generalise the study. There is no evidence that the 68 series are representative of time series in general, or with respect to the mix of characteristics within the sample. Before the results could be generalised it would be necessary to replicate the study with different time series.

The second limitation affects the internal validity of the study. There is no guarantee that the metrics considered here are the most significant metrics. It is also possible that metrics identified in the study are not causally related to the relative accuracy of the forecasting methods, but are correlated to other metrics that are so causally related.

3.6 REFERENCES

Armstrong, J.S., "Forecasting by extrapolation: conclusions from 25 years of research", *Interfaces* 14, (1984), 52-66.

Armstrong, J.S., *Long-range Forecasting from Crystal Ball to Computer* 2nd edn. Wiley, New York, (1985).

Dalrymple, D.J., "Sales forecasting practices: Results from a United States Survey", *International Journal of Forecasting*, forthcoming (1986).

Eggleton I.R.C. "Patterns, prototypes, and predictions: an exploratory study", Selected Studies on Human Information Processing in Accounting, *Supplement to Journal of Accounting Research* 14: (1976), 68-131.

Eisenbeis, R.A. "Pitfalls in the application of discriminant analysis in business, finance and economics", *Journal of Finance*, (1977) 32 : 875-897.

Fildes, R., and Lusk, E.J., "The choice of a forecasting model", *Omega*, 12, (1984), 427-435.

Fildes, R., "Gains through univariate forecasting model selection", *Fifth International Forecasting Symposium*, Montreal, (1985).

Hays, W.L., *Statistics*, Holt Saunders, New York, (1981).

Lachenbruch, P.A. *Discriminant Analysis*, Hafner, New York, (1975).

Lawrence M.J., "An exploration of some practical issues in the use of quantitative forecasting models", *Journal of Forecasting,* 2 (1983), 169-179.

Lawrence, M.J., Edmundson, R.H. and O'Connor, M.J. "An examination of the accuracy of judgmental extrapolation of time series", *International Journal of Forecasting* 1 (1985), 25-35.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. "The accuracy of extrapolative (time series) methods : results of a forecasting competition", *Journal of Forecasting*, Vol 1, no 2, (1982).

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. *The Forecasting Accuracy of Major Time Series Methods*, Wiley, Chichester, (1984).

Makridakis, S., and Hibon, M., "Accuracy of forecasting: an empirical investigation", *Journal of the Royal Statistical Society, Series A*, 142, (1979), pp97-145

Makridakis, S., Wheelwright, S.S., and McGee, V.E., *Forecasting: Methods and Applications*, 2nd ed., Wiley, New York, (1983).

Mosteller, F. and Tukey, J.W. *Data Analysis and Regression*, Addison-Wesley, Reading, Mass., (1977).

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., Bent, D.H *Statistical Package for the Social Sciences*, McGraw Hill, New York, (1975).

Schnaars, S.P., "Situational factors affecting forecast accuracy", *Journal of Marketing Research*, vol xxi, (1984), 290-297.

APPENDIX 3A

A1 SEASONALITY METRICS

A1.1 METRIC: SEAS

Seasonal factors for each time series had been computed by the authors of the M-Competition [23]. The seasonal factors were used to compute a seasonality metric by summing the absolute differences of the factors from 1. This produces a simple measure of the magnitude of the seasonality of the series. It does not reflect the variation in the seasonal and it fails to distinguish between series that exhibit high seasonality in a few periods and those that have somewhat less variation but are affected for many periods.

A1.2 METRIC: SESDEV

As an indicator of the magnitude of the variation in the seasonal a crude seasonal index was computed for each period. The index was the fraction of the annual value of the time series contributed by each month. The 'sesdev' metric was then derived from the absolute differences of the monthly indices from the mean of the indices for that month. The metric was expressed as the average of the differences.

---

[23] The method of calculation used was based on the calculation of the ratio of the month to a centered moving average

A1.3 METRIC: SYN

A lack of stability in the seasonal pattern in a time series may be observable by considering the relationship between successive observations, and those of the same period in the previous year. For each pair of adjacent periods taken in chronological order the sign of the difference between them was compared with the sign of the pair for the preceding year, as far as the data extended. If the two signs were not the same the metric was incremented by 1/n, where n was the number of comparisons made in the time series.

A1.4 METRIC: MAG

This metric follows the same form as 'syn' but takes into account the signed magnitude of the differences between successive periods.

A2 TREND METRICS

A2.1 METRIC: TREND

There is a fundamental problem in generating a metric for trend because the metric is likely to be affected by the scale of the series. The approach adopted here was to fit a regression line to the last two years of the deseasonalised history data. The tangent of the angle of the regression line was divided by the value of the regression function for the twelfth period in order to eliminate the effect of scale. This metric will not return the same result for two time series that have similar slopes but are displaced relative to one another by a fixed amount.

A2.2 METRIC: TURN

It was felt that the forecasting methods may be differently affected if a turning point occurred with the final six months of the history (model fit) data. As an attempt to identify such a turning point the sum of the signed deviations of the final six observations from the regression line derived in determining trend was computed and divided by the value of the regression function for the final period to account for scale. The metric was expressed as an average monthly figure. The effect of this is to generate relatively higher values if the final six observations occur more consistently on one side of the regression line.

A3 NOISE METRICS

Three alternative metrics for noise were developed.

A3.1 METRIC: RSQ

As a first pass at a metric for noise the Coefficient of Determination (R-Square) for the regression line developed in the trend metric was used. This measure would probably be confounded by trend.

A3.2 METRIC: NOISE

In an attempt to avoid the confounding effect of trend the absolute deviations from the regression line were summed for the two years covered. This sum was divided by the value of the regression function for the twelfth period to account for the effect of scale and by 24 to reduce the metric to a monthly average.

A3.3 METRIC: AUTORSQ

In addition to a regression line fitted to the deseasonalised data an auto-regression of lag one was generated. The rationale for this lay in the assumption that the graphical judgmental forecasts would be heavily anchored on the final value. An R-Square for the auto-regression function was computed.

In order to permit consideration of any interaction of time series characteristics some compound metrics derived from the product and the quotient of the above metrics were tested. For instance, it was thought possible that the judgmental detection of a weak seasonal signal in the presence of high noise would be difficult. The compound metrics tested were:

A4 COMPOUND METRICS

The construction of the individual metrics described above does not allow for a multiplicative model to be developed in the discriminant function. The function would only take the form of a weighted additive model. There was sufficient indication in the human information processing literature reviewed in chapter 4 to warrant the consideration of the interaction of signal type metrics such as trend and seas with noise metrics. Many compound metrics were evaluated, and the following were selected by pilot studies, as the most promising.

A4.1 TREND BASED COMPOUND METRICS

noise * trend

noise * turn

A4.2 SEAS BASED COMPOUND METRICS

seas / syn

seas / mag

seas / noise

seas / sesdev

APPENDIX 3B

B1 FISHER FUNCTIONS

The Fisher discriminant functions for each of the three effective cases are presented below. In order to determine which method should be used to forecast a time series, the value of the discriminating metrics are used in the pair of functions, and the method chosen that has the higher outcome.

B1.1 GR/DSE1-6

GRAPH FUNCTION:

12.69(autrsq) + 0.06(seas/syn) + 0.82(noise) - 7.17

DSE FUNCTION

6.14(autrsq) + 1.20(seas/syn) + 0.51(noise) - 2.03

B1.2 GR/BJ1-6

GRAPH FUNCTION

1.72(noise) - 3.49

BJ FUNCTION

0.66(noise) - 0.73

B1.3 GR/BJ7-12

GRAPH FUNCTION

6.36(autrsq) - 3.09

BJ FUNCTION

3.37(autrsq) - 1.12

4. REVIEW OF THE HUMAN INFORMATION PROCESSING LITERATURE

4.1 INTRODUCTION

The objective of this chapter is to review the human information processing literature with a view to establishing the context for the design and testing of the forecasting decision aid reported in the later chapters. The issues addressed include aspects of human judgment, task decomposition and the impact of cue form and presentation.

The literature review is presented according to the following structure:

Section 4.2 Reviews the general H.I.P. literature that has, until recently, been the basis for opinions regarding the performance of human judgment, including the accuracy of judgmental forecasting.

Section 4.3 Considers the forecasting specific implications of the literature.

Section 4.4 Looks at the theoretical basis for the two strategies adopted in this dissertation to improve judgmental forecasting, namely the decomposition of the decision process and the provision of graphical data presentation.

4.2 GENERAL H.I.P. LITERATURE

The literature reviewed below points to several possible shortcomings in human judgment that have been reported as affecting outcome performance. The literature has its base in the 1970's. Since that time it has not been uncommon to find the results of the early experiments generalised into untested areas, such as forecasting, with little hesitation. It will be shown that the supposed shortcomings of judgment or intuitive data processing that could adversely affect the accuracy of a forecast are mainly:

* limitations in parameter estimation,
* inconsistent application of model computation rules,
* poor perception of randomness.

These biases are considered in section 4.3 as they relate to forecasting.

4.2.1 INTUITIVE DATA PROCESSING

The literature concerning the experimental study of human rationality is extensive, and the implications drawn from that literature have been common currency. For instance, Slovic, Fischhoff and Lichtenstein (1977) concluded that human judgment fails to take account of the basic principles when dealing with probabilistic tasks, and that people "lack the correct programs for many important judgmental tasks". The adoption of the outcomes of the early experiments is evidenced in the acceptance and quotation of the so called Kahneman and Tversky biases and heuristics, some of which are mentioned below:

REPRESENTATIVENESS: Kahneman and Tversky (1972)

AVAILABILITY: Tversky and Kahneman (1973)

ANCHORING AND ADJUSTMENT: Tversky and Kahneman (1974)

VERIFICATION: Johnson-Laird and Watson (1970)

Hogarth and Makridakis (1979), in a table made more generally available in Hogarth (1980), illustrated the many biases and heuristics, and their meaning. The biases that they reported as operating in human information processing were:

* Inconsistency in the application of a strategy over several cases,
* Conservatism in the face of new data
* Inability to extrapolate exponential growth processes,
* Application of heuristics to reduce mental effort:

  rules of thumb

  anchoring and adjustment

  representativeness

  law of small numbers

  justifiability

  regression bias

  best guess strategy
* Complexity, information overload and stress affect judgment reliability
* Data gathering for comfort rather than improved performance.

Hogarth (1980) did issue a caveat at the end of the chapter containing the table, in the following terms:

> "Despite the somewhat authoritative manner in which this book is written, it should be made clear that we know relatively little about how the human mind works and its influence on behaviour. The research to date has only scratched the surface....the range of conditions under which human judgment has been examined remains small" Hogarth (1980, p 179).

This is illustrative of current concern to limit the findings of the human rationality literature to task types and environments for which they have been shown to hold. It is amusing to conjecture that the ready acceptance of the literature by applied social science writers may in itself be a justification of the research results. Recency, saliency, concreteness, selective perception, and availability biases could all be implicated in any mistaken adoption of the literature outside the acceptable bounds of its applicability.

In the last five years, there has been evidence in the literature of a counter movement. Comments on the limited external validity of the experimental results have generally been received without fuss:

> "We are sympathetic to a number of points made by Cohen: the importance of context in assessing error in judgment..." Einhorn and Hogarth commentary on Cohen (1981, p 334).

In contrast, the attacks (typified by Cohen 1981) on the internal validity and conclusions drawn from the experiments has given rise to lively and robust comment and counter comment. Cohen (1981) classified the results of the human rationality experiments as belonging to at least one of four categories:

* Cognitive illusions,

* Tests of intelligence or education,

* Misapplication of appropriate normative theory, and

* Application of inappropriate normative theory.

This prompted Kahneman (1981) to the following rebuttal in his comment on the paper:

> "Cohen has nothing of substance to offer on these difficult issues, beyond a vague message of faith, charity, and authority." Commentary on Cohen (1981 p 340).

It is not the intention to comment here on the dialogue concerning the external and internal validity of the findings of the experiments on human rationality. It is intended merely to draw from that dialogue the inference that it is a valid pursuit to address the question of human competence in applied fields, including the extrapolative forecasting task. The literature does not point to a conclusive failure to make rational decisions in all task types and settings. For instance, it has been shown that changing the task setting from abstract (Watson 1966) to concrete (Watson and Shapiro 1971) resulted in improved performance.

Whether or not the biases identified in the literature reviewed above exist within the forecasting task there is no indication of any outcome effect on the accuracy of forecast. Christensen-Szalanski (1986) criticise existing research in the field of judgment for its lack of practical significance and commented that:

> "..the information most needed by the practitioner to improve the quality of his or her thinking (is) information that evaluates the effect of these cognitive biases on the quality of the resulting outcome" Christensen-Szalanski (1986, p 399)

Thus it remains to be seen whether the effects of the biases are to be found in judgmental extrapolation of real economic time series, and the extent that those biases are dysfunctional. For instance, anchoring and adjustment might prove to be a successful strategy with highly autocorrelated data, but it is not known whether the level of adjustment would be appropriate.

4.3 FORECASTING H.I.P. LITERATURE

The psychological literature reviewed in section 4.2 above examined the possible reasons for the failure of judgment to outperform decision models. That literature has been adopted by authors such as Makridakis and Hibon (1979) despite the lack of clear evidence that the human judge is a less accurate extrapolator than statistical forecasting methods, and despite the differences between the tasks on which that literature is based and forecasting. Thus a view has been encouraged that the human forecaster is adversely affected by factors such as anchoring, regression bias, representation bias, and the failure to apply proper principles and consider relevant data.

Despite the views of the human as a poor intuitive data processor, there is some evidence in the literature

(Winkler and Makridakis 1983; Moriarty and Adams 1984; Lawrence, Edmundson, and O'Connor 1986) of an interest in using the special attributes of the human judge, to capture information not addressed by the statistical methods, in combinations of forecasts. It remains, however, to determine the extent to which the human judge, when extrapolating economic time series, is prone to the effects of the biases mentioned in the psychological literature. It is also of interest to determine whether any such effects might be mitigated by the support of a decision aid.

4.3.1 INTUITIVE FORECASTING

The perceived shortcomings in human intuitive information processing described above may be categorised as failures in either model building, or in the application of that model. Both types of activity take place in extrapolation, and it is therefore possible that the effects of the biases might be observed in that task setting. Wagenaar and Timmers (1978) described subjective extrapolation as a two stage process in which time series properties are first identified, and then the discovered rules are applied to generate the predicted values. Eggleton (1982) suggested a slightly different process depending upon model identification, parameter estimation and application of the applied rules. The two structures are equivalent in practical terms since the parameter estimation processes depend on the model selected. Consider the estimation of a trend component, for all practical purposes in the short term it is unlikely whether the outcome would be noticeably affected if either a linear model or a curvilinear model were selected as long as the parameter estimates were appropriate. This result flows from the observation of many economic time series, and the absence of series exhibiting such severe curvilinear characteristics that, for forecasts up to 12 months ahead, a linear approximation would be entirely inappropriate.

In model fitting for extrapolative purposes, classical theory identifies a number of sub-tasks. Those relevant to short term extrapolation are:

* Identification of the seasonal model,
* Identification of the trend model, and
* Extrapolation from the residual (noise) process.

Each sub-task is addressed in that sequence in the following sections. There is little literature that directly addresses the issues of judgmental identification of the above models so it is necessary to attempt to draw inferences from peripheral research.

4.3.1.1 SEASONAL MODEL IDENTIFICATION

Eggleton (1976), in reviewing the literature, commented that there was evidence that the human judge would have difficulty separating the random and non-random cues. This was confirmed by his findings that subjects were unable to distinguish between alternating sequences and random sequences in contrived time series data. The link between "alternating sequences" and seasonal patterns is somewhat strained, but it provides the nearest approach available for the consideration of seasonality.

Eggleton's (1976) result is somewhat at odds with the folklore that ascribes good pattern processing capabilities to mankind, and with the psychological and sociological literature he reviewed that illustrates pattern recognition as a survival skill [1]. The results that Eggleton (1976) reported must be placed in context. The decision environment in which he placed his subjects was not reconcilable with reflective forecasting decisions. The data was displayed for very short time periods. Eggleton (1976) was primarily interested in the immediate intuitive response to data in a decision environment requiring "spot" decisions. This experimental design might cause the results to have limited applicability to the forecasting decision as a reflective, cognitive task rather than as a reflex task.

The issue of subjects being confused by noise when attempting to identify seasonality was mentioned briefly by Lawrence, Edmundson, and O'Connor (1985) who commented that, in forecasting real time series, subjects appeared to attempt to identify seasonal

---

[1] See for example Simon H.A. and Sumner R.K.(1968).

patterns from random oscillations in the data. This opinion was not subjected to empirical tests at that time. The observation was derived from considering the efforts of subjects required to decompose the decision task, but unable to decompose the cue data because of the hard copy graphical data display provided. This raised several issues:

* Is the observation that judges impute seasonal pattern to noise verifiable?
* Would the provision of the capability to decompose the cue data in some way influence any confusion on the part of forecasters?
* Is the use of statistical techniques such as identification of seasonality on the basis of the "ratio to centered moving average" method adequate?

The latter issue goes to the heart of the interpretation of many investigations. There is no standard and absolute method to determine the "seasonal" or the "trend" components of time series data. These are rather loose concepts that have been found to be helpful in breaking up the overall task of time series analysis. It is quite possible that the actions of subjects that caused Lawrence, Edmundson, and O'Connor (1985) to make their comments were based on information in the series that was not detected by the statistical analysis. Further, even in the case of the Eggleton data, which was mathematically contrived and therefore the underlying process of which was known, it is possible that random events might, for some series, have simulated the effects of a pattern process in the short term. The problem is compounded in the applied setting of forecasting economic time series because the underlying process may not remain constant long enough for short term effects to wash out.

There is no conclusive evidence that human judges perform well or poorly in seasonal pattern identification as a result of the biases or heuristics said to apply to human rationality. It therefore remains to examine the performance of judges in that respect, such an analysis is reported in chapter 7.

4.3.1.2 TREND MODEL IDENTIFICATION

The second aspect of judgmental forecasting that is to be considered is the identification of the trend in the time series. Again, Eggleton (1982) has indicated that the human forecaster may be affected by shortcomings in the identification of trend. Eggleton (1982) found that subjects underestimated outcomes in rising trend series, and considered this as evidence of either a parameter error or of an heuristic error. There is doubt about the generalisability of this result because of the experimental design:

> "...it is not believed his results are generally applicable due to the short duration projection method used to present information to research subjects" Lawrence and Makridakis (1986, p 2).

Contrary to the Eggleton (1976) findings, Mosteller et al (1981) found that subjects were able to perform well in the task of fitting a straight line to a scatter plot of data. Once more the task is somewhat different to identifying trend in a time series, but there are some parallels. Mosteller et al (1981) found a slight tendency to overstate the slope of the line (cf.Lawrence and Makridakis 1986), but this was not reported as a significant effect. In the case of data exhibiting very high scatter, Mosteller et al (1981) called this the "fat" plot, there was some evidence that subjects did not fit the usual regression line by minimising the "y-wards" deviation, but seemed to fit a line by minimising the principal component (normal to the fitted line). The ability to fit a good straight line to a scatter plot is an indication that subjects might perform well in identifying trend in a time series. The actual cue design used in the decision aid was not a scatter plot, but a line graph. Shutz (1961) concluded that line graphs were the best form of graphical representation for estimating trend.

Lawrence and Makridakis (1986) conducted a study to determine whether judgmental trend identification was accurate. They presented subjects with 7 points that they advised were drawn from an annual time series of sales expressed in units. The trend model that the subjects applied to the forecasting task was then estimated by comparing each of two point forecasts with the regression line fitted to the data. It was found that the subjects understated the slope of the series, especially for downward sloping series. This was attributed to a "rational" decision that management action would intervene to arrest a decline. There was no empirical basis for this observation, it was merely a rationalisation of the observed results. The design of the experiment does not permit any statement to be made concerning the model of the trend component fitted to the cue data rather than the model used in extrapolation. The results are interpretable as:

* Subjects failed to correctly estimate the trend parameter, understating the slope in the cue data, or
* If subjects correctly identified the correct trend model they applied rules to damp the effect of that model for extrapolative purposes.

The results of Lawrence and Makridakis (1986) are of interest to the extent that if it can be shown that subjects do damp the trend component in extrapolation they are conforming to procedures that are now being used in sophisticated forecasting packages. Gardner and McKenzie (1985) showed that damping trend is a successful strategy for longer term forecasts, and that it is a factor in the success of the Parzen and the Lewandowski forecasting techniques.

The applicability of the Lawrence and Makridakis (1986) study to short term forecasting of monthly data is uncertain however. Two factors in the design of the experiment indicate that the results may not be generalisable to short term forecasting.

1) The use of seven "annual" cue points.

2) The use of a minimum forecast horizon of three years.

The first of these is a complex problem. If the subjects responded to the time scale then the results would certainly be limited to the consideration of annual series. If, however, the subjects merely responded to the plotted cues they might have little confidence in the constancy of the model fitted. The difficulty is that it is not known what is the effect on the subjects of describing data as "annual".

The second problem, concerning the forecast horizon, was that the shortest forecast horizon in the study was almost 50% of the history period given to the subjects. It is possible that subjects would be less prepared to maintain a trend model in their forecasts three years out on the basis of seven years data than they would for 1 to 12 months on the basis of, say, four years data.

Without being conclusive by any means, the literature does appear to point to relatively good performance in the detection of trend, or slope, in data. It is possible that human judges may account for regression to the mean, despite the implications to the contrary from the human rationality literature. It remains to be determined the extent to which this occurs, this is discussed in chapter 8.

4.3.1.3 EXTRAPOLATION OF THE RESIDUAL SERIES

The final aspect of human extrapolation to be considered is that of extrapolating the residual series that remains after the seasonal and trend signals have been removed. There is little literature to be found on this topic, but some inferences may be drawn from tangential literature. For instance, the task might be construed as a special case of smoothing a noisy series.

Spencer (1961) obtained estimates of the means of sets of 10 and 20 random numbers presented as lists, and in graphs. He showed that human judges were capable of good performance, with high consistency:

> "The mean errors....are very small, and the general overall accuracy of the judgments must be regarded as remarkably high" Spencer (1961, p 318).

When the subjects were asked to forecast the next point expected, treating the numbers as a series, Spencer found that the results he obtained:

> "...provides an example of a complex judgment situation in which different people tended to give similar answers to a problem for which there is no 'correct' or absolute answer" Spencer (1961, p323).

The results showed that the subjects were sensitive to trend cues that appeared by chance, and were prepared to be influenced to some extent by "trend" cues in the final 3 to 4 points. In this respect the term "trend" was not meant to imply that the subjects identified a long term effect, merely that they were sensitive to local "runs", perhaps in a similar way to high order smoothing algorithms. The subjects were seen to give less weight to "rogue" events when processing data with outliers. This result offers some positive support for judgmental smoothing of the residual.

Rouse (1976) performed a study in which he provided subjects with a signal comprised of points sampled from a sine wave with white noise added. He required the subjects to infer the underlying process, and to smooth the points to conform to that process (removing the noise). He considered the effect of the number of cue points that the subjects used, and the effect of the noise level. His results indicated that subjects used few points in making any single smoothing decision, and that at very high noise to signal ratios the subjects were not able to improve the series by smoothing. The application of these results is doubtful in the absence of details on the levels of noise in the experiment, and in economic time series. However it is worth noting that Rouse in commenting on the suboptimal performance of judgment as a smoothing strategy, when compared with the most applicable statistical algorithm, wrote:

> "There is no single universal algorithm that can handle all linear, nonlinear, stationary, and nonstationary problems, especially when there is no a priori knowledge about the class to which a particular problem belongs. Yet, humans do handle such a wide class of problems.." Rouse (1976, p337).

As with the previous two tasks, seasonal and trend modelling, there is no direct evidence available concerning the effect on outcome accuracy of judgmental extrapolation from a residual series. In chapter 3 it was shown that the human judge, supported only by a hard copy plot of the original series, appeared to perform better than exponential smoothing for "high noise" series. That result is of course limited to the particular definition of noise used there.

In chapter 9 the accuracy of the human judges in extrapolating the residual series is examined with respect to the use of simple smoothing methods.

4.4 STRATEGIES TO IMPROVE JUDGMENTAL FORECASTING

The review of the forecasting literature in chapter 2 did not provide conclusive evidence of the accuracy or lack of accuracy of the human forecaster relative to competing processes. There was an indication from the work of Lawrence, Edmundson and O'Connor (1985) that the human judge did not seem to be a poorer extrapolator on average than the statistical processes examined. Despite this, there is evidence from the psychology literature that certain biases in human judgment may exist, though their effect in the forecasting task is unknown. The design of the forecasting decision aid reported in chapter 5 was intended to address those issues. The primary objective was to use screen based data displays to permit the decomposition of the decision, with appropriate decomposition of the cue data. The secondary objective was to provide spatial or graphical expression of models, with all computational operations carried out by the computer. The following two sections review the literature that considers first the effect of decomposition of the decision, and then the effect of graphical data presentation, on the forecast decision.

4.4.1 DECOMPOSITION OF THE DECISION

Slovic, Fischhoff and Lichtenstein (1977) stated the case for decomposition as follows:

> "Most of these decision aids rely on the principle of divide and conquer. This decomposition approach is a constructive response to the problem of cognitive overload. The decision aid fractionates the total problem into a series of structurally related parts, and the decision maker is asked to make subjective assessments for only the smallest components." Slovic, Fischhoff and Lichtenstein (1977, p17).

There is considerable literature to support the intuitively appealing proposition that it is likely that the division of a complex task into simpler sub-tasks will result in improved performance. For instance, Armstrong, Deniston, and Gordon (1975), in a study of almanac type problems found that, in twelve out of thirteen estimates, the decomposed decision was more accurate. Armstrong et al (1975) suggested that subjects were able to consider more factors, and that the decision may be improved if the subjects provide the data for formal analysis rather than performing judgmental analysis.

This finding was not entirely supported by Lyness and Cornelius (1982) who considered:

1) an holistic judgment,

2) a decomposed judgment with algorithmic synthesis, and

3) a decomposed judgment with judgmental synthesis, in a performance rating setting.

They found that although the algorithmic synthesis outperformed the judgmental synthesis, it did not outperform the holistic judgment. Their results may not be generalizable because of the simulated decision setting used. They point out that further studies are required in more realistic settings, and that decomposition may be more applicable as the decision process becomes more complex. There was no guidance on the question of how to assess whether a problem is complex enough for decomposition to have an effect. If Lyness and Cornelius (1982) are correct in their conclusions that an holistic judgment may be as accurate as a decomposed judgment except in complex circumstances, then there is the possibility that a circular argument may evolve. Is a decision environment complex only if decomposition is effective?

A more direct problem exists in applying the conclusions reported above to the time series extrapolation problem. The type of decision, and the nature of the cues are very different. Consider the Armstrong et al (1975) research for instance; in that experiment subjects were asked for solutions to "almanac" type questions such as 'how many families live in the USA?'. Obviously the subjects would either have some data on that issue, or they would not. The form of the decomposition approach used was to attempt to bring to bear data that would more probably be available to the subjects....

'How many people live in the USA?'

'How many people are there in an average family?'

For this type of question there is scope for the quality of the decision outcome to be improved by making sure that it is, at least, not inconsistent with known, related data. In the time series extrapolation task the decision cues are provided, in the time series history. The cues, supposing that they be trend, cycle, seasonal, and noise, are complex and they are combined in a series of point values. There is no access to other related data when considering the outcome of the extrapolation. Therefore, any conclusions from the prior research must be limited to "persuasive" authority.

There is some limited support for the use of decomposition in judgmental extrapolation. That support may be found in the results of Lawrence, Edmundson and O'Connor (1985). Those authors reported that extrapolations made by both experienced and inexperienced forecasters were as accurate as those made using statistical techniques. This was the first indication of such success, and it is worth noting that the authors required their subjects to consider the time series first as generating a trend line, then to consider the seasonality in the series separately. No other study of judgmental extrapolation described such a process. That is not to say that subjects used by, say Carbone and Gorr (1984) did not use such a technique, but it was not reported. It may be said, from the foregoing, that decomposition of the judgmental extrapolation decision has lead to acceptable outcomes, even if it may not be concluded that the decomposition actually lead to improvement in the outcome. It should be noted that the decision in Lawrence, Edmundson and O'Connor (1985) was arrived at by judgmental synthesis.

The advantages envisaged as flowing from task decomposition are:

1) Subjects may be encouraged to view each task separately, and to develop appropriate decision rules for each. Dickhaut and Eggleton (1975) conclude that subjects develop decision rules early in the decision and tend to stick to them throughout the whole process. Decomposition may reduce that tendency.

2) The subjects may be encouraged to observe a greater number of cues than they would in an holistic judgment. This follows the findings of Phelps and Shanteau (1978) who showed that judges of livestock considered a greater number of cues in a decomposed process than they did holistically.

3) There is the expectation that the variance in the error may be reduced.[2]

4) It is possible that mathematical synthesis of the components may give rise to improvement in the outcome.[3]

The provision of an interactive graphical decision aid for the extrapolation task gives rise to new opportunities to test decomposition. As described above, the subjects in Lawrence, Edmundson and O'Connor (1985) were required to decompose the task according to the structure of classical time series analysis. They were not able to review the outcome of their decision for each component, and they were not able to remove the cues component cues from the data. That is, once a trend line had been determined, the subjects were required to identify seasonality from a series that still had the trend cues embedded. Thus, to determine the seasonal character of any particular month, over the history of

---

[2] This follows from Armstrong (1985), who showed that there would be expected improvement in the variance of the outcome if the components were relatively independent, of roughly equal importance, and were based on reliable data.

[3] For instance see Einhorn (1972) and Gettys et al (1973), but consider the different outcome in Lyness and Cornelius (1982).

the data, the impact of trend had to be judgmentally eliminated. If, as supposed by Miller (1956), there is a finite limit to the data load that a judge can handle, it is possible that the removal of the identified trend cues from the data might advantage the decision maker in subsequently identifying seasonality. This would fit the extrapolation task more closely to the nature of the tasks studied in the "decomposition" literature by allowing a subsequent component (seasonality) of the decision to be considered in isolation, even if the prior component (trend) could not.

A further possibility arising from the nature of an interactive interface is the provision of mathematical recombination of the parts of the decision. Indeed, the decomposition may be similarly supported. That is, any trend model may be mathematically removed from the cue data before the seasonal model is identified, and the recombination of the trend and seasonal models might be accomplished automatically. This capability is inextricably bound to the interactive interface. It is not currently possible to provide real time decision support of that nature except in a screen based interactive system, and it does not seem to be feasible or sensible to decompose the decision without mathematical decomposition and recomposition.

The literature cited above is not conclusive concerning the effect of decomposition, and barely touches on decomposition of the cues in judgmental extrapolation. It is, therefore necessary to examine whether decomposition of the cues leads to improvement in the accuracy of the extrapolation. This issue is not addressed in isolation in this dissertation. In chapter 6 the accuracy of the extrapolation made with the support of a decision aid is considered. A major feature of the decision aid is the serial decomposition of the cue data, but this is achieved by using a sophisticated interactive graphical data interface. As described in chapter 6, there is no practical means to consider the effects of decomposition and graphics separately. Therefore, the results of the investigations reported in chapter 6 are confounded to some extent by the simultaneous effects of decomposition and graphics.

4.4.2 FORM AND PRESENTATION OF DATA.

There is evidence in the literature that information use can be affected by data presentation (Tversky and Kahneman 1981), however, the manner and outcome of that effect is indeterminate for graphical data presentation as an alternative to a tabular form despite fairly extensive investigation, much of which was centred on the so-called Minnesota experiments [4].

---

[4] See for instance Chervany and Dickson (1974); Dickson, Senn and Chervany (1977); Schroeder and Benbasat (1975)

There may be found many expressions of unsubstantiated opinions [5] proposing the benefits of graphical displays.

There are also some conflicting results from empirical work. Lucas (1981) provides a summary of the literature, indicating that, for instance, Benbasat and Schroder (1977) found that graphical data presentation improved decision outcome in an inventory management experiment (though the improvement was not statistically significant), while Lucas' (1981) own experiments using a task of similar complexity indicated no improvement with the use of graphics, a result that confirmed the findings of Lusk and Kersnick (1979). Recently Remus (1984) reported that in a simulated production scheduling experiment, tabular data presentation was not different to graphical data presentation, but when the subjects decisions were modeled and erratic components removed tabular data presentation was better. Lucas (1981) concluded that the <u>problem type and structure</u> may have an important effect upon the results of using graphics.

Reporting on a carefully executed series of experiments considering the effect of the task type on the effectiveness of data displays, Dickson, DeSanctis and McBride (1986) found that there was a task related effect. They were unable to propose a theory to explain the effect however, leaving the consideration of display design still an ad hoc exercise.

---

[5] See Paller A.T.(1981).

Within the time series extrapolation field there has been only one study impinging on data presentation. Lawrence, Edmundson and O'Connor (1985) used both hard copy plots and tabular data displays of time series in their experiment. They found that there was no significant difference in the outcome between the two forms of display.

The literature illustrates that there are uncertainties in the effect of graphics, and that the results of research cannot yet be generalized. This dissertation considers an application area and task that differs from most of the prior research, and extends the nature of the graphics support provided. The time series extrapolation task considered here was only peripherally addressed in the Lawrence, Edmundson and O'Connor (1985) study. Further, the prior research on the effect of data presentation is concerned with static displays of data, which may be likened to providing hard copy reports. This dissertation includes the consideration of a data interface which allows manipulation of graphical plots using a 'mouse' driven cursor. Thus the cue decomposition described in the preceding section is achieved using graphical models, and the outcome of applying the models is displayed graphically. The data interface differs from those considered above in the following respects:

1) The display is dynamic, permitting the effect of decision outcomes to be reviewed immediately,

2) the interface is interactive, thus decision outcomes are expressed in the same form as the

cues, and the components of the decision are entered directly on the display,

3) different aspects of the decision may be supported by displays designed to suit the task, and the data presented at each stage reflects the outcome of prior stages.

The lack of a theory to predict the effect of a particular data interface design (Dickson et al 1986), coupled with the great differences between the design of the experimental extrapolation decision aid and the displays studied to date indicates the need to evaluate the particular interface. The studies reported in chapter 6 address this issue in conjunction with the effect of decomposition.

4.5 CONCLUSION

Although the two characteristics of the data interface, namely the ability to decompose the decision and the data and the graphical nature of the interface, have been considered separately in this chapter they are not separable in examination. The decomposition aspect relies upon the interactive graphical interface for feasibility[4] . Therefore the results of the research will not be attributable directly to either the decomposition of the decision or to the graphical nature of the data interface.

---

[4] Although it would be possible to construct an interactive tabular decision model, based on spreadsheet technology, it would be somewhat cumbersome.

4.6 REFERENCES

Armstrong J.S., *Long-Range Forecasting from Crystal Ball to Computer*. 2nd Edn. Wiley, New York, (1985).

Armstrong J.S., Denniston W.B. Jr., and Gordon M.M., The use of the decomposition principle in making judgments" *Organizational Behavior and Human Performance*, 13, (1975), 257-263.

Benbasat I. and Schroder R., "an experimental investigation of some MIS design variables" *MIS Quarterly* 1,1 (1977), 37-49.

Carbone, R., & Gorr, W.L., "Accuracy of judgmental forecasting of time series" *Decision Sciences*, vol 16, (1984), 153-160.

Chervany, N.L., & Dickson, G.W., "Experimental evaluation of information overload in a production environment", *Management Science*, 20, (1974), 1335-1344.

Christensen-Szalanski J.J., "Improving the practical utility of judgment research", in *New Directions In Research On Decision Making*, North Holland, (1986), p 383-410.

Cohen L.J., "Can human rationality be experimentally demonstrated?", *The Behavioral and Brain Sciences*, 4, (1981), p317-370.

Dickhaut J.W., and Eggleton I.R.C., "An examination of the processes underlying comparative judgments of numerical stimuli." *Journal of Accounting Research*, Spring (1975), 38-72.

Dickson, G.W., DeSanctis, G., & McBride, D.J., "Understanding the effectiveness of computer graphics for decision support: accumulative experimental approach", *Communications of the ACM*, 29,1, (1986), 40-47.

Dickson, G.W., Senn, J.A., & Chervany, N.L., "Research in management information systems- Minnesota experiments", *Management Science*, 23, (1977), 913-923.

Eggleton I.R.C. "Intuitive Time Series Extrapolation", *Journal of Accounting Research*, 20,1. (1982), 68-102.

Eggleton I.R.C. "Patterns, Prototypes, and Predictions: An Exploratory Study " *Selected Studies on Human Information Processing in Accounting*, (1976). Supplement to *Journal of Accounting Research* 14: 68-131

Einhorn H.J., "Expert measurement and mechanical combination", *O.B.H.P*. 7, 1972, 84-106.

Gardner E.S. & McKenzie E. "Forecasting trends in time series", *Management Science*, 31,10, (1985), 1237-1246.

Gettys C.F., Michel C., Steiger J.H., Kelly C.W. III., Peterson C.R., "multiple-stage probabilistic information processing", *O.B.H.P*. 10, 1973, 374-387.

Hogarth R.M., *Judgment and Choice*, Wiley, Chichester, (1980).

Hogarth R.M., and Makridakis S., "Forecasting and planning: an evaluation", Unpublished manuscript, INSEAD, Fontainebleau, France, (1979).

Johnson-Laird P.N. & Watson P.C., "A theoretical analysis of insight into a reasoning task", *Cognitive Psychology*, 1, (1970), 134-148.

Kahneman D. & Tversky A., "Subjective probability: a judgment of representativeness", *Cognitive Psychology*, 3, 1972, p430-454.

Kahneman D. & Tversky A., "On the psychology of prediction", *Psychology Review*, 80, (1973), 327-351.

Lawrence, M.J., Edmundson, R.H., & O'Connor, M.J., "An examination of judgmental time series extrapolation", *International Journal of Forecasting*, 1, (1985), 25-36.

Lawrence, M.J., Edmundson, R.H., & O'Connor, M.J., "The accuracy of combining judgmental and statistical forecasts", *Management Science*, 32,12, (1986), 1521-1532.

Lawrence, M.J., & Makridakis, S., "Human judgment in extrapolation", *International Forecasting Symposium*, Paris, 1986.

Lucas H.C., "an experimental investigation of the use of computer-based graphics in decision making." *Management Science*, 27,7. (1981), 757-768.

Lusk, E.L., & Kersnick, M., "The effect of cognitive style and report format on task performance: the MIS design consequences", *Management Science*, 25, (1979), 787-798.

Lyness K.S. and Cornelius E.T. III., "A comparison of holistic and decomposed strategies in a performance rating simulation" *Organizational Behavior and Human Performance*, 29, (1982), 21-38.

Makridakis, S., and Hibon, M., "Accuracy of forecasting: an empirical investigation", *Journal of the Royal Statistical Society, Series A*, 142, (1979), 97-145.

Miller G.A.,"The magic number seven, plus or minus two: some limitations on our capacity for information processing." *Psychological Review* 63, (1956), 81-97.

Moriarty, M.M., and Adams A.J., "Management judgment forecasts, composite models, and conditional efficiency", *Journal of Marketing Research*, vol xxi, (1984), 239-250.

Mosteller, F., "Innovation and Evaluation" *Science*, 211, (1981), 881-886.

Paller A.T., "Improving management productivity with computer graphics." *Computer Graphics and Applications*. (1981), 9-16.

Phelps R.H., and Shanteau J., "Livestock judges: how much information can an expert use?" *Organizational Behavior and Human Performance*, 21, (1978), 209-219.

Remus W., "An Empirical Investigation of the Impact of Graphical and Tabular Data Presentations on Decision Making." *Management Science*, 30,5 (1984), 533-542.

Rouse W.B., "A model of the human as a suboptimal smoother", *IEEE Transactions on Systems, Man, and Cybernetics,* smc-6, 5, (1976), 337-343.

Schroeder, R.G., & Benbasat, I., "An experimental evaluation of the relationship of uncertainty in the environment to information used by decision makers", *Decision Sciences*, 6, (1975), 5560-5567.

Sekular R.W. and Abrams M. "Visual sameness: a choice time analysis of pattern recognition processes." *Journal of Experimental Psychology*. 77,2 (1968), 232-238.

Shutz, H.G., "An evaluation of formats for graphic trend displays", *Human Factors*, 3, (1961), 99-107.

Simon H.A. and Sumner R.K. "Pattern in Music" in *Formal Representation of Human Judgment.* ed Kleinmuntz, New York. Wiley (1968), 219-250

Slovic P. and Lichtenstein S.,"Comparison of Bayesian and regression approaches to the study of information processing in judgment." *O.B.H.P.*, 6 (1971), 649-744.

Slovic P., Fischhoff B., & Lichtenstein S., "Behavioral Decision Theory." *Ann. Rev. Psychol*. 28 (1977), 17

Spencer J., "Estimating Averages" *Ergonomics*, 4, (1961), 317-328.

Tversky A. & Kahneman D., "Availability: a heuristic for judging frequency and probability" *Cognitive Psychology*, 5, (1973), 207-232.

Tversky A. & Kahneman D., "Judgment under uncertainty: heuristics and biases", *Science*, 185, (1974), 1124-1131.

Tversky A. & Kahneman D., "The framing of decisions and the psychology of choice", *Science*, 211, (1981), 453-458.

Wagenaar W.A. and Timmers H., "Extrapolation of exponential Time Series is not Enhanced by Having More Data Points." *Perception and Psychophysics*. 24,2 (1978), 182-184.

Wagenaar W.A., "Generation of random sequences by human subjects: a critical survey of literature." *Psychological Bulletin*. 77,1. (1972), 65-72.

Watson P.C. "Reasoning" in *New Horizons in Psychology*, vol 1, , ed. B. Foss. Harmondsworth: Penguin (1966).

Watson P.C. & Shapiro D. "Natural and contrived experience in a reasoning problem", *Quarterly Journal of Experimental Psychology*, 23, (1971), 63-71.

Winkler, R.L., and Makridakis, S., "The combination of forecasts", *Journal of the Royal Statistical Society*, 146 Part 2, (1983), 150-157.

5. DESIGN OF THE FORECASTING DECISION AID

5.1 INTRODUCTION

This chapter reports on the design, development and features the experimental forecasting decision aid called GRAFFECT[1]. The current version is implemented in TURBO PASCAL and the user manual is presented in appendix A, and the source code is contained in appendix B.

The design of the decision aid was influenced by the results of the discriminant analysis study reported in chapter 3, and the literature reviewed in chapters 2 and 4.

It was shown in chapter 3 that the intuitive extrapolation processes are reproducible, and that they have properties different to the statistical processes. Ideally, the different abilities of human and machine would be traded off according to the characteristics of the time series. This view accords with that of Felsen (1976):

> "..so computers are used to augment rather than replace human judgment when solving investment decision problems. Fortunately, it can be observed that people and computers have complementary information processing capabilities and therefore significant advantages may be gained by building decision systems containing both." Felsen (1976, p110)

The trade off might operate at one or both of two levels:

1) Computational support of the judgmental processes,

---

[1] Graphical Aid For the Forecasting of Economic Time series.

2) Replacement of some of the judgmental processes.

Before the second of those strategies is implemented, it is important to determine the extent to which the intuitive processes might be improved by the provision of a supporting decision environment. In chapter 3 some limited rules for the selection of extrapolation method were developed that would lead to expected improvement in forecast outcome. Those rules represent extreme strategies in trading off between methods, that is the total replacement of one method by another.

In the commercial environment it is doubted whether judgment will be entirely replaced in the forecasting task. In most companies there are time series that are critical enough to warrant judgmental review of the forecasts, and other series that are so influenced by external factors that no statistical forecasting methodology alone could produce a reliable forecast. It is therefore anticipated that for many time series the trade off between the human judge and the machine based capabilities will lie between the extremes of total replacement of one method by another.

It is the <u>objective of the decision aid, the development of which is reported here, to facilitate research to evaluate the performance of the human judge when supported by a simple, interactive graphical environment</u>. It is the further objective to permit research on issues associated with decision support system and decision aid design, and the task allocation decisions taken in such a design:

1) What task allocation is optimal?

This refers to the allocation of tasks within the extrapolative process to either the decision maker or the machine. For instance, the task of identifying the seasonal pattern might be performed by either.

2) Is the decision aid viable?

There are two aspects to this question. The first is whether the decision aid improves the accuracy of the forecast. The second aspect concerns the extent to which the decision aid affects the time taken to generate the forecast.

3) Does the data interface have any effect?

This is an attempt to throw light on the effect of using interactive graphics for data display. It will not be possible entirely to isolate the effect of the interactive graphics capabilities of GRAFFECT. However, it is of interest to determine, for say the detrending and the deseasonalising tasks, whether the decisions made differ if paper and pencil are used instead of the screen graphics.

5.2 DECOMPOSITION OF THE DECISION

The decision aid was designed to enforce a decomposition of the extrapolation task along the line of classical time series analysis:

* Identify the trend model
* Identify the seasonal model
* consider the residual (noise) component.

The particular design decisions for each module are described below.

Although the literature reviewed in chapter 4 did not point unequivocally to the success of decomposition per se, and there was no real examination of the use of decomposition in time series extrapolation it was felt that there were sufficient indications in favour of decomposition to warrant its use. The major justifications for the use of decomposition were:

1) To take advantage of any possible reduction in cognitive load.

2) To permit the consideration of the replacement of judgment, module by module.[2]

The decomposition of the decision, and of the time series cue data, also permitted the provision of computational support both in parameter estimation in model building and in the recomposition of the decision outcome. The mathematical recombination has received general support in the literature, and in this case was adopted as a necessary factor in the design of the system. It was, after all, impractical to have the judge mentally multiply the residual extrapolation by the monthly factors for trend and seasonality (estimated mathematically from spatial relationships determined by the judge).

---

[2] Many statistical extrapolation processes are decomposed along the same lines as that described. It is therefore possible to substitute the ratio to centred moving average process for the judgmental process in seasonal identification for instance.

5.3 THE USE OF GRAPHICS

The literature on the effect of graphics reviewed in chapter 4 is possibly less consistent in its conclusions than the decomposition literature. The one consistent message appears to be that testing is still required in each task environment because of the lack of a general theory. As discussed in chapter 4, the literature that does exist is concerned with the use of static graphics, therefore any conclusions that were drawn in that literature would have limited relevance to the use of interactive graphics.

The major impetus for the use of graphics in the decision aid stemmed from the desire to permit the decision maker to analyse the data and produce the decision in the same terms. That is, the decision maker was to be presented with graphical cues, could formulate graphical models, and have the decision outcome presented in graphical form without the need to translate in and out of the number system with the possibilities of error that that entails.

5.4 TREND IDENTIFICATION

In chapter 3 it was shown that human judges were more accurate than deseasonalised single exponential smoothing for series with high lag one autocorrelation characteristics. Although the correlation between the metric for that characteristic and the metric developed for trend was less than 0.4 it is possible that the advantage did arise to some extent from the judgmental handling of trend. There is an a priori reason for such a conclusion. Deseasonalised single exponential smoothing has no explicit trend handling capabilities, and even with a very high "alpha" will lag a trend in the first period out and get progressively worse for longer time horizons. The judgmental process included specific consideration of the trend term, and therefore should have a definite advantage. The lack of such a definite advantage points to either poor handling of the trend by the judges, or to an effect due to changes in the test time series in the validation period. It is necessary for that issue to be examined. It was therefore determined that the data interface for trend identification would present the same cues used in the hard copy graph. The subjects would be able to fit a trend model to the history of the time series, and if they wished to provide an extrapolation with a different slope. Thus it would be possible to determine whether:

1) the judges could fit a trend line acceptably, and
2) whether they showed any inclination to damp the trend when extrapolating.

The latter point was addressed to the issue raised in the review of the work of Lawrence and Makridakis (1986).

For that reason, it was determined that trend identification would be supported, as a first pass, by the classical serial plot of monthly data. It was left for future developments to examine alternative data presentations such as:

* plots aggregated by different horizons (quarterly, or yearly),

* monthly data with additional data superimposed: monthly average for the year,

* a regression fit to the monthly data.

5.5 SEASONAL IDENTIFICATION

The pattern processing capabilities of the human that were discussed in chapter 4 seemed to be somewhat supported by the results in chapter 3. There it was shown that the human judge had an advantage over deseasonalised single exponential smoothing for low seasonal series in the presence of high noise in the seasonal. The relatively poorer result in the case of high, stable seasonal patterns in the presence of low noise is interesting, and the reason for it can only be a matter of conjecture at this stage. The most obvious explanation seems to be that the ratio to centered moving average process for seasonal identification in the deseasonalised single exponential smoothing method has a precision advantage, rather than a signal detection advantage. This implies that the judge is not incapable of modelling the seasonal pattern, but that errors may be made in determining the average (or other) model of the seasonal pattern. If that were the case, then it might be possible to assist the judge towards greater precision. The ratio to centred moving average process works, conceptually, as follows:

1) For each month, of each year, the effect of the seasonal process on that month is estimated,

2) An average of those effects is determined across the equivalent months in each year. That averaging process may take account of outliers.

Presentation of the monthly data, net of any trend identified, on the same 12 month grid would assist the judge to perform a similar operation. The judge would have a picture of the effect on equivalent months that would be more easily processed than it would be from a serial plot of the data.

The provision of the data in this form, a "parallel" plot of the years, did not seem to jeopardise the detection of subtle seasonal patterns. Indeed, it was hoped that any effect would be advantageous.

As described later, and illustrated in figure 5.1 below, it was determined to provide such a "parallel" plot of the data to support the seasonal identification process, and the mode of model entry was to be designed to reduce the necessary amount of data transformation or transposition. The input of the model was permitted to take place on the actual plot of the history data. This was to further assist in providing greater precision.

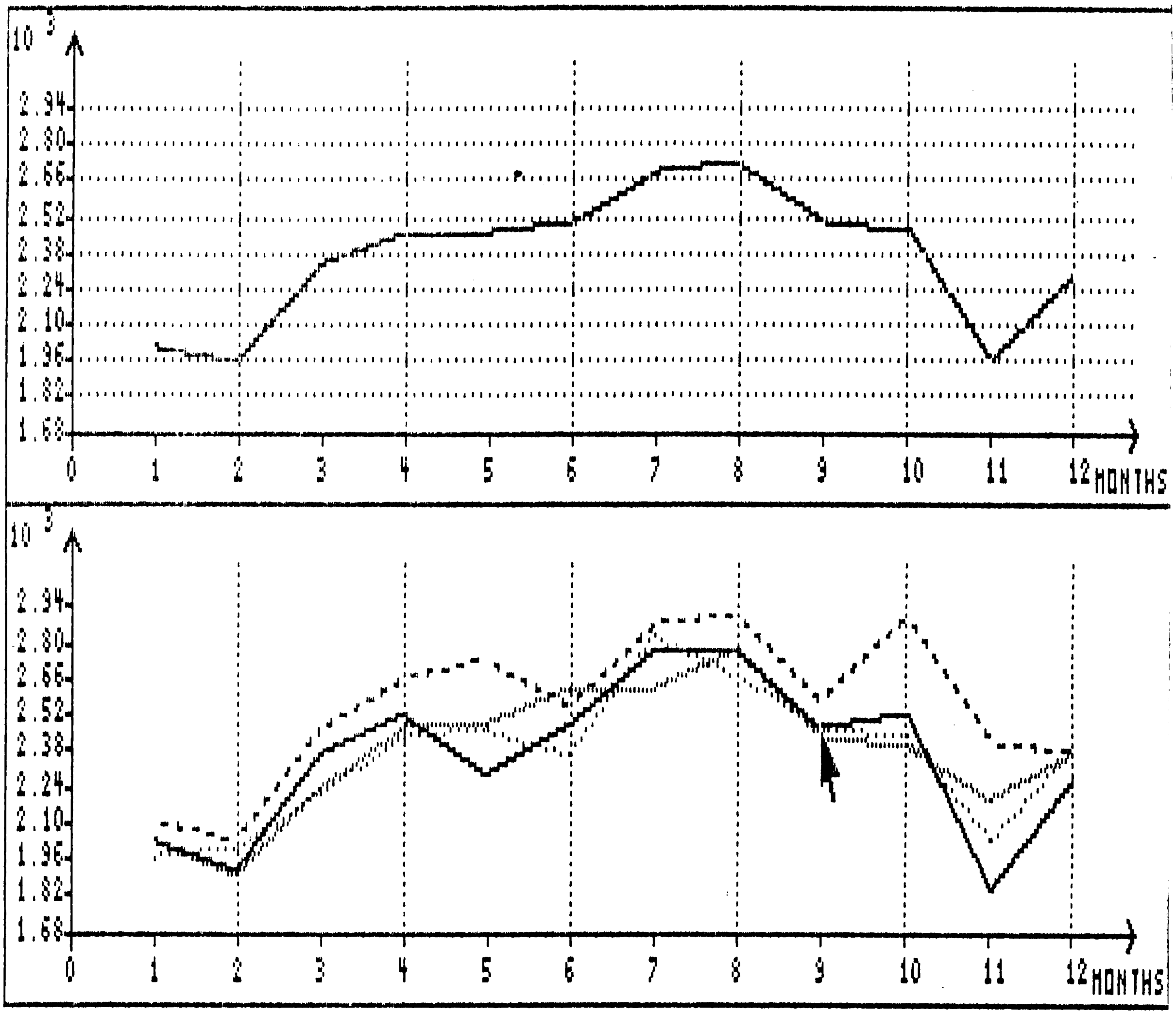


Figure 5.1 Seasonal Identification Screen.

The seasonal model is more complex to identify than the trend model. This is reflected in the number of parameters (11[3]) that must be estimated. Although the general functioning of the proposed decision aid was intended to provide a simple iterative capability, it was determined that in the case of the seasonal model identification that iterative ability was to be enhanced. This was achieved by permitting the effects of the decision process to be simply reviewed at intermediate stages. this

---

[3] In terms of the formal seasonal model, the sum of the monthly seasonal factors is 12, by definition. Therefore, estimation of any 11 of the parameters determines the twelfth.

was unlike the trend identification module in which the effect could be reviewed after the model had been completed, and a new model built if the review was unfavourable. The approach for the seasonal module was to permit the judge to review the effect any number of times, as the seasonal patterns for each of the months were considered.

5.6 THE NOISE COMPONENT

Perhaps the strongest signals from the H.I.P. literature were the problems that judges have with the concept of randomness. Those findings had little relevance to the extrapolation of a random sequence. In chapter 3 it was reported that series with a high value for the "noise" metric were handled rather well by human judges. In the anticipated decision aid environment it was expected that the extrapolation of the noise component would take place using data that had been "washed" clean of trend and seasonal signal. The resultant series would therefore be relatively stationary. The possible strategies for judgmental extrapolation of such a series are to use a smoothed value such as the mean of the series (or part of the series) or to use the latest value of the series, as an indicator of the most probable next value(s).

The relatively good performance of the judges prompted the decision to provide only the serial plot of the resultant series, as the first pass. This was intended to avoid masking any advantage that the judgment processes had in dealing with series with high autocorrelation characteristics. Additional cues might be examined in the future, such as the provision of summary statistics on the resultant.

The provision of interactive graphical facilities permitted the removal of a step in the extrapolation process, that of extracting the desired value from the display. The means of data capture was made independent of the actual values, therefore reducing the chance of error in reading the display, and eliminating any dysfunctional "post-processing" of the desired model in the light of the number values read. There is no indication that such post-processing does occur, but it is at least possible that in extracting data from a plot rounding to whole numbers or adjusting to appealing numbers (say from 97 to 100) might occur. The omission of the data extraction step avoided that risk entirely.

5.7 DESCRIPTION OF THE DESIGN OF THE DECISION AID

The human performance literature and the recent forecasting literature reviewed in chapters 2 and 4 shows the need to construct and evaluate a decision aid for forecasting. The objectives of such a system would be to take advantage of the respective capabilities of the decision maker and computerised processes in order to produce more accurate forecasts.

Although the forecasting literature shows that the opportunity exists for improvement by combining the man and the machine it does not identify the particular attributes to be combined, nor how they would interact. Similarly, the human performance literature points to perceived failings in human judgment, but is silent on the question of how a decision support system or a decision aid might be constructed to alleviate those shortcomings.

The DSS literature is also lacking in specifics as far as research findings on the question of the construction of decision aids. In his paper reporting an investigation of graphical and tabular data presentation Remus (1984) comments:

> "Although there have been major efforts to develop decision support systems to aid managerial decision making, there has been little research to suggest how best to accomplish this." Remus (1984, p 533).

Kosaka and Hirouchi (1982) comment that the research on decision support has paid little attention to architecture but has concentrated on data enlargement and model refinement. They propose a structure based upon what they call a 'model unit' which comprises a single problem solving function. The decision maker invokes model units iteratively to approach a solution to the whole problem. This strategy has been adopted for the experimental forecasting decision aid, because it permits the desirable flexibility of operation for the decision maker, and it facilitates flexibility in the features offered in GRAFFECT.

The individual model units are designed to place responsibility for data extraction upon the decision maker. This reflects the views of Felsen (1976):

> "Experimental evidence indicates that for many types of complex and abstract patterns, human analysts are more effective at feature extraction and numerical encoding of observations than current mechanical measurement procedures (Fenker and Evans, 1971). ...substituting subjective human judgments for the objective measures which would ordinarily be used." Felsen (1976, p117)

Once the encoding, in terms of spatial relationships rather than direct numerical coding, is complete, the machine performs the necessary computational tasks. This reflects the findings of the human performance literature that man is a poor intuitive numerical data handler, and seeks to reduce the cognitive load. The model units are designed to:

1) allocate model recognition tasks to the decision maker,

2) have parameter estimation carried out between the decision maker and the machine. The decision maker proposes and reviews pattern shapes, and the machine generates the numerical parameter values.

3) have computational and display tasks carried out by the machine.

The scope of the decision aid has not been limited to the experimental requirements of the research. An attempt has been made to design a decision aid that is complete, and that could be placed in the field for practical evaluation. In 1977 Slovic et al remarked upon the need to research the effect of design decisions upon the decision maker and in doing so indicated the need for the experimental decision aid to approach 'reality':

> "With systems designed for research purposes, a critical issue is the tradeoff between realism and generality. One strategy is to design systems whose complexity begins to approach that found in the real world-at the risk of investing too much of available resources in the machine and too little in understanding how people us it. Some human factors questions worth studying are (a) how do variations in the basic system (e.g. different instructions or information displays) affect people's performance? (b) how do person and machine errors interact? (c) how should machine output be adjusted to different decision makers' cognitive styles and work paces? and (d) when do people heed the machine's advice?" Slovic et al (1977, p 26).

5.8 DESCRIPTION OF THE EXPERIMENTAL INSTRUMENT.

The Graffect decision aid was developed and implemented by an incremental process over a period of four years, with each intermediate stage being tested for user acceptability and convenience. The fundamental design principles did not change during that process, but the user interface was refined, and some functions were expanded incrementally. For instance, initially there was no provision for a damped trend extrapolation to be prescribed. All testing reported in this dissertation was carried out using the version of the software described in the user documentation.

The forecasting process has been decomposed along the lines of the recognised classical decomposition method. Thus, long-term trend and cycle, short-term cycle (seasonal), and the residual (noise) are considered separately. That is :

$$X_t = f(I_t, T_t, C_t, E_t)$$

where:

X_t is the time series value at t
I_t is the seasonal component at t
T_t is the trend component at t
C_t is the cycle component at t
E_t is the random component at t

This function could be implemented as either an additive or a multiplicative model. The multiplicative model may be expressed as:

$$X_t = I_t * T_t * C_t * E_t$$

There is no evidence in either the research or the normative literature that one model outperforms the other, and therefore no current basis for choosing. The model adopted as the basis for Census II decomposition is the ratio to moving averages, or multiplicative model. Makridakis and Wheelwright (1978) make the comment that this model provides the "much preferable approach" but offer no evidence for that opinion. In the absence of evidence regarding which model to use, the multiplicative model has been adopted for the Graffect decision aid. This decision was based on a desire to use conventional seasonal factors, a desire to be consistent with other major systems such as Census II, and the ease of construction of the model. In truth, the differences between multiplicative and additive models would only be felt in the interaction of the error terms implied in each of the components of the models. It is a matter for further research to determine whether there is any practical difference between the models.

5.9 ILLUSTRATIVE EXAMPLE OF THE USE OF GRAFFECT

The entry screen of GRAFFECT only permits the loading of a time series for analysis. Once a series has been loaded, the Main Menu, shown in figure 5.2, screen displays upto the last 48 observations of the series.

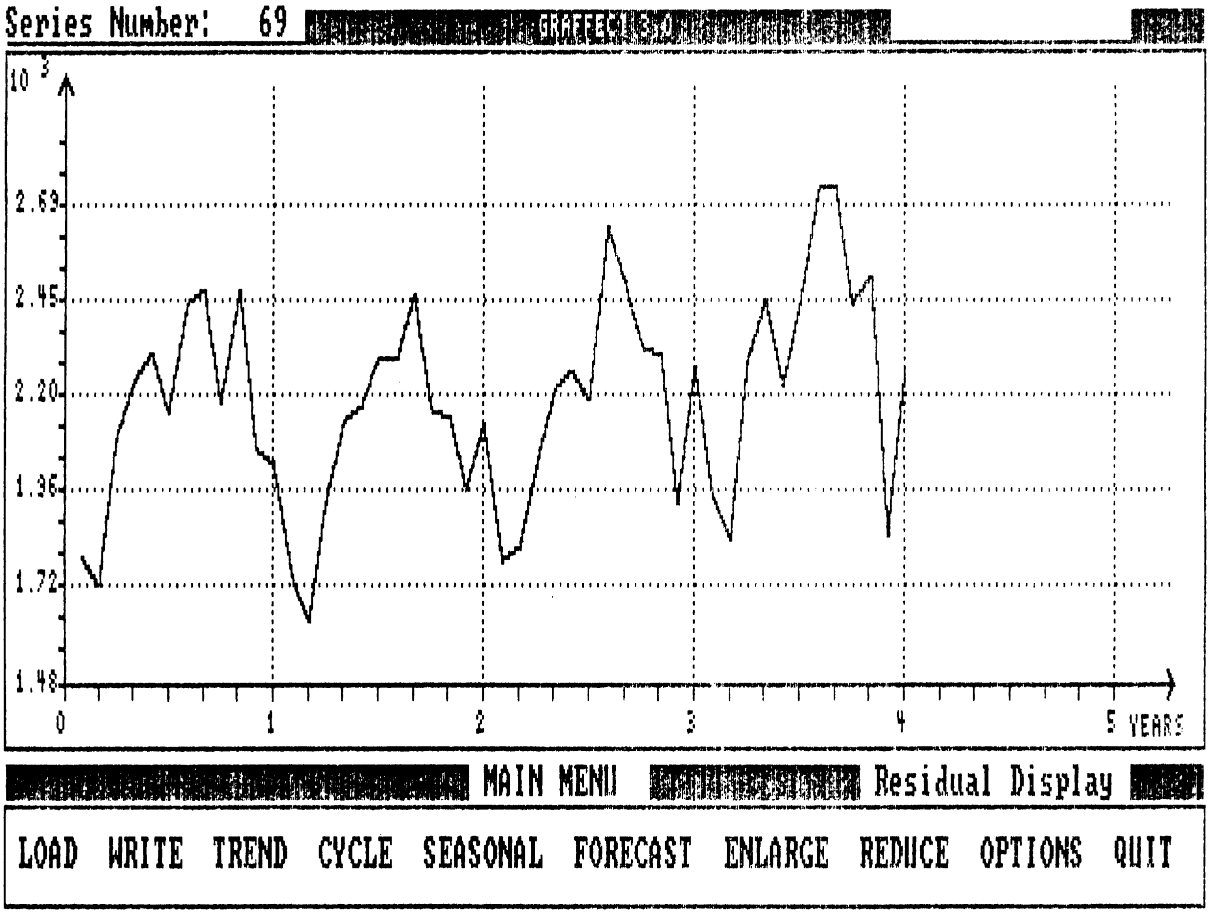


Figure 5.2 Main Menu Screen

Decomposition of the cue data may be carried out in any sequence, but the more difficult task of seasonal identification is easier if carried out after any necessary detrending. Figure 5.3 illustrates the Trend identification screen. It shows that, in this case, a trend line has been fitted to the data from about month 12 to month 48.

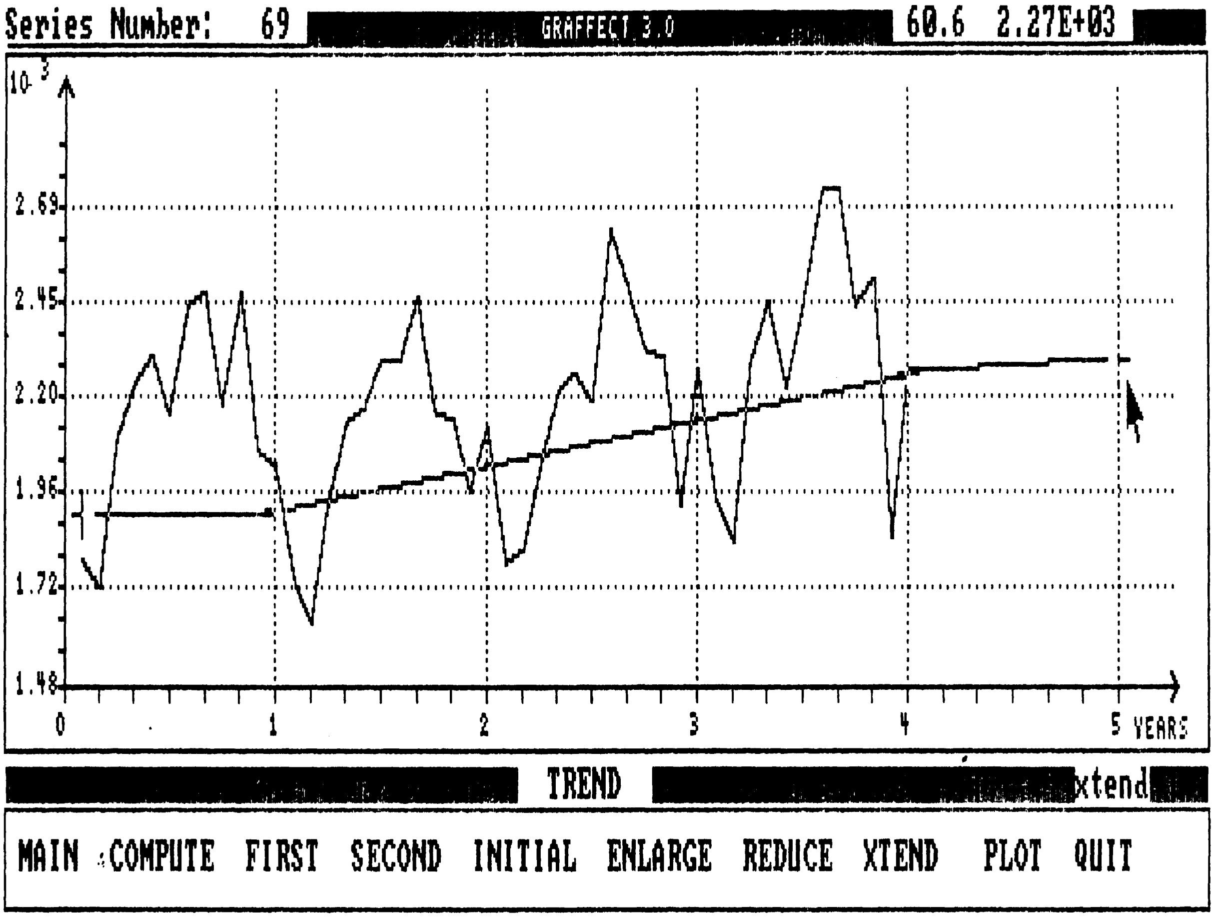


Figure 5.3 Trend Identification Screen

The trend lines[4] fitted to the historical observations (months 1 to 48) are used by GRAFFECT to remove the trend cues from the data. In the absence of an explicit extension of trend into the forecast area, the chronologically later trend line is extrapolated by GRAFFECT. In the illustration there is an explicit extension of the trend model into months 49 to 60. That model, with a lower slope, is thus the trend model that will be used in the final recomposition of the forecast.

---

[4] One or two lines may be fitted, thus taking account of one turning point in the time series.

After computation of the trend factors and the return to the Main Menu, the "residual" time series is displayed as transformed to remove the trend model. Figure 5.4 shows the result.

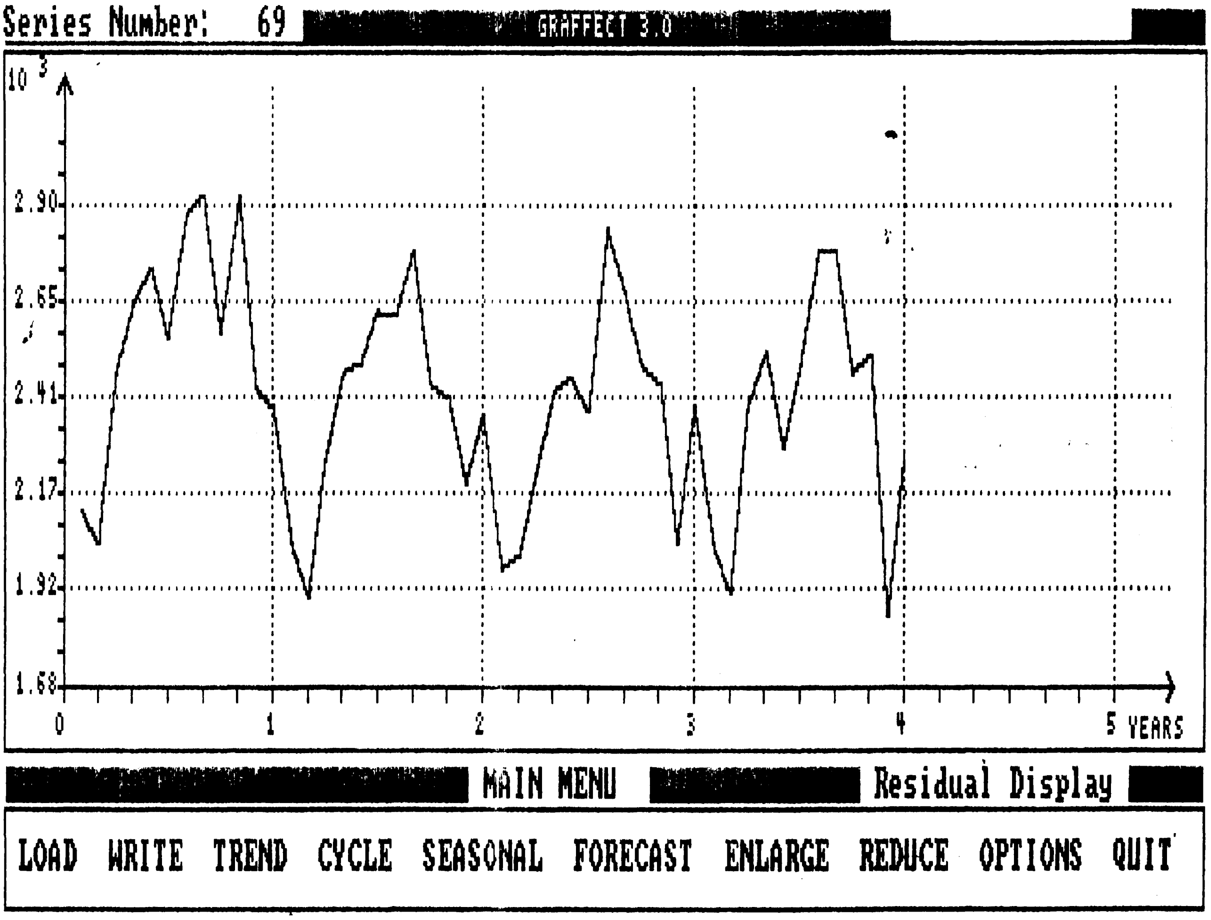


Figure 5.4 Residual Series after Detrending.

Since the residual time series shows signs of a seasonal pattern, the Seasonal identification screen is entered. Figures 5.5 and 5.6 illustrate the operation of this screen. In figure 5.5 the seasonal pattern in the data has been modelled by pointing to the selected position for each month with the mouse. The figure shows the entry of a point for month 9 in the lower part of the screen, and the shape of the identified seasonal model in the upper.

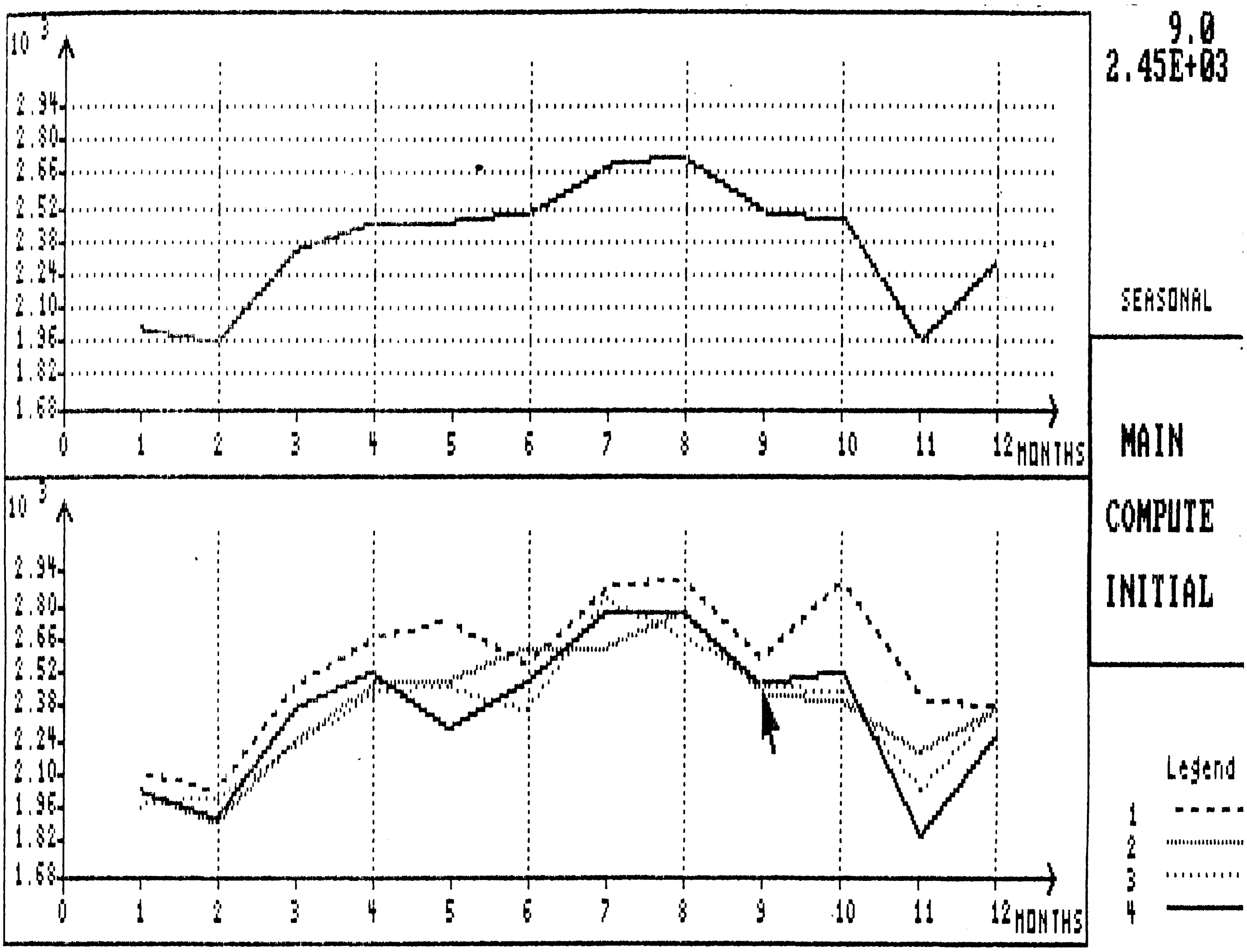


Figure 5.5 Seasonal Modelling

Figure 5.6 illustrates the display resulting from computation of the seasonal factors, and removal of the model from the data. The identification activity continues until the judge is content that there is no more signal to be detected in the residual shown in the lower panel.

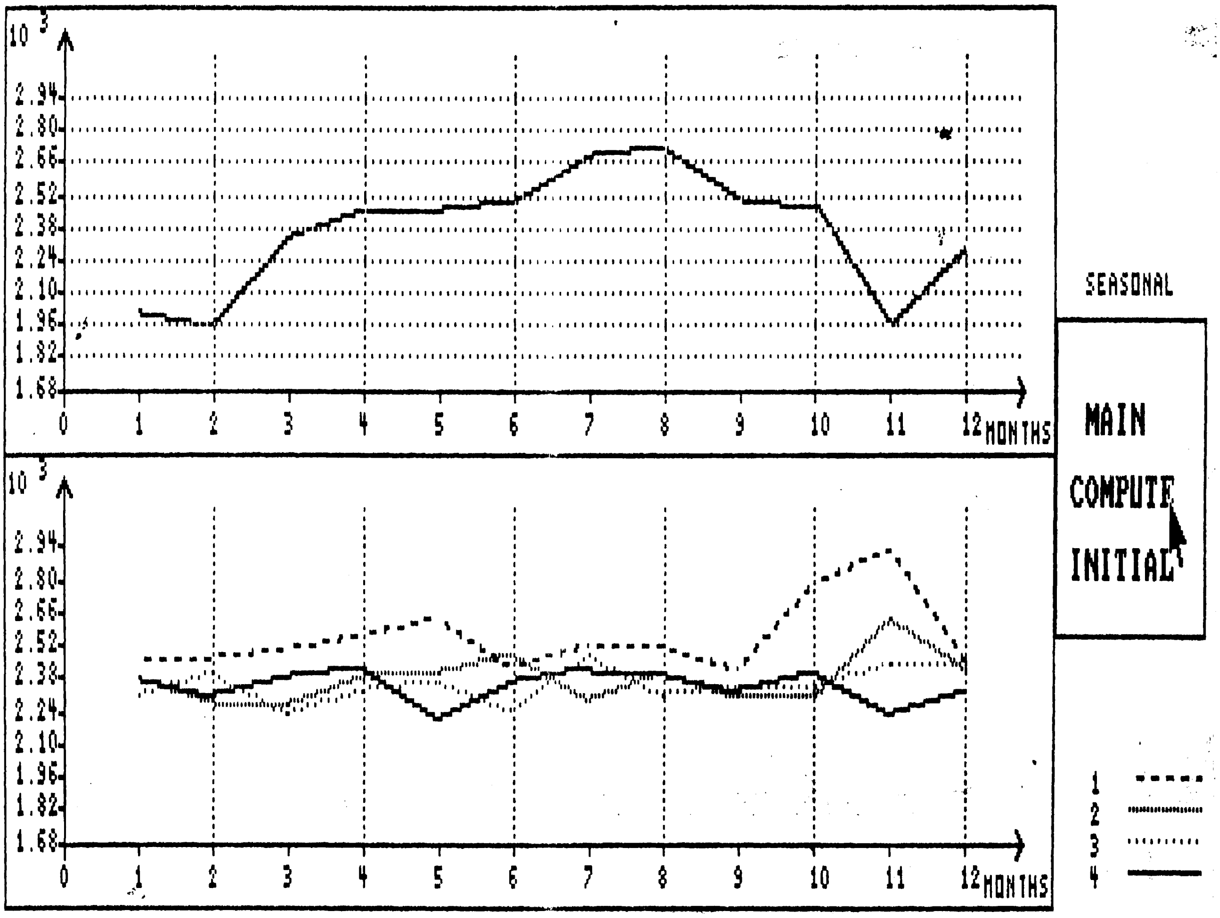


Figure 5.6 Removal of the Seasonal Model

Having returned to the Main Menu, and being satisfied with the results of the trend and seasonal modelling, the forecast module may be entered. Figure 5.7 shows the extrapolation line from the residual time series.

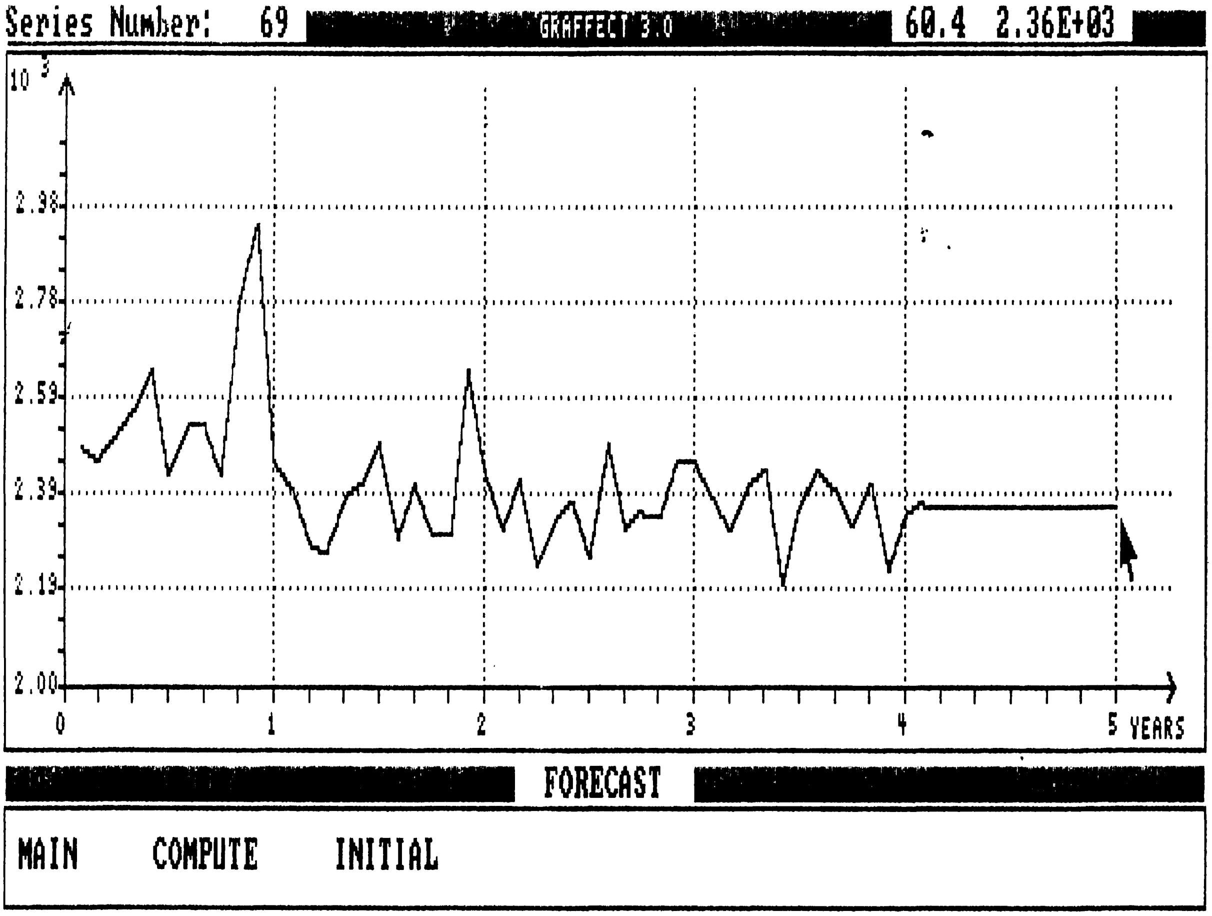


Figure 5.7 Extrapolation from the Residual Series

Extrapolated values may be entered for individual months, or for groups of months. In the latter case, the value for the first and last months of the group are entered, and GRAFFECT interpolates the values for the intermediate months. In the illustration, an horizontal extrapolation is illustrated, reflecting the forecaster's belief that there is no signal remaining in the residual series. After computation of the extrapolation factors and return to the Main Menu the screen displays the final forecast, as shown in figure 5.8.

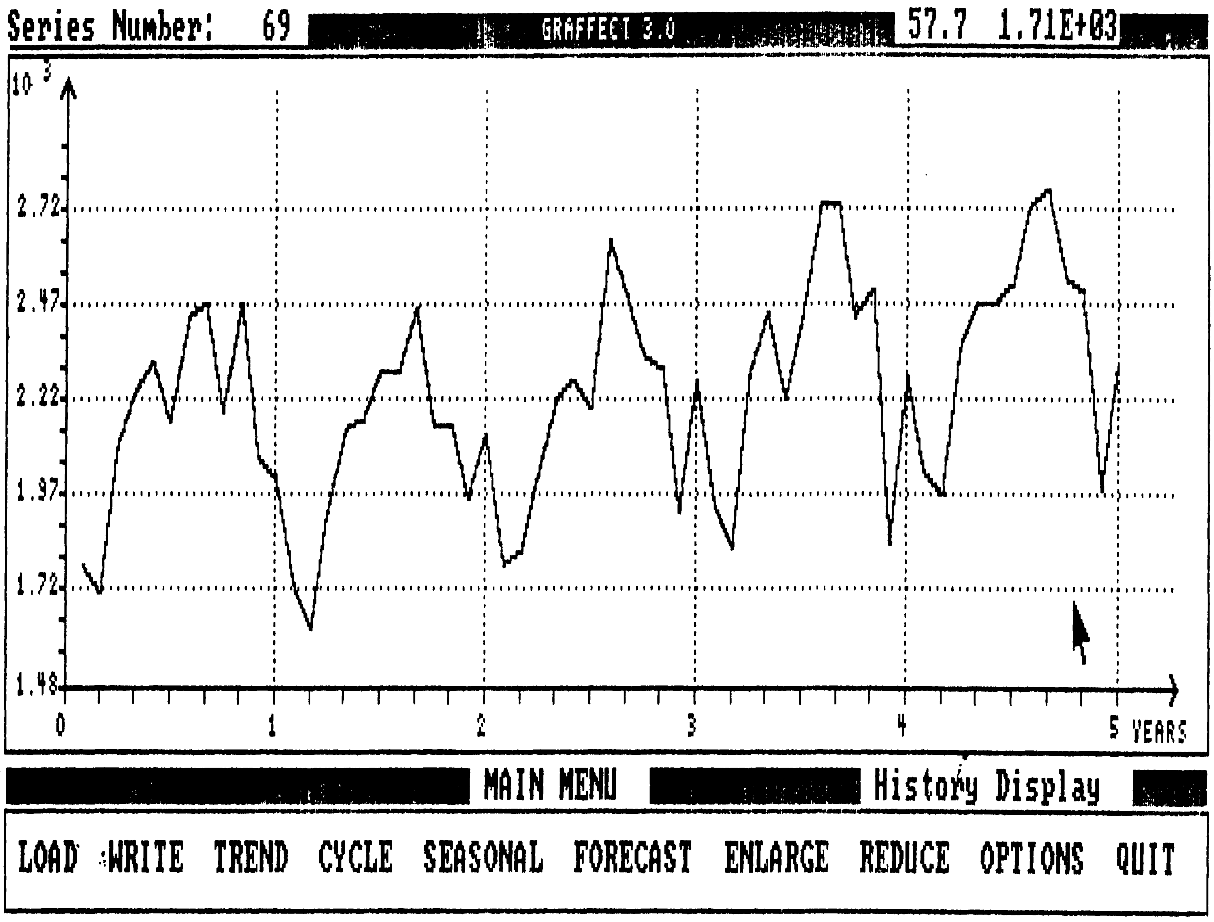


Figure 5.8 The Forecast

The whole operation, or any part, may be repeated until the judge is satisfied with the forecast. Naturally, GRAFFECT stores the models constructed and the extrapolation on disk for analysis or future use.

5.10 REFERENCES

Felsen J., "A Man-Machine Investment Decision System.", *Int. Man-Machine Studies*, 8,(1976), 169-179.

Fenker I., and Evans S.H., "A model for optimizing the effectiveness of man-machine decision making in a pattern recognition system. *Report AD-730-944.* Springfield, Va: National Technical Information Service. (1971).

Kosaka T. and Hirouchi T. "An effective architecture for decision support systems" *Information and Management*. 5, (1982), 7-17.

Lawrence, M.J., & Makridakis, S., "Human judgment in extrapolation", *International Forecasting Symposium*, Paris, (1986).

Makridakis S., and Wheelwright S.C., *Forecasting Methods and Applications*. Wiley. Santa Barbera. (1978).

Remus W., "An Empirical Investigation of the Impact of Graphical and Tabular Data Presentations on Decision Making." *Management Science*, 30,5 (1984), 533-542.

Slovic P., Fischhoff B., & Lichtenstein S., "Behavioral Decision Theory." *Ann. Rev. Psychol*. 28 (1977), 1-39.

APPENDIX A

GRAFFECT USER DOCUMENTATION

INTRODUCTION.

This document describes the operation of the GRAFFECT package. An example of the use of the package is presented in section 5.9. That section contains illustrations of the screens presented.

After a description of the hardware requirements, and general comment on running the package, the documentation is organised to correspond to the control structure of the package.

HARDWARE REQUIREMENTS.

In order to operate the package you will require an IBM Personal Computer with two disk drives, graphics card and monitor, a MICROSOFT MOUSE, and at least 128K of memory.

The software has been written in Turbo Pascal, and requires a number of library functions to be present on the default disk.

RUNNING THE SOFTWARE.

For GRAFFECT to operate, the mouse driver software must be resident before the Graffect package is invoked. At start-up the package provides a data entry screen that permits the user to enter an identifying name, and to reset the default drive. These housekeeping functions are only available at start-up. After completion of the start-up data entry the user enters the functional phase of the package.

THE STRUCTURE OF THE PACKAGE.

OVERVIEW.

The package provides a screen oriented interface to the time series data. Each screen has two areas, in one the user is offered a menu of GRAFFECT functions appropriate to that screen. The other area contains displays of the data, and is used for interactive data manipulation.

THE CONTROL STRUCTURE OF THE PACKAGE.

The various functions provided by GRAFFECT are arranged in a series of independent modules callable from a 'MAIN' module. The 'MAIN' module also provides certain services such as data loading and scaling. The overall structure is as shown in figure 5A.1.

```
                      MAIN
                        |
    |------------|---------|---------|--------------|
    |            |                   |              |
DE-TRENDING  DE-CYCLING  DE-SEASONALISING  FORECASTING
```

Figure 5A.1 GRAFFECT Control Structure

Upon entry to GRAFFECT the user is placed in the main module. The user will probably wish to 'LOAD' a time series of interest, and initially no other function will work until a time series is loaded. As with all GRAFFECT operations, 'LOAD' may be invoked in either of the following ways:

1. Pressing the initial letter of the desired operation as shown in the menu.
2. Placing the head of the 'MOUSE' cursor on the selected operation, and pressing the LEFT button.

If additional parameters are required for the operation, the user will be prompted to enter them. As described in more detail below, interactive data manipulation is performed as follows:

1. Placing the head of the 'MOUSE' cursor in the desired position within the data display or model display areas of the screen and pressing the LEFT button.

MAIN MODULE.

Offers the other modules, and provides certain core operations such as data loading, plotting etc.

THE OPERATIONS OFFERED.

LOAD This permits a new time series to be loaded from the file. The user will be prompted for the series number to be loaded. If a time series is currently loaded the user will be required to confirm.

WRITE This operation is not available in this version of the software.

TREND Enters the de-trending module.

CYCLE Enters the de-cycling module.

SEASONAL Enters the de-seasonalising module.

FORECAST Enters the forecasting or extrapolation module.

ENLARGE This operation changes the scale factor for the main, trend, cycle, and forecast screens, to display the plot larger on the screen.

REDUCE This operation reduces the size of the plot, and is the converse of enlarge.

QUIT Exits the package. The user will be requested to confirm the wish to quit.

OPTION This operation provides several optional functions including producing a plot of the main screen data display on an Epson dot matrix printer, and changing the data presentation in the main screen.

TREND MODULE.

Permits the user to determine the trend component in the time series. The trend is entered interactively by placing a line through the plotted time series. The module permits the user to set the start (left hand) point for the trend line, thus taking account of a turning point in the series. A second line may be entered to reflect the trend component prior to the turning point. Finally, the user may wish to model a different trend for the extrapolation, and this may be carried out using the Extend function. The user may review the trend components identified, and either adjust them or abandon them.

THE OPERATIONS OFFERED.

MAIN Returns to the main module. If a 'compute' operation has not been performed the user is requested to confirm the return to main. Failure to 'compute' will lose the effect of any manipulations carried out.

COMPUTE Determines the trend component factors and adjusts the 'user view' of the data to wash out the trend. The screen plot of the data does not change until the main module is entered.

FIRST This permits the user to choose to enter or adjust the primary trend line. That line is the trend line that terminates on the most recent observation month of the time series. See below for details on data entry.

INITIAL Re-initialises the trend component factors to '1' to allow the user to abandon any work carried out. This operation will effectively reverse the 'compute' operation, and return the data to the state pertaining before the trend module was entered.

ENLARGE This operation is not available in this version of the software.

REDUCE This operation is not available in this version of the software.

QUIT Exits the package. The user will be requested to confirm.

LINEAR This operation is not available in this version of the software.

SECOND Permits the user to enter or adjust the secondary trend line. That line is the trend line prior to a turning point, and it terminates at the primary trend line. Note that reselecting 'first' eliminates any existing secondary trend line, which may subsequently re-entered.

XTEND Permits the user to enter or adjust the extrapolation model of the trend line

INTERACTIVE DATA ENTRY.

The user indicates the trend component to be modelled by plotting trend lines on the data display part of the screen. The right hand end of the primary (first) trend line is anchored at the mean of the three most recent observations. The left hand end of the proposed trend line may be placed at any point in the grid to the left of the anchored end. The line may be repositioned at will until a 'compute' is invoked.

THE CYCLE MODULE

Permits any cyclic component to be washed out of the time series. The screen, and functionality for this module is identical to the TREND module. For a description of the module please refer to the description of the trend module.

THE SEASONAL MODULE.

Permits the identification and elimination of the seasonal component of the time series. It should be noted that the 'seasonal' screen appears to differ from other screens in that there are seemingly three areas. However, the display area is merely divided into a data display area, and a model display area. The shape of the seasonal process may be input by "clicking" the mouse at an appropriate point in either of the display areas.

Note that executing the 'calculate' command results in the re-plotting of the time series in the data display area, with the identified seasonal component eliminated.

THE OPTIONS OFFERED.

MAIN Returns to the main menu. If a 'compute' operation has not been invoked the user will be required to confirm.

COMPUTE Calculates the values of the seasonal component factors, and washes the seasonal component out of the time series.

INITIAL Re-initialises the seasonal factors to 1, negating the effect of a previous compute.

THE FORECAST MODULE.

Permits the extrapolation of the residual time series after the elimination of the trend, cycle, and seasonal components. The screen displays the resultant series, and provides for the entry of twelve extrapolation points. Note that it is necessary to forecast all 12 points, and that the software will not permit you to proceed unless all 12 have been entered. Entry of a data point will result in the extrapolation line being plotted. If you choose to enter a point other than the next in sequence then the values for the intermediate points will be interpolated.

THE OPTIONS OFFERED.

MAIN Returns to the main menu. If a 'compute' operation has not been invoked the user will be required to confirm.

COMPUTE Captures the values of the extrapolated series and computes the values of the forecast series (with the trend, cycle, and seasonal factors included). On return from the forecast function, the composite forecast will be displayed, and the history data will contain all information (ie. it will be a plot of the original data). The display on the screen may be toggled between this "history" display, and the "residual" display by using the options function.

INITIAL Not available in this version.

THE OPTION MODULE.

Offers two modes of main module screen display, and output to the printer. The data display is as for the main module. The menu offers the options described below.

OPTIONS OFFERED.

MAIN Returns to main menu.

PLOT Outputs an image of the main screen data to the printer.

DISPLAY Toggles either 'history' time series or 'residual' time series to be displayed in the main module.

APPENDIX B

```
program graffect;


(*


   Pascal
   Graffect  Date
   Version   written    Author        Remarks
   ------------------------------------------------------------------------
   3.0 P     14-3-1986  Re-write by   Monochrome GRAPHIX Toolbox version.
                       R.H.Edmundson

                                      Requires data files of the form
                                      tseries.nnn in the active directory




   -------------------------------------------------------------------------
   PLEASE NOTE:

   This version of Graffect has used the updated version of Graphix Toolbox
   which can use exclusive-or mode of DrawLine, Gaxis.hgh and DrawLinInwi.i.

   The .PA comment are for Source Lister to perform
   page advances and .I- comment to turn off include file listing.
   ------------------------------------------------------------

*)
{.I-}
{$I typedef.sys}
{$I graphix.sys}
{$I kernel.sys}
{$I windows.sys}
{$I gaxis.hgh}
{$I polygon.hgh}
{$I mousey.i}
{$I drwlinwi.i}
type

     { These Types are for easy reference when writing the program }

     { Main menu }
     menutype= (load,   print,   trend, cycle, seasonal, forecast, enlarge,
                reduce, options, quit,  null);

     { Trend and Cycle both use this menu }
     tremenu = (main,   compute, first, second, initial, tenlarge, treduce,
                xtend, plot,    bye,   nul);

     { Seasonal and Forecast both use this }
     sfmenu  = (smain, scompute, sinitial, snull);

     { Options uses this }
     omenu   = (omain, oplot, display, dhardcopy, onull);
```

```
  { SERIES NUMBER TYPE }

  sertype = array [1..60] of real;


  { Determine whether mousey returns a location in the graph }

  returntype = ( screenloc, graph ); { function MgetMouse }

  PrintString = String[10];



var

  number,                       { Number of series entered           }
  x, y,                         { Screen location return by Mousey   }
  FirstS, LastS,                { Upper and Lower bound of the series }
  startseries : integer;        { Starting Pointer of the series     }
  mbutton, mdummyx, mdummy : integer; {main module mouse variables}
  Rx, Ry : real;                { Graph values converted from MgetXY  }

  option : menutype;            { Main menu Option selected          }
  toption : tremenu;
  sfoption : sfmenu;
  ooption : omenu;

  computed, loaded, buttonpress, disphist: boolean;     { Flag indicators          }

  History,                      { History  Multiplier                }
  trendx,                       { Trend    Multiplier                }
  cyclex,                       { Cycle    Multiplier                }
  seasx,                        { Seasonal Multiplier                }
  residx : sertype;             { Residual Multiplier                }


  ActiveD, INfile,              { Active Directory, Input file name  }
  UserFileName,
  UserName:string[18];          { User surname and output file       }

  ch,DriveNumber : char;        { Disk Drive for output data         }

  Outfile    : text;

  Scratchpad: plotarray;


(* These are used to store the screen location for the various menus *)


const

  menusel : array [menutype] of integer =          { Main Menu Location }

          (21,74,129,184,253,331,407,479,547,605,0);
```

```
menutre : array [tremenu] of  integer =         { Trend/Cycle Location }

         (21,84,145,205,273,347,415,479,536,581,0);

menuf   : array [sfmenu] of   integer =         { Forecast Location }

         (30,116,210,0);

menus   : array [sfmenu] of   integer =         { Seasonal Location }

         (90,105,121,0);

menuo   : array [omenu] of    integer =         { Option Location }

         (45,110,187,266,0);


function time: PrintString;
type
  regpack = record
              ax,bx,cx,dx,bp,di,si,ds,es,flags: integer;
            end;

var
  recpack:         regpack;             {assign record}
  ah,al,ch,cl,dh:  byte;
  hour,min,sec:    string[2];

begin
  ah := $2c;                            {initialize correct registers}
  with recpack do
  begin
    ax := ah shl 8 + al;
  end;
  intr($21,recpack);                    {call interrupt}
  with recpack do
  begin
    str(cx shr 8,hour);                 {convert to string}
    str(cx mod 256,min);                      { " }
    str(dx shr 8,sec);                        { " }
  end;
  time := hour+':'+min+':'+sec;
end;


function Date: PrintString;
type
  regpack = record
              ax,bx,cx,dx,bp,si,ds,es,flags: integer;
            end;

var
  recpack:       regpack;               {record for MsDos call}
  month,day:     string[2];
  year:          string[4];
  dx,cx:         integer;

begin
  with recpack do
```

```
      begin
        ax := $2a shl 8;
      end;
      MsDos(recpack);                   { call function }
      with recpack do
      begin
        str(cx,year);                   {convert to string}
        str(dx mod 256,day);                { " }
        str(dx shr 8,month);                { " }
      end;
      date := day + '/' + month + '/' + year;
    end;

function MgetXY:returntype;

{
  Mgetxy decodes the last button down position.
  It returns two sets of values:

  Rx and Rx are the real values in the graph
  x and y are absolute screen location for the menus

  When the button is not pressed, the actual values of Rx and Ry
  are displayed on the upper right corner of the screen.
}

var
    button, dummyx, dummy : integer;
    temp: returntype;

begin
     repeat
          GetMousePos(button,x,y);
          gotoXY(60,1);
          if (Y>10) and (Y<150) and (x>29) then begin
              MgetXY := graph;
              with world[1] do begin
                  Rx := (x-30)/589*(X2-X1)+X1;
                  Ry := (149-y)/135*(Y2-Y1)+Y1;
                  write(Rx:5:1,'  ',Ry:8)       { display on upper right }
              end
          end
          else begin
              write('                 ');       { clear if not in graph }
              MgetXY := screenloc;
          end;

              buttonpress := false;

              if Keypressed then
              begin
                  buttonpress := true;
                  MgetXY := screenloc;
                  button := 1;
                  read(kbd,ch);
                  case ch of
                     'L','l':begin
                                 option := load;
                             end;
```

```
                    'P','p':begin
                                option := print;
                                toption := plot;
                                ooption := oplot
                            end;
                    'T','t': option := trend;
                    'S','s':begin
                                 option := seasonal;
                                 toption := second;
                            end;
                    'F','f':begin
                                 option := forecast;
                                 toption := first;
                            end;
                    'E','e':begin
                                option := enlarge;
                                toption := tenlarge;
                            end;
                    'R','r':begin
                                option := reduce;
                                toption := treduce;
                            end;
                    'O','o': option := options;
                    'Q','q': option := quit;
                    'I','i':begin
                                 sfoption := sinitial;
                                 toption := initial;
                            end;
                    'M','m': begin
                                 ooption := omain;
                                 option := cycle;
                                 sfoption := smain;
                                 toption := main;
                             end;
                    'C','c': begin
                                 sfoption := scompute;
                                 toption := compute;
                             end;
                    'D','d': ooption := display;
                    'H','h': ooption := dhardcopy;
                    'X','x': toption := xtend;
                 end;
             end;

     until button <> 0;


     (* Software KeyDebounce: prevent multiple entry *)

     repeat GetMousePos(button,dummyx,dummy) until button = 0

end;


procedure makemenu; { Paint menu window }
begin
     SelectWindow(2);
     SelectWorld(2);
     SetBackground(0);
```

```
     Drawborder;
     gotoXY(32,22); write(' MAIN MENU  ');
     if disphist then
     begin
        gotoXY(58,22);write(' History Display  ')
     end;

     if not disphist then
     begin
         gotoXY(58,22);write(' Residual Display ')
     end;
     gotoXY(2,24);
     write('LOAD  WRITE  TREND  CYCLE  SEASONAL  FORECAST  ENLARGE  REDUCE  OPTIONS  QUIT')
END;



procedure init;      { Initialises the parameters }


var
       i : integer;
      t1 : menutype;
      t2 : tremenu;
      t3 : sfmenu;
      t4 : omenu;
    para : text;

begin
     assign(para,'params.dat');
     reset(para);
     read(para,FirstS,LastS);
     close(para);
     loaded := false;
     for i := 1 to 60 do
     begin
          trendx[i] := 1;
          cyclex[i] := 1;
          seasx[i]  := 1
     end
end;



var npts          : integer;    { Number of points limit      }
    scalefactor   : real;       { Scale factor for the Y-Axis }



procedure CalResidual;

{ Computes Residuals }


var
     i : integer;

begin
```

```
     for i := 1 to 48 do
         residx[i] := History[i] * trendx[i] * cyclex[i] * seasx[i];
         if residx[i] < 0 then residx[i] := 0;
     for i := 49 to 60 do
         begin
         History[i] := residx[i] / trendx[i] / cyclex[i] / seasx[i];
         end;
end;


procedure FindMaxmin(N:integer;var Max,Min:real; var SER:sertype);

{ A Linear search for the maximum and minimum values in an array   }

var
     i, fstpt, lstpt : integer;

begin
     Max := -9.9e32;
     fstpt := startseries;
     if fstpt < 1 then fstpt := 1;
     if N < 48 then lstpt := 48 else lstpt := N;
     Min := SER[fstpt];
     for i := fstpt to lstpt do begin
         if SER[i] > Max then Max := SER[i];
         if SER[i] < Min then Min := SER[i];
     end;
     Max := Max * 1.1;
     Min := Min * 0.9;

end;


procedure GraphResidual(npts:integer);

{ Draw Axis and graph residuals or history depending on flag disphist}


var Max,Min,temp,temp2:real;
    R:PlotArray;
    i, fstpt, lstpt: integer;
begin
     HideMouse;
     CalResidual;
     SelectWindow(1);
     if not disphist then
     begin
          FindMaxmin(npts,Max,Min,residx)
     end;
     if disphist then

     begin
         FindMaxmin(npts,Max,Min,History)
     end;

     temp2 := (Max - Min) / 2 + Min;
     temp := (Max - Min)/2 * scalefactor;   { Expand world according
```

```
                                    to scale factor        }
with world[1] do begin
     Y1 := temp2 - temp;
     Y2 := temp2 + temp;  { Due to the BUG in GRAPHIX }
     X1 := 0;
     X2 := 63.37
end;


SelectWorld(1);
Setbackground(0);
SetHeaderOn;
DrawBorder;
DrawAxis(2,7,0,0,0,0,0,0,$03);

SetLinestyle($AA);
                                    { Request }
for i := 0 to 5 do begin
  drawLineInWin(i*12,world[1].Y1,i*12,world[1].Y2);
  setwindowmodeoff;
  drawtext(i*112+28,157,1,chr(i+ord('0')));
  setwindowmodeon
end;

setLineStyle(0);

setwindowmodeoff;
for i := 1 to 30 do
  drawline(round(i*18.666)+31,151,round(i*18.666)+31,154);
drawtext(600,158,1,'YEARS');
setwindowmodeon;

fstpt := startseries;
if fstpt < 1 then fstpt := 1;
if npts <= 48 then lstpt := 48 else lstpt := npts;

if not disphist then
begin
    for i := 1 to lstpt do begin
        R[i,1] := i;
        R[i,2] := residx[i]
        end;
end;

if disphist then
begin
    for i :=  1 to lstpt do begin
        R[i,1] := i;
        R[i,2] := History[i]
        end;
end;

DrawPolygon(R,fstpt,-lstpt,0,0,0);

gotoXY(1,1); write('Series Number: ', number:4,' ');
ShowMouse;
end;
```

```
procedure GraphResidSeas(npts:integer);
begin
end;



procedure Doenlarge;
begin
     gotoXY(1,22); write('Enlarging');
     scalefactor := scalefactor / 1.3;
     GraphResidual(npts)
end;



procedure Doreduce;
begin
     gotoXY(1,22); write('Reducing');
     scalefactor := scalefactor * 1.3;
     Graphresidual(npts)
end;



function decode(x,y:integer):menutype;  { What option is this for Main Menu }
var
     i : menutype;

begin
     if (y<190) and (y>180) then
     begin
        i := load;
        while not (abs(x-menusel[i])<15) and (i < null) do
              i := succ(i);
        decode := i
     end
     else decode := null
end;



procedure Doload;     { Load a series from the input file }
var
     ch        : char;
     i, sernum, filnum  : integer;
     Numstring : String[5];
     datafile  : text;
     tmp: real;

begin
     HideMouse;
     SelectWindow(2);
     SelectWorld(2);
     SetBackground(0);
```

```
if loaded then     { Query Load }
begin

     gotoXY(2,23); writeln('ALREADY LOADED WITH A SERIES');
     write('Load a new Series? (y-n)');
     read(kbd,ch);
     loaded := Upcase(ch) <> 'Y';
     setbackground(0)
end;
if not loaded then
begin
     loaded := true;
     disphist := false;
     for i := 1 to 60 do
     begin
          trendx[i] := 1;
          cyclex[i] := 1;
          seasx[i] := 1
     end;
     gotoXY(1,22);
     write('LOADING');
     scalefactor := 1;
     repeat
           gotoXY(2,24);
           write('Series Number:         ');
           gotoXY(17,24);
           read(number);
           setbackground(0);
           if (number<FirstS) or (number>LastS) then
              write('NUMBER OUT OF RANGE')
     until (number>=FirstS) and (number<=LastS);

    if number > 49 then INfile := '5.dat' else INfile := '4.dat';
    if number > 59 then INfile := '6.dat';
    if number > 69 then INfile := '7.dat';
    if number > 79 then INfile := '8.dat';
    if number > 89 then INfile := '9.dat';
    if number > 99 then INfile := '10.dat';
    if ActiveD = '\' then INfile := ActiveD + 'tser' + INfile else
     INfile := ActiveD + '\tser' + INfile;


     assign(datafile,INfile);
     reset(datafile);


     while (number<>sernum) and not eof(datafile) do
     begin
          readln(datafile,sernum,npts);
          if npts>48 then npts := 48;
          startseries := 49 - npts;

          for i := 1 to 48 do
              if i<startseries then History[i] := 0
              else read(datafile,History[i]);
              if eoln(datafile) then readln(datafile);
     end;
     number := sernum;
     close(datafile);
```

```
        GraphResidual(npts);
        str(number,NumString);

        UserFileName := DriveNumber+':'+UserName+Numstring+'.'+'dat';
        Assign(outfile,Userfilename);
        rewrite(Outfile);
        writeln(Outfile,'Graffect ',date);
        close(Outfile)
    end;
    ShowMouse
end;




procedure Doprint;
begin
    gotoXY(1,22);
    write('WRITE');
end;




var trendline          :integer;
    ytemp1,
    ytemp2,ytemp3,ytemp4     :real;

    xtemp1,
    xtemp2,xtemp3,xtemp4,
    XMAX,XMIN,monthist,
    month2x,month3x,month4x   :integer;
    tr1,tr2,tr3               :boolean;
const
    trendkind : array[1..3] of String[8] = (' First  ',' Second ','xtend');




procedure drawLin_cross(x2,y2,x1,y1:real);

{  Draw a line with crosses at the end points.

   This is a fake drawline routine because of the BUG,
   the aim is to purposely fool DrawPolygon into thinking
   that there are three points to be plotted. In this way
   the crosses at the end point can also be drawn }

var a:plotarray;
begin
    resetaxis;
    a[1,1] := x1;
    a[2,1] := x1;
    a[3,1] := x2;
    a[1,2] := y1;
    a[2,2] := y1;
    a[3,2] := y2;
```

```
    drawpolygon(a,1,3,1,5,0)    { 1,5 -> crosses @ scale of 5 }
end;



procedure movepoint;

begin
     computed := false;
     SelectWindow(1);
     SelectWorld(1);
     HideMouse;
     XorLine := true;

     if trendline = 1 then
      begin
          DrawLin_cross(xtemp2,ytemp2,xtemp1,ytemp1);
          xtemp2 := round(rx);
          month2x := monthist;
          ytemp2 := ry;
          xtemp3 := xtemp2;
          ytemp3 := ytemp2;
          month3x := monthist;
          drawLin_cross(xtemp2,ytemp2,xtemp1,ytemp1)
     end;

     if trendline = 2 then
      begin
          DrawLin_cross(xtemp3,ytemp3,xtemp2,ytemp2);
          xtemp3 := round(rx);
          ytemp3 := ry;
          month3x := monthist;
          DrawLin_cross(xtemp3,ytemp3,xtemp2,ytemp2);
          if xtemp3 = xtemp2 then tr2 := false else tr2 := true;
     end;

     if trendline = 3 then
      begin
          DrawLin_cross(xtemp4,ytemp4,xtemp1,ytemp1);
          xtemp4 := round(rx);
          ytemp4 := ry;
          month4x := monthist;
          DrawLin_cross(xtemp4,ytemp4,xtemp1,ytemp1);
          if xtemp4 = 48 then tr3 := false else tr3 := true;
     end;

     ShowMouse;
     XorLine := false;
end;



procedure SettrendPar;
var
    sum   : real;
    month : integer;

begin
```

```
     sum := 0;
     monthist := 48;
     for month := 45 to 48 do sum := sum + residx[month];
     ytemp1 := sum /4;
     ytemp2 := ytemp1;
     ytemp3 := ytemp1;
     ytemp4 := ytemp1;
     xtemp1 := 48;
     xtemp2 := 48;
     xtemp3 := 48;
     xtemp4 := 48;
     XMIN   := startseries;
     XMAX   := 60;
     tr2 := false;
     tr3 := false;
end;




function tdecode(x,y:integer):tremenu; { Use by both Trend and Cycle }
var
     tt : tremenu;

begin
     if (y<190) and (y>180) then
     begin
          tt := main;
          while not (abs(menutre[tt]-x)<15) and (tt < nul) do
                tt := succ(tt);
          tdecode := tt
     end
     else tdecode := nul
end;




procedure PrintVector(vector: sertype);
var i,j:integer;
begin
   for i := 0 to 9 do
     begin
        for j := 1 to 6 do
          write(OUTfile,vector[i*6+j]:13:6);
        writeln(OUtfile)
     end
end;



procedure CalTrendCycle(var vector:sertype);
var
  gradient : real;
  i        : integer;

begin
```

```
  computed := true;
  gotoXY(1,22); write('Computing...');

  Gradient := (ytemp1-ytemp2)/(48-month2x);

  for i :=  1 to 60 do
    vector[i] := ytemp2 + ((i-month2x) * gradient);

if tr2 then
 begin
    gradient := (ytemp2-ytemp3)/(month2x-month3x);

  for i := 1 to month2x do
      vector[i] := ytemp2 + ((i-month2x) * gradient);
 end;

if tr3 then
 begin
    gradient := (ytemp4-ytemp1)/(month4x-48);

  for i := 49 to 60 do
      vector[i] := ytemp1 + ((i-48) * gradient);
 end;

  for i := 1 to 60 do
      vector[i] := ytemp1 / vector[i];

  assign(OUTfile,UserFileName);
  Append(Outfile);
  writeln(Outfile);
  if option = trend then writeln(OUTfile,'Trend ',time,' ',month2x:3)
  else writeln(OUTfile,'Cycle ',time,' ',month2x:3);

  PrintVector(vector);

  close(OUTfile);
end;



procedure FirstTrend;
var
     i : integer;

begin
     HideMouse;
     SelectWindow(1);
     SelectWorld(1);
     gotoXY(1,22); write('First');
     ShowMouse;
     if computed then begin
        for i := startseries to 60 do
            if option = trend then trendx[i] := 1
            else cyclex[i] := 1;
        computed := false
     end;
     if trendline = 2 then begin
        trendline := 1;
        xorline := true;
```

```
        drawlin_cross(xtemp3,ytemp3,xtemp2,ytemp2);
        xorline := false;
        xtemp3 := xtemp1; ytemp3 := ytemp1;
        monthist := month2x
     end
end;


procedure redraw;
begin
    selectwindow(1);
    selectworld(1);
    Xorline := true;
    if trendline = 1 then
       drawLin_cross(xtemp2,ytemp2,xtemp1,ytemp1)
    else
      drawlin_cross(xtemp3,ytemp3,xtemp2,ytemp2);
    xorline := false;
end;


procedure SecondTrend;
begin
  trendline := 2;
  redraw;
end;

procedure XtendTrend;
begin
     trendline := 3;
end;

procedure test;
var
   ch   : char;

begin
     SelectWindow(2);
     SetBackground(0);
     HideMouse;
     setforegroundcolor(12);
     gotoXY(1,24);
     write('Not computed. REALLY QUIT?');
     read(kbd,ch);
     if Upcase(ch) <> 'Y' then toption := nul;
     setforegroundcolor(11);
     ShowMouse
end;




procedure Dotrend;
var
     corr  : boolean;
     i     : integer;
```

```
begin
    trendline := 1;
    computed := true;
    SetTrendPar;
    redraw;
    corr := true;
    repeat
         if corr then begin
            HideMouse;
            SelectWorld(2);
            SelectWindow(2);
            SetBackground(0);
            DrawBorder;
            gotoXY(72,22);
            write(trendkind[trendline]);
            gotoXY(35,22);
            write('  TREND  ');
            gotoXY(2,24);
            write('MAIN  COMPUTE  FIRST  SECOND  INITIAL  ENLARGE  REDUCE  XTEND   PLOT  QUIT');
            corr := false;
            ShowMouse
         end;
         toption := nul;
         if MgetXY = screenloc then
         begin
            if not buttonpress then toption := tdecode(x,y);
            corr := true
         end
         else
         begin

              toption := nul;
              i := round(rx);
              if trendline <> 3 then
                 begin
                     if (i < 49) or ((trendline=2) and (i<xtemp2)) then
                        begin
                          monthist := i;
                          movepoint
                        end;
                 end
              else
                 begin
                     if (i <= 60) and (i > 48) then
                        begin
                          monthist := i;
                          movepoint;
                        end;
                 end;
         end;
         case toption of

              main:     if not computed then test;

              compute:  if not computed then CaltrendCycle(trendx);

              first:    FirstTrend;
              second:   SecondTrend;
              initial:   begin
```

```
                        for i := 1 to 60 do trendx[i] := 1;
                        trendline := 1;
                        computed := false;
                        Graphresidual(npts);
                        redraw
                      end;

            tenlarge: begin
                          Doenlarge;
                          redraw
                      end;

            treduce: begin
                         Doreduce;
                         redraw
                     end;
            xtend  : XtendTrend;
            plot:   if computed then graphresidual(npts)
        end
    until toption=main;
    GraphResidual(npts);
end;



procedure Docycle;
var
     corrupt : boolean;
     i       : integer;



begin
     computed := true;
     corrupt := true;
     SetTrendPar;
     Trendline := 1;
     redraw;
     repeat
           if corrupt then begin
               HideMouse;
               corrupt := false;
               SelectWindow(2);
               SelectWorld(2);
               SetBackground(0);
               DrawBorder;
               gotoXY(35,22);
               write('  CYCLE  ');
               gotoXY(72,22);
               write(trendkind[trendline]);
               gotoXY(2,24);
               write('MAIN  COMPUTE  FIRST  SECOND  INITIAL  ENLARGE  REDUCE  LINEAR  PLOT  QUIT');
               ShowMouse;
               toption := nul;
           end;
```

```
        if MgetXY = screenloc then begin
            if not buttonpress then toption := tdecode(x,y);
            corrupt := true
        end
        else begin
            toption := nul;
            i := round(rx);
            if (i < 49) or ((trendline=2) and (i<=xtemp2)) then
             begin
               monthist := i;
               movepoint
             end
        end;
        case toption of
             main:    if not computed then test;
             compute: if not computed then CalTrendCycle(cyclex);
             first:   FirstTrend;
             second:  SecondTrend;
             initial: begin
                          for i := 1 to 60 do cyclex[i] := 1;
                          trendline := 1;
                          computed := false;
                          GraphResidual(npts);
                          redraw;
                      end;
             tenlarge: begin
                            Doenlarge;
                            redraw
                       end;
             treduce:  begin
                            Doreduce;
                            redraw
                       end;
             xtend  :  ;
             plot:    if computed then graphresidual(npts)
        end
    until toption = main;
    GraphResidual(npts)
end;



function SgetXY:returntype;

{ This is slightly different from MgetXY. It can return two graphical
  screen values so that a point can be reflected to the top           }

var
   i, button, dummyx,dummy: integer;
   temp:returntype;

begin
     repeat
         GetMousePos(button,x,y);
         temp := screenloc;
         for i := 3 to 4 do
             if (y<window[i].y2-14) and (y>window[i].y1+3) and
                (x>30) and (x<512) then
```

```
            begin
              temp := graph;

              rx := (x - 30) / 481 * 12;

              ry := (window[i].y2 - y - 15) / 82 *
                    (world[3].y2 - world[3].y1) + world[3].y1

            end; {phew!}

        gotoXY(73,1);
        if temp = graph then begin
            write(Rx:6:1);
            gotoXY(72,2); write(Ry:8)
        end else
        begin
            write('      ');
            gotoXY(72,2); write('        ')
        end;

            buttonpress := false;

            if Keypressed then
            begin
                buttonpress := true;
                temp := screenloc;
                button := 1;
                read(kbd,ch);
                case ch of
                   'I','i':begin
                               sfoption := sinitial;
                               toption := initial;
                           end;
                   'M','m': begin
                               ooption := omain;
                               sfoption := smain;
                               toption := main;
                            end;
                   'C','c': begin
                               sfoption := scompute;
                               toption := compute;
                            end;

                end;
            end;

    until button <> 0;

    SgetXY := temp;

    (* ditto *)
    repeat GetMousePos(button,dummyx,dummy) until button = 0;

end;
```

```
function fdecode(x,y:integer):sfmenu;
var
     sf : sfmenu;

begin
     if (y<190) and (y>180) then
     begin
          sf := smain;
          while not (abs(x-menuf[sf])<15) and (sf<snull) do
               sf := succ(sf);
          fdecode := sf
     end
     else fdecode := snull
end;


function sdecode(x,y:integer):sfmenu;
var
     sf : sfmenu;

begin
     if (x<625) and (x>565) then
     begin
          sf := smain;
          while not (abs(y-menus[sf])<5) and (sf<snull) do
               sf := succ(sf);
          sdecode := sf
     end
     else sdecode := snull
end;




procedure plot4year;
var
    R:plotarray;
    i,j,k:integer;

const    bites : array[0..3] of byte=($0F,$AA,$88,$FF);

begin
   selectworld(3);
   selectwindow(4);
   if npts >= 48 then k := 0;
   if npts < 48 then k := 1;
   if npts < 36 then k := 2;
   for i := 1 to 12 do
    R[i,1] := i;

   DrawText(600,160,1,'Legend');

   for i := k to 3 do
     begin
       for j := 1 to 12 do
         R[j,2] := residx[j+i*12];
       setlinestyle(bites[i]);
```

```
        SetWindowModeOff;
        SetClippingOff;


        DrawLine(600,i*7+170,639,i*7+170);

        DrawText(580,i*7+170,1, chr(i+ord('1')));

        SetClippingOn;
        SetWindowModeOn;

        resetaxis;
        DrawPolygon(R,1,-12,0,0,0);
      end;
    setlinestyle(0)
end;



procedure Stest;
var
  ch   : char;

begin
  Selectwindow(5);
  HideMouse;
  SetBackground(0);
  setforegroundcolor(12);
  Drawtext(565,80,1,'Not Computed');
  Drawtext(565,90,1,'Really Quit?');
  read(kbd,ch);
  if Upcase(ch) <> 'Y' then sfoption := snull;
  ShowMouse;
  setforegroundcolor(11)
end;





procedure Doseasonal;
var
     i,j,k     : integer;
     seas    : array [1..12] of real;
     sumseas : real;
     mstring : String[2];

begin
     HideMouse;

     ClearScreen;

     world[3].y2 := world[1].y2;
     world[3].y1 := world[1].y1;

     for i := 1 to 12 do
       begin
```

```
    Scratchpad[i,2] := (world[3].Y2 - world[3].Y1) / 3 + world[3].y1;
    Scratchpad[i,1] := i;
  end;

SelectWorld(3);
SelectWindow(4);

DrawBorder;
DrawAxis(2,9,0,0,0,0,0,0,$01);

Selectworld(3);
SelectWindow(3);

DrawBorder;
DrawAxis(2,9,0,0,0,0,0,0,$07);

setwindowmodeoff;
setclippingoff;
for i := 0 to 12 do
  begin
    j := i * 40 + 32;
    if not odd(i) then
      begin
        setlinestyle($AA);
        drawline(j,85,j,10);
        drawline(j,110,j,185)
      end;
    setlinestyle(0);
    drawline(j,  85, j,   88);
    drawLine(j, 185  , j, 188);

    str(i,mstring);
    drawtext(j-2,  91, 1, mstring);
    drawtext(j-2, 191, 1, mstring);

  end;
drawtext(523, 93,1,'MONTHS');
drawtext(523,193,1,'MONTHS');
setwindowmodeon;
setclippingoff;
setlinestyle(0);


selectwindow(5);
copyscreen;

selectwindow(3);
selectworld(3);
resetaxis;
XorLine := true;
DrawPolygon(Scratchpad,1,-12,0,0,0);
XorLine := false;

plot4year;       computed := true;

sfoption := snull;
repeat
    SelectWindow(5);
    SelectWorld(2);
```

```
SetBackground(0);
DrawBorder;

SetClippingOff;
DrawText(576,60,1,'SEASONAL');
SetClippingOn;

gotoXY(73,12);        write('MAIN');

gotoXY(72,14);        write('COMPUTE');

gotoXY(72,16);        write('INITIAL');

ShowMouse;
sfoption := snull;
if SgetXY = screenloc then
     begin if not buttonpress then sfoption := sdecode(x,y)
end
else begin
   Selectworld(3);
   SelectWindow(3);
   HideMouse;
   sfoption := snull;
   XorLine := true;
   ResetAxis;
   DrawPolygon(Scratchpad,1,-12,0,0,0);
   ResetAxis;
   Scratchpad[round(Rx),2] :=  Ry;
   DrawPolygon(Scratchpad,1,-12,0,0,0);
   ShowMouse;
   XorLine := false;
   computed := false
end;
case sfoption of
     smain:    if not computed then Stest;
     scompute: if not computed then
                 begin
                    HideMouse;
                    swapscreen;  copyscreen;
                    gotoXY(72,14);   write('COMPUTE');

                    assign(outfile,UserFileName);
                    append(outfile);
                    writeln(outfile);

                    sumseas := 0;
                    for i := 1 to 12 do
                      sumseas := sumseas + scratchpad[i,2];
                    for i := 1 to 12 do
                      seas[i] := sumseas / scratchpad[i,2] / 12;
                    gotoXY(22,10);
                    writeln(outfile,'Seasonal: ',' 0 ',time);
                    writeln(outfile);
                    for i := 0 to 4 do
                      for j := 1 to 12 do
                        if i*12+j >= startseries then
                          seasx[i*12+j] := seas[j];
                    PrintVector(seasx);
                    computed := true;
```

```
                        close(outfile)
                     end
                     else sfoption := snull;
         sinitial: if computed then
                     begin
                        HideMouse;
                        swapscreen;
                        copyscreen;
                        gotoXY(72,16);
                        write('INITIAL');
                        for i := startseries to 60 do seasx[i] := 1;
                        for i := 1 to 12 do
                          Scratchpad[i,2] := (world[3].y2-world[3].y1)/3+world[3].y1;
                     end
                   else sfoption := snull
      end;
      if sfoption in [scompute,sinitial] then
        begin
          calResidual;
          selectwindow(3);
          selectworld(3);
          resetaxis;
          XorLine := true;
          DrawPolygon(Scratchpad,1,-12,0,0,0);
          XorLine := false;
          plot4year;
        end
   until sfoption = smain;
   selectworld(1);
   selectwindow(1);
   clearscreen;
   Graphresidual(npts)
end;




procedure Doforecast;
type
     ForecastPointer = ^forecastLogic;
     forecastLogic = record
                       xcoord,ycoord:real;
                       next : ForecastPointer
                     end;

var
     i,j        : integer;
     corrupted,
     found, cor : boolean;

     Sentinel,SentinelPrev,TempLogic,ForecastList:ForecastPointer;

     Gradient : real;


begin
     HideMouse;
```

```
if npts=60 then Graphresidual(48);
npts := 48;
SelectWindow(2);
SelectWorld(2);
new(ForecastList);
with ForecastList^ do begin
    xcoord := 48;
    ycoord := residx[48];
    next   := nil
end;
computed := true;
setbackground(0);
copyscreen;

ShowMouse;
sfoption := snull;
repeat
     HideMouse;
     Selectworld(2);
     Selectwindow(2);
     SetBackground(0);
     DrawBorder;
     gotoXY(35,22);
     write('  FORECAST  ');
     gotoXY(2,24);
     write('MAIN     COMPUTE     INITIAL');
     ShowMouse;
     sfoption := snull;
     if MgetXY = screenloc then
     begin
         if not buttonpress then sfoption := fdecode(x,y);
     end
     else if (Rx>48) and (Rx<=60.4) then
       begin
         HideMouse;
         SelectWorld(1);
         SelectWindow(1);
         sfoption := snull;
         swapscreen;
         copyscreen;

         SentinelPrev := ForecastList;
         Sentinel := ForecastList^.next;
         found := false;

         i := Round(Rx);
         while not found and (Sentinel <> nil) do
           begin
             found := sentinel^.xcoord >= i;
             if not found then begin
               SentinelPrev := Sentinel;
               Sentinel := Sentinel^.next
             end
           end;

         cor :- Sentinel <> nil;
         if cor then cor := sentinel^.xcoord > i;

         if cor or (Sentinel = nil) then begin
```

```
      new(TempLogic);
      SentinelPrev^.next := TempLogic;
      with TempLogic^ do begin
        next := Sentinel;
        Ycoord := Ry;
        Xcoord := i
      end;
    end
    else Sentinel^.ycoord := Ry;

    i := 0;
    Sentinel := ForecastList;
    while Sentinel <> nil do begin
      i := succ(i);
      scratchpad[i,1] := Sentinel^.xcoord;
      scratchpad[i,2] := sentinel^.ycoord;
      Sentinel := Sentinel^.next
    end;
    computed := false;
    resetaxis;
    if i = 2 then
      DrawLineInwin(scratchpad[1,1],scratchpad[1,2],
                    scratchpad[2,1],scratchpad[2,2])
    else
      DrawPolygon(scratchpad,1,-i,0,0,0);
    ShowMouse
  end;
ShowMouse;
case sfoption of
     smain:    if not computed then begin
                 SetBackground(0); HideMouse;
                 SetForegroundcolor(12);
                 gotoXY(1,24);
                 write('Not Computed. Really goto MAIN?');
                 read(kbd,ch);
                 if Upcase(ch) <> 'Y' then sfoption := snull;
                 setforegroundcolor(11);
                 ShowMouse;
               end;
     scompute: if not computed and (scratchpad[i,1]=60) then
                 begin
                   computed := true;
                   gotoXY(1,22); write('Computing.... ');
                   assign(OUTfile,Userfilename);
                   append(OUTfile);
                   writeln(Outfile);
                   writeln(OUTfile,'Forecast: 0 ',time);

                   j := 48;
                   sentinel := ForecastList^.next;
                   while Sentinel <> nil do
                     begin
                       j := succ(j);
                       if sentinel^.xcoord > j then
                        begin
                         i := pred(j);
                         Gradient := (sentinel^.ycoord-residx[i])/(sentinel^.xcoord-i);
                         while j < sentinel^.xcoord do
                           begin
```

```
                                   residx[j] := residx[i]+(j-i)*gradient;
                                   j := succ(j)
                                 end
                              end;
                             residx[j] := Sentinel^.ycoord;
                             sentinel := sentinel^.next
                           end;
                         PrintVector(residx);
                         close(OUTfile);
                         CalResidual;
                         npts := 60
                       end;
            sinitial: if computed then begin end;     { It was not there }
        end
   until sfoption = smain;
   GraphResidual(npts)
end;




function odecode(x,y:integer):omenu;
var
     oo : omenu;

begin
     if (y>180) and (y<190) then
     begin
          oo := omain;
          while not (abs(menuo[oo]-x)<15) and (oo < onull) do
                oo := succ(oo);
          odecode := oo
     end
     else odecode := onull
end;




procedure Dohardcopy;
begin
     HideMouse;
     SelectWindow(2);
     SelectWorld(2);
     SetBackground(0);
     gotoXY(1,24); write('User: ',username);
     HardCopy(false,1);
     ShowMouse
end;

procedure Dodisplay;
begin
     if disphist then disphist := false
     else disphist := true;
end;
```

```
procedure Dooptions;

begin
     HideMouse;
     SelectWindow(2);
     SelectWorld(2);
     repeat
           SetBackground(0);
           DrawBorder;
           gotoXY(35,22);
           write('  OPTIONS  ');
           gotoXY(5,24);
           write('MAIN    PLOT    DISPLAY  HARD-COPY');
           ShowMouse;
           ooption := onull;
           if MgetXY = screenloc then
           begin
               if not buttonpress then ooption := odecode(x,y);
           end;
           case ooption of
                oplot: ;
                display:Dodisplay;
                dhardcopy:Dohardcopy
           end
    until ooption = omain;
    GraphResidual(npts)
end;



procedure Hello;
var
   ch : char;
begin
    UserName    := 'AA';
    DriveNumber := 'C';
    ActiveD     := '\TURBO';
    repeat
        clrscr;
        textcolor(12);
        gotoXY(20,1); write('Graffect Forecasting System');
        gotoXY(20,2); write('***************************');

        textcolor(14);
        gotoXY(1,5);  write('D');
        gotoXY(1,7);  write('U');
        gotoXY(30,5); write('A');

        gotoXY(16,7);  write(UserName);
        gotoXY(16,5);  write(DriveNumber);
        gotoXY(49,5);  write(ActiveD);

        textcolor(7);
        gotoXY(2,5);    write('efault Drive: ');
        gotoXY(2,7);    write('serName     : ');
        gotoXY(31,5);   write('ctive Directory : ');

        textcolor(10);
```

```
      read(kbd,ch);          gotoXY(1,12);
      case ch of
       'A','a': begin
                      write('New Directory: ');
                      read(ActiveD)
                end;
       'U','u': begin
                      write('UserName: ');
                      read(UserName)
                end;
       'D','d': begin
                  DriveNumber := succ(DriveNumber);
                  if DriveNumber > 'D' then DriveNumber := 'A'
                end
      end
    until ch = ^M;
end;



(**********************************************************************)
(*                                                                    *)
(*              M A I N     P R O G R A M                             *)
(*                                                                    *)
(**********************************************************************)

begin
     LeaveGraphic;
     Hello;

     InitGraphic;
     Mousereset;
     setforegroundcolor(11);

     DefineWorld(3,0,100,12.74,0);    { For Seasonal }


     { Top }
     DefineWindow(3, 0,            0 , XMaxGlb-10,  YMaxGlb div 2  );

     { Bottom }
     DefineWindow(4, 0, YMaxGlb div 2 , XMaxGlb-10,  YMaxGlb        );



     { small window for seasonal }
     DefineWindow(5, XMaxGlb-9, YMaxGlb -130 , XMaxGlb, YMaxGlb - 60 );


     { Menu Window for the rest }
     DefineWorld(2,0,0,639,199);
     DefineWindow(2,0,YMaxGlb-32,XMaxGlb,YMaxGlb);
     DefineHeader(2,' ');


     { Main graph }
     DefineWorld(1,0,0,1000,1000);
     DefineWindow(1,0,0,XMaxGlb,YMaxGlb-35);
```

```
    SetHeaderOn;
    DefineHeader(1,'    GRAFFECT 3.0    ');
    selectworld(1);
    selectwindow(1);

    init;


    makemenu;

    repeat
       ShowMouse;
       option := null;
       if MgetXY = screenloc then
         begin
            if not buttonpress then option := decode(x,y);
         end;

         if option = load then  Doload
         else if loaded then
         case option of
               print    : Doprint;
               trend    : Dotrend;
               cycle    : Docycle;
               seasonal : Doseasonal;
               forecast : Doforecast;
               enlarge  : Doenlarge;
               reduce   : Doreduce;
               options  : Dooptions
         end;
         HideMouse;
         makemenu
    until (option = quit);
    leavegraphic
end.
```

6. ACCURACY OF COMPUTER ASSISTED FORECASTING

6.1 INTRODUCTION

In the first study reported in this chapter the accuracy of forecasting using the GRAFFECT decision aid is evaluated. The decision aid was intended for use by line, and general management in forecasting business time series. Case study work carried out by Edmundson, Lawrence, and O'Connor (1987) indicated that it was unlikely that many such managers would possess any training in time series analysis. In light of the above it was deemed necessary to consider the effect of providing the decision aid to persons without, as well as to those with time series analysis skills.

In addition to the examination of the accuracy of forecasts generated using GRAFFECT it was fitting to consider the time taken to produce those forecasts. The second study reported in this chapter addresses this issue.

Finally, a replication of the discriminant analysis study reported in chapter 3 is presented. This study was intended to throw light on any changes in the characteristics of the forecasting process resulting from the use of GRAFFECT. The possibility of developing a decision rule to discriminate time series better forecast using the GRAFFECT decision aid from those better forecast using deseasonalised single exponential smoothing[1] was considered.

6.2 FORECAST ACCURACY STUDIES

6.2.1 THE APPROACH TO EXPERIMENTAL DESIGN

In making comparisons of forecast accuracy a number of techniques are candidates for examination. As discussed in section 6.2.2 below, deseasonalised single exponential smoothing (DSE) is a suitable representative statistical technique and the hard copy graphical method (GRAPH) described by Lawrence, Edmundson and O'Connor (1985) is the base line judgmental method. Thus it would appear that three methods should be evaluated, in the hands of subjects with two differing levels of time series analysis experience.

At first glance, the questions addressed by this chapter appear to fall into a structure suitable for an Analysis of Variance study, with two dimensions, as shown in Figure 6.1.

---

[1] Similar to the decision rule developed in chapter 3 to distinguish between "use of hard copy" and deseasonalised single exponential smoothing.

	SUBJECTS	
F'CAST METHOD	EXPERIENCED	NOVICE
GRAPH	A	B
GRAFFECT	C	D
DSE	E*	E*

* DSE is deterministic and does not vary with experience

Figure 6.1 Dimensions of the Studies

Two features diminish the usefulness of the ANOVA approach for the matters under examination in this chapter. The first is that deseasonalised single exponential smoothing is a deterministic process, and does not fall properly into the structure depicted above. A more telling problem is that the error levels for the time series used in the study vary greatly. For instance, deseasonalised single exponential smoothing exhibits a range of MAPE errors over the sample from about 1 to 78. Analysis of the Lawrence, Edmundson and O'Connor (1985) data showed that tests based on the pooled errors are dominated by a few series.

In order to account for the large variation in MAPE errors across the sample, it was determined to test the experimental effects by t-tests paired on time series. The disadvantages of this approach were that any interaction effects across the two dimensions would not be detectable, and that it would be necessary to control for multiple testing.

The results of pilot studies conducted during the development of GRAFFECT indicated that novice and experienced subjects performed with similar accuracy. This coupled with the results from Lawrence, Edmundson and O'Connor (1985) which showed that there was no difference between experienced and novice subjects using hard copy graphs indicated that there was little likelihood of an interaction effect. In the absence of an unusual result, the "contrasts" of interest from the structure displayed in figure 6.1 are:

C/A GRAFFECT vs. GRAPH with Expert subjects

C/E GRAFFECT with Expert subjects vs. deseasonalised single exponential smoothing.

C/D Expert vs. Novice subjects using GRAFFECT

The threat to the type I error from a multiple use of the data gathered from expert subjects using GRAFFECT might be addressed by either adjusting the significance levels of the tests (using a Bonferroni type adjustment[2]), or by conducting a number of independent tests. As described in section 6.2.4 below, there was opportunity to gather sufficient data for independent tests to be conducted, therefore this was the approach adopted.

6.2.2 JUSTIFICATION OF THE STUDIES

In chapter 2 there was evidence given for the proposition that forecasting is an important management function. Any increase in forecast accuracy can have a significant effect on critical business variables such as

---

[2] See Hays (1981) at page 435.

inventory levels, leading to lower holding costs or reducing "stock-out" costs.

The Lawrence, Edmundson, and O'Connor (1985) study showed that judgment was a viable method for extrapolation, especially for critical time series for which it was necessary to avoid very large errors. The standard deviation of forecast error for the judgmental methods was shown to be approximately 50% of that for statistical methods. Any improvement in judgmental forecasting would be to the advantage of business, especially if the improvement was not at the cost of the standard deviation of forecast error. The benchmark that Lawrence, Edmundson, and O'Connor (1985) used was deseasonalised single exponential smoothing, which performed well in the Makridakis et al (1982) "M-Competition", and that same benchmark was adopted for these studies.

The studies were intended to determine whether:

a) the use of GRAFFECT leads to an improvement in forecast accuracy over the "Graph" based method of Lawrence, Edmundson, and O'Connor (1985),

b) the use of GRAFFECT leads to an improvement in forecast accuracy over deseasonalised single exponential smoothing,

c) there is a difference in the levels of accuracy achieved by persons experienced in time series analysis and those inexperienced.

6.2.3 RESEARCH HYPOTHESES

There was a prior expectation that the use of GRAFFECT would not give rise to errors higher than either the GRAPH or DSE methods. That expectation arose form the results obtained by Lawrence, Edmundson, and O'Connor (1985) and pilot studies conducted with early versions of the GRAFFECT which indicated a likely improvement in accuracy over forecasts supported by hard copy plots. It was therefore possible to state the research hypotheses as:

H6.1 Forecasts made using the GRAFFECT decision aid will not be more accurate than those produced using GRAPH.

H6.2 Forecasts made using the GRAFFECT decision aid will not be more accurate than those produced using DSE.

The comparative accuracy of experienced and novice subjects has implications for the use of GRAFFECT, and for the development of the tool for in-experienced users. The work of Lawrence, Edmundson and O'Connor (1985) did not find a significant difference in the accuracy of experienced and novice forecasters using the GRAPH method. Again, pilot studies conducted during development indicated that novice forecasters did not perform differently to experienced forecasters. The research hypothesis is therefore:

H6.3 Experienced subjects will perform no better using the GRAFFECT decision aid than in-experienced subjects.

6.2.4 DESCRIPTION OF THE EXPERIMENT.

6.2.4.1 THE TIME SERIES DATA

As with the other studies reported in this dissertation, the time series used came from the database of time series provided by the authors of the "M-Competition". In this case, all 68 of the monthly series from that database were used.

In addition to the time series, the authors of the "M-Competition" also provided the forecasts for the statistical methods reported. This chapter draws on that data for the deseasonalised single exponential smoothing forecasts used as a comparison base, and for the other methods for illustrative purposes.

The judgmental forecasts obtained by Lawrence, Edmundson, and O'Connor (1985) were also made available for this analysis, and the "graph" results are taken from that study.

6.2.4.2 DATA COLLECTION

As mentioned above, the forecasts for deseasonalised single exponential smoothing were obtained from the "M-Competition" and the judgmental forecasts using hard copy graphical data presentation were obtained from the Lawrence, Edmundson, and O'Connor (1985) study.

The "experienced" subjects used in the studies were three post graduate students with a knowledge of time series analysis, and with prior experience in the use of GRAFFECT. The subjects had undertaken post graduate courses in operations research, and had general commercial experience. Each subject forecast all 68 time series in the database, and their forecasts for each were randomly assigned (without replacement) to the three studies conducted using a computer program. Thus three sets of experimental data were derived. Each set had forecasts of all 68 series, and the subjects were randomly "scattered" throughout each set.

The novice subjects were 35 post graduate students with no prior experience of time series analysis, but they had general commercial experience. The subjects were subjects enrolled in a decision support systems course at the University of New South Wales. The subjects were volunteers who agreed to use the GRAFFECT decision aid following a brief presentation of the tool as part of their course work.

The subjects were given an introduction to the terms, and basic concepts of time series analysis. That is, the idea that a time series is expressible as trend, cycle and noise was introduced. The tool was demonstrated to the subjects. The instruction and demonstration was carried out in a single session of twenty minutes.

It was explained that the results of the exercise was to be used in this dissertation and the subjects were asked to use the GRAFFECT decision aid to attempt to forecast their given time series as accurately as possible. No course credit was given for the exercise, and subjects were advised that they were not to feel any coercion to take part.

The 68 time series were allocated to the 35 students taking part according to the following rules:

* Five time series were allocated to each subject,
* It was ensured that every time series occurred at least once in the set of time series numbered 3 to 5 in the subject's lists.
* subject to the above, and the requirement that no series would occur more than once in each list, the time series were "randomly"[3] allocated to the lists of five.

The subjects forecast the series in the order in which they appeared on their particular "assignment" list, the lists were randomly assigned to subjects.

The procedure ensured that there was little or no bias effect expected as a result of the position of time series in the sequence of the forecasts. The first two forecasts from each subject were eliminated as practice, and the remainder of the completed forecasts were accepted as candidate forecasts. The file of candidate forecasts contained a minimum of one

---

[3] The process was performed using the RND function in BASIC.

forecast for each series, and for some series there were up to four forecasts.

In order to permit the use of a paired t-test a single forecast attempt was required for each series. Multiple observations for individual time series were eliminated by random selection.

The subjects followed the procedure laid down in the GRAFFECT User Manual, forecasting each series by first considering the trend component, then the seasonal component, and finally forecasting from the residual noise series. This was similar to the sequence used by Lawrence, Edmundson, and O'Connor (1985), who required subjects to first consider the trend component followed by the seasonal pattern.

No time restriction was placed on the completion of the task. There was no outcome feedback provided until all series had been forecast. As with the other studies reported in this dissertation, and in line with the Lawrence, Edmundson, and O'Connor (1985) approach, the subjects were not given any information outside the values in the time series. Thus, the subjects had no indication of the source or nature of the time series, and they did not know the time periods from which the experimental series were drawn. This placed the subjects in the identical position, with regard to information about the time series, as the subjects in the Lawrence, Edmundson, and O'Connor (1985) study. It also placed the decisions of those subjects on the same footing as the extrapolations of the statistical methods which cannot take any external data into account.

6.2.4.3 ANALYSIS OF THE DATA

Following Lawrence, Edmundson, and O'Connor (1985) the data was manipulated to generate the Mean Absolute Percentage Error of forecast (MAPE) for the two forecast horizons, months 1-6 and 7-12. MAPE was chosen as the error measure in order to ensure consistency with Lawrence, Edmundson and O'Connor (1985), and because it has been shown that measures based on percentage error are widely accepted[4]. The use of MAPE also avoids the heavy emphasis that a squared error measure such as Mean Squared Error places on extreme errors. Such emphasis may be warranted in particular commercial circumstances, but not for the purposes of the current investigations. The results of Lawrence, Edmundson and O'Connor (1985) indicated that judgmental forecasts avoided the very large errors associated with statistical processes applied to certain series. To apply a squared measure in those circumstances would be seen to give advantage to the judgmental processes under investigation.

In order to eliminate the individual time series effect that was evident in the previous studies on this database, it was necessary to use a "paired"

---

[4] See Armstrong (1985) at page 360.

method of comparison. The hypotheses, in the light of prior results, had been stated in "one-tailed" form. Thus the method of analysis chosen was the one-tailed, paired t-test. This form of testing provides a most sensitive test for differences in samples. As described above, multiple testing on a single set of observations was avoided in the case of the major aims of the studies.

6.2.5 RESULTS

The forecasts of the experienced subjects had been randomly assigned to three data sets. The first set was to be used for significance testing in the comparison with GRAPH, the second with DSE, and the third with the forecasts from the novice subjects. Table 6.1 contains the MAPE's for each of the data sets.

DATA SET	MAPE for MONTHS		
	1-6	7-12	1-12
GRAFFECT vs GRAPH set	10.8	13.7	12.3
GRAFFECT vs DSE set	9.5	12.7	11.1
EXPERT vs NOVICE set	9.9	13.4	11.7
MEAN ERROR	10.1	13.3	11.7

Table 6.1 GRAFFECT Error Rates for the Three Samples

The results show that human judgment using the GRAFFECT decision aid provides error rates lower than any reported single forecasting method. Table 6.2 displays the error rates for the methods under test, and for several other statistical methods reported in the "M-Competition". For simplicity, the error rate for expert forecasters using GRAFFECT is reported as the mean of the rates reported in table 6.1.

METHOD	MAPE for months		
	1-6	7-12	1-12
GRAFFECT			
EXPERT (mean)	10.1	13.3	11.7
NOVICE	10.0	14.0	12.0
DSE	11.0	14.2	12.6
GRAPH	11.6	16.5	14.1
BOX - J	11.3	16.3	13.8
D ARR EXP	10.8	13.8	12.3
BAYES F	10.7	14.5	12.6

Table 6.2 Average MAPE over 68 Time Series

As can be seen, the MAPE for GRAFFECT was approximately 1 percentage point lower than deseasonalised single exponential smoothing (when rounding errors are taken into account), for both the 1-6 and the 7-12 forecast horizons. In the "M-Competition" the deseasonalised single exponential smoothing method was one of the best methods overall, however, for the 1-6 horizon it was bettered by deseasonalised adaptive response rate exponential smoothing and the Bayesian F method. GRAFFECT shows a 6% improvement in the 1-6 case, and a 4% improvement in the 7-12 case, over the best previously reported methods for each forecast horizon. Those results must be interpreted in the light of table 6.3 and the discussion that follows.

It is clear from Table 6.2 that provision of the GRAFFECT decision aid did lead to a dramatic improvement in performance in judgmental forecasting. The improvement obtained by providing experienced subjects with GRAFFECT over the results from similarly experienced subjects with GRAPH is 13% for the 1-6 case and 20% for the 7-12 case.

The table also indicates that there is little difference between experienced and novice subjects for months 1-6, but that the novice subjects were somewhat less accurate over the longer forecast horizon.

The paired t-tests carried out are reported in table 6.3

	ONE-TAILED PROBABILITY	
	1-6	7-12
GRAFFECT EXPERT with		
GRAPH	0.005	0.001
DSE	0.060	0.110
GRAFFECT NOVICE	n.s.	n.s.

Table 6.3 Paired T-Test Results

Given the above, there is no problem in rejecting Hypothesis H6.1 with a confidence level of 99% or better for both time horizons considered. That is, the GRAFFECT decision aid clearly produces results that are better than the Graph technique by 13% or more, and that the improvement is statistically significant.

The comparison with deseasonalised single exponential smoothing was not quite so clear. The level of confidence with which hypothesis H6.2 could be rejected (94%) bordered on a commonly adopted lower bound of significance of 95%. For the 7-12 case there was no opportunity to reject hypothesis H6.2 with a reasonable level of confidence (the significance level being 89%). The sample data set used for the examination of hypothesis H6.2 had the lowest error rate of the samples. This observation would tend to indicate that claims that GRAFFECT has a lower error rate than DSE would be suspect. On the other hand, all three sample data sets had lower MAPE's than DSE, and this would tend to lend support to the supposition that GRAFFECT is more accurate than DSE. The difficulty in interpreting the results of the study without equivocation centres on the normative question of what level of significance should be adopted. The conservative view of the results would be that any improvement was not statistically significant, but that view generates a real possibility that the null hypothesis is mistakenly accepted in marginal cases such as this.

Hypothesis H6.3 cannot be rejected. There is no statistical difference between the results obtained by experienced and novice forecasters. The higher error rate of the novice forecasters over the longer forecast horizon, though not statistically significant, raised an implication that they might not be dealing with trend in the same way as the experienced forecasters. Post hoc analysis revealed that in only two of the 68 series did the subjects damp the trend functions that they had identified[5]. As reported in chapter 8, the experienced forecasters damped the trend in 10 series. The question of damping trend is fully discussed in chapter 8, but it should be pointed out here that there is opportunity for potential improvement in the accuracy of novice forecasters arising out of further support in the decision aid, or in education programs.

The analysis of the error data revealed that the GRAFFECT method obtained its low levels of MAPE error without losing the advantage of the relatively low standard deviation of error that judgmental forecasting has exhibited. Figure 6.2 shows the histograms of the frequencies of error rates for intervals of "4" on the MAPE scale[6]. It can be seen that comparing GRAFFECT with the other methods:

* there is greater mass at the low end of the scale,
* the high outlier is well inside the deseasonalised single exponential smoothing outlier, and slightly lower than the Graph outlier.

---

[5] The two series with damped trend came from the same subject. Examination of the data revealed that the subject had damped the trend for the second to fifth series in the sequence forecast. This might therefore have been associated with experimenting with the tool rather than a real decision to damp the trend.

[6] The data set was that used in the comparison with DSE. The other data sets exhibited very similar standard deviations of MAPE'S, as shown later in table 6.4

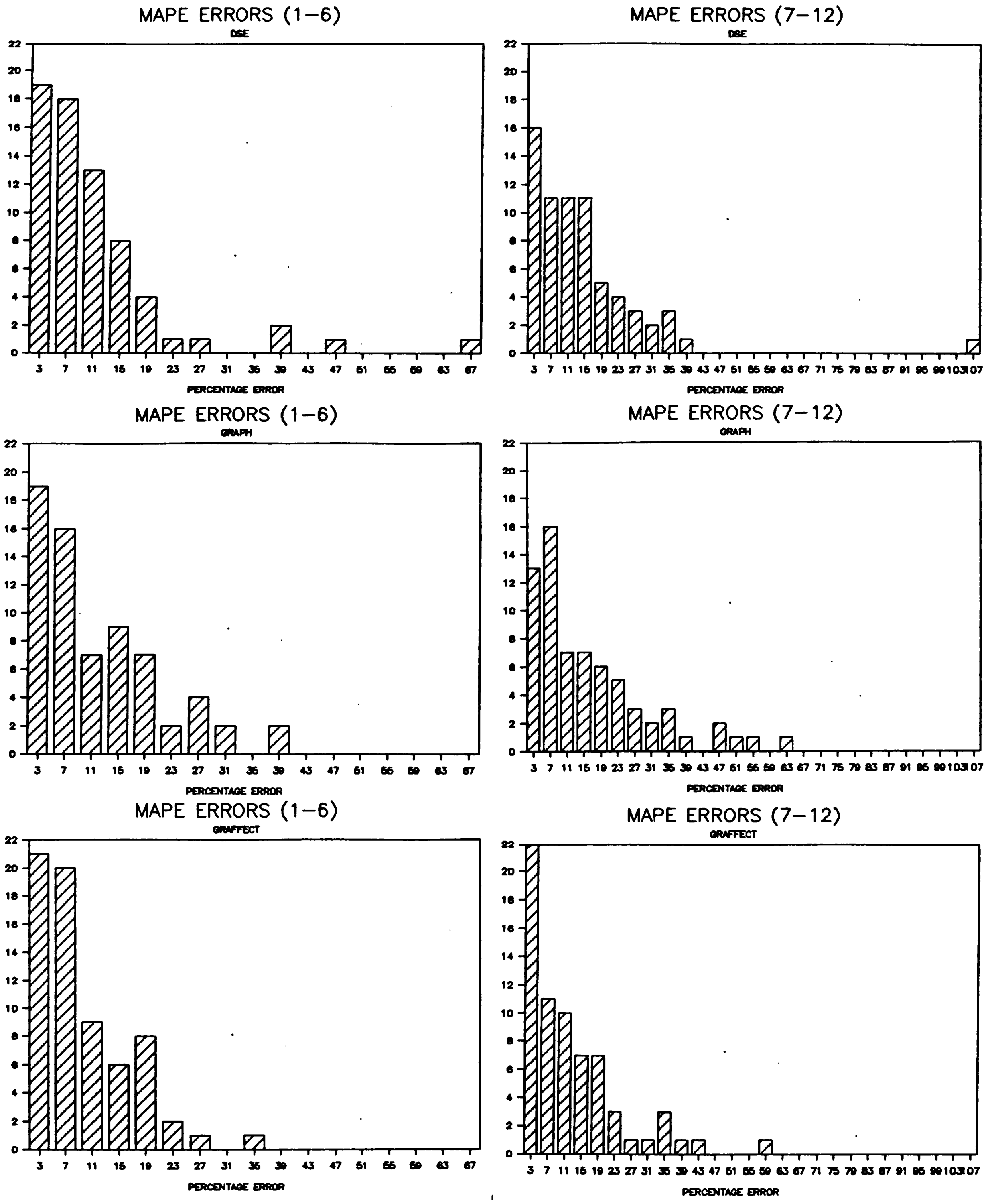


Figure 6.2 Error Distributions

The above conclusions are supported by the statistics. Table 6.4 displays the standard deviations of error for the three GRAFFECT data sets, DSE, and GRAPH for the two forecast horizons.

METHOD	1-6	7-12
}	0.074	0.112
GRAFFECT }	0.079	0.113
}	0.085	0.112
DSE	0.115	0.150
GRAPH	0.095	0.142

Table 6.4 Standard Deviations of MAPE Error

This would indicate that there may be incentive to use judgmental techniques, and especially methods such as GRAFFECT, for the forecasting of critical time series in which there is a desire to avoid very large error rates. The incidence of extreme error would be expected to be lower in that case.

6.2.6 DISCUSSION

This study shows that the GRAFFECT decision aid is a viable tool for both experienced forecasters and novice users. There is no evidence that the forecasts made by in-experienced forecasters are less accurate than those made using deseasonalised single exponential smoothing, or those made by experienced forecasters. Post hoc analysis of the data failed to reveal a systematic explanation of the possible difference between experienced and in-experienced forecasters over the longer forecast horizon, though there was less evidence of damping the trend by in-experienced forecasters. It is concluded that in-experienced users, with a minimum of instruction, may forecast well. The provision of more detailed instruction regarding trend, or the provision of an advisory signal dependent on the number of points in the trend line[7], to encourage damping may be of benefit.

The results of the comparison of the use of GRAFFECT with the use of GRAPH show that there are real advantages to be obtained from the use of GRAFFECT. Therefore, if the forecasting task calls for the use of judgmental extrapolation because of the need to limit the risk of large errors, or to permit the consideration of data external to the time series, then the use of a tool such as GRAFFECT is indicated. The improvement in accuracy cannot be simply explained. It might have arisen for a number of reasons:

a) The cue data presented by GRAFFECT for seasonal identification, and for the extrapolation from the noise residual was less cluttered with other cues.

b) The structure of the decision was enforced. The subjects were required to consider trend, seasonal and noise characteristics separately for both judgmental methods, but in the GRAPH method they had the opportunity to avoid that strategy.

c) The automatic recomposition of the components may have contributed greatly to the result.

Further research is required to evaluate the relative contribution of the explanations to the outcome.

The comparison of the accuracy of GRAFFECT and DSE leads to some interesting observations. The first of

---

[7] See the discussion of judgmental damping of trend in chapter 8.

those concerns the longer time horizon, in which GRAFFECT did not gain the anticipated benefit of being able to model trend. The DSE forecast has no trend component, thus implying that extrapolation of trend, even over a relatively short period, may not be an optimum strategy. This issue requires closer examination, because it is possible that the cause was associated with inappropriate trend identification rather than extrapolation of the model. This issue is examined in chapter 8 of the dissertation.

6.3 EVALUATION OF THE EFFICIENCY OF FORECASTING

6.3.1 INTRODUCTION

The costs of forecasting are not limited to the costs of the errors in the forecast. The use of resources must also be considered in determining the desirability of a particular forecasting method. In that light, it was necessary to establish what effect the use of the GRAFFECT decision aid had on the time taken to forecast a time series.

It is not possible to compare the use of resources in the deseasonalised single exponential smoothing method with that in a judgmental method. Deseasonalised single exponential smoothing is a simple, computer based method which may utilise little or no human resources for the extrapolation phase. The trade off between human and computer resources is complex, and must include consideration of a number of qualitative factors such as the implications for management understanding of the results and the effect on the political scenario.

This study was directed, primarily, at the evaluation of the decision aid against the hard copy graphical technique. The study was limited to data from experienced subjects in order to avoid confounding the results with effects from any learning curve experienced by novice subjects.

6.3.2 RESEARCH HYPOTHESIS

There was no a priori basis for assuming that the GRAFFECT decision aid would permit faster or slower forecasting than would the provision of hard copy plots. The hypothesis addressed is therefore expressed in "two-tailed" form:

H6.4 The use of the GRAFFECT decision aid will have no effect on the time taken to forecast time series.

6.3.3 DESCRIPTION OF THE STUDY

The GRAFFECT decision aid was constructed to record the time taken for each activity in the forecast, and therefore the data collected for the accuracy experiments contained the time taken for each forecast. Unfortunately, the Lawrence, Edmundson, and O'Connor (1985) study was deficient in that there was no measurement of the time taken for the forecasts. It was therefore necessary to replicate a part of the Lawrence, Edmundson, and O'Connor (1985) study in order to obtain data on time taken.

A random sample was taken of 20 time series from the 68 monthly series used above, and from that 20, ten were randomly assigned to each of two subjects. Graphs for those series were prepared, with the historical data numbered for ease of seasonal identification, and the series were forecast exactly as described in Lawrence, Edmundson, and O'Connor (1985).

The subjects used had taken part in the accuracy study reported above. Some six months had elapsed since they had taken part in that exercise. Never the less, it was considered that their previous involvement would render any analysis of the error rates doubtful, though any effect arising from that prior experience would have been to shorten the time taken to forecast the series. Although it is doubted that such an effect could take place over the large number of series handled, it was possible that the subjects could recognise any series that they had seen before, and therefore process the series faster. The same subjects were used in an attempt to obtain forecasts from judges of the same level of experience and motivation.

The subjects forecast the series in two batches of five series, and were asked to record how long each time series took to forecast. They were not directly aware of the objective of the study, and would have assumed that it was another accuracy trial, with additional timing data being gathered. It was considered equally likely that any bias would be to understatement as overstatement of the time taken. The subjects were monitored, and it was established that the total time taken for each of the batches of five forecasts was close to the sum of the times recorded for the individual forecasts.

6.3.4 RESULTS

The times from the GRAFFECT decision aid, and the hard copy graph forecasts were expressed in seconds, and a two tailed t-test conducted, paired on time series. Table 6.5 summarises the results which revealed that the time taken with the GRAFFECT decision aid was about 60% of that taken with the hard copy graph. The difference between the methods was significant at the p=0.000 level.

METHOD	MEAN TIME	STD DEV.
GRAFFECT	161.7	42
Graph	265.4	62

Table 6.5 Time for Forecast

The hypothesis H6.4 can be rejected with high confidence, and the conclusion made that the data presentation methods in GRAFFECT resulted in substantial time savings.

6.3.5 DISCUSSION

The savings in time taken to forecast a series using GRAFFECT, over the time taken using hard copy graphs understates the advantages of the GRAFFECT decision aid. On completion of the forecast the GRAFFECT decision aid has the forecast data captured in machine readable form. This saves the rather onerous, and error prone, tasks of extracting the data from the hard copy graphs, and keying that data to a file. Perhaps of equal importance is the fact that GRAFFECT not only records the final forecast, but also the trend and seasonal models for the time series. This would enable the forecaster to:

* Re-use the trend and seasonal models for future forecasts, saving a considerable part of the forecast time, until it was appropriate to re-estimate the models.

* Perform sensitivity trials, and simulation exercises upon the models to evaluate the scope and effect of management action on the business variable forecast.

As described above, there is no possible rational analysis of the resource utilisation vis a vis deseasonalised single exponential smoothing. However, to the extent that the forecaster would wish to amend the deseasonalised single exponential smoothing forecast to take account of additional knowledge[8] then the time spent on a judgmental forecast would need to be offset against the time taken to judgmentally evaluate the deseasonalised single exponential smoothing forecast.

The other, commonly recognised forecasting procedure that has been considered elsewhere in this dissertation is the Box Jenkins method. In the "M-Competition" the Box

---

[8] Fildes 1979 describes such an amendment as usual practice.

Jenkins forecasts were produced manually by Alan Andersen, then of Sydney University. He reported that each forecast took in the region of 60 minutes to produce. Once the Box Jenkins model had been produced, it could be used again until it was appropriate to re-estimate the model. On that basis, the Box Jenkins model would have to be much more robust and longer lasting than the judgmental model (which takes one twentieth the time to build), for it to be a viable alternative

6.4 DISCRIMINATING BETWEEN THE GRAFFECT AND DSE

6.4.1 INTRODUCTION

The GRAFFECT decision aid was developed in the light of the discriminant analysis study reported in chapter 3. It was found that it was possible to discriminate between series that were forecast well over the 1-6 month forecast horizon by judges using hard copy and deseasonalised single exponential smoothing. The discriminant function included terms based on the lag one autoregressive nature of the series, the ratio of seasonality to instability in the seasonal, and noise. There was no direct term based on trend, though it might have been expected that judgment would have an advantage over deseasonalised single exponential smoothing in the presence of trend.

The grounds for conducting the discriminant analysis study reported in chapter 3 were:

1) to attempt to discover a reliable rule for the selection of forecasting method that might lead to lower forecast errors, and

2) to discover the relative strengths and weaknesses of judgment with respect to deseasonalised single exponential smoothing, in order to identify possible means to improve the performance of judgment.

In part, each of those aims was fulfilled. In particular, the resulting design of the GRAFFECT decision aid has been shown to improve the accuracy of judgmental forecasting. That being the case, it is necessary to replicate the discriminant analysis. It is still possible that improvement in overall accuracy might be achieved by accurate selection between judgmental forecasting using GRAFFECT and deseasonalised single exponential smoothing. It is also possible that further developments in the decision aid might be indicated as a result of such an analysis.

6.4.2 ANALYSIS OF THE DATA

The procedures reported in chapter 3 were fully replicated, for the comparison of judgment and deseasonalised single exponential smoothing, with the GRAFFECT supported forecasts replacing the hard copy supported forecasts. For the purposes of this analysis the GRAFFECT forecasts used were those from expert forecasters that had been used in the accuracy comparison with DSE. All metrics were again considered, not only those that were found to have an effect in chapter 3.

It was found, in chapter 3, that in classifying the series as either better forecast by judgmental hard copy or by deseasonalised single exponential smoothing the former class was about half the size of the latter. This relationship changed with the use of the GRAFFECT decision aid for judgmental forecasting. The distribution of series between the two classes (better forecast by GRAFFECT decision aid or deseasonalised single exponential smoothing) was equal. Figure 6.3 displays the distribution of the series by difference in 1-6 MAPE between the methods, according to which method has the lower error.

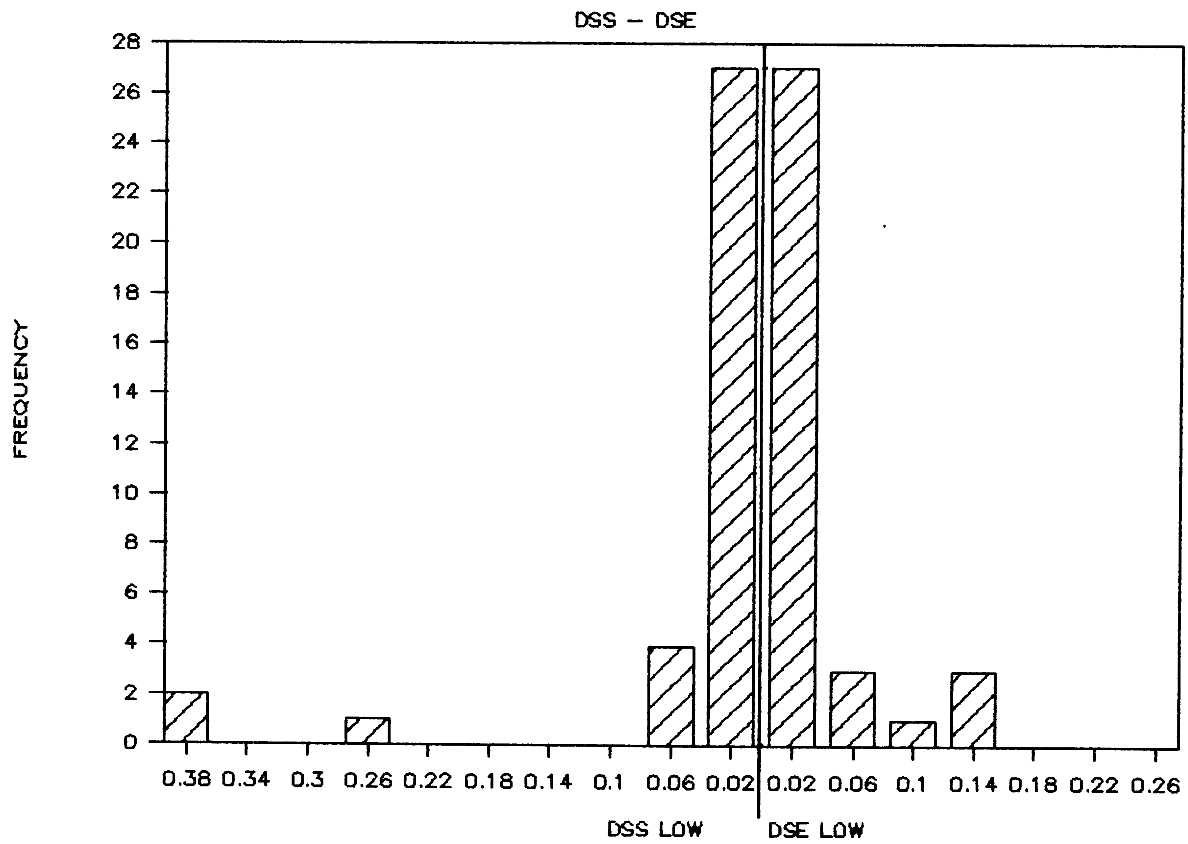


Figure 6.3 Distribution of Differences in MAPE, Months 1-6

The close similarities in performance for many of the series caused difficulties in performing discriminant analyses. Indeed, it was not possible to identify a reliable discriminant function for either the 1-6 or the 7-12 forecast horizons.

This provides some evidence that the strategies adopted in the design of the GRAFFECT decision aid had an effect. The disadvantage to the judgmental method of the high seasonal signal in the presence of low noise was eliminated. This was achieved at the loss of the corresponding advantage in the presence of low seasonal signal and high noise. The implication to be drawn from this is that the change in the data interface may have caused the judges to more closely emulate the ratio to centred moving average process in identifying seasonality. The identification of the seasonal component is examined more fully in chapter 7.

The failure of the judges to achieve an advantage based on the trend in the series is cause for some concern. It would have been expected that the absence of any capability to model trend in deseasonalised single exponential smoothing would lead to considerably poorer performance in the presence of trend. The identification of trend is covered in chapter 8

Apart from the observation that the average error arising from the use of judgment decreased it was found that the number of series for which the judgmental process is absolutely more accurate increased. This indicated that the improvement in accuracy did not occur only by improvement in series on which judgment already performed well.

The implications of the above finding are that the decision to adopt a judgmental approach to extrapolation for reasons such as the reduced standard deviation of error, has a lower chance of being dysfunctional in terms of error rates for the particular series.

6.5 CONCLUSIONS

The foregoing analysis has demonstrated that the provision of interactive graphical support for the forecasting decision can improve judgmental forecasting by between 13% and 20% depending upon the forecast horizon. There is also evidence that a similar, but smaller, improvement is obtainable over deseasonalised single exponential smoothing. There is no evidence, however, that this improvement is statistically significant, especially for the 7-12 forecast horizon.

Again, judgmental processes have been shown to be somewhat more controlled (see also Lawrence, Edmundson, and O'Connor 1985) in that they have lower standard deviations of error.

The results obtained in the studies reported raise further questions of interest. The major issues arising concern the process by which the improvement in judgment came about, and the effect of trend on the accuracy of judgmental extrapolations. Preliminary examinations of those issues are reported in the following three chapters.

6.6 LIMITATIONS

The time series used in this study were drawn from a database of time series used in the "M-Competition". The fact that 68 series were used in the study gives some strength to the study, but it is not possible to assume that the 68 series reflect the general population of time series. Thus, the results must be interpreted as indicating that for the series tested, the GRAFFECT decision aid was eminently successful, but that this provides only persuasive evidence for the generalisability of that success.

Comments on the more skewed distribution of error for deseasonalised single exponential smoothing are similarly constrained by the characteristics of the sample. It is clear that relatively few series contributed heavily to that skewedness. To the extent that the sample reflected the general population in the frequency of such series, the results are generalisable. Again, there is no evidence to indicate that such an assumption holds.

6.7 REFERENCES

Armstrong, J.S., *Long-Range Forecasting from Crystal Ball to Computer*, 2nd Edn. Wiley, New York, (1985).

Edmundson, R.H., Lawrence, M.J., & O'Connor, M.J., "The use of non time series information in sales forecasting : a case study", *Information Systems Research Reports*, University of New South Wales, (1987).

Fildes, R., "Quantitative forecasting - the state of the art: extrapolative models", *J. Opl. Res. Soc.*, 30,8 (1979), 691-710.

Lawrence, M.J., Edmundson, R.H., & O'Connor, M.J., "An examination of judgmental time series extrapolation", *Journal of Forecasting*, 2,2 (1985), 169-179.

Hays, W.L., *Statistics*, Holt Saunders, New York, (1981).

Makridakis S., Andersen,A., Carbone,R., Fildes,R., Hibon,M., Lewandowski,R., Newton,J., Parzen, E., and Winkler,R. "The accuracy of extrapolative (time series) methods : results of a forecasting competition", *Journal of Forecasting*, Vol 1, no 2, (1982).

7. SEASONAL PATTERN IDENTIFICATION

7.1 INTRODUCTION

In chapter 5 the design of the GRAFFECT decision aid was described, and in chapter 6 its overall accuracy was examined. The conclusion from that chapter was that there was an increase in accuracy attributable to GRAFFECT, but the reason(s) for that improvement were not identified. This chapter reports the first of a number of studies to evaluate aspects of model identification in an attempt to throw light on the extent to which the improvement in accuracy arose from cue processing rather than the benefits of automatic recombination of the component parts of the decision. In this case, the seasonal identification process is considered, both with a view to evaluate the effectiveness of the decision aid, and of judgmental seasonal pattern identification per se.

Later chapters address similar issues concerning modelling the trend component and handling the residual (noise) component.

This chapter also reports a trial conducted to assess the effect of the provision of automatic deseasonalising procedures in the place of the judgmental deseasonalising procedures.

7.1.1 OBJECTIVES OF THE INVESTIGATIONS

The GRAFFECT decision aid, based on conventional time series analysis, comprises modelling the time series in the form of seasonal, trend, and random components [1]. The major objective of this chapter is to examine the first of those modelling functions:

a) Does the use of the GRAFFECT forecasting decision aid described in chapter 5 give rise to the identification of seasonal patterns different to those identified using hard copy plots?

A secondary objective of the chapter is to consider the question of judgmental pattern identification in the presence of randomness. The results of the major study reported in this chapter, and a brief analysis of the data gathered for chapter 6 provide some insights to, though not a controlled study of:

b) the characteristics of the judgmental seasonal identification process, and whether it is affected by variations in the strength of the seasonal metric *SEAS* and the noise metric *SYN* described in chapter 3.

Finally, the effect of replacing judgment in the seasonal identification process is examined to determine:

c) does the provision of automatic deseasonalising capability affect the accuracy of the forecast?

7.1.2 JUSTIFICATION OF THE STUDIES

In developing and proving a decision aid for time series forecasting it is relevant to determine whether the use of judgmental methods is justified, and whether

---

[1] As described in chapter 2, there is no capacity to consider cycle for short term forecasting in the absence of data external to the time series numbers.

the mode of data presentation and manipulation can affect the outcome of the judgmental process. These two factors are linked, in that it is not possible to determine the former without resorting to some mode(s) of data cue handling. For instance, _if_ an examination of judgmental seasonal identification using a tabular presentation of raw data as a cue found that the seasonal models identified were considerably poorer than the ratio to centred moving average method, that could not be taken as an indictment of judgment per se. A different form of cue might give rise to excellent results from a judgmental process.

In the development of a forecasting decision aid it is important to determine which method of performing the sub-tasks gives rise to the "best" outcome. In this case three strategies have been identified for the task of modelling the seasonal pattern:

1) Using judgment supported by a conventional plot of the time series,

2) Using judgment supported by the data interface within the decision aid, as described in chapter 5, and

3) Using a statistical process.

Apart from the pragmatic objective of moving towards more accurate means of forecasting, the testing of those strategies will provide some input to the discussion on issues raised in chapter 4:

* Is the outcome of the judgmental seasonal identification process sensitive to the form and content of the data display?

* Are there implications arising from the use of interactive graphics in this task setting?

* Are there characteristics of human judgment that indicate the use or rejection of judgmental processes for this task setting?

It is not suggested that any contribution to those areas of academic discussion will bring the discussion to a head, or give rise to the development of a general theory, merely that another "observation" may be established for general consideration.

The human information processing literature described in chapter 4, the work of Eggleton (1982), and the results obtained in chapter 3 indicate that judgmental identification of seasonal pattern might be influenced one way or the other by noise. It is therefore of interest to consider that effect in the results of forecasts using either of the cue forms (GRAPH and GRAFFECT) examined here.

7.2 SEASONAL IDENTIFICATION USING GRAFFECT

7.2.1 INTRODUCTION

This section considers the evaluation of the seasonal identification processes of the GRAFFECT decision aid relative to "pencil and paper" techniques. The objective was to determine whether the design of the data display and the ability to decompose the data in line with the decomposition of the decision task changes the outcome of the decision.

In chapter 3 it was shown that the ratio of the metrics *SEAS* and *SYN* [2] acted as a discriminator between relatively good and poor performances of judgment. The effect of those metrics on the identification of the seasonal pattern is also examined in this chapter.

7.2.2 RESEARCH HYPOTHESES

Although there was some general support found in the literature reviewed in chapter 4 for the use of a data interface that permitted the decomposition of the decision (see section 4.4.1) and the spatial expression of the model (see section 4.4.2) with automatic parameter estimation, there were no clear findings to suggest what the effect of such a facility would be. Indeed it is questionable whether it would have any effect at all. The first hypothesis addressed in this study is therefore:

H7.1 The seasonal models identified using GRAFFECT will not not be different to those identified using a hard copy plot of the series (GRAPH).

In chapter 3 it was shown that for forecast months 1-6 the ratio of the seasonal metric *SEAS* to the noise metric *SYN* discriminated, in part, between good judgmental forecasts generated using GRAPH and good

---

[2] *SYN* provides a simple measure of the variability of the seasonal signal in the series.

forecasts generated by deseasonalised single exponential smoothing. The design of the data display for seasonal identification was intended to assist in the estimation of an accurate "average" seasonal under conditions of high *SEAS* and low *SYN*. The studies reported in chapter 6 showed that the metrics did not act as discriminators between GRAFFECT and deseasonalised single exponential smoothing. It is therefore of interest to determine whether the change was due to seasonal model identification:

H7.2 The influence of the metrics *SEAS* and *SYN* will not be affected by the use of the GRAFFECT decision aid.

The use of the GRAFFECT decision aid was shown to give rise to an improvement in forecast accuracy (a matter reported in chapter 6). As a guide to future developments in forecasting decision aids it is necessary to determine whether the improvement was related in an improvement in seasonal pattern identification. From the forecasting viewpoint, improvement may only be realistically measured in terms of the fit of the model to the forecast period, not to the model fitting data. The problem of such an evaluation lies in finding the seasonal component in the forecast period. The best that can be achieved in the short term is to accept the actual pattern in the validation data as a surrogate for the true seasonal, although that pattern is obviously affected by the noise in the series. The Third hypothesis is therefore:

H7.3 The use of the GRAFFECT decision aid will not lead to identification of a seasonal model with a better fit to the pattern in the validation data than that identified using GRAPH.

7.2.3 THE TIME SERIES DATA.

Time series were selected from a database of 68 monthly series of real economic data used by Makridakis et al (1982) for a forecasting competition and made available by the authors. The series were selected for three levels (called high, medium and low) of the seasonality metric *SEAS* and three levels of an unsophisticated randomness metric *SYN* [3]. Thus, nine time series were selected to fill the 3*3 matrix. As far as possible the series were matched on other metrics described in chapter 3.

7.2.4 CUE PRESENTATION.

Two forms of data presentation were used in the experiment, and the subjects were required to provide a forecast using an appropriate decomposed decision strategy as described below. The forms of data presentation were:

GRAPH: A hard copy plot of the time series.

GRAFFECT: A screen based display of the data using the experimental GRAFFECT decision aid. Chapter 5 contains a full description of GRAFFECT.

---

[3] The metrics are fully described in chapter 3, and the appendix to chapter 3.

In each treatment the subjects were required to pay initial attention to trend components, then to model the seasonal component, and finally to generate the forecast.

Data capture for subjects with the GRAPH data display was an exact replication of the Lawrence, Edmundson, and O'Connor (1985) study reported in chapter 2. The subjects extrapolated an eyeballed trend line upon which they would later impose their seasonal pattern. They were then required to consider the seasonal pattern in the data by numbering the months to assist with seasonal identification. The subjects then sketched the perceived seasonal pattern onto the trend line to form their forecasts.

The subjects with the GRAFFECT Decision aid fitted eyeballed trend lines to the data on the screen and established an extrapolated trend. The decision aid then recorded the trend model and removed its effects from the data. Seasonal identification took place using a display that placed all four years of data (net of any trend identified) on the same, 12 month wide, grid. The years were coded for identification purposes.

The subjects with the GRAFFECT decision aid therefore had the "cleanest" view of the seasonal cues (being net of any trend cues that they had identified in the previous step), and had the advantage of the patterns being superimposed to highlight similarities and differences across the years. These subjects were also able to test their patterns by washing the identified seasonals out of the data and reviewing the results. This should have enabled the subjects to identify any pattern remaining in the data and to adjust the model accordingly.

7.2.5 DATA COLLECTION.

The subjects were ten postgraduate students and staff in the Faculty of Commerce at the University of New South Wales. No time restriction was placed upon completing the task. No subject used more than one form of data presentation, to which they had been randomly assigned. This approach was adopted to avoid influencing the strategies used by subjects with the conventional, GRAPH data display by exposing them to the screen based system. Five cases were obtained for each time series with each form of data display, each subject contributed a forecast for each of the time series.

7.2.6 ANALYSIS METHOD

The forecasts obtained using each presentation form were manipulated to generate forecast seasonal factor values for the forecast year:

1) The trend component in the forecast was identified by fitting a regression line, and this component was then removed. This permitted the comparison of the identified seasonal components relatively free of effects of trend and anchoring of the forecast

2) The forecast seasonal factors (FF's) were computed for each month as a ratio of the monthly

value to the total for the year. The ratios were scaled to sum to 12, and were thus directly comparable to the seasonal factors computed by the centered moving average process (MF's) used in deseasonalised single exponential smoothing and the computation of the *SEAS* metric (see chapter 3).

For the purposes of examining hypotheses H7.1 and H7.2 the difference between the forecast seasonal factors (FF) and the moving average factors (MF) were considered. For each of the 90 cases (being 9 time series * 5 observations per series * 2 methods of data presentation) a score was developed to indicate the difference of the seasonal pattern from the moving average pattern (MADIF):

$$\text{MADIF} = \text{Sum}_{n=1 \text{ to } 12} \left(\text{Abs}((FF_n - MF_n) / MF_n) \right)$$

For the examination of hypothesis H7.3 "actual seasonal factors" (AF) [4] were derived from the values of the time series used as validation data for the forecasts. Those factors were computed in exactly the same way as the factors from the forecasts, and a score of difference from the actual (ACTDIF) was derived as with MADIF:

$$\text{ACTDIF} = \text{Sum}_{n=1 \text{ to } 12} \left(\text{Abs}((FF_n - AF_n) / AF_n) \right)$$

[4] As previously described, these factors were affected by the noise in the series, and there was no guarantee that either the seasonality or the noise remained constant into the forecast period. Given that caveat, this was the closest estimation possible to the seasonal pattern in the forecast period.

7.2.7 RESULTS

7.2.7.1 EXAMINATION OF H7.1 AND H7.2

Table 7.1 contains the analysis of variance summary for the ANOVA run on the difference from the moving average seasonal (MADIF). It shows all effects to be significant at better than 0.01

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects					
SEAS	4.408	2	2.204	42.401	.000
SYN	.664	2	.332	6.389	.003
METHOD	2.431	1	2.431	46.769	.000
2-way Interactions					
SEAS SYN	1.300	4	.325	6.251	.000
SEAS METHOD	.738	2	.369	7.103	.002
SYN METHOD	.524	2	.262	5.041	.009
3-way Interactions					
SEAS SYN METHOD	.813	4	.203	3.908	.006
Explained	10.879	17	.640	12.310	.000
Residual	3.743	72	.052		
Total	14.621	89	.164		

Table 7.1 ANOVA on MADIF

The results show that there is a strong effect on "method" (p=.000). Across the whole sample the variation from a moving average seasonal pattern for the GRAFFECT decision aid was almost half that of the hard copy graphical technique (GRAFFECT= 0.36 GRAPH= 0.69).

From the foregoing it is possible to reject hypothesis H7.1 that the provision of the interactive graphical display would have no effect on seasonal pattern identification. The GRAFFECT decision aid enables the human judge to identify a seasonal pattern that deviates less from a moving average seasonal than the pattern identified using a conventional hard copy plot of the data. Therefore, the provision of the decision aid does influence the seasonal pattern identified.

The significant effect found on *SEAS* reflects increasing difference from the moving average seasonal with rising seasonality. This matter is further considered below.

The significant effect on *SYN* is harder to interpret, it reflects an increase in difference with the change from low to medium *SYN*. There is no difference in means between medium and high *SYN* however, this matter is also further considered below.

The lower deviation from the moving average seasonal exhibited by GRAFFECT was also found in the two way interactions between method and *SEAS* (p=0.002) and method and *SYN* (p=0.009). Figure 7.1 displays the cell means for the method - seasonality interaction. It shows that for all levels of seasonality the GRAFFECT decision aid has a lower deviation from the moving average process than the hard copy graphical method. The relationship is not a simple one, with the difference being wider at low and high seasonality level than at medium. This might be somewhat explained by the supposition that at low seasonality[5] subjects with the hard copy cues imputed pattern from randomness in the absence of a seasonal cue. It is also possible that unidentified individual time series effects caused this result. This matter is pursued later in section 7.4.

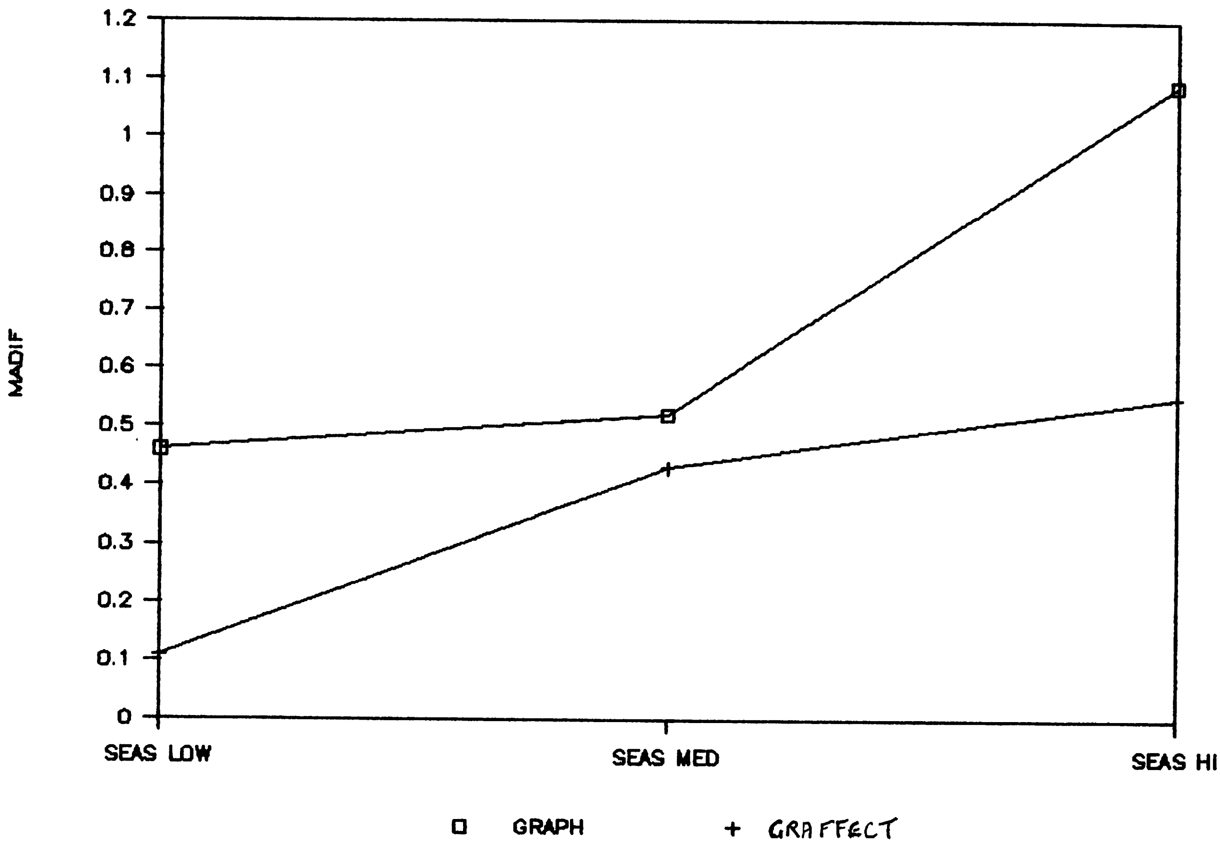


Figure 7.1 Cell Means for Method/SEAS Interaction

Both GRAPH and GRAFFECT methods show an increasing tendency to deviate from the moving average seasonal as the seasonality increases. Post hoc analysis to test for the significance of this effect,

[5] low seasonality in this case is the class of series for which the moving average process detected no significant seasonal factors.

using the rather conservative Scheffe rule, showed (Table 7.2) that the observed tendency was significant in each case:

METHOD	F	SIGNIFICANCE
GRAFFECT	32.705	better than 0.01
GRAPH	12.500	better than 0.025

Table 7.2 Scheffe analysis of Method/*SEAS* Contrast

The analysis of the method/*SYN* interaction is somewhat similar. Again, as shown in figure 7.2, the GRAFFECT decision aid exhibits a lower deviation from the moving average seasonal pattern at all levels of the *SYN* noise metric.

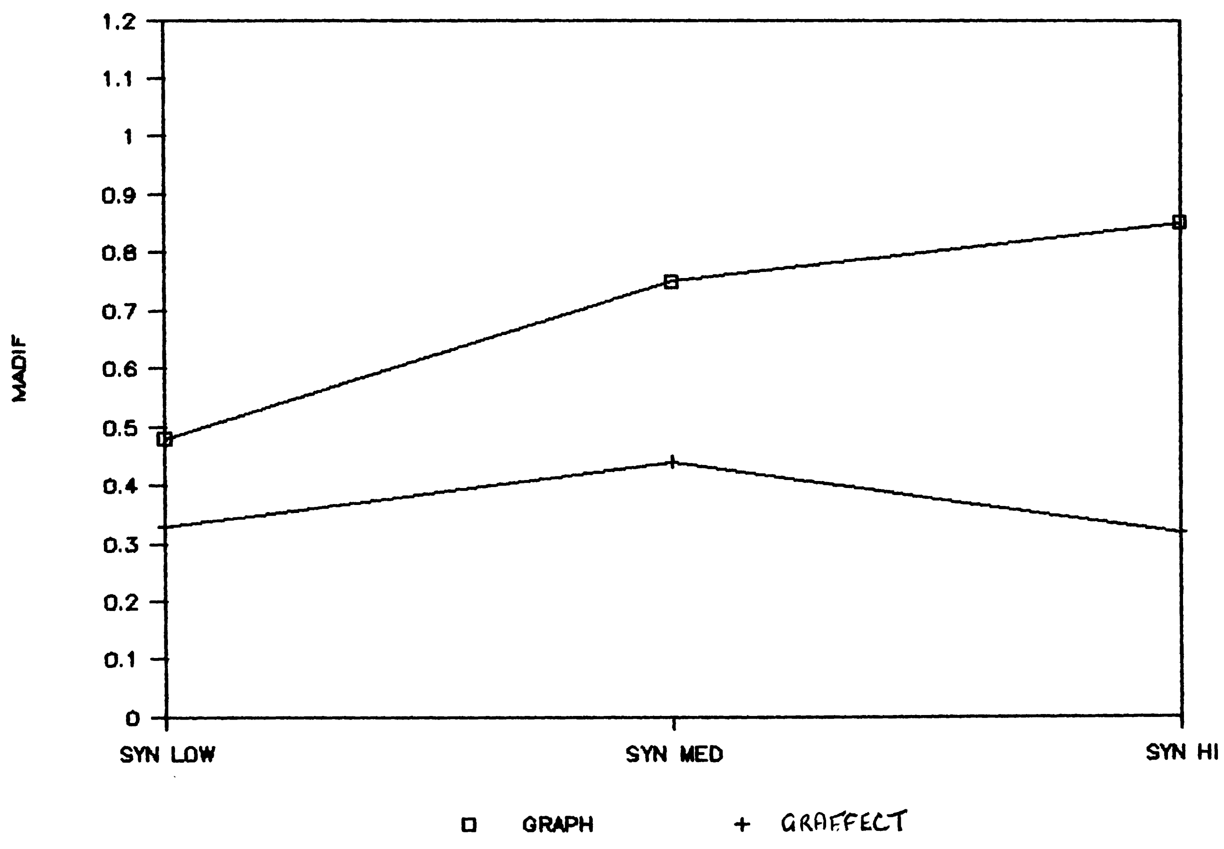


Figure 7.2 Cell Means for the Method/*SYN* Interaction

Figure 7.2 shows that increasing *SYN* results in increasing deviation from the moving average seasonal for the GRAPH method, whereas the GRAFFECT method shows a higher deviation for medium than for either high or low *SYN*. These results indicate that subjects using GRAPH were less reluctant to discount the influence of monthly movements than the subjects using GRAFFECT[6].

The above analyses of the two way interactions between Method/*SEAS* and Method/*SYN* throws a little light on the issue addressed in hypothesis H7.2. The effect of those metrics that was found in chapter 3 was that as the ratio of *SEAS* to *SYN* fell so the GRAPH method gained an advantage over deseasonalised single exponential smoothing. The introduction of GRAFFECT caused the subjects to identify a seasonal pattern that was closer to that of deseasonalised single exponential smoothing than was identified using GRAPH. Therefore, the effect of GRAFFECT was to gain ground with series with a high *SEAS/SYN* ratio, but to lose ground for low ratio series. This raises the question of whether overall gains could be achieved by either using different displays dependant on the characteristics of the time series, or whether a

---

[6] The influence of a high *SYN* characteristic in a series would be observed in the fact that one or more observations would appear to move in the "wrong" direction when compared with the same month in other years. In a moving average calculation this would result in a damping of any "seasonal" signal in the other years, and a lower seasonal factor.

different design of screen altogether could gain on the "swings" as well as the "roundabouts". From this alone it is not possible to predict the overall effect of the introduction of GRAFFECT. However, coupled with the outcome reported in chapter 6 that no discriminator could be found, it is not unreasonable to conclude that the differences in the support for seasonal identification did in part contribute to the change, and that hypothesis H7.2 may be rejected.

7.2.7.2 EXAMINATION OF HYPOTHESIS H7.3

This hypothesis concerns the accuracy of detection of the seasonal pattern as measured in terms of the sum of the absolute differences of the seasonal factors from the factors imputed from the validation data (ACTDIF). Table 7.3 contains the analysis of variance summary.

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects					
SEAS	21.510	2	10.755	230.061	.000
SYN	12.038	2	6.019	128.755	.003
METHOD	0.156	1	0.156	3.348	.071
2-way Interactions					
SEAS SYN	45.438	4	11.360	242.997	.000
SEAS METHOD	0.100	2	.050	1.073	.348
SYN METHOD	0.047	2	.024	0.504	.606
3-way Interactions					
SEAS SYN METHOD	.205	4	.051	1.095	.366
Explained	79.494	17	4.676	100.029	.000
Residual	3.366	72	.047		
Total	82.860	89	.931		

Table 7.3 ANOVA on ACTDIF

In consideration of hypothesis H7.3 the result of interest is the main effect on method, this shows that the GRAFFECT decision aid has a lower "error" (GRAFFECT=1.32, GRAPH=1.40), and that the difference approaches significance at p=0.071. Not only was the "error" lower for the GRAFFECT decision aid, but the standard deviation of the error was also marginally lower at 0.958 compared with 0.980 for the GRAPH method.

The results indicate that the hypothesis H7.3 may be rejected with better than p=0.1 confidence, and that it is thus probable that the special data presentation and manipulation characteristics provided in GRAFFECT give rise to improvement in forecasting the seasonal component of time series. Therefore the improvement in accuracy reported in chapter 6 did not arise entirely from the automatic recombination of the components of the forecast. In part, the improvements are attributable to differences in the seasonal modeling process.

The analysis did not reveal significant interactions between method and either *SEAS* or *SYN*, though both *SEAS* and *SYN* had significant main effects. Figure 7.3 displays the cell means for the *SYN* dimension for both GRAFFECT and GRAPH methods.

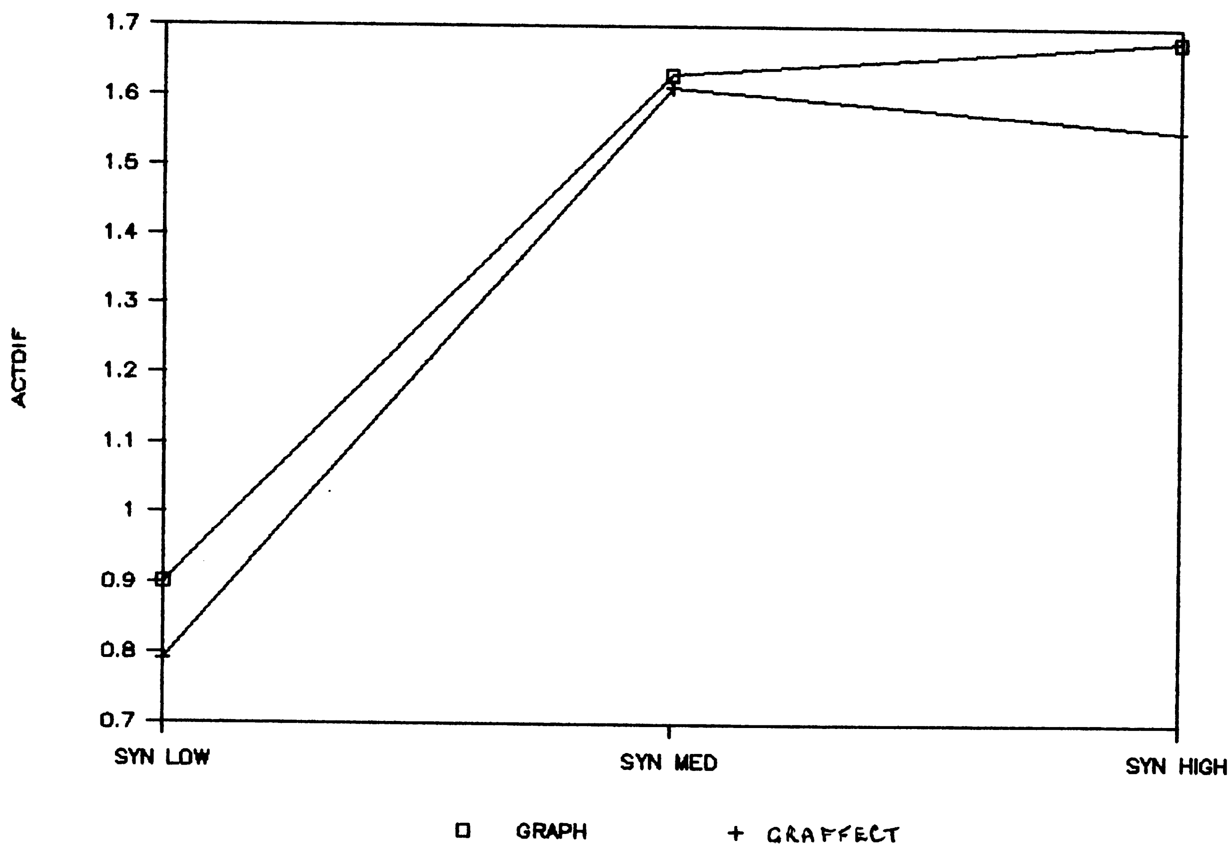


Figure 7.3 Cell Means for Method/*SYN* Contrast

This indicates that the effect of noise on the seasonal identification process is not linear, but that very low levels of the noise metric *SYN* is associated with a high level of accuracy not achieved at higher levels of *SYN*.

The effect of *SEAS* is almost the reverse of that of *SYN*. Figure 7.4 shows the cell means.

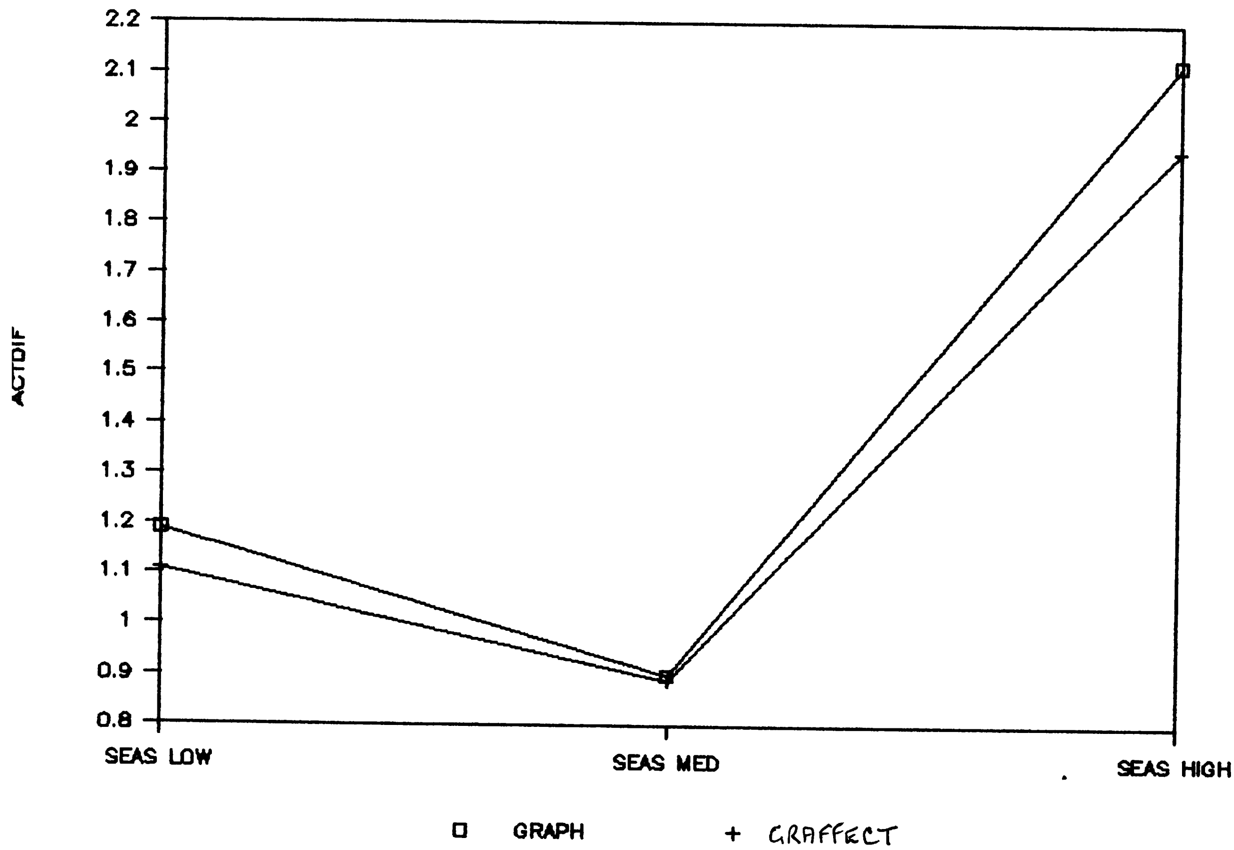


Figure 7.4 Cell Means for Method/*SEAS* Contrast

The figure indicates that high levels of seasonality gives rise to increased error. This cannot be explained merely as a scale effect because ACTDIF was computed as a percentage variation. As with the *SYN* effect above this aspect of the determination of seasonal patterns is an attractive candidate for further investigation.

7.2.8 DISCUSSION

The foregoing results are highly encouraging. They indicate that the form of data presentation does influence the decision outcome, and that it might, therefore, be possible to obtain further increases in accuracy by improvements to the decision aid.

The seasonal factors identified by subjects using the GRAFFECT decision aid were consistently closer to the moving average seasonals than those of subjects using hard copy graphs. Although subjects with both methods did move further away from the moving average seasonals as seasonality increased, the effect was less pronounced for the GRAFFECT decision aid. This effect operated to the advantage of the judgment method in that the results from chapter 3 indicated that deseasonalised single exponential smoothing was to be preferred in cases of high seasonality. The change in the decision outcome caused by introducing GRAFFECT was not entirely positive. The display form influenced the judges to be less "adventurous" in their treatment of atypical months in constructing the seasonal. A possible explanation is that the plot of the annual seasonal patterns, all on the same 12 month x-axis, somewhat masked the sequence of the plots. The lines were coded, as exhibited in chapter 5, to show the sequence, but the extraction of that information from the plot required specific attention.

Any loss in the perception of temporal aspects of the seasonality cues would lead to difficulty in applying appropriate weight to outliers. It would also make it difficult to detect any tendency for the seasonality to change over time. Proposals for development of the display for seasonal identification, and the testing of the conjectures mentioned are contained in chapter 10.

The improvement, due to GRAFFECT, in the fit between the identified seasonal pattern and the actual pattern in the forecast period was shown to be a contributor to the improvement in the overall accuracy reported in chapter 6. Given the findings of Lawrence (1983) that commercial forecasters commonly use judgmental methods for forecasting this points to the potential to improve commercial forecasts by improving data presentation methods. It should be noted that the improvement in seasonal pattern identification cannot be ascribed to the form of data display alone. The decision aid allowed the judges to remove the effects of trend before seasonal identification took place, and this might have had an influence in the improvement.

7.3 HUMAN INFORMATION PROCESSING ISSUES

7.3.1 INTRODUCTION

The discussion of the "HIP" literature in chapter 4 leads to the conclusion that human judges might be adversely affected by the presence of randomness in the seasonal pattern recognition process. That conclusion was based on research that was not directly applicable to the task setting addressed in this dissertation however. The general H.I.P. literature that formed the basis of many former opinions concerning judgment in forecasting did not include evaluation of the forecasting task at all. The nearest approach to the task setting was found in Eggleton (1976), in that research mathematically generated series were displayed for very short periods, and it was found that subjects were confused by noise. In an un-tested opinion Lawrence, Edmundson and O'Connor (1985) stated that subjects using the GRAPH method appeared to be attempting to fit seasonal pattern to noise sequences.

The results of the discriminant analysis reported in chapter 3 did not unequivocally reveal such a process. Although the presence and sign of the discriminant based on the autoregression coefficient of lag one might be interpreted as an implication that judgment was contra-indicated in the case of unstable series, the opposite implication may be derived from the *SEAS/SYN* discriminant. In that case, high values of the noise metric *SYN* indicated the use of judgment.

The design of the investigations carried out, and reported, in this dissertation does not permit issues of human information processing to be directly addressed. The data used is real data, and it is thus not possible to determine the exact nature of the "driving process" that generated the data. For instance, although it is of interest to examine the issue of judgmental seasonal pattern identification in the presence of randomness, the precise seasonal pattern in the data, and the precise nature of the randomness is not determinable. Despite this it would be unfortunate if the opportunity to present circumstantial evidence of rational judgment were allowed to pass.

7.3.2 RESEARCH HYPOTHESES

Although this issue is not capable of controlled statistical evaluation, for the sake of clarity two specific hypotheses are considered. In the absence of literature that considered the identification of seasonal pattern in economic time series, the first issue addressed is the applicability of the general H.I.P literature, and the results of Eggleton (1976). thus the first hypothesis addressed here is:

H7.4 Judges attempt to impute seasonal patterns from noise components in time series.

The GRAFFECT decision aid described in chapter 5 was designed to support the judge by allowing for decomposition of the time series data in line with the decomposition of the forecasting task. It was considered that it might be possible to reduce the cognitive load on the judge, giving rise to improved decisions. There is no a priori reason to suppose that this effect occurred in the task of seasonal identification. The second hypothesis addressed here is therefore:

H7.5 The use of the GRAFFECT decision aid does not affect the effect of noise on seasonal pattern identification.

7.3.3 DESCRIPTION OF THE STUDY

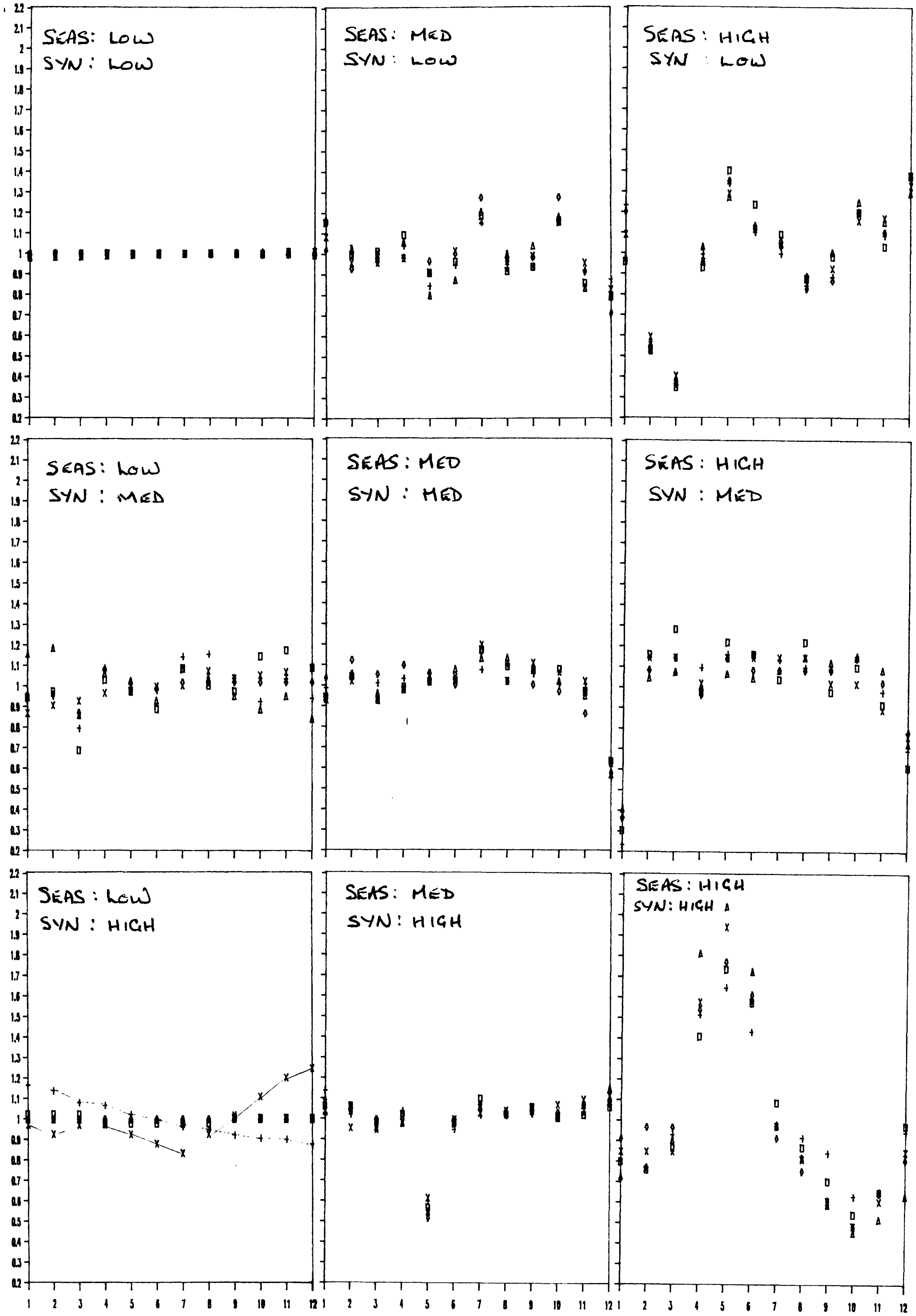
Initially, the data from the first study in this chapter was examined for evidence of poor judgmental performance, or systematic bias. Finally, data from a study reported in chapter 6 was considered. This gave a different perspective by providing few observations from several non-seasonal series in comparison with the data above which provides relatively more observations for fewer series.

Direct statistical analysis of the seasonal patterns poses certain difficulties. The patterns, by definition, have identical means, and there is no commonly accepted single metric to describe a pattern of twelve values in ordinal or better terms. It was assumed here that the seasonal pattern could be described by separately considering the shape of the pattern, and the amplitude of the pattern about the mean.

7.3.4 RESULTS

Figure 7.5, overleaf, shows scatter diagrams of the seasonal patterns identified by the subjects using the GRAPH display.

Figure 7.5 Plots of "GRAPH" Seasonal Patterns



Eyeballing the plots reveals reasonable consistency in the patterns. There is some evidence that subjects were prepared to identify seasonal patterns in series classed as non-seasonal by techniques based on the ratio to moving average. However, there is some doubt that the effect is systematic. Certainly no seasonal was imputed to the LOW *SEAS*/LOW *SYN* series, while series with higher values for the *SYN* noise metric did have some seasonal identified by the subjects. The LOW *SEAS*/HIGH *SYN* case has two observations where subjects identified a seasonal component, one of these (plotted as a +) may reflect slight trend [7] rather than seasonality. Thus there is some doubt that subjects imputed seasonality from noise in the LOW *SEAS*/HIGH *SYN* case. If there is doubt in that case, there can be no doubt that all the subjects identified a seasonal in the LOW *SEAS*/MED *SYN* case. The seasonal patterns all have a similar shape, but vary in amplitude and some seem to be possibly confounded with slight trend. It is not possible to interpret this as an imputation of seasonality from noise, in fact, the similarity in the shapes of the seasonal patterns would tend to deny this. The similar shapes could have been caused by:

* the subjects reacting to a seasonal cue to which the moving average process is not sensitive,

* a common heuristic being adopted in the face of the particular noisy signal, or

[7] a slight trend may not have been detectable using regression techniques on the forecast data.

* by chance.

It is not possible to determine which of the above holds from the plots themselves. Further, for the cases with a "moving average" seasonal, examination of the plots cannot reveal a systematic effect. The shapes of the plots are remarkably similar, though some vary in amplitude.

Since the visual examination of the plots of the seasonal factors indicated that all subjects identified very similar shapes it was assumed that any differences in the seasonal models might be observable in the amplitude of the fitted pattern.

The seasonal patterns identified were analysed for variation in amplitude by generating a score (AMPL) comprised of the sum of the absolute deviations of the factors from 1[8] :

$$AMPL = Sum_{n=1\ to\ 12}(Abs(FF_n - 1))$$

Analysis of variance produced the results displayed in Table 7.4

[8] the mean of the seasonal factors is 1, by definition

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects					
SEAS	109.834	2	54.917	793.188	.000
SYN	5.885	2	2.942	42.498	.000
METHOD	0.107	1	0.107	1.542	.218
2-way Interactions					
SEAS SYN	20.482	4	5.121	73.958	.000
SEAS METHOD	0.899	2	.450	6.492	.003
SYN METHOD	0.299	2	.149	2.159	.123
3-way Interactions					
SEAS SYN METHOD	.314	4	.078	1.133	.348
Explained	137.820	17	8.107	117.093	.000
Residual	4.985	72	.069		
Total	142.805	89	1.605		

Table 7.4 ANOVA on AMPL

The significant main effect on *SEAS* is obvious, since the value of the *SEAS* metric is determined by the amplitude of the "moving average" seasonal factors. The lack of a significant effect on method would seem to indicate that the provision of the decision aid did not affect any confusion that noise caused judges in identifying seasonal pattern.

In terms of the questions addressed in this section, the effect of the noise metric *SYN* is of major interest. The main effect on *SYN* is significant, though the reason for this was less obvious and required some post hoc examination. According to the conservative Scheffe rules, the effect of *SYN* on AMPL for low seasonal series was not significant, although the F value was relatively high (F=5.835). The only contrast to reveal a significant post hoc result was the effect of *SYN* on high *SEAS* series. Even this contrast did not reveal a linear effect, with the cell means being:

low *SYN*:- 2.71, med *SYN*:- 1.94, and high *SYN*:- 4.13.

This might imply that there are some individual time series effects confounding the analysis, but that there is a probable link between the amplitude of a judgmentally identified seasonal pattern and *SYN* for high seasonality series. The results of the analysis do not, unfortunately, shed any real light on the issue of the attempt to identify seasonality in noise sequences. In the event it is only possible to draw indirect inferences on this issue.

In an attempt to provide some insight on judgmental identification of seasonal patterns, the data from a separate study (reported in chapter 6) was reviewed. In that study 68 monthly time series were each forecast by the GRAPH and the GRAFFECT methods by persons experienced in time series analysis. According to the moving average process 17 of those series were non-seasonal. The forecasts for those 17 series were detrended using a regression process, and the amplitude of the imputed seasonal patterns in the forecast computed (as with AMPL above). Ignoring values less than 0.3 [7] as too small to be significant in the light of the computations carried out the following results were observed:

---

[7] this represents an average deviation (from 1) of less than 0.025 for each of the 12 seasonal factors

* Eleven of the series were treated as non-seasonal by every judge,

* Six series were treated as seasonal by every judge,

* Three of the above six series were found to have significant autocorrelation coefficients of lag 12, indicating the presence of a statistically detectable seasonal component.

The foregoing does not provide evidence that judges are confused by noise to any great extent in the seasonal forecasting task. There is also no indication that the provision of the GRAFFECT decision aid affects any such confusion. Where seasonality was observed, it was consistently observed by all judges. This is not to indicate that the observed seasonality was true seasonality. The difficulty here was adverted to in chapter 4 in the discussion of the work by Eggleton (1976), that there is in fact no absolute definition of what is seasonality. The presence of a similar pattern in certain months of a time series, over a few years, could be the result of chance. Similarly, a recent change in market conditions that generated a seasonality that did not previously exist will not be reflected in a seasonal signal detectable from the time series alone.

7.3.5 DISCUSSION.

The results obtained here conflict with those obtained in previous studies in that it was shown that human judges were not confused by noise when identifying seasonal patterns in real time series. The subjects achieved remarkable consistency in their results, with a variety of data display treatments and at varying levels of cue signal to noise ratios. This result explains in part the success of human forecasters that Lawrence, Edmundson, and O'Connor (1985) report.

The results indicate that overall the human judge is able to perform a seasonal pattern identification exercise on commercial time series data with consistency, regardless of the noise levels in the series. If, as seems likely, the consistency reflects the detection of a "true" seasonal then it has been shown that the human judge may be more sensitive to seasonal pattern in time series than commonly used statistical processes.

There is encouragement in the results for future research covering the effect of noise on seasonal pattern identification in series exhibiting noise and seasonal characteristics closely matched to real economic series.

7.4 AUTOMATIC DESEASONALISING

7.4.1 RESEARCH HYPOTHESES

In a prior, un-reported study, Lawrence, Edmundson, and O'Connor had determined that subjects presented with data that had been deseasonalised using the ratio to centered moving average process performed relatively poorly in the forecasting task. It was considered that this result was caused by an unforeseen problem of subject motivation, and the results were not reported.

However, the poor result may have been caused, at least in part by problems associated with the interaction of the deseasonalising process and the forecaster. For instance, the judges may have had difficulty in determining whether there was any seasonal signal remaining in the time series, and if so how that was to be handled. In a study primarily aimed at the investigation of judges' confidence in their forecast O'Connor (1987) examined the forecast accuracy and confidence levels of forecasts performed on three series with medium levels of seasonality. The study considered three experimental conditions:

1) Three levels of noise
2) Three levels of scale used in plotting the series
3) Cue data presented raw, or deseasonalised using the ratio to centred moving average.

O'Connor (1986) reported that deseasonalising had no impact on the confidence that the judges felt in the forecast, but that there was a significant (p=0.017) decrease in accuracy as a result of automatic deseasonalising. The mean accuracy over the three series was 15.8% poorer as a result of presenting deseasonalised data. This study leaves open two questions that are important to the development of the forecasting decision aid:

1) What would be the effect of automatic deseasonalising in the decision aid environment?

2) Is the decrease in accuracy related to the level of seasonality in the series?

In the first issue above there are a number of factors that might cause a difference between forecasts based on a hard copy data presentation and those from GRAFFECT. The subjects with the hard copy plots of deseasonalised data did not see the original series, this might have affected their approach to the extrapolation for instance by causing them to seek signal in what would have been generally a noise sequence. The users of the decision aid would deal with the original series in determining any trend component, and in determining whether to invoke the deseasonalising process. Thus the position that the judge would face after automatic deseasonalising would be closely analogous to the position after judgmental deseasonalising. There would remain a difference, however, that in the former case the judge would have no explicit knowledge of the nature of the model fitted.

The second issue is of interest because it was shown in chapter 3 that the level of seasonality in the series probably had some effect on the accuracy of judgment relative to deseasonalised single exponential smoothing for the 1-6 month forecast period.

7.4.2 DESCRIPTION OF THE STUDY

7.4.2.1 THE SUBJECTS

Five subjects were selected for this trial, all were enrolled in a postgraduate course in operations research at the University of New South Wales. They had completed seven weeks of study of time series analysis, and in that time had undertaken forecasting exercises with a wide range of forecasting techniques. Their participation in the study was optional, and carried no course credit. An atmosphere of amicable competition was engendered, and a small prize was offered for the student achieving the best overall results in the study.

7.4.2.2 CUE PRESENTATION

A version of the GRAFFECT decision aid was produced that invoked an algorithm[10] to deseasonalise the data instead of the special screen presentation for judgmental deseasonalising. Thus, subjects were able to decide whether to deseasonalise, and they had seen the original data containing the seasonal pattern.

Each subject used both the original and the modified decision aid. The forecasts by each subject were made in three "sittings" one week apart, and the sequence of the data cues was randomly assigned,

---

[10] ratio to centered moving average.

without replacement, for the sittings. This is more fully explained in the next section when dealing with the assignment of time series to each sitting.

7.4.2.3 THE TIME SERIES

Nine series were selected, from the "M-Competition" database of monthly series, that exhibited three levels of seasonality based on the seasonal factors developed by the authors of the "M-Competition". The seasonal factors were manipulated as described in chapter 3 to produce a metric called *SEAS* for each series. The levels of seasonality selected were:

1) LOW: series that exhibited no seasonal pattern, that is with a *SEAS* metric value of zero.

2) MEDIUM: series with a *SEAS* metric from the mid-range of the sample.

3) HIGH: series with a *SEAS* metric near the top of the sample distribution.

Each series was forecast by every subject, with each of the two experimental instruments. Thus each subject forecast six series in each of three sittings. No series was forecast more than once in a subject sitting, but with that proviso the series were randomly assigned to the three sittings. This, and the random assignment of the order of the use of the two experimental instruments, was achieved by the random allocation to the "subject-sittings" of six "instrument-series" pairs, without replacement, from the 18 possible combinations. In making that allocation any duplication of a time series in a sitting was rejected.

The MAPE errors of the forecasts were computed for forecast periods 1 to 6 and 7 to 12. The results analysed using analysis of variance.

7.4.3 RESULTS

Table 7.5 contains the results of the analysis of variance on month 1 to 6. It shows that the only significant main effect was due to seasonality, and that no interactions were significant.

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects					
METH	1.282	1	1.282	.021	.886
SEAS	1206.068	2	603.034	9.672	.000
2-Way Interactions					
METH SEAS	30.006	2	15.003	.241	.787
Explained	1237.356	5	247.471	3.969	.003
Residual	5237.155	84	62.347		
Total	6474.511	89	72.747		

Table 7.5 Months 1-6

Figure 7.6 illustrates the cell means, it clearly shows that the error rate rises with increasing seasonality, for both the GRAFFECT (GRF) and the Moving Average (GMA) methods. The figure shows that the error rate for the low seasonality series was lower for the GMA method, and that the lower error for high seasonality series was the GRF method. Post hoc analysis showed that this effect was not significant.

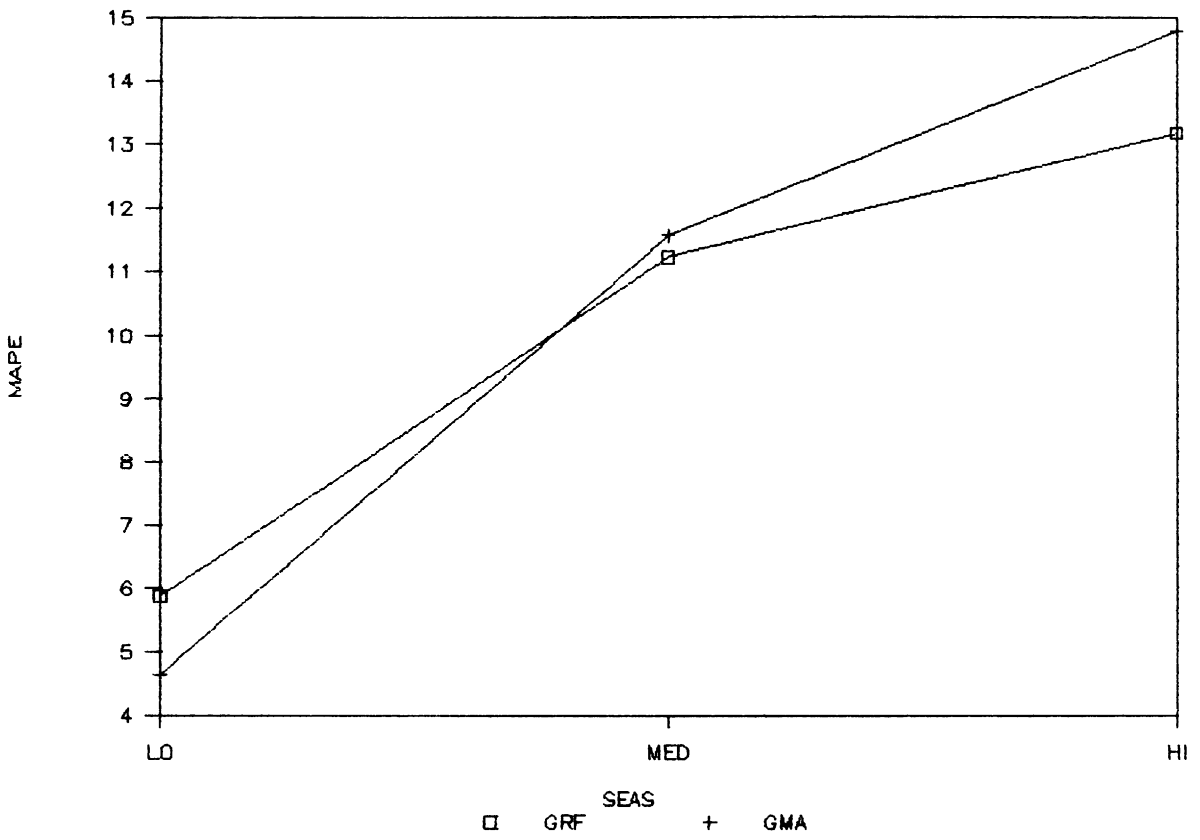


Figure 7.6 Cell Means for Method/*SEAS* Interaction

The analysis of the errors for months 7 to 12 failed to reveal any significant main or interaction effects. The GRF method was seen to have a marginal advantage in error rate at each level of seasonality, but that this was not statistically significant.

7.4.4 DISCUSSION

The results obtained in this study are different to those of O'Connor (1986). He found that deseasonalising the data prior to presentation to the subjects caused them to forecast significantly less accurately. The same result had been observed by Lawrence, Edmundson and O'Connor in an un-reported study. The results here show that subjects who are permitted to see the original series, and as part of a decomposed strategy are allowed to invoke an automatic deseasonalising procedure, do not lose accuracy in the same way.

The results provide another confirmation of those reported earlier in this chapter. The particular data interface used in GRAFFECT seems to have modified the seasonal identification decisions of the forecasters to make them behave more like the Moving Average process used in the automatic deseasonalising.

The above would have significant implications for the future design of forecasting decision aids. However, before any conclusion that judgmental seasonal identification could be replaced by automatic procedures it would be necessary to test alternative designs of the data interface. It might be recalled that there was an implication in chapter 3 that judgment supported by hard copy plots had an advantage in the case of low seasonal signal. That advantage appears to have evaporated with the introduction of the data interface used in GRAFFECT. Different designs of the interface discussed in chapter 10 might restore that advantage. In that case it might be possible to adopt a strategy of prompting judges to use either the moving average automatic process or judgmental deseasonalising depending on the level of seasonality seen in the series.

7.5 SUMMARY

Overall, the results reported in this chapter indicate that judges are capable of performing relatively well in the task of deseasonalising time series. There is a strong implication that the accuracy of the extrapolative task is highly dependent on the ability to identify the seasonal pattern, and that further attention to this aspect of the decision aid is warranted. It has been shown that the decision outcome in this task setting is altered by changes to the form of data presentation, and that this points to the possibility of gaining further accuracy in judgmental extrapolation.

The results raise the possibility of replacing judgmental seasonal identification in some circumstances because the data interface used in GRAFFECT does not suffer from the loss in overall accuracy reported for automatic deseasonalising using hard copy plots. The desirability of such a process would depend on a trade off between the effort required to perform the judgmental task, the relative accuracies of the processes, and any requirement for the forecast to be reviewed judgmentally to take account of market factors. The data captured shows that judgmental identification and modelling of the seasonal component of a time series takes approximately 100 seconds. If it were necessary to carry out the task for every extrapolation then considerable effort could be saved by providing automatic support. The importance of this would diminish however if, as seems probable, the seasonal model would have greater currency than one extrapolation.

7.6 LIMITATIONS.

The data used in this experiment was drawn from a data base of 68 monthly time series. It is not possible to determine the extent to which the series reflect the general population of business time series. This constitutes a threat to the external validity of the results. It should be said, however, that the series were selected from a very large database, and that the academic forecasting community has eagerly adopted the database as the most significant heterogeneous database available.

The time series used were real series, therefore it was not possible to determine with certainty the true seasonal component in the series. Initially, it was assumed in the latter part of the chapter concerning human information processing, that a seasonal pattern identified using a centered moving average calculation reflects the true seasonal pattern. The results obtained here have been interpreted to indicate that the human judge is capable of identifying the seasonal component of a time series, though that component may be different from the moving average seasonal. It is also possible to interpret the results to the effect that the human judge is capable of generating a consistent seasonal, but that this might not be the true seasonal pattern.

It is not possible to determine whether the assumption of constancy held for the time series selected, and if not, the results might be confounded by the subjects responding to turning point cues.

7.7 REFERENCES

Eggleton I.R.C. "Intuitive Time Series Extrapolation", *Journal of Accounting Research*, 20,1. (1982), 68-102.

Eggleton I.R.C. "Patterns, Prototypes, and Predictions: An Exploratory Study " *Selected Studies on Human Information Processing in Accounting*, (1976). Supplement to *Journal of Accounting Research* 14: 68-131

Lawrence M.J. "An Exploration of Some Practical Issues in the Use of Quantitative Forecasting Models", ***Journal of Forecasting***, 2 1983, 169-179.

Lawrence M.J., Edmundson R.H., and O'Connor M.J. "An Examination of the Accuracy of Judgment Extrapolation of Time Series" *International Journal of Forecasting* 1 (1985), 25-35

Makridakis S., Andersen,A., Carbone,R., Fildes,R., Hibon,M., Lewandowski,R., Newton,J., Parzen, E., and Winkler,R. "The accuracy of extrapolative (time series) methods : results of a forecasting competition", ***Journal of Forecasting***, Vol 1, no 2, 1982.

O'Connor, M.J. "An examination of the accuracy and determinants of confidence levels in judgmental time series forecasting", Ph.D. Dissertation, University of New South Wales, (1986).

8. IDENTIFICATION OF TREND

8.1 INTRODUCTION

As described in chapter 5, traditional time series analysis, and the structure of the GRAFFECT decision aid, is based on decomposition of the time series in terms of seasonal, cycle, trend and noise components. In chapter 7 the identification of seasonal characteristics was examined in the attempt to see if the differences in cue presentation contributed to the improvement in accuracy achieved by GRAFFECT, and to throw light on other aspects of seasonal identification. In this chapter trend identification is considered. However, the design of the cue presentation display for trend identification was no different in GRAFFECT than in the judgmental hard copy method GRAPH. In making their forecasts, all subjects were required to consider trend first, therefore the data displayed was also the same for GRAPH and GRAFFECT. Therefore this chapter is concerned solely with human information processing aspects of trend identification.

This study is aimed at the evaluation of the capabilities of the human judge to identify trend in a time series. As discussed in chapter 7, there is no absolute definition of the components identified in classical time series analysis. Thus, it is not possible to determine what is the true trend in a real time series. However, there has been common acceptance that the least squares fit provides a reasonable estimator for trend. Judgmental capability to fit a trend line to a time series is compared with the "regression" fit, and the following questions are addressed:

a) Is the trend identified by eye-balling a time series similar to the trend identified by fitting a regression line?

b) Is it possible to improve accuracy of forecast by substituting a trend derived using a regression fit, or by artificially damping the judgmentally identified trend?

8.2 JUSTIFICATION OF THE STUDY

It has been shown that the accuracy of forecast is significantly affected by trend identification, especially beyond the first few months of forecast (see Fildes 1983, Gardner and McKenzie 1985). As described in chapter 4, judgmental fitting trend lines to scatter diagrams is an acceptably accurate process. Mosteller et al (1981) showed that students were able to accurately fit a regression line to a scatter diagram, but that they exhibited a slight tendency to exaggerate the slope of the line, and to adopt a slightly different strategy for what Mosteller et al (1981) called "fat" plots [1].

There has been some slight doubt as to whether the results of Mosteller et al (1981) are directly applicable to time series extrapolation. Lawrence and Makridakis (1986), in an examination of the extrapolation of a time series of seven events, found that subjects damped the slope of the

[1] The subjects fitted a line by minimising the principal component. This finding probably has little importance to time series forecasting.

trend line, especially for downward sloping series. The subjects had been informed that the series represented annual unit sales, and Lawrence and Makridakis (1986) conjectured that the damping was a result of the forecasters anticipating managerial action to stop the decline. There was no evidence for this supposition [2], and the study left open the question of forecasting monthly time series.

Gardner and McKenzie (1985) showed that by damping the trend component of a Holt forecasting model the long term accuracy of the forecast was improved. That improvement appeared to commence at about the eighth month of the forecast horizon. The algorithm that they used damped the trend more if the "trend data was erratic". The proprietary forecasting method of Parzen uses a conceptually similar method, damping according to the "length of memory" of the series.

For the development of judgmental forecasting, and of forecasting decision aids in particular, it is of interest to determine how human judges fit a model, and then apply that model in the extrapolation. It is also of interest to determine whether the forecast can be improved by an automatic damping of the trend depending upon the stability of the trend model.

---

[2] It could equally be argued that if the management was not able to affect a decline of seven years standing then there was little evidence that they would be more successful in the future.

8.3 DESCRIPTION OF THE STUDY

8.3.1 RESEARCH QUESTIONS

This study comprises the exploratory analysis of data captured for the "accuracy" study reported in chapter 6. The analysis is intended to throw light on the following research questions:

Q8.1 Does the eye fitted trend line approximate a regression line fitted to the same data points?

Q8.2 Is there evidence of the judges modifying the fitted line when extrapolating by, say, damping the trend in a regression to the mean?

Q8.3 Is the forecast improved by damping the judgmentally determined trend extrapolation?

8.3.2 THE METHOD

The data examined in this chapter comes from the experiments reported in chapter 6. In that experiment, three subjects used the GRAFFECT decision aid to forecast time series from the 68 monthly series from the "M-Competition". The data examined here is that used in the comparison of accuracy with DSE in chapter 6. The subjects were skilled in time series analysis, and the use of the GRAFFECT decision aid, but they had not had the opportunity to read the Gardner and McKenzie (1985) study at the time the data was captured. Thus, the subjects were not directly aware of the implications for accuracy of damping the trend model.

The GRAFFECT decision aid was designed to capture the modelling data as well as the extrapolation data. Therefore, the data from that experiment permits a review of the trend handling strategies of the subjects.

Of the 68 monthly time series forecast, the subjects identified a model fit trend in 37 series. This means that the subjects found trend in the history data of those series. This chapter examines those 37 series, and the trend lines that the subjects fitted.

8.3.3 RESULTS

8.3.3.1 MODEL FITTING, EYE VERSUS REGRESSION

Figure 8.1 illustrates the capabilities of the trend identification function in GRAFFECT. It shows that subjects may model the trend in the history data with an optional single turning point, and to then extrapolate the trend at any desired angle.

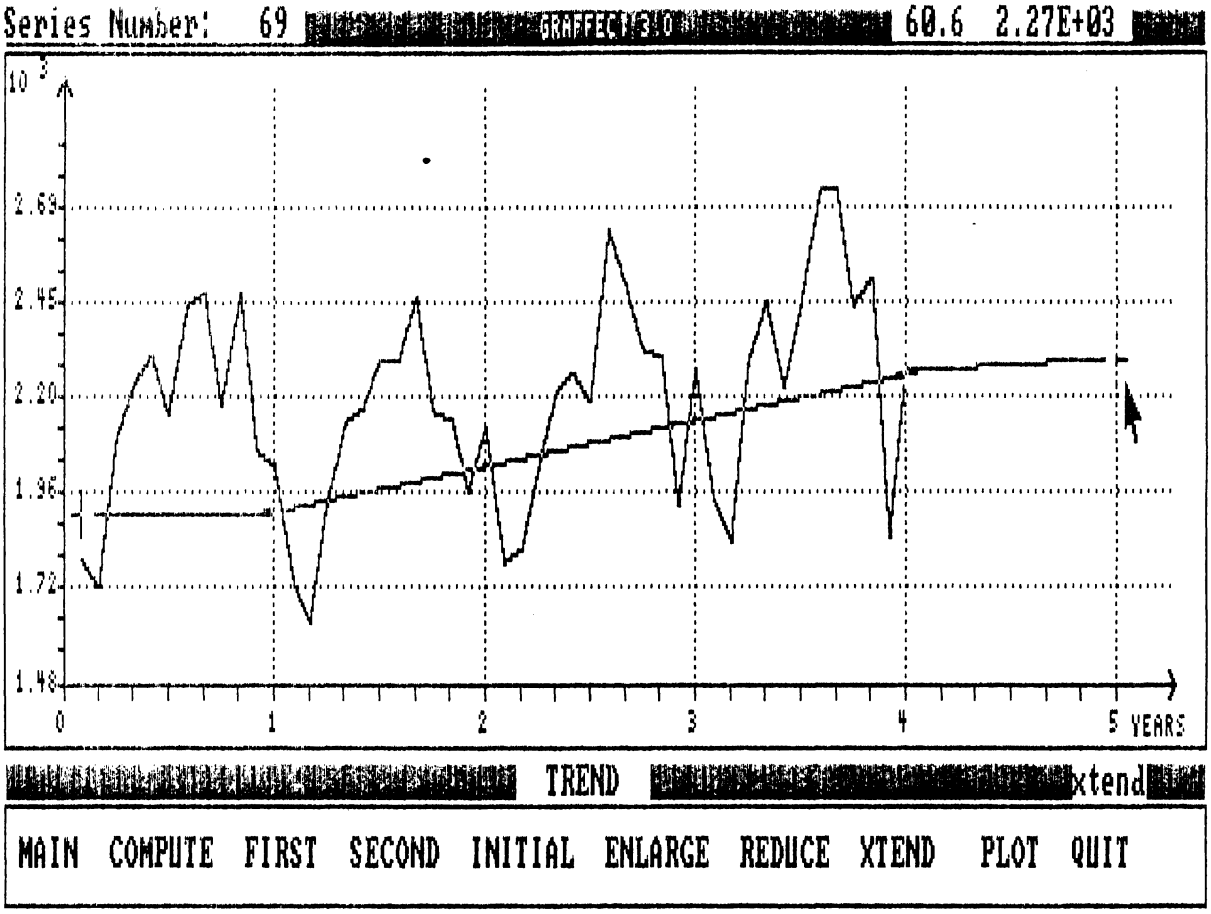


Figure 8.1 Illustration of Trend Identification Screen

The capability of including a turning point in the model trend enables the subjects to identify the most recent trend model. The ability to extrapolate a different trend line allows forecasters to anticipate that the trend model will not continue in the model fit form. In the following analysis only the "most recent" trend line in the history data[3], and the extrapolated trend line are examined. In fitting a regression line to the data for comparison purposes, only the points

[3] The other trend line that might be identified in the history data was intended to permit the removal of any trend effect in that data so that identification of the seasonal could proceed.

corresponding to the most recent eyeball fit were included. Figure 8.2 displays the results of comparing the "regression coefficients" derived from the regression fit and the eyeball fit. It shows the distribution of the differences between the regression and the eye fitted "X" coefficients.

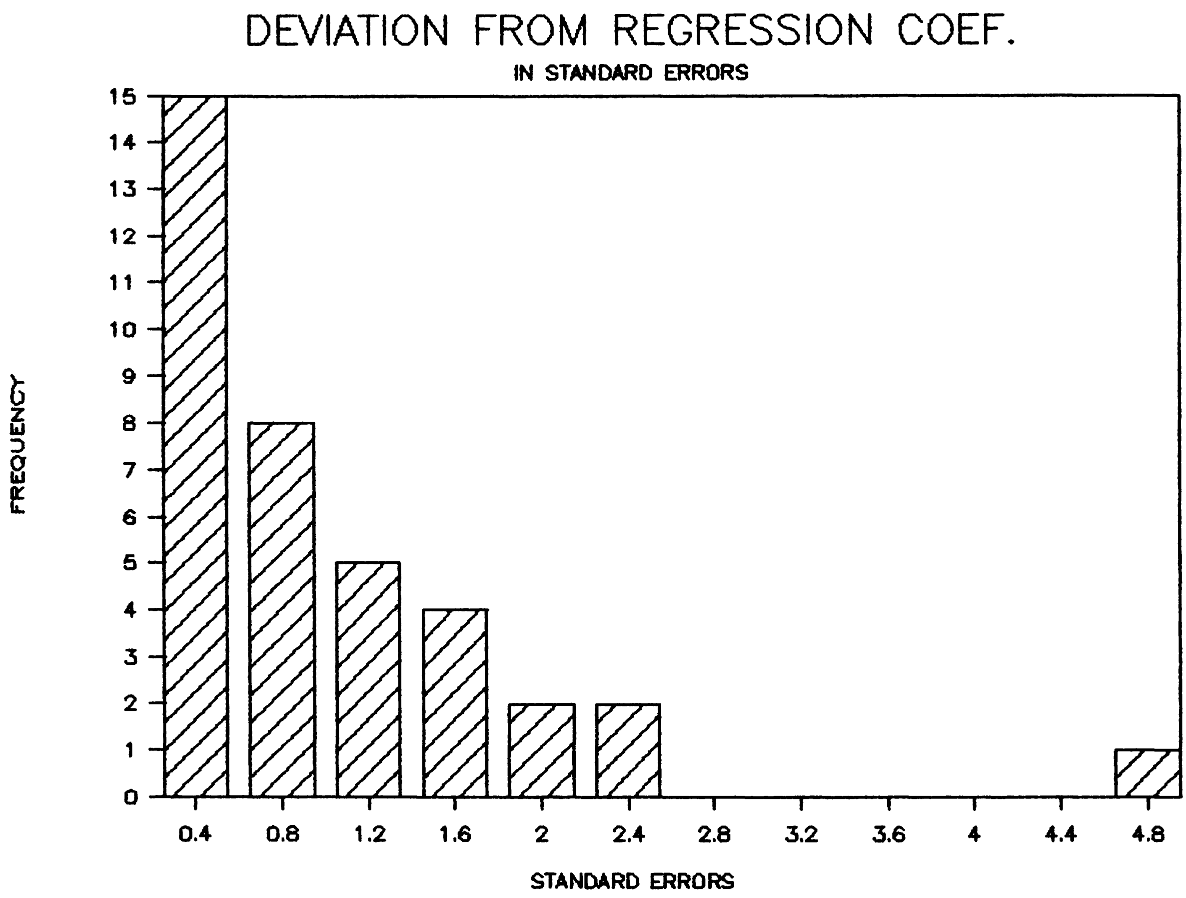


Figure 8.2 Distribution of Model Errors.

The histogram indicates that, with one exception, the eye fitted trend model is the same as the regression line, with greater than 90% confidence. A paired t-test showed that the two data sets did not come from different populations, and that they were correlated at 0.995 (p=.000).

The results from model fitting confirm the results of Mosteller et al (1981) that human judges are capable of fitting a good regression line by eye. The exception noted by Mosteller et al (1981), that with what they called a "fat" data set the subjects appeared to fit a line by minimising the principal component, had no relevance in this study. The time series data most closely followed their "standard" data set, though even that data set was a scatter diagram, and had several observations at each "x" value.

The results of this study show that the Mosteller et al (1981) results are generalisable beyond the fitting to scatter diagrams, and that human judges fit "good" regression lines by eye to time series data. The results are limited to the extent that only the slope of the lines was considered, the GRAFFECT decision aid constrained the positioning of the line by anchoring the right hand end of the line. Thus the results do not address the issue of the estimation of the "Y" intercept.

8.3.3.2 JUDGMENTAL DAMPING OF TREND

As described above, subjects were able to extrapolate the trend at a different slope to that of the "most recent model" of trend in the history data. In 10 of the 37 series examined, the subjects had chosen to utilise this feature, and had damped the trend in each of those cases. None of the damped series had more than 25 observations in the "most recent model", and no non-damped series had fewer than 18. It appears that a major determinant of whether the subjects damped the trend was the number of observations in the "most recent model". Table 8.1 reports the incidence of series that were damped and not damped, by the number of months covered by the "most recent model" of the historical trend.

Number of Series	Length of the "Most Recent" Historical Trend Model in Months <18	18	19	20	21	22	23	24	25	>25
DAMPED	6		1		1				2	
NOT DAMPED		1		1	2		1		3	19

Table 8.1 Effect of Number of Points on Damping

There was no determinable criterion for damping or not damping for the series with 18 to 25 observations in the modelled trend line. The small number of such series precluded the use of sophisticated statistical tests to determine any criteria.

The results indicated that the human judges were inclined to damp modelled trend, especially in cases where the modelled trend was not of long duration. Since the subjects had not been exposed to the literature on damping trend, this probably reflected the level of confidence in the model rather than an attempt to reduce the long term error rate.

These results are not inconsistent with those of Lawrence and Makridakis (1986). In that case the subjects were given contrived time series with limited numbers of observations. If the subjects in that study had viewed the time series merely as a sequence of seven points, and did not take the extra step of considering that the series did in fact represent seven years of data, then it would be expected that the subjects would damp the trend, in line with the results reported above.

8.3.3.3 AUTOMATIC DAMPING OF TREND

Each of the 37 series was placed into one of four classes according to the following subjectively determined criteria:

a) series for which the forecast trend was an accurate reflection of the trend in the "actual" data, that is no further damping was required beyond any damping already applied by the subjects,

b) series for which only very minor further damping was indicated,

c) series for which further damping was clearly indicated, and

d) series for which the slope of the trend line would require increasing in order to obtain a better forecast.

This was carried out by inspection of the time series and forecast plots from the data described above, with the actual values superimposed. In doing this, particular attention was paid to the 7-12 month forecast horizon, but the assessment made was intended to preserve the accuracy of the 1-6 horizon forecast. Table 8.2 summarises the results of this process, and indicates the numbers in each class that had been damped by the subjects when they made the forecast.

CLASS	NUMBER OF SERIES IN THE CLASS	
a) NO FURTHER DAMPING REQUIRED	25	(of which 7 had been damped by the subjects)
b) MINOR DAMPING	2	(of which 1 had been damped by the subjects)
c) DAMPING REQUIRED	8	(of which 1 had been damped by the subjects)
d) INCREASE, NOT DAMP	2	(of which 1 had been damped by the subjects)

Table 8.2 Classification of Series by Effect of Damping Trend

The table 8.2 shows that damping, whether by the subjects or by another process, would have been beneficial in a total of 17 cases out of the 37 examined (being all the cases in classes b) and c) and the 7 cases in class a) that had been damped by the subjects). The benefit would have arisen from the mitigation of the effect of an apparent turning point having occurred in the trend component of the time series.

As can be seen, of the 17 series for which damping would be of benefit the subject forecasters had already applied some damping in 9 cases. Two of the cases judgmentally damped could have benefited from some further damping. These results indicate that there were only 10 of the 37 time series that might have been forecast better had an automatic damping process been in place in the decision aid, provided that such a process did not cause unintended side effects in the judgmental processes. For 27 cases the forecast would not have improved, and may well have deteriorated. In itself, this low probability of obtaining a benefit (p=.27) does not rule out the effectiveness of a damping mechanism. It is possible that the benefit to be obtained in the 27% of cases is sufficient to outweigh the costs in the remaining 73%. This is evaluated below.

Both from the above results, and in the light of the Gardner and McKenzie (1985) results, there would be no justification in applying the same damping factor to all series. Damped forecasts were therefore produced from the forecast data generated by the subjects by taking into account the stability of the trend model. Thus trend that had been sustained over more periods in the history data was less damped than that over shorter periods. A simple, linear method of damping was adopted. The method was to forecast the monthly event as the weighted average of the judgmental forecast as it was made, and the judgmental forecast without the trend component included. The seasonal component was therefore not affected by the data manipulation. The weights adopted reflected the proportion of the possible 48 months of history data to which the trend model was fitted. For instance, a trend model that was fitted to 48 points was not damped, while one fitted to 24 points had half of its effect removed.[4]

As might have been expected from the results of the visual examination of the plots of the data, there was no overall benefit obtained by the damping exercise. Paired t-tests on the 37 series examined failed to reveal a significant difference between the original forecast and the damped forecast, though the damped forecast was always inferior in terms of the MAPE at both the 1-6 and 7-12 forecast horizons. It is possible that a more sophisticated, non-linear damping mechanism might have revealed an advantage. However, the risk of the subjects not understanding the implications of such a mechanism, and reacting unpredictably would be high. Therefore such a process would require full examination in another experiment rather than a post experimental examination.

8.3.3.4 COMPARISON OF ACCURACY

It was somewhat surprising that the "trend" metric developed in chapter 3 failed to act as a discriminator between judgmental forecasts and deseasonalised single exponential smoothing. This could have been caused by:

1) the failure of judgment to handle trend adequately,

2) the inadequacy of the metric to indicate the presence of trend, or

---

[4] Several weightings were assessed by reducing the maximum of points considered from 48.

3) the high correlation of the metric with another that had entered the analysis previously.

In the light of the prior analysis contained in this chapter, it would seem unlikely that the first of these possibilities would hold. As reported in chapter 3, "trend" is highly correlated with "noise" at 0.7. This confusion of the metrics might have masked any true trend related effect. Indeed it was not possible to fully determine whether the entry of noise into the analysis was mainly as a result of a trend effect. If either the trend or noise metrics had measured the true trend in the operative part of the series, then one of them should have appeared in the discriminant function in chapter 6. The failure of either to do so indicates that the trend characteristic was not adequately captured by the metrics[5].

In order to establish whether there might be a judgmental advantage in the presence of trend the data from chapter 6 was analysed further. There were 37 series that had a trend judgmentally identified, and 31 series with no trend, in that study. The accuracy of the forecast, measured in MAPE, was compared with deseasonalised single exponential smoothing for each of the two groups. In order to determine the accuracy of the judgmental forecast relative to a simple

[5] An alternative explanation would be that across the sample of series there was no relationship between trend in the history data and that in the validation data.

statistical method that accommodated trend a comparison was also made with the Holt forecast from the "M-Competition".

Table 8.3 indicates that for the series that had a judgmental trend identified the GRAFFECT based forecasts were more accurate than either the deseasonalised single exponential smoothing or the Holt forecasts. The improvement was statistically significant at better than 0.01 for the months 1-6 forecast horizon, and better than 0.05 for the months 7-12 forecast horizon.

Method	MAPE	Improvement In MAPE	Significance of Improvement
months 1-6			
GRAFFECT	6.69		
DSE	8.29	1.6	0.007
HOLT	8.45	1.76	0.003
months 7-12			
GRAFFECT	8.40		
DSE	10.50	2.1	0.03
HOLT	14.35	5.95	0.024

Table 8.3 Improvement in MAPE for Series with Trend

The results of the comparison for the 31 series that did not exhibit trend revealed that although the average MAPE achieved by the use of GRAFFECT was approximately one percentage point lower than that of deseasonalised single exponential smoothing, and somewhat more than that below HOLT, the improvement was not statistically significant. The results are reported in table 8.4 below.

Method	MAPE	Improvement In MAPE	Significance of Improvement
months 1-6			
GRAFFECT	12.92		
DSE	14.20	1.28	0.552
HOLT	15.90	2.98	0.303
months 7-12			
GRAFFECT	17.89		
DSE	18.71	0.82	0.759
HOLT	21.63	3.74	0.381

Table 8.4 Improvement in MAPE for Series without Trend

The results above indicate that there is no difference in forecast accuracy between the judgmental and statistical methods for series without trend. However, for those series in which the judges found trend the GRAFFECT method proved to have an advantage in average accuracy of approximately 19% or better of the MAPE of the statistical forecast. This would strongly suggest that the derivation of the metric for trend was lacking. It should prove possible to derive a metric that would indicate the use of judgmental forecasting, for both the forecast horizons considered. In the absence of such a metric, it would appear that a visual examination of a plot of the data should lead to a reliable determination of whether judgment would have an advantage.

8.4 DISCUSSION

The results obtained here show that subjects who are familiar with data analysis are prepared to damp the trend component in some time series. That damping takes place primarily in the case of the trend model being of short duration. This conforms to the conceptual basis of the Gardner and McKenzie (1985) work, though the subjects were not aware of that work at the time of data capture. It is possible that a higher proportion of series would have been damped had the subjects received instruction based on the Gardner and McKenzie (1985) results.

The design of the study precludes a statistical analysis to determine whether there are any systematic effects such as the heavier damping of downward sloping series (Lawrence & Makridakis 1986), or of the over estimation of the slope (Mosteller et al 1981). Of the 37 series exhibiting trend 29 had an upward slope. The slope identified by the judges was greater than the fitted regression line slope for 14 of the 29, and less than the regression slope for 15 of the 29. This would indicate that it was unlikely that the judges were systematically biased to over estimate an upward sloping trend line. The picture was somewhat different for the downward sloping series. Only eight of the series examined did show a downward slope. In seven of the eight cases the judges fitted a line that was less steep than that of the regression line. In the remaining case the difference was negligible. With such a small number it is not possible to interpret the results with confidence. It is possible, however, that the results of Lawrence and Makridakis (1986) could be explained either in terms of the down slope affecting the damping, or of it affecting the model fitted to the data. Further experiments would be required to determine which of the two explanations (or both) holds.

The results do not show that an automatic damping mechanism would be likely to benefit the forecaster. The damping technique adopted in this study was very simple, and it is possible that a more sophisticated algorithm would have been of benefit. This would have caused a difficulty in the interface with the subject. If the subject did not fully understand the damping mechanism then the outcome of the subject attempting to balance the damping desired against that provided by the automatic mechanism would be indeterminate.

Finally, it was shown that in cases where the judge identifies a trend, there is good reason to forecast the series judgmentally. If, as discussed in chapter 10, it proves possible to construct a metric that correlates highly with judgmentally identified trend, further developments in the decision aid would be possible. Provision of a mechanism to identify series for which judgment would probably prove more accurate would enable large databases of time series to be screened automatically. The decision aid could then extract those series for which judgmental extrapolation was indicated.

8.5 REFERENCES

Fildes R. "An evaluation of Bayesian forecasting", *Journal of Forecasting*, 2, (1983), 137-150.

Gardner E.S. & McKenzie E. "Forecasting trends in time series", *Management Science*, 31,10, (1985), 1237-1246.

Lawrence, M.J., & Makridakis, S., "Human judgment in extrapolation", *International Forecasting Symposium*, Paris, 1986.

Mosteller F., Siegel A.F., Trapido E., & Youtz C. "Eye fitting straight lines", *The American Statistician*, 35,3, (1981), 150-152

9. THE RESIDUAL NOISE IN THE TIME SERIES

9.1 INTRODUCTION

This chapter completes the examination of the three forecasting sub-tasks by reporting on the extrapolation of the residual series after trend and seasonality have been judgmentally removed.

There are two major objectives addressed in this chapter:

a) To determine the extent to which the subject forecasters succeeded in removing the information content of the series by their seasonal and trend modelling activities,

b) To examine how well the judges performed in the extrapolation in comparison to alternative techniques.

9.1.1 JUSTIFICATION OF THE STUDY

It was shown by Spencer (1961) that subjects were able to perform well in the estimation of averages from both symbolic and graphical data, and that they performed consistently in extrapolating a number sequence. However, it has not been shown that judgmental extrapolation is rational for economic time series that have been transformed to obtain stationarity of the mean. Even the result obtained in chapter 3, pointing to an advantage to judgment for series with a high "noise" characteristic was not clear out. As discussed in chapter 3, and in chapter 8 there is a high correlation between the "noise" and "trend" metrics.

Although the de-trending and de-seasonalising processes may be well suited to the use of judgment it is possible that the extrapolation of the resultant series might be performed better by statistical procedures. In that case, the separation of the extrapolation from the other processes would not be un-natural in the commercial setting. It is quite possible that the identification of trend and seasonal characteristics of a time series would take place irregularly, and less frequently than would extrapolation.

9.2 DESCRIPTION OF THE STUDY

9.2.1 RESEARCH QUESTIONS

This study is intended to throw light on the following research questions:

Q9.1 Is there any evidence in the residual series that the subjects failed to model the trend and seasonal components adequately?

Q9.2 Would the overall forecast have been improved if the extrapolation of the residual series had been performed using statistical extrapolative processes?

9.2.2 THE METHOD

The data examined in this study was derived from the data captured in the studies reported in chapter 6. In that study, two sets of subjects were used to obtain forecasts using the GRAFFECT decision aid. One set of subjects comprised 3 post graduates experienced in time series analysis, and would have been expected to appreciate the implications of extrapolating a noise sequence over a twelve month forecast horizon. That is, after they had removed all the signal that they could detect in the time series, there would be no reason to deviate from a straight, horizontal line in extrapolating the residual.

The 35 post graduates subjects in the second set were not experienced in time series analysis, but they had received minimal instruction in the concept before attempting the forecasting task. It was considered that many managers that make forecasts in commerce would have a similar level of knowledge. The results from the two studies have been referred to here as "EXPERT" and "NOVICE" respectively:

EXPERT: three subjects used the GRAFFECT decision aid to forecast the 68 monthly time series from the "M-Competition". The data examined here was that used for comparison of accuracy with DSE in chapter 6.

NOVICE: 35 subjects forecast the 68 time series using the GRAFFECT decision aid. These subjects were relatively inexperienced in time series extrapolation. .

In making their forecasts, the subjects judgmentally decomposed the model fit data, removing the trend and seasonal components. This generated a transformed series that has been called the "residual" in the following description. The residual should be noise if the trend and seasonal components are adequately removed.

The residual series was then forecast by the subjects, and their forecasts have been compared with those generated by various statistical extrapolation methods operating on the residual.

9.3 EXAMINATION OF THE RESIDUALS

9.3.1 SIGNAL IN THE RESIDUAL

Each series of residuals generated by the judgmental removal of the trend and seasonal signals was examined to determine the effectiveness of the judgmental processes, and to gain insight to the nature of the series.

Before the residual data was examined, a preliminary test of each of the residual series was carried out to determine whether they approximated the normal distribution. According to the Kolmogorov-Smirnov goodness of fit test, of the 68 series only 6 from the EXPERT data and 5 from the NOVICE data did not appear to be normally distributed, (with better than 90% confidence). Four series were common to both, and these exhibited significant turning points within the four years of data. This could explain the failure to fit the normal distribution in those cases, and chance alone could explain the remainder.

On the basis of the above finding it was reasonable to assume that the residual series were approximately normally distributed. That being the case, the series could validly be screened for information content using the Durbin-Watson statistic d. This indicated that all the residual series contained autocorrelations. This does not indicate that the original series were similarly autocorrelated, it is possible that serial correlations were generated in the residual series by the modelling transformations [1]. Indeed, the process of washing out the trend could have induced autocorrelations since the deviations from the trend line were scaled (up or down depending on the direction of the trend) by ever increasing amounts [2] with increasing forecast horizon.

The autocorrelation and partial-autocorrelation coefficients for each of the series were examined [3] judgmentally. This examination showed that the series, were stationary about the mean, and that no detectable trend or seasonal pattern remained in the residual series. Thus it was confirmed that the modelling exercise for seasonality and trend had removed most, if not all, the signal related to those items. Also, the modelling had not generated any spurious trend or seasonal signals.

---

[1] See Kendall (1973) page 164.

[2] See chapter 2 for a discussion of the effect of the multiplicative model on the variance of the residual series.

[3] This was carried out using the Box Jenkins "Identify" procedures in SPSSX.

The analysis of the EXPERT data showed that 14 of the 68 residual series appeared to be noise, that is there were no significant correlation coefficients to be detected [4]. This was contrary to the indications of the Durbin-Watson statistic, however this difference may simply arise from the different assumptions on which the processes are based. The remaining 54 series had a significant auto-regressive characteristic, many appeared to be AR(1) processes, though as many as 12 may have been ARIMA processes [5].

9.4 EXAMINATION OF THE JUDGMENTAL FIT

9.4.1 BASES OF COMPARISON

The judgmental extrapolation of the residual was compared to the extrapolation using various statistical methods:

a) Single exponential smoothing (Exp); it was anticipated that this would be a successful method of extrapolation since there was no judgmentally identifiable trend or seasonal component in the residual. The best "alpha" was determined on the model fit residual data by exhaustive trial.

---

[4] That is, the values of the auto-correlation coefficients did not lie outside two standard errors. The standard errors computed by SPSS are based on the assumption that the series are white noise

[5] For a discussion of integrated auto-regressive moving average processes see Makridakis, Wheelwright & McGee (1983) page 352 following.

b) The mean of the residual series (M48, M12); this might provide a good estimate also as the process was stationary in the mean. In case the modelling activities involved in the judgmental extraction of trend and seasonality had generated instability in the residual for the early observations, the mean was computed for the whole residual series (M48), and for the most recent 12 observations (M12).

c) The naive forecast (Nve); this forecast adopts the value of the most recent observation as the predictor of the first value in the forecast period. This technique is recommended for "drunkards-walk" series.

d) Box-Jenkins (BJ); this method was included as a reflection of its reputation, and to permit any autoregressive component of the series to be appropriately modelled.

9.4.2 ANALYSIS

The direct comparison of the "accuracy" of the methods is not possible. The "series" forecast by the above methods were residuals. There is no corresponding "residual-actual" value in the forecast period with which to compare these forecasts. In the place of such an actual, accuracy has been determined with reference to the withheld actual values of the time series. Those values were adjusted by the seasonal and trend models identified in the judgmental forecast. Thus, the "accuracy" of the extrapolation of the residual is determined in terms of the the accuracy of the overall forecast based, in part, on that extrapolation. This is not strictly the same as the accuracy of forecast of the residual series, however, it is a fine theoretical point that does not affect the objectives of the study. After all, the objective in examining this component of the overall forecast is to consider the accuracy of that overall forecast.

9.4.3 RESULTS

Examination of the extrapolations made by the subjects revealed that in no case did the experienced subjects put any pattern in the extrapolation. Of the 35 novice subjects, 12 produced extrapolations that deviated from a straight, horizontal line. Only one carried this tendency into the "non-practice" series examined, and for only one series. The rest of the subjects did this only in their practice attempts and this perhaps reflected attempts to exercise the capabilities of the software. This result was clearly linked to the brief instruction and demonstration session conducted with the subjects.

As anticipated, initial analysis clearly indicated that the M48 extrapolation, based on the mean of the whole series, was afflicted with problems arising from turning points early in the series. Error rates for the M48 method were approximately double those of the other methods, and the method is only mentioned here for the sake of completeness. In the following, the results for M48 have been omitted in the interests of conciseness

The most notable outcome is reported in Table 9.1 which shows the correlations between the error of the EXPERT extrapolation and those of the statistical methods. This shows that there was a correlation of approximately 0.9. All the correlations were highly significant (p=.000).

PERIOD	EXP	M12	NVE	BJ
1	.912	[illegible]	.895	.878
1-6	.934	.936	.910	.903
7-12	.924	.909	.904	.844
1-12	.928	.920	.907	.868

Table 9.1 Correlations with EXPERT

As can be seen, the correlation with the BJ error rate was the lowest at each time horizon considered. There is little to be drawn from this observation. The difference in levels of the correlation coefficients is very small. Never-the-less, the effect might be explained by the influence of the early data points on the BJ forecast. As previously mentioned, some series were unstable in the very early periods. The other extrapolative methods would not have been as affected, if at all, by those periods. Clearly, the M12 and the Nve methods avoided the problem by considering only the last 12 and the final point respectively. The Exp method would only have been influenced by the early observations in the determination of the optimum smoothing factor. For the factors actually used, which lay between 0.3 and 0.5, the influence of the early periods would have all but disappeared.

As with the correlations, the actual error rates were also very similar, as reported in Table 9.2.

PERIOD	EXPERT	NOVICE	EXP	N12	NVE	BJ
1	10.37	10.67	10.61	10.52	10.16	9.46
1-6	10.10	10.00	10.38	10.51	10.30	10.45
7-12	13.30	14.00	13.24	13.67	14.18	13.61
1-12	11.70	12.00	11.81	12.09	12.24	12.03

Table 9.2 MAPE Error Rates

The similar means, coupled with the high correlations lead to an expectation that there would be no significant differences between the methods. Certainly, no such significance was detectable using ANOVA, since the pooling that takes place in ANOVA diminishes the sensitivity of the test. In the circumstances, the most sensitive test of significant difference would be the paired t-test, if the effect of performing multiple tests were accounted for. Such a multiple paired t-test was carried out between pairs comprising EXPERT and each of the others. Only in two cases was a difference found with a level of confidence exceeding 85%. The method involved in both cases was the Nve method, for the 7-12 month time horizon there was a two tailed significance of p=.015, and for the 1-12 horizon p=.048. In the light of the number of tests carried out it is not possible to rely on those levels of significance.

The conclusion to be drawn from the above testing is that there is no evidence whatsoever that the substitution of one of the tested extrapolation methods for judgment would result in an improved forecast. The only hint of a significant difference detected was that Nve might prove to be a less successful method.

9.5 DISCUSSION

The results obtained in this study give further indication that the de-seasonalising and de-trending processes are adequately performed judgmentally using the GRAFFECT decision aid. The series resulting from such judgmental transformations was found to be free from detectable signals of either component. This was true both for relatively expert and novice subjects.

The results also show that there was no advantage to be gained in the substitution of selected statistical extrapolation methods for the judgmental extrapolation. This is not to say that there is not a statistical process that would improve the forecast. The methods tested:

* Single exponential smoothing,
* 12 period Moving average,
* Naive forecast, and
* Box-Jenkins,

cover a wide spectrum in terms of sophistication, and include methods shown to be successful in the M-Competition.

A common criticism of judgmental forecasting processes is that they are time consuming. This was discussed in chapter 6, where it was shown that the time taken for the whole extrapolation was a fraction of the time for a Box - Jenkins forecast. The time taken in extrapolating the residual series is a factor to be considered in the design of a forecasting decision aid. The lack of an advantage to the judgmental process indicated by the analysis performed in this chapter raises the issue of replacing judgment by a simple method such as exponential smoothing. To put the matter in perspective, the time taken for the extrapolation by the experienced subjects averaged 23.75 seconds. The maximum time was 58 seconds, and the time had a standard deviation of 14 seconds. This is not a substantial use of human resource, per series. The decision to use judgment or not therefore would depend on other factors:

* the number of series to be forecast,
* the perceived stability of the trend and seasonal patterns in the series to be forecast, and
* the extent to which the forecasts are to be modified judgmentally to take account of external market factors.

Obviously, some organisations require to forecast huge numbers of series for, say, inventory control purposes. In that case there is a strong argument for the final extrapolation to be performed automatically. The identification of the trend and seasonal models could be carried out judgmentally according to perceived need. Perhaps a quality control procedure based on error rate would provide a trigger to such action, apart from some regular review. For series that are subject to regular judgmental review either because the trend and seasonal models are dynamic, or to include extra information, there would be little point in automating the extrapolation process.

The external validity of the results is threatened to the extent that the 68 series included in the study do not reflect the general population of time series.

9.6 REFERENCES

Kendall M.G., *Time-series*, Griffin, London, (1973)

Spencer J. "Estimating Averages", *Ergonomics* 4, (1961), 317-328.

Makridakis S., Wheelwright, S.S., and McGee, V.E., *Forecastings methods and applications*, 2nd ed., Wiley, New York, (1983).

10. SUMMARY AND CONCLUSIONS

10.1 SUMMARY OF THE STUDIES

This thesis reports on a project directed at the development and evaluation of a tool, called **GRAFFECT**, to assist in the forecasting of economic time series. The tasks carried out in the project are summarised below.

Title : INVESTIGATION OF TIME SERIES CHARACTERISTICS

Chapter : 3

Objective : To identify characteristics of time series that acted as discriminators between series that were forecast better by unaided judgment, and those forecast better by statistical processes. This provided two opportunities, the first was the development of an a priori rule for forecast method selection. The second was to develop means to support judgment in a way to overcome the identified deficiencies.

Title : DESIGN OF THE FORECASTING DECISION AID

Chapter : 5

Objective : To create an experimental tool that was capable of supporting forecasters in the extrapolation phase of the forecast. The decision aid enforced a structure on the decision process, and presented decision cues in forms that might overcome some of the shortcomings of judgment identified in chapter 3 and the human information processing literature.

<u>Title</u> : ACCURACY OF COMPUTER ASSISTED FORECASTING

<u>Chapter</u> : 6

<u>Objective</u> : To determine whether the use of the decision aid affected the accuracy of the forecast, or the time taken to generate the forecast. The relative accuracy achieved by novice and experienced forecasters was examined. In addition, a means to discriminate between series forecast better by supported judgment and those better forecast by statistical methods was sought.

<u>Title</u> : SEASONAL PATTERN IDENTIFICATION

<u>Chapter</u> : 7

<u>Objective</u> : To examine how well the subjects performed in the sub-task of seasonal pattern identification, including the possibility of improving the forecast by providing automatic functions.

<u>Title</u> : IDENTIFICATION OF TREND

<u>Chapter</u> : 8

<u>Objective</u> : To examine how well the subjects performed in the sub-task of trend identification, including consideration of judgmental or automatic damping of the trend model.

<u>Title</u> : THE RESIDUAL NOISE IN THE TIME SERIES

<u>Chapter</u> : 9

<u>Objective</u> : To examine how well the subjects performed in the extrapolation from the noise residual in the time series, including the effect of providing automatic functions.

10.2 SUMMARY OF FINDINGS

10.2.1 INVESTIGATION OF TIME SERIES CHARACTERISTICS

10.2.1.1 MAJOR RESULTS

Discriminant analyses were carried out using a number of metrics that were developed from classical decomposition theory. A function was generated that distinguished between "GRAPH" series and "deseasonalised single exponential smoothing" series for a short forecast horizon (months 1-6), with an accuracy of 74%. A significant improvement in accuracy was achieved by using the discriminant function. It was not possible to generate a similar function for months 7-12.

The function developed for the short forecast horizon showed that for series in which the ratio of the seasonality to a simple noise metric was low judgment performed well relative to deseasonalised single exponential smoothing, and vice versa. Unexpectedly, judgment was not adversely affected by noise and did not have an advantage in the presence of trend as those characteristics were measured by the metrics developed.

10.2.1.2 SECONDARY RESULTS

In a preliminary study it was shown that the judgmental extrapolations were sufficiently similar in replication to warrant their study.

10.2.2 ACCURACY OF COMPUTER ASSISTED FORECASTING

10.2.2.1 MAJOR RESULTS

The use of GRAFFECT gave rise to a significant improvement in accuracy of extrapolation, both for experienced and novice users. The error rates were significantly lower than those of deseasonalised single exponential smoothing for the short forecast horizon. The time taken to make judgmental extrapolations was approximately 60% lower using GRAFFECT.

10.2.2.2 SECONDARY RESULTS

For the longer forecast horizon the improvement in accuracy over deseasonalised single exponential smoothing did not achieve statistical significance.

No discriminant functions could be generated between GRAFFECT and deseasonalised single exponential smoothing for either the short or the long forecast horizons. This cast doubts over the usefulness of the trend metrics in particular.

10.2.3 SEASONAL PATTERN IDENTIFICATION

10.2.3.1 MAJOR RESULTS

It was concluded that the form of the cue display had changed the decision outcome. The differences between the seasonal patterns identified, and those produced using a ratio to centered moving average were greatly reduced. This change did not come without cost, the advantage that judgment had shown in extrapolating series with low seasonality and high noise was lost. This probably arose from the masking of temporal data concerning seasonality. Alternative cue displays have been designed to attempt to recover the lost advantage.

10.2.3.2 SECONDARY RESULTS

There were strong indications that judges were not adversely affected by noise, especially in attempting to identify seasonal patterns. There was evidence of consensus in the identification of seasonal patterns.

The provision of automatic deseasonalising functions in the decision aid did not lead to an significant change in accuracy. This indicates that provision of such functions is desirable, because they would save effort. It is necessary, however, to evaluate new cue designs intended to obtain the benefits achieved by GRAFFECT, without the loss associated with low seasonality series with high noise levels.

10.2.4 IDENTIFICATION OF TREND

10.2.4.1 MAJOR RESULTS

Judges were able to model the trend in the historical data with a high level of accuracy. The trend models that experienced time series analysts used for the forecast period was damped for certain of the series. The evidence suggested that the damping was related to the number of periods over which the historical trend model had been stable.

Various damping automatic mechanisms were evaluated, but none were found to be of advantage. It is possible that a curvilinear damping mechanism would be of benefit, this remains to be tested.

The trend metric from chapter 3 appeared to be deficient because it was not associated with an advantage to the judgmental method over deseasonalised single exponential smoothing which has no mechanism for trend modelling. This was confirmed by the analysis performed, which showed a distinct advantage to the judgmental method for those series in which the judges identified trend. New trend metrics are under development, which may enable useful discriminant functions to be generated. Such functions would be of major benefit in identifying those series which are currently being forecast by statistical approaches, perhaps because of the number of series to be forecast, which would benefit from judgmental forecasting.

10.2.5 RESIDUAL NOISE IN THE TIME SERIES

10.2.5.1 MAJOR RESULTS

Analysis of the data from prior studies showed that there was no detectable signal left in the series after the removal of the trend and seasonal models identified by the subjects. Statistical procedures for performing the final extrapolation were tested but these did not achieve better results than the subjects.

10.2.6 OVERALL CONCLUSIONS

The most important result achieved in this project is the demonstration that judgment, when suitably supported, is more accurate in certain circumstances, and otherwise is at least as accurate as the better statistical approaches to extrapolation. The importance of this finding is not limited to the provision of yet another viable method of extrapolation, it widens the scope of future developments in forecasting to include judgmental processes as a core feature.

A secondary, but never-the-less important, outcome is the addition to the accumulation of knowledge on the effect of the use of graphical support for judgment. The results show that the effect of graphical data presentation changes with the design of the display. The objective of determining whether graphical or tabular displays are best, even for a specific class of decision, over simplifies the problem and probably could not be achieved given the almost limitless variety of possible display designs. What has emerged, however, is that it may be possible to take advantage of display design features to support particular decision strategies with consequent improvement in decision outcome.

There are several supplementary conclusions which have more specific application to forecasting. The practicality of an a priori rule for selection of forecasting method was illustrated. With new metrics, especially for trend, the decision aid could be expanded to flag or identify series for which judgmental or statistical processes are suited. The selective intuitive damping of trend carried out by experienced forecasters was shown to be successful, and this might also find application in future decision aids in a mechanism to propose damping in appropriate circumstances.

Finally, there was the interesting observation that subjects who chose to apply an automatic deseasonalising procedure were not disadvantaged in the same way as subjects who had the deseasonalised data presented to them. It is not possible to draw firm conclusions from this without further study. However, it appears to point to an aspect of judgment calling for some understanding of prior processes if subsequent process are not to be adversely affected.

10.3 RESERVATIONS AND LIMITATIONS

In each empirical chapter there is a discussion of the threats to external validity arising from the selection of subjects, and the selection of time series. Any single study carries the risk that subjects and stimuli are not typical of the real world, and that is true in this case. Care was taken to try to ensure that the knowledge levels and motivation of subjects was not atypical. The time series were also selected from a large data base of real series, but the mix of series types could not be controlled.

A serious limitation of the studies was the abstraction of the task to be merely an extrapolation from a set of time series cues. This was done to assist in internal control of the studies, so that conclusions could be drawn about the relative accuracies of extrapolation. The real task of forecasting depends on data outside the time series values alone. Nothing can be said about the accuracy of forecasts that have such data taken into consideration either holistically or by way of adjustment to a prior extrapolation.

Finally, the accuracy levels achieved in the project reflect the particular data interface design used. Other designs may give different results, and work is proceeding on this aspect.

10.4 FUTURE DIRECTIONS

10.4.1 DEVELOPMENT OF THE DECISION AID

There remains considerable scope for new features of the decision aid to be developed and tested:

* Pre-whitening of the time series: this would provide the decision maker to model the effects of disturbances to the driving process underlying the time series. Preliminary design work has been undertaken that would allow the forecaster to consider the effect of a perturbation, such as a work stoppage, on the series including any "shadow" that the event cast on subsequent observations of the series.

* Provision of statistical processes for guidance. As described in chapter 5, the cues presented to the forecaster were limited to time series values. Possible changes to that strategy include the provision of, say, the seasonal model identified by a moving average process as an anchor for the subsequent judgment.

* Conferencing, for consensus or averaging. Initial work is proceeding addressed at the support of more than one decision maker.

* provision of simulation facilities, perhaps via an interface with a spreadsheet. This would allow the extension of the scope of the decision to include the effects of the forecast on business plans. Iteration between forecast development and planning considerations would be greatly enhanced.

10.4.2 FURTHER TESTING

As described in section 10.3, the testing that has been carried out has concentrated on extrapolation, in laboratory settings. It now remains to carry out long term, real time testing in the commercial environment:

* testing the effect of separating the modelling of seasonality and trend from the "rolling" forecasting activity. This testing may be achieved in a laboratory setting, in which case the judges would only have the time series values from which to determine whether new seasonal and trend models were required. In the field, this would be greatly supplemented by additional cues from the market. The advantage of this additional information would be at the cost of substantially lengthening the time taken to perform the study.

* Field studies to consider the effect of market and socio-economic data on the forecast. The issues here are whether inclusion of the extra data by way of adjustment of a prior extrapolation, holistically with the extrapolation, or by some combination of separate forecasts is most effective.

Adam E.E., and Ebert E.R., "A comparison of human and statistical forecasting", *AIIE Transactions*, 8,1, (1976), 120-127.

Armstrong J.S., *Long-Range Forecasting from Crystal Ball to Computer*. 2nd Edn. Wiley, New York, (1985).

Armstrong J.S., "Forecasting by extrapolation: conclusions from 25 years of research", *Interfaces* 14:6, (1984), 52-66.

Armstrong J.S., & Lusk E.J., Commentary on the Makridakis Time Series Competition (M-Competition), *Journal of Forecasting*, 2, (1983), 259-311.

Armstrong J.S., Denniston W.B. Jr., and Gordon M.M., The use of the decomposition principle in making judgments" *Organizational Behavior and Human Performance*, 13, (1975). 257-263.

Ayres R.U., Commentary on Armstrong J.S., "Forecasting by extrapolation: conclusions from 25 years of research", *Interfaces* 14:6, (1984), 61-62

Benbasat I. and Schroder R., "an experimental investigation of some MIS design variables" *MIS Quarterly* 1,1 (1977), 37-49.

Brown L.D., and Rozeff M.S., "The superiority of analyst forecasts as measures of expectations: evidence from earnings", *Journal of Finance*, 33, (1978), 1-16.

Carbone R., and Armstrong J.S., "Evaluation of extrapolative forecasting methods: results of a survey of academicians and practitioners", *Journal of Forecasting*, 1, (1982), 215-217

Carbone R., and Gorr W.L. "The accuracy of judgmental forecasting of time series.", *Decision Sciences*, 16, (1985), 153-160.

Carbone R., Andersen A., Corriveau Y., and Corson P.P., "Comparing for different time series methods the value of technical expertise, individual analysis, and judgmental adjustment" *Management Science* 29, (1983), 559-566.

Carbone, R., & Gorr, W.L., "Accuracy of judgmental forecasting of time series" *Decision Sciences*, vol 16, (1984), 153-160.

Cerullo M.J., and Avila A., "Sales forecasting practices: a survey" *Managerial Planning*, 24, (1975), 33-39.

Chatfield C., and Prothero D.L., "Box-Jenkins seasonal forecasting: problems in a case study", *Journal of the Royal Statistical Society, Series A*, 136 part 3, (1973), 295-336.

Chervany, N.L., & Dickson, G.W., "Experimental evaluation of information overload in a production environment", *Management Science*, 20, (1974), 1335-1344.

Christensen-Szalanski J.J., "Improving the practical utility of judgment research", in *New Directions In Research On Decision Making*, North Holland, (1986), p 383-410.

Cohen L.J., "Can human rationality be experimentally demonstrated?", *The Behavioral and Brain Sciences*, 4, (1981), p317-370.

Dalrymple D.J., "Sales forecasting practices in business: results from a 1983 U.S. survey." Working paper, Graduate School of Business, Indiana University.

Dalrymple D.J., "Sales forecasting practices: results from a United States survey" *International Journal of Forecasting*, forthcoming (1987).

Dickhaut J.W., and Eggleton I.R.C., "An examination of the processes underlying comparative judgments of numerical stimuli." *Journal of Accounting Research*, Spring (1975), 38-72.

Dickson, G.W., DeSanctis, G., & McBride, D.J., "Understanding the effectiveness of computer graphics for decision support: accumulative experimental approach", *Communications of the ACM*, 29,1, (1986), 40-47.

Dickson, G.W., Senn, J.A., & Chervany, N.L., "Research in management information systems- Minnesota experiments", *Management Science*, 23, (1977), 913-923.

Ebbesen E.B., and Konecni V.J., "On the external validity of decision making research: what do we know about decisions in the real world?", in *Cognitive Processes in Choice and Decision Behaviour*, ed. Wallsten T.S., LEA, (1980)

Edmundson R.H. "Metrics for a priori selection of forecasting methods: a preliminary investigation." *Fourth International Symposium on Forecasting*, London (1984).

Edmundson R.H., Lawrence M.J.,, & O'Connor M.J., "The use of non time series information in sales forecasting : a case study." *Information Systems Research Reports*, University of New South Wales. (1987).

Edwards W., "Human cognitive capabilities, representativeness, and ground rules for research", in *Analysing and Aiding Decision Processes*, ed Humphreys P., Svenson O., and Vari A., Akademiai Kiado, Budapest, (1983), p 507-513.

Eggleton I.R.C. "Intuitive Time Series Extrapolation", *Journal of Accounting Research*, 20,1. (1982), 68-102.

Eggleton I.R.C. "Patterns, Prototypes, and Predictions: An Exploratory Study " *Selected Studies on Human Information Processing in Accounting,* (1976). Supplement to *Journal of Accounting Research* 14: 68-131

Einhorn H.J., "Expert measurement and mechanical combination", *O.B.H.P.* 7, 1972, 84-106.

Eisenbeis, R.A. "Pitfalls in the application of discriminant analysis in business, finance and economics", *Journal of Finance*, (1977) 32 : 875-897.

Felsen J., "A Man-Machine Investment Decision System.", *Int. Man-Machine Studies*, 8,(1976), 169-179.

Fenker I., and Evans S.H., "A model for optimizing the effectiveness of man-machine decision making in a pattern recognition system. *Report AD-730-944.* Springfield, Va: National Technical Information Service. (1971).

Fildes R. "An evaluation of Bayesian forecasting", *Journal of Forecasting,* 2, (1983), 137-150.

Fildes R., "Gains through univariate forecasting model selection" *Fifth International Symposium on Forecasting*, Montreal, (1985).

Fildes R., and Howell S., "On selecting a forecasting model", *TIMS Studies in the Management Sciences* 12, (1979), 297-312.

Fildes, R., "Quantitative forecasting - the state of the art: extrapolative models", *J. Opl. Res. Soc.*, 30,8 (1979), 691-710.

Fildes, R., and Lusk, E.J., "The choice of a forecasting model", *Omega*, 12, (1984), 427-435.

Gardner E.S. & McKenzie E. "Forecasting trends in time series", *Management Science,* 31,10, (1985), 1237-1246.

Gettys C.F., Michel C., Steiger J.H., Kelly C.W. III., Peterson C.R., "multiple-stage probabilistic information processing", *O.B.H.P*. 10, 1973, 374-387.

Hays, W.L., *Statistics*, Holt Saunders, New York, (1981).

Hogarth R.M., and Makridakis S., "Forecasting and planning: an evaluation", Unpublished manuscript, INSEAD, Fontainebleau, France, (1979).

Hogarth R.M., and Makridakis S.. "Forecasting and planning: an evaluation", *Management Science*, 27,2 (1981), 115-138.

Hogarth R.M., *Judgment and Choice*, Wiley, Chichester, (1980).

Jenkins G.M., "Some practical aspects of forecasting in organizations", *Journal of Forecasting* 1, (1982), 3-21.

Johnson-Laird P.N. & Watson P.C., "A theoretical analysis of insight into a reasoning task", *Cognitive Psychology,* 1, (1970), 134-148.

Kahneman D. & Tversky A., "On the psychology of prediction", *Psychology Review*, 80, (1973), 327-351.

Kahneman D. & Tversky A., "Subjective probability: a judgment of representativeness", *Cognitive Psychology*, 3, 1972, p430-454.

Kendall M.G., *Time-series*, Griffin, London, (1973)

Klein L.R., & Burmeister, E,. *Econometric model performance*, University of Pennsylvania Press, Philadelphia, (1976).

Kosaka T. and Hirouchi T. "An effective architecture for decision support systems" *Information and Management*. 5, (1982), 7-17.

Lachenbruch, P.A. *Discriminant Analysis*, Hafner, New York, (1975).

Lawrence M.J., "An exploration of some practical issues in the use of quantitative forecasting models", *Journal of Forecasting,* 2 (1983), 169-179.

Lawrence M.J., Edmundson R.H, and O'Connor M.J., "The accuracy of combining judgmental and statistical forecasts", *Management Science*, 32,12, (1986), 1521-1532.

Lawrence M.J., Edmundson R.H., and O'Connor M.J. "An Examination of the Accuracy of Judgment Extrapolation of Time Series" *International Journal of Forecasting* 1 (1985), 25-35

Lawrence, M.J., & Makridakis, S., "Human judgment in extrapolation", *International Forecasting Symposium,* Paris, 1986.

Lucas H.C., "an experimental investigation of the use of computer-based graphics in decision making." *Management Science*, 27,7. (1981), 757-768.

Lusk, E.L., & Kersnick, M., "The effect of cognitive style and report format on task performance: the MIS design consequences", *Management Science*, 25, (1979), 787-798.

Lusted L.B., Roberts H.V., Wallace D.L., Lahiff M., Edwards W., Loop J.W., Bell R.S., Thornbury J.R., Seale D.L., Steele J.P., and Fryback D.B., "Efficacy of diagnostic radiological procedures" in *Practical Evaluation: Case Studies in Simplifying Complex Decision Problems*, ed Snapper K., Information Resources Press, Washington, (1982) Lyness K.S. and Cornelius E.T. III., "A comparison of holistic and decomposed strategies in a performance rating simulation" *Organizational Behavior and Human Performance*, 29, (1982), 21-38.

Makridakis S., "Forecasting accuracy and the assumption of constancy", *Omega* 9:3, (1981), 307-311

Makridakis S., and Hibon, M., "Accuracy of forecasting: an empirical investigation", *Journal of the Royal Statistical Society, Series A*, 142, (1979), 97-145

Makridakis S., and Wheelwright S.C., *Forecasting Methods and Applications*. Wiley. Santa Barbera. (1978).

Makridakis S., Andersen,A., Carbone,R., Fildes,R., Hibon,M., Lewandowski,R., Newton,J., Parzen, E., and Winkler,R. "The accuracy of extrapolative (time series) methods : results of a forecasting competition", *Journal of Forecasting*, Vol 1, no 2, (1982), 111-153.

Makridakis S., Andersen,A., Carbone,R., Fildes,R., Hibon,M., Lewandowski,R., Newton,J., Parzen, E., and Winkler,R. *The forecasting accuracy of major time series methods*, Wiley, Chichester, (1984).

Makridakis S., Wheelwright, S.S., and McGee, V.E., *Forecasting: methods and applications*, 2nd ed., Wiley, New York, (1983).

Makridakis, S., and Hibon, M., "Accuracy of forecasting: an empirical investigation", *Journal of the Royal Statistical Society, Series A*, 142, (1979), 97-145.

Mclaughlin R.L., "Forecasting models: sophisticated or naive?" Commentary on the Makridakis Time Series Competition (M-Competition), *Journal of Forecasting*, 2, (1983), 259-311.

Mentzer J.T., and Cox J.E., "Familiarity, application, and performance of sales forecasting techniques" *Journal of Forecasting* 3, (1984), 27-36.

Miller G.A.,"The magic number seven, plus or minus two: some limitations on our capacity for information processing." *Psychological Review* 63, (1956), 81-97.

Moriarty M.M., and Adams A.J., "Management judgment forecasts, composite models, and conditional efficiency", *Journal of Marketing Research*, vol xxi, (1984), 239-250.

Mosteller F., Siegel A.F., Trapido E., & Youtz C. "Eye fitting straight lines", *The American Statistician*, 35,3, (1981), 150-152

Mosteller, F. and Tukey, J.W. *Data Analysis and Regression*, Addison-Wesley, Reading, Mass., (1977).

Mosteller, F., "Innovation and Evaluation" *Science*, 211, (1981), 881-886.

Murphy A.H., and Winkler R., "Reliability of subjective probability forecasts of precipitation and temperature", *Journal of the Royal Statistical Society*, Series C, 26, (1977), 41-47

Newbold P., and Granger C.W.J., "Experience with forecasting univariate time series and the combination of forecasts", *Journal of the Royal Statistical Society, Series A*, 137 part 2, (1974), 131-164.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., Bent, D.H *Statistical Package for the Social Sciences*, McGraw Hill, New York, (1975).

O'Connor, M.J. "An examination of the accuracy and determinants of confidence levels in judgmental time series forecasting", Ph.D. Dissertation, University of New South Wales, (1986).

Paller A.T., "Improving management productivity with computer graphics." *Computer Graphics and Applications*. (1981), 9-16.

Phelps R.H., and Shanteau J., "Livestock judges: how much information can an expert use?" *Organizational Behavior and Human Performance*, 21, (1978), 209-219.

Phillips L.D., "A theoretical perspective on heuristics and biases in probabilistic thinking" in *Analysing and Aiding Decision Processes*, ed Humphreys P., Svenson O., and Vari A., Akademiai Kiado, Budapest, (1983), p 525-543.

Reid D.J., "A comparative study of time series prediction techniques on economic data" Ph.D. Thesis, University of Nottingham (1969).

Reid D.J., "Forecasting in action: a comparison of forecasting techniques in economic time series", *Joint Conference of O.R. Society's Group on Long Range Planning and Forecasting*. (1971).

Remus W., "An Empirical Investigation of the Impact of Graphical and Tabular Data Presentations on Decision Making." *Management Science*, 30,5 (1984), 533-542.

Rothe J.T., "Effectiveness of sales forecasting methods" *Industrial Marketing Management*, 7, (1978), 114-118.

Rouse W.B., "A model of the human as a suboptimal smoother", *IEEE Transactions on Systems, Man, and Cybernetics*, smc-6, 5, (1976), 337-343.

Schnaars, S.P., "Situational factors affecting forecast accuracy", *Journal of Marketing Research*, vol xxi, (1984), 290-297.

Schroeder, R.G., & Benbasat, I., "An experimental evaluation of the relationship of uncertainty in the environment to information used by decision makers", *Decision Sciences*, 6, (1975), 5560-5567.

Sekular R.W. and Abrams M. "Visual sameness: a choice time analysis of pattern recognition processes." *Journal of Experimental Psychology*. 77,2 (1968), 232-238.

Shutz, H.G., "An evaluation of formats for graphic trend displays", *Human Factors*, 3, (1961), 99-107.

Simon H.A. and Sumner R.K. "Pattern in Music" in *Formal Representation of Human Judgment.* ed Kleinmuntz, New York. Wiley (1968), 219-250

Slovic P. and Lichtenstein S.,"Comparison of Bayesian and regression approaches to the study of information processing in judgment." *O.B.H.P.*, 6 (1971), 649-744.

Slovic P., Fischhoff B., & Lichtenstein S., "Behavioral Decision Theory." *Ann. Rev. Psychol*. 28 (1977), 1-39.

Sparkes J.R, and McHugh A.K., "Awareness and use of forecasting techniques in British industry" *Journal of Forecasting*, 3, (1984), 37-42.

Spencer J., "Estimating Averages" *Ergonomics*, 4, (1961), 317-328.

Tversky A. & Kahneman D., "Availability: a heuristic for judging frequency and probability" *Cognitive Psychology*, 5, (1973), 207-232.

Tversky A. & Kahneman D., "Judgment under uncertainty: heuristics and biases", *Science*, 185, (1974), 1124-1131.

Tversky A. & Kahneman D., "The framing of decisions and the psychology of choice", *Science*, 211, (1981), 453-458.

Wagenaar W.A. and Timmers H., "Extrapolation of exponential Time Series is not Enhanced by Having

More Data Points." *Perception and Psychophysics*. 24,2 (1978), 182-184.

Wagenaar W.A., "Generation of random sequences by human subjects: a critical survey of literature." *Psychological Bulletin*. 77,1. (1972), 65-72.

Watson P.C. "Reasoning" in *New Horizons in Psychology*, vol 1, , ed. B. Foss. Harmondsworth: Penguin (1966).

Watson P.C. & Shapiro D. "Natural and contrived experience in a reasoning problem", *Quarterly Journal of Experimental Psychology*, 23, (1971), 63-71.

Welsch G.A., *Budgeting*: *Profit Planning and Control*, 3rd edn, Prentice Hall, Englewood Cliffs N.J., (1971).

Winkler R., and Murphy A.H., "Experiments in the laboratory and real world", *Organisational Behaviour and Human Performance*, 10, (1973), 252-270.

Winkler, R.L., and Makridakis, S., "The combination of forecasts", *Journal of the Royal Statistical Society*, 146 Part 2, (1983), 150-157.