FORECASTING OF INTERMITTENT DEMAND

by

Argyrios Syntetos

A Thesis submitted for the degree of Doctor of Philosophy

Business School

Buckinghamshire Chilterns University College

Brunel University

July, 2001

Summary

This thesis explores forecasting for intermittent demand requirements.

Intermittent demand occurs at random, with some time periods showing no demand. In addition, demand, when it occurs, may not be for a single unit or a constant size. Consequently, intermittent demand creates significant problems in the supply and manufacturing environment as far as forecasting and inventory control are concerned.

A certain confusion is shared amongst academics and practitioners about how intermittent demand (or indeed any other demand pattern that cannot be reasonably represented by the normal distribution) is defined. As such, we first construct a framework that aims at facilitating the conceptual categorisation of what is termed, for the purposes of this research, “non-normal” demand patterns.

Croston (1972) proposed a method according to which intermittent demand estimates can be built from constituent elements, namely the demand size and inter-demand interval. The method has been claimed to provide unbiased estimates and it is regarded as the “standard” approach to dealing with intermittence. In this thesis we show that Croston’s method is biased. The bias is quantified and two new estimation procedures are developed based on Croston’s concept of considering both demand sizes and inter-demand intervals. Consequently the issue of variability of the intermittent demand estimates is explored and finally Mean Square Error (MSE) expressions are derived for all the methods discussed in the thesis.

The issue of categorisation of the demand patterns has not received sufficient academic attention thus far, even though, from the practitioner’s standpoint it is appealing to switch from one estimator to the other according to the characteristics of the demand series under concern. Algebraic comparisons of MSE expressions result in universally applicable (and theoretically coherent) categorisation rules, based on which, “non-normal” demand patterns can be defined and estimators be selected.

All theoretical findings are checked via simulation on theoretically generated demand data. The data is generated upon the same assumptions considered in the theoretical part of the thesis.

Finally, results are generated using a large sample of empirical data. Appropriate accuracy measures are selected to assess the forecasting accuracy performance of the estimation procedures discussed in the thesis. Moreover, it is recognised that improvements in forecasting accuracy are of little practical value unless they are translated to an increased customer service level and/or reduced inventory cost. In consequence, an inventory control system is specified and the inventory control performance of the estimators is also assessed on the real data. The system is of the periodic order-up-to-level nature. The empirical results confirm the practical validity and utility of all our theoretical claims and demonstrate the benefits gained when Croston’s method is replaced by an estimator developed during this research, the Approximation method.

CONTENTS

List of Figures

List of Tables

Acknowledgements

I would like to express my wholehearted gratitude to John Boylan, my supervisor, for his patience and his constant support and guidance throughout the period of my Ph.D. study.

I am also grateful to Jonathan Hill, my mentor, for everything he has done for me during the last four years.

I would like to thank my employers, Buckinghamshire Business School, for the financial assistance provided and for allowing me the time to do my research. I would also like to thank Mr Tony Abuzolof, from Unicorn Systems (UK) Ltd., for authorising a financial support for my Ph.D. project and Dr John English (from the same organisation) for our very interesting meetings and for helping me to obtain the empirical data that was used for my research.

I am indebted to Mr John Croston (independent systems and statistical consultant) for his constructive comments and criticism regarding the material presented in chapter 6 of the thesis. I would also like to thank Mr Andrew Eaves from AEA Technology Rail management consultancy and Mr Bill Sichel from the Monet Group, USA, for their interest in my work and for their helpful comments regarding material presented in the empirical part of the thesis.

Finally I would like to thank the following:

My friends in Athens and my friend Paul for all their help during the last four years.
My colleagues (and friends) Branko, Sven, Broder, Sara and Natasha for making everyday life look normal.
Lucia, for all the colours that she brought in my life.
Vana, my mother, and Maria-Helen, my sister, for all their love and encouragement, and for always supporting my choices (conventional or not).

For all the people who live on the first floor
24 Parmenidou street, Pagrati 116 – 36, Athens, Greece
with all my love.

CHAPTER 1

INTRODUCTION

1.1 Business context

Intermittent demand appears at random, with some time periods having no demand at all. Moreover, demand, when it occurs, is not necessarily for a single unit or a constant demand size. In the academic literature, intermittent demand is often referred to as lumpy, sporadic or erratic demand. Intermittent demand items may be engineering spares (e.g. Mitchell, 1962; Hollier, 1980; Strijbosch et al, 2000), spare parts kept at the wholesaling/retailing level (e.g. Sani, 1995), or other items within the range of products offered by all organisations at any level of the supply chain (e.g. Croston, 1972; Willemain et al, 1994).

Williams (1984) analysed the demand categorisation system employed by a public utility. The study covered approximately 11,000 products and the author found that only 5% of them were classified as "non-sporadic". Vereecke and Verstraeten (1994) described the algorithms developed for the implementation of a new computerised inventory management system for spare parts in a chemical plant, in Belgium. The spare parts inventory contained about 34,000 different types of items and had a total value of approximately £17,740,000. Ninety per cent of the items were classified as lumpy. Similar figures have been reported elsewhere in the academic literature (e.g. Johnston, 1980; Dunsmuir and Snyder, 1989; Kwan, 1991). According to the "Pareto law" approximately 80% of the stock-keeping units (SKUs) contribute to approximately 20% of the sales. Moreover, these slower moving items are usually supported by more than 60% of the value of stock (see, for example, Johnston and Boylan, 1996).

Intermittent demand creates significant problems in the manufacturing and supply environment as far as forecasting and inventory control are concerned. Silver (1970) noted that "*Most useable inventory control procedures are based upon assumptions about the demand distribution (e.g., unit sized transactions or normally distributed* *demand in a replenishment lead time) that are invalid in the case of an erratic item. If this is not the case, the procedures tend to be computationally intractable* (p. 87)". In addition, forecasting for intermittent demand has long been recognised, in the academic literature (e.g. Croston, 1972), as a very difficult business task. It is not only the variability of the demand size but also the variability of the demand pattern that make intermittent demand so difficult to forecast.

The empirical data to be used for this research has been provided by Unicorn Systems (UK) Ltd. The organisation is a forecasting and inventory control software package manufacturer. Discussions with representatives from this company revealed the practitioners' concern for better algorithms to deal with intermittence.

In conclusion, managing intermittent demand items is a significant organisational problem. In an industrial context, the proportion of the stock range that is devoted to intermittent demand items is often considerable. Therefore, small improvements in a company's system, regarding the intermittent demand items, may be translated to substantial cost savings.

1.2 Theoretical background and research objectives

Despite the significant benefits that small improvements in managing intermittence could offer to organisations, this area has attracted limited academic attention over the years. There is a considerable volume of literature on slow moving items. Little though has been written on improving the management of intermittent demand items. In addition, most papers that appear in this area focus on the control of inventories of intermittent demand SKUs, assuming that an appropriate estimator is used to forecast future demand requirements (e.g. Silver, 1970; Ward, 1978; Schultz, 1987; Watson, 1987; Dunsmuir and Snyder, 1989; Segerstedt, 1994). A limited amount of research has been conducted in the area of forecasting for intermittent demand. Few substantial contributions have been made in this area since Croston's work in 1972. This research aspires to take forward the current state of knowledge on forecasting intermittent demand. It is recognised, though, that improved forecasting practices are not necessarily translated in an improved inventory control performance, which is what matters from a practitioner's perspective. Therefore, we[1] wish, at the empirical part of the thesis, to assess not only the validity but also the utility of our findings, the latter referring to the inventory control implications of the theory developed in this research.

Croston (1972) suggested building intermittent demand estimates from constituent elements, namely the inter-demand interval and the size of demand, when demand occurs. Croston's concept has been claimed to be of great value to manufacturers forecasting intermittent demand. Despite the theoretical superiority of such an estimation procedure, empirical evidence suggests modest gains in performance when compared with simpler forecasting techniques; some evidence even suggests losses in performance.

Croston (op. cit.) was concerned with estimating, in an unbiased manner, the mean intermittent demand level. However, as Johnston and Boylan (op. cit.) pointed out, he did not go further, to produce an estimate of the variability of demand which is equally important for inventory control purposes.

It is common practice in industry to categorise demand patterns considering certain properties of the demand series, and consequently to select the forecasting method and inventory control model that perform "best" in this particular category. These categorisation schemes, though, are arbitrary in nature, and the values assigned to the criteria used are selected on a purely subjective basis. Some work in this area has been simulation-based (Johnston and Boylan, 1996) and lacking empirical testing of categorisation rules. Other work has been related to a specific industrial situation (Williams, 1984) and lacking universal applicability.

[1] The use of the word "we" throughout the thesis is purely conventional. The work presented in this Ph.D. thesis is the result of research conducted by the author alone, albeit with support from an academic institution and a commercial organisation (see sub-section 1.4.1).

Finally, a problem that has received very limited attention in the academic literature is that of obtaining reliable forecasting accuracy results in an intermittent demand context. It is the very nature of intermittent demand data, and in particular the existence of some zero demand time periods, that creates considerable difficulties in selecting an appropriate accuracy measure. Nevertheless, those special properties of intermittent demand series seem to have been underestimated or ignored in the past by both practitioners and academicians.

Given this background, the six main objectives of this research have been formulated as follows:

1. To identify some of the causes of the unexpected poor performance of Croston's method
2. To develop new intermittent demand estimation procedures
3. To derive results for the mean squared forecast error of a range of intermittent demand estimates
4. To propose theoretically coherent categorisation rules that distinguish between intermittent and non intermittent demand
5. To identify appropriate accuracy measures for application in an intermittent demand context
6. To test the empirical validity and utility of the theoretical results on a large set of real world data.

The conclusions of our research are discussed in chapter 12 of the thesis.

1.3 Structure of the thesis

This thesis is structured as follows:

In *chapter 2* the literature on managing intermittent demand items is overviewed, the detailed assumptions and methods employed in this research are discussed and some information regarding our empirical data is presented. Detailed review of the literature follows in the relevant chapters.

Some confusion is shared amongst academics and practitioners about how intermittent demand (or indeed any other demand pattern that cannot be reasonably represented by the normal distribution) is defined. Therefore, in *chapter 3*, we construct a framework to categorise "non-normal" demand patterns.

Croston's method has been widely implemented in practice. It has been claimed to provide unbiased estimates of the mean demand, when demand is intermittent. In *chapter 4* we show that Croston's method is biased. The bias is quantified and two new, approximately unbiased, estimation procedures are developed, based on Croston's idea of considering demand sizes and inter-demand intervals separately.

In *chapter 5* the issue of variability of the intermittent demand estimates is explored in detail. The variance of the estimates is derived for all the methods considered at that stage of the thesis.

In *chapter 6*, Mean Square Error (MSE) expressions are derived for all estimators. The MSE is similar to the statistical measure of the variance of forecast errors but not identical since bias is also taken into account. Algebraic comparisons of MSE expressions result in categorisation rules. According to these rules, different "non-normal" demand patterns can be defined and estimators be selected.

In *chapter 7* a simulation experiment using theoretically generated data is developed to check the validity of all our approximate theoretical results.

In *chapter 8* the most commonly used accuracy measures are discussed. The most appropriate ones are selected for application in an empirical intermittent demand context.

In *chapter 9* an inventory control model is developed to be used later in the thesis (chapter 11) to assess the empirical utility (inventory control implications) of our theoretical findings.

The empirical validity and utility of our theoretical findings are analysed in *chapters 10* and *11* respectively. Finally, in *chapter 12* the conclusions, recommendations and possible extensions of the thesis are discussed.

All appendices appear at the end of the thesis.

1.4 Methodology

1.4.1 The context for academic research

This Ph.D. has been supported financially by Buckinghamshire Chilterns University College and by Unicorn Systems (UK) Ltd. Despite the fact that an industrial organisation was involved in this project, the research cannot be characterised as "applied" (Saunders et al, 1997) in the sense that our main objectives have neither been negotiated with the software package manufacturer, nor they have been modified by them.

The research carried out could be perceived as solely academic. However, it is argued that the conceptualisation, discussions and outcomes provide practical benefits to industry. In the development of forecasting methods and categorisation schemes, the focus has been on the production of universally applicable results within industry. Therefore, such procedures do have some relevance to the sponsoring organisation and it is anticipated that they will be utilised by them in the near future.

1.4.2 A deductive approach

Our methodological approach is purely deductive. Theoretical results are first generated upon well-specified assumptions. The approximate nature of the theoretical results necessitates the assessment of their validity on theoretically generated data. Upon verification of the good performance of our approximations, the results are checked on empirical data.

The structure of the thesis is presented schematically in figure 1.1.

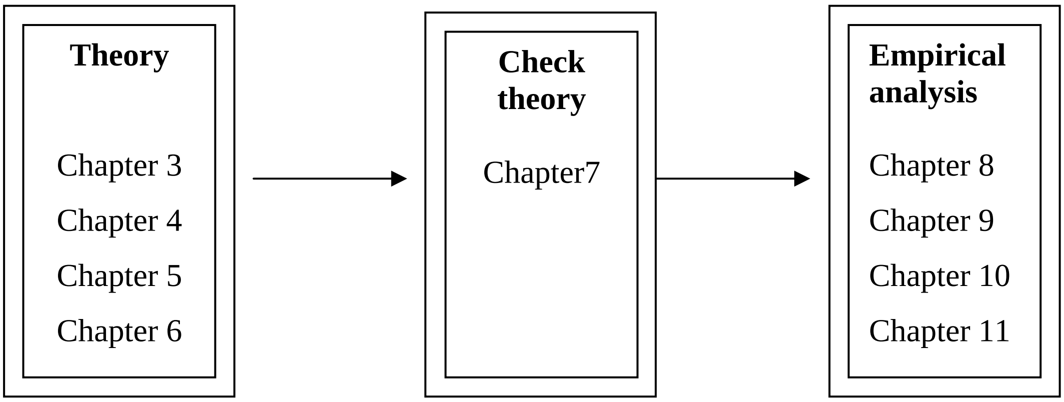

Figure 1.1. Structure of the thesis

1.5 Conclusions

Managing intermittent demand items is a significant academic and business problem. From an industrial perspective there is great potential for improvements and subsequent cost savings in this area. A considerable amount of research has been conducted in the area of inventory control for intermittent demand items. Not much, though, has been done in improving the related forecasting practices. The purpose of this Ph.D. is to take forward the current state of knowledge on forecasting intermittent demand.

In this chapter the main objectives of our research have been discussed. The structure of the thesis has also been presented along with a brief reference to the content of the eleven chapters that follow this introductory section. Finally, the methodological approach to meeting our objectives has been introduced.

CHAPTER 2

METHODS AND LITERATURE OVERVIEW

2.1 Introduction

The purpose of this chapter is to provide an overview of the literature on forecasting intermittent demand. Detailed review of the literature, on specific research areas (including stock control), follows in the relevant chapters. This chapter also serves the purpose of clarifying the details of our methodological approach. In the previous chapter, broad methodological issues were discussed. In this chapter, our specific methods and assumptions are presented in detail. Finally, a brief introduction is made to the organisation that provided the empirical data to be used in this research.

Unless demand occurs at every inventory review period and is of a fairly constant size, it always creates problems in the manufacturing and supply environment as far as forecasting and inventory control are concerned. Demand series of this type are often referred to as intermittent.

Intermittent demand items are not necessarily low demand items, as the latter demand category implies low demand sizes (or very often unit-sized transactions), when demand occurs. A considerable amount of research has been conducted on managing low demand items. This category of demand is not addressed explicitly in the thesis, since different approaches may be needed for non-intermittent low demand items.

This research focuses on improving current forecasting practices to deal with intermittent demand. It is assumed that a suitable inventory control model is in place, so that inventory control issues are not explicitly addressed in the thesis from a theoretical perspective. We do, though, consider the inventory control implications of our theoretical findings in chapter 11 of the thesis.

Inventory control systems can broadly be categorised as periodic (re-order interval) or continuous (re-order level) systems. Implicit reference to inventory control applications is made throughout the thesis by researching forecasting issues related to both types of stock control systems. Nevertheless, it has been argued in the academic literature (see for example Sani, 1995; Silver et al, 1998), that periodic systems are the most suitable for intermittent demand items (see also chapter 9). Therefore, in this thesis, we focus on the implications of forecasting on periodic inventory control.

This research refers to any stage of a given supply chain, focusing on single Stock Keeping Unit (SKU) forecasting processes. Aggregate or group forecasting issues are not discussed in the thesis.

Moreover, we focus on medium and short-term forecasting applications[1]. Issues related to long-range forecasting are not addressed in the thesis.

In a manufacturing context, no interactions are explored between the production planning process and the capacity requirements plans. Moreover only Make To Stock (MTS) manufacturing environments are considered, in which case the lead time is limited to the time that is required to deliver the products under concern. Further assumptions made throughout the thesis are the following:

- Demand is assumed to be independent. We refer only to level 0 in the Bill of Materials (BoMs). Higher levels in the BoMs are not considered, with the obvious exception of the spare parts
- Intermittent demand items do not attract the attention of the Marketing and Sales activities
- It is cost-beneficial to keep the SKUs under concern in stock. The issue of whether or not an item should be kept in stock has received considerable attention in the academic literature (see for example Smith and Vemuganti, 1969; Croston, 1974; Tavares and Almeida, 1983; Kwan, 1991) but is not pursued in this thesis
- The data is not trended and has no seasonal components
- There are no explanatory variables (e.g. atmospheric conditions) upon which we may rely to estimate the future behaviour of the demand data
- Time is treated as a discrete variable.

[1] A medium or short-range forecast generally has a time span of up to 18 months (see for example Slack et al, 1998).

2.2 Definition of intermittent demand

Silver et al (1998) defined intermittent demand as "*infrequent in the sense that the average time between consecutive transactions is considerably larger than the unit time period, the latter being the interval of forecast updating* (p. 127)". Intermittent demand is often referred to as sporadic. Johnston and Boylan (1996) have provided an "operationalised" intermittent demand definition: inter-demand interval greater than 1.25 forecast review periods[2]. Therefore intermittence (sporadicity) refers to the demand incidences and not to the demand size when demand occurs. Nevertheless sporadicity has often been associated with lumpiness in the academic literature (Ward, 1978; Schultz, 1987; Dunsmuir and Snyder, 1989). Lumpiness refers to intermittence coupled with erratic (irregular) demand sizes when demand occurs.

An intermittent demand definition should not introduce any restrictions as to how demand sizes are distributed. Demand, when it occurs, may be unit-sized (low demand), constant (clumped demand) or highly variable (lumpy demand). These issues are discussed in chapter 3, where the literature is also reviewed in detail. For the moment it is sufficient to say that we treat intermittent demand as sporadic with no reference to the distribution of the demand sizes.

2.3 Forecasting for intermittent demand items

2.3.1 The stationary mean model (SMM)

Under the Stationary Mean Model (SMM) assumption the mean demand level does not change over time and the optimum estimate of demand in any future time period is a simple arithmetic average of all previous demand data periods. The SMM is described by the following equation:

[2] Operationalisation is the translation of abstract concepts into indicators or measures, enabling observations to be made (Popper, 1972).

$$Y_t = \mu + \varepsilon_t \tag{2.1}$$

where

t	is the current time period, $t = 1, 2, \ldots\ldots$
Y_t	is the observed demand in the current time period
μ	is the constant mean demand level
ε_t	is a random disturbance assumed to be drawn from a distribution with zero mean, $\mathrm{E}(\varepsilon_t) = 0$, and constant variance, $Var(\varepsilon_t) = V$.

This is also called the "global" constant mean model. To forecast Y_t we shall have to assign a value to μ. The estimate of the level of demand that minimises the Mean Square forecast Error (MSE) is the following:

$$Y'_t = \frac{\sum_{j=1}^{t-1} Y_j}{t-1} \tag{2.2}$$

(i.e. the sample mean based on all available data)

where Y'_t is the (unbiased) estimate of the level of demand, for any future time period, made at the end of period $t-1$.

The Exponentially Weighted Moving Average (EWMA) and Moving Average (MA) estimators, to be discussed in the following sub-section, are also unbiased for the SMM but are not optimal.

2.3.2 The steady state model (SSM)

Under the Steady State Model (SSM) assumption the mean demand level varies stochastically through time and the optimum estimate of demand in any future time period is a simple Exponentially Weighted Moving Average (EWMA) of all previous demand data periods. The model starts with an observation equation:

$$Y_t = \mu_t + \varepsilon_t \tag{2.3}$$

where

t , Y_t , ε_t	are as defined in the previous sub-section
μ_t	is the unobserved mean demand level of the process at time t.

The process level develops stochastically according to the system equation

$$\mu_t = \mu_{t-1} + \gamma_t \tag{2.4}$$

with $\mathrm{E}(\gamma_t) = 0$, $Var(\gamma_t) = W$ (constant value).

We also assume that

$\mathrm{E}(\varepsilon_t \varepsilon_s) = 0 \qquad \forall\, t \neq s$

$\mathrm{E}(\gamma_t \gamma_s) = 0 \qquad \forall\, t \neq s$

i.e. no auto-correlation of the error terms

$\mathrm{E}(\varepsilon_t \gamma_s) = 0 \qquad \forall\, t\,,\ \forall\, \mathrm{s}$

i.e. no cross-correlation of the error terms.

The above-described model is often referred to as the local constant mean model. It corresponds to a (0,1,1) model in the ARIMA (Autoregressive Integrated Moving Average) terminology:

$$Y_t = Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{2.5}$$

where θ_1 is a constant, -1< θ_1 <1.

The following updating procedure (EWMA or exponential smoothing) can be shown to minimise the one step ahead MSE for the SSM (Harrison, 1967):

$$Y'_t = \alpha Y_{t-1} + (1-\alpha) Y'_{t-1} = Y'_{t-1} + \alpha \varepsilon_{t-1} \tag{2.6}$$

where

α is a smoothing constant, $0 \le \alpha \le 1$, and
$\varepsilon_{t-1} = Y_{t-1} - Y'_{t-1}$ (or the error at the observation level in period $t-1$).

For very low α values the estimator implies stationarity whereas for very high values (say greater than 0.3) a higher level model (e.g. with a trend term) may be more appropriate.

Subsequently the current level of demand is estimated as:

$$\hat{\mu}_t = Y'_t \tag{2.7}$$

The application of EWMA assuming the SSM has been studied extensively over the years (Harrison, 1967; Johnston and Harrison, 1986; Johnston, 1993; Johnston and Boylan, 1994).

Another approach, according to which there is a perfect correlation between the error terms at the observation and system levels (Snyder et al, 1999 a, b; Snyder et al, 2000), is the following:

Observation equation, $Y_t = \mu_t + \varepsilon_t$

System equation, $\mu_t = \mu_{t-1} + \alpha \varepsilon_t$.

In the above set of equations there is a single source of error (SSOE) as opposed to the SSM where there are multiple sources of error (MSOE). The SSOE approach enables one to see directly the relevance of EWMA. Nevertheless, it has been criticised as not being a model, in the OR sense[3] (Johnston, 2000), since it confuses the representation of the generation process with the estimation of the parameters in the model.

Another estimator that is often used in practice, assuming a local constant mean model, is a N period simple moving average (MA). A suitable length of the average, for typical data, employing monthly revision, might be somewhere between 3 and 12 points, and if using weekly revision, between 8 and 24 points (Johnston et al, 1999a).

$$Y'_t = \frac{\sum_{j=t-N}^{t-1} Y_j}{N} \tag{2.8}$$

Even though the estimator implies that the level of demand varies through time, it was not until recently that its properties were explored under the SSM assumption (Johnston et al, 1999b).

2.3.3 Modelling demand from constituent elements

For situations without growth or seasonality the SSM has been claimed to be the simplest realistic model that can be used (Johnston et al, 1999b). It is difficult to think of a real situation where the mean demand level does not change over time. The widespread use of EWMA and MA is consistent with such a time-varying mean. Both methods are used extensively in practice to deal with intermittent demand.

[3] A model, in the OR sense, is not just a set of equations but rather a mechanism based on which the future behaviour of a variable can be studied in detail. A model should offer the opportunity to gain insight on real data and derive properties of various estimators. It should provide the means of understanding the process and enable further developments to be made (see Johnston et al, 1986).

Exponential smoothing and moving averages consider the aggregate demand (demand per unit time period) and estimate how that moves through time. Both methods have been shown to perform well on empirical intermittent demand data (see chapter 4). Nevertheless, the “standard” forecasting method for intermittent demand items is considered to be Croston’s method (Croston, 1972; see for example Silver et al, 1998). Croston built demand estimates from constituent elements, namely the demand size, when demand occurs, and the inter-demand interval.

Croston’s estimator is, intuitively at least, superior to EWMA and MA. This particular method is currently used by best-selling statistical forecasting software packages (e.g. Forecast Pro) and it has motivated a substantial amount of research work over the years, not only in the area of forecasting (e.g. Schultz, 1987; Segerstedt, 1994; Willemain et al, 1994; Johnston and Boylan, 1996) but also in the area of inventory control for intermittent demand items (e.g. Dunsmuir and Snyder, 1989; Janssen, 1998; Strijbosch et al, 2000).

Croston (1972) showed that the EWMA method produces biased estimates of the mean level of demand in an intermittent demand context (see chapter 4). The model used by Croston in developing his method is based of the following assumptions:

1. SMM for the demand sizes
2. SMM for the inter-demand intervals
3. No cross-correlation between demand sizes and inter-demand intervals
4. Geometrically distributed inter-demand intervals
5. Normally distributed demand sizes

The last assumption is the only one that can be “relaxed”, in the sense that it does not affect the results given by the forecasting method.

The implications of assuming the SMM and SSM in an intermittent demand context are discussed in sub-section 2.3.4.

Croston suggested updating separately the demand sizes and inter-demand intervals with EWMA and he claimed that the ratio of the two estimates provides an unbiased estimate of the local constant level of demand. An identical estimator, generated upon the assumption that the demand process can be modelled as Poisson rather than Bernoulli, has been presented by Johnston and Boylan (1996). The method is called the Size-Interval method and is discussed in detail in chapter 4.

Croston's method has been claimed to be of great value to organisations forecasting intermittent demand. Nevertheless, empirical evidence (Willemain et al, 1994) suggests modest gains in performance when compared with less sophisticated techniques; some evidence even suggests losses in performance (Sani and Kingsman, 1997). An effort will be made in this research to identify some of the causes of this unexpected poor performance. For this to be done Croston's method needs to be revisited, examined and tested in detail.

2.3.4 The smoothing constant value

For the greater part of this thesis, demand will be theoretically modelled based on the first four of Croston's assumptions. One of the objectives of this thesis is to explore possible ways of improving the current standard practice (Croston's method) in forecasting intermittent demand. Nevertheless, there is an inconsistency between Croston's model and Croston's method that needs at this stage to be discussed. Croston assumed a SMM but he proposed updating the size and interval estimates using EWMA. As stated at end of sub-section 2.3.1, EWMA is unbiased for the SMM but it is not the optimal predictor.

Three arguments may be put forward to demonstrate that this inconsistency is not very restrictive (at least within the context of our research): a practical, a statistical and a methodological argument.

A practical argument

By using the EWMA estimator the forecast is always alert for any changes in the situation, which would reveal themselves through the forecast error (Johnston and Boylan, 1994). This is exactly the feature required if we are to forecast a constant mean model which is valid locally rather than globally.

It is clear that unless we can be absolutely sure that the SMM has applied in the past, and will continue to apply in the future, then it will be safer to use the weighted mean (EWMA) rather than the ordinary mean. Small smoothing constant values cause minor deviations from the SMM.

A statistical argument

By using EWMA we account for any small undetected changes in level. The changes cannot be detected because of the very nature of the data (few demand occurrences). To strengthen our point: unless we have demand histories of, say, 10 years, we cannot detect a SSM. That is, we cannot tell that the mean changes over time because we can hardly see that mean. Use of EWMA ensures that any real changes in the underlying demand level that cannot be practically detected (and theoretically be accounted for), because of the scarcity of data, will be reflected in the demand estimates.

A methodological argument

Theoretical results will be generated in the thesis, assuming that EWMA methods are used in conjunction with the stationary mean model assumption. The results will be tested on stationary theoretically generated data (chapter 7) but they will also be tested on real intermittent demand data that may not be stationary (chapter 10). The empirical results will allow us to gain insight as to what extent the model-estimator inconsistency is reflected in real world situations.

The first argument necessitates the use of low smoothing constant values in this thesis. Burgin and Wild (1967) found that $\alpha = 0.2$ is suitable for most weekly data but they implicitly recommended lower α values for slow movers. Croston (1972) recommended the use of α values in the range 0.05 - 0.2, when demand is intermittent. He also demonstrated, numerically, that for deterministic demands of size μ occurring every p (constant) review intervals, the EWMA forecast error reduces with the smoothing constant value. Finally he suggested that higher values of α, in the range of 0.2 - 0.3, may be found necessary only if there is a high proportion of items that is known to be non-stationary. Willemain et al (1994) conducted a forecasting accuracy comparison exercise between Croston's method and EWMA on empirical and theoretically generated data. The simulation results reported were based on $\alpha = 0.1$. Sani (1995) conducted research with the purpose of identifying the best periodic inventory control model and best estimator in an intermittent demand context. EWMA and Croston's method were simulated on real demand data using $\alpha = 0.15$. Johnston and Boylan (1996) simulated the performance of Croston's method and EWMA for the purpose of establishing regions of superior forecasting accuracy. Results were reported for $\alpha = 0.15$.

In an intermittent demand context, low smoothing constant values are recommended in the literature. Smoothing constant values in the range 0.05 - 0.2 are viewed as realistic. From an empirical perspective, this range covers the usual values of practical usage whereas from a theoretical perspective it does not introduce significant inconsistencies between the EWMA and SMM. In consequence, this is the range of values that we focus upon, during this Ph.D. research.

One way suggested in the literature to define an exponential smoothing system that is equivalent to an N period moving average, is to select the smoothing constant in such a way so that the estimates have the same age, or lag (Brown, 1963):

$$\frac{1-\alpha}{\alpha} = \frac{N-1}{2} \text{ or } \frac{2}{N+1} = \alpha \tag{2.9}$$

It is important to note that the equivalence given in equation (2.9) holds only under the stationary mean model assumption. If the SSM is assumed to best represent the underlying demand pattern then the equivalence is different (Johnston et al, 1999b).

2.4 Alternative approaches to dealing with intermittence

In addition to the forecasting methods discussed in the previous section, three other approaches to forecasting intermittent demand have been presented thus far in the academic literature. The first approach relies on the past data and is restricted to a specific planning and control environment (Early Sales method). The second approach incorporates subjective elements in the forecasting process, requires close departmental co-operation and is restricted to small industrial markets (Order

Overplanning). The last approach is the most computationally demanding and is based on employing parametric bootstrapping methods.

2.4.1 The Early Sales method

The "Accurate/Quick Response strategy" and the corresponding approach to forecasting, which is the Early Sales (EaSa) method, were developed in the early 90's in order to formulate the production planning decisions (for reduced mark-downs/stockouts and increased profit) in the apparel industry (Fisher et al, 1994; Fisher and Raman, 1996). The concepts apply also to any product line with similar characteristics (Kerke et al, 1990; Kurawarwala and Matsuo, 1996), namely: short product life cycle, long total production lead time, lack of historical demand data availability and lumpy demand because of sales occurring in a concentrated season.

The EaSa method aims at exploiting information from actual orders that have already been received for future delivery. At the moment of generating the forecasts only demands of those customers having long delivery lead times are known with certainty. The balance of the demand is unknown. However it can be estimated provided some degree of correlation exists between the known and the unknown portion of the demand. The technique applies in contexts with seasonal demand, where early buyers provide important information on fashion and trends in the market, or in the case that sales of a volatile product occur in a concentrated season, which means that a manufacturer would need an unjustifiably large capacity to be able to make goods in response to actual demand. The method applies only to the above specified planning and control environments with no application to other lumpy demand situations.

2.4.2 A "subjective" approach to forecasting intermittent demand

The Make To Order (MTO) manufacturing environment is known to be characterised by high levels of lumpiness (e.g. Marucheck and McClelland, 1986). The Order Overplanning (OrOv) method has been reported to assist MTO manufacturers in dealing with lumpy demand. OrOv (Bartezzaghi and Verganti, 1995; Bartezzaghi et al, 1996; Verganti, 1996) is a forecasting method that aims at fully exploiting early information that the prospective and regular customers generate during their purchasing process. OrOv uses as forecasting unit each single customer order instead of the overall demand for the Master Production Schedule (MPS) unit. So the forecast unit is distinguished from the MPS planning unit. The expected requirements for a module (that belongs to a particular order) are overestimated (because of the sources of uncertainty within the planning horizon, namely: order acquisition, actual due date, system configuration (number and types of apparatus) and apparatus configuration (modules)) by implicitly[4] incorporating in them the slack necessary to handle those uncertainties, i.e. by means of introducing redundant configurations, so as to satisfy any request that may actually be received. The demand forecast for the MPS unit is obtained by adding up the requirements included in the individual forecast orders. In particular, if Q_{ij} is the estimated requirement for a module in a forecast order j that is likely to be due in period i, and if this module belongs to m_i different orders then the total forecast Q_i for period i simply equals:

$$Q_i = \sum_{j=1}^{m_i} Q_{ij} \tag{2.10}$$

In OrOv, forecasting is not the numerical result of an algorithm for smoothing historical data, but is an organisational process, closely linked to the purchasing practices of the customer. In fact the method relies upon the capabilities of Sales to anticipate future requirements by continuously gathering information from customers and to exchange this subjective information with Manufacturing.

As shown in figure 2.1, inventory costs and obsolete parts are reduced since physical buffers are located at the lower level of BoMs and there is no need to stock end items/modules unless they have been ordered by the customers.

[4] The size of the slack cannot be directly estimated but is an output variable together with the service level achieved. That is, the slack can be quantitatively defined only after the method has been implemented.

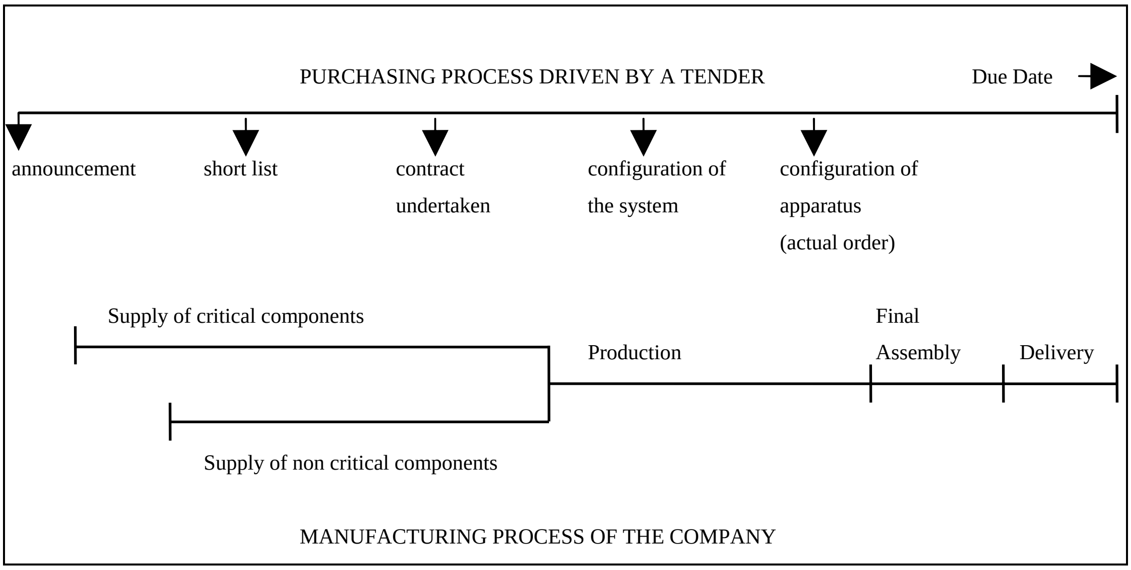

Figure 2.1. Order Overplanning (flow of the early information)

Using OrOv means that exploitation of information on future requests that otherwise would be lost, now is feasible. Uncertainty variability is closely followed over time and slacks are planned only when specific market opportunities are foreseen. When periods without likely requests occur, neither requirements nor buffers are set.

The benefits associated with the use of this method can be realised only in an industrial MTO context, when (a) there is a certain amount of information on customers future requests, and (b) there is a certain pattern of the information provided by the customers during their purchasing process.

2.4.3 Parametric bootstrapping

The bootstrap method introduced by Efron (1979) is a resampling method whereby information in the sample data is recycled for estimating the mean, variance, confidence intervals and other statistics. The bootstrap exploits the similarities of the sample to the population. In fact, it just reconstructs an approximate population by replicating the sample as many times as the computing capacity allows.[5] Equivalently, the original sample is viewed as the population and a sampling process with replacement is introduced.

[5] It has been suggested that the number of bootstrap replications required to obtain reasonable estimates is in the range 50 – 200 for most applications (Efron and Tibshirani, 1986).

Suppose we have a sample $x = (x_1, x_2, ..., x_n)$ which has been drawn randomly from an unknown distribution F (x is independent and identically distributed variable, iid). The problem is to estimate the unknown population parameter y_F. A bootstrapped sample is drawn with replacement from the original observations and the parameter of interest is estimated, $\hat{y}_{F,1}$. This procedure is repeated k times and finally we approximate the distribution of the estimates of y_F, $\hat{y}_F$, by the bootstrap distribution $(\hat{y}_{F,1}, \hat{y}_{F,2}, ..., \hat{y}_{F,k})$. The bootstrap point estimate for the mean and standard error (s.e.) of the parameter of interest to us can then be calculated as follows:

$$\bar{y}_F = \frac{\sum_{j=1}^{k} \hat{y}_{F,j}}{k} \qquad (2.11)$$

$$s.e.(\hat{y}_F) = \left[\frac{\sum_{j=1}^{k} \left(\hat{y}_{F,j} - \bar{y}_F \right)^2}{k-1} \right]^{1/2} \qquad (2.12)$$

Even though the bootstrap sampling distribution often suffers from not being centred at the same place as the true sampling distribution, it displays about the same variability as the true sampling distribution. It is this remarkable property that permits the bootstrap to make confidence intervals of about the correct width.

Few parametric bootstrapping approaches have appeared in the academic literature to deal with intermittent demand (e.g. Snyder et al, 2000; Willemain et al, 2000). Snyder (1999) used a parametric bootstrap method to approximate the lead time demand distribution. Assuming that the underlying demand pattern can be reasonably represented by the SSOE “model”, the following procedure may be applied:

1. Calculate the variability of the intermittent demand sample data (σ^2)
2. Use the sample data to optimise the smoothing constant value and the seed value of the level of demand

3. Use Monte-Carlo random number generation methods to obtain values for the errors $\varepsilon_t \sim N(0, \sigma^2)$
4. Use the SSOE "model" to generate a realisation $Y_{t+1}, \ldots\ldots\ldots, Y_{t+L}$ of future series values (where L is the lead time)
5. Calculate the lead time demand: $\sum_{i=1}^{L} Y_{t+i}$
6. Repeat steps no. 3, 4 and 5 many times to approximate the lead time demand distribution
7. Use the mean and the variance of the approximate lead time demand distribution for inventory control purposes.

Two main drawbacks can be identified in the above approach:

- The sampling error in estimating the model parameters is not taken into account. Therefore the bootstrap procedure is likely to yield values for the replenishment level or re-order point below those actually required. Snyder (1999) calculated replenishment levels for a specified fill-rate (proportion of demand satisfied directly from stock) using the bootstrap method discussed above. Theoretical levels were also calculated using the gamma probability distribution approach to inventory control (Snyder, 1984). For all four SKUs that were examined in detail, the "bootstrap" levels were considerably lower than the theoretically correct ones.

- The computation time increases significantly. The bootstrap procedure discussed above is repeated, presumably, at the end of the following demand data period (for a periodic review system), or at the following demand occurring period (for a continuous review system) to update our estimates in the light of new information. No results have been presented in Snyder (1999) on the percentage increase in the computation time as compared to more standard approaches to updating estimates.

2.5 Data used

Unicorn Systems (UK) Ltd. is a software package manufacturer specialising in Forecasting and Inventory Management and Transportation Logistics. The company was established in 1989 under the name Unicorn Systems AB. In 1995 Unicorn purchased the MURCO division of Data Sciences and in 1997 added *OpenSlim* (software) to its portfolio with the purchase of Slimstock Systems. Today it is part of Syncron Supply International (http://www.unicorn.se). Unicorn's best sellers are *Paragon* (vehicle routing and scheduling) and *MURCO* (forecasting and inventory control). *MURCO* is primarily based upon algorithms developed by the late Professor J. Murdoch of the Cranfield Institute of Technology and is the software of interest to us in the context of this research.

The monthly data provided by Unicorn Systems (UK) Ltd. come from the automotive industry and cover the demand history of 3,000 SKUs over a two year period (24 demand data periods). All SKUs are treated as "single" units as opposed to being sold in packs of a certain size.

All data files provided by Unicorn Systems have been treated in practice as intermittent. This does not mean that there are not significant differences between them with respect to the scale of the demand data, when demand occurs, and the length of the inter-demand intervals. The statistical properties of the empirical data used for this research are discussed in detail in chapter 10 of the thesis.

The forecasting method currently employed by the software manufacturer, to deal with intermittence, is a 13 period simple moving average (MA). The method has been found, by the software manufacturer, to provide the best results on empirical data. No theoretical results will be derived in this research for the MA method. Nevertheless the 13 period MA will be tested on empirical data in chapters 10 and 11, as it is a commonly used method for forecasting intermittent demand. The forecasting and inventory control performance obtained for the MA will be viewed as a benchmark for the purpose of analysing the simulation results (see also chapter 9).

From equation (2.9), under the stationary mean model assumption, a 13 period moving average corresponds to an EWMA procedure with $\alpha = \frac{2}{14} \approx 0.145$. In chapter 6 of the thesis Mean Square Error (MSE) pair-wise comparisons will result in rules that indicate under what circumstances one estimator performs better than another. Results will be generated for α values in the realistic range 0.05 - 0.2 (see also sub-section

2.3.4). Nevertheless, one smoothing constant value only will need to be considered for the purpose of illustrating numerically the behaviour of the various estimators. The value chosen for that purpose is 0.145.

2.6 Conclusions

In this chapter an overview of the literature on forecasting intermittent demand requirements has been presented. Detailed review and critique of the literature in various research areas (including inventory control for intermittent demand items) will follow in the relevant chapters. Four main approaches to forecasting intermittent demand have been examined thus far in the academic literature:

1. EWMA methods (including Croston’s method)
2. Early Sales (EaSa)
3. Order Overplanning (OrOv)
4. Parametric Bootstrapping.

The first approach is the one to be considered in this thesis. The application of EaSa and OrOv is restricted to specific planning and control manufacturing environments. The application of bootstrap methods becomes easier every year. Nevertheless, and despite the tremendous current computing capabilities, the method is, time-wise, extremely demanding.

The assumptions to be made throughout this project have been explicitly stated in this chapter and specific methodological issues have also been addressed. We have discussed the theoretical inconsistency between Croston’s method and Croston’s model and we have selected the range of smoothing constant values that we focus upon during this Ph.D. research.

Finally, an introduction has been made in this chapter to the company that provided the empirical data to be used in this research. The nature of the data has also been, briefly, discussed. Details regarding the data series and their statistical properties follow in chapter 10 of the thesis.

CHAPTER 3

The Categorisation of "Non-Normal" Demand Patterns

3.1 Introduction

Williams (1984) noted that "*ever since 1915*[1]*, work has progressed on the control of stock of products whose demand is smooth and continuous; it has been found however that this work is not applicable to products whose demand is very sporadic or lumpy* (p. 939)". In addition, a similar statement could be made for the work that has been performed in the area of forecasting. Unless demand for an item occurs at every inventory review period and is of a fairly constant size, it is, in the majority of cases[2], expected to cause significant problems as far as forecasting and inventory control are concerned. Infrequent demand occurrences and/or irregular demand sizes, when demand occurs, do not allow demand per unit time period or lead time demand to be represented by the normal distribution and demand in these cases will be referred to as *non-normal* for the purpose of this research. By using this nomenclature we do not mean to imply that all regular, non-sporadic, non-lumpy, demands are necessarily normally distributed. A different distribution may be more appropriate, although it is true to say that the normality assumption should reasonably cover the majority of real world cases (see for example Silver et al, 1998).

A rather modest part of the Operational Research literature has been devoted to exploring non-normal demand patterns, modelling them under well stated assumptions and proposing certain theoretical solutions in order to overcome the associated irregularity (i.e. forecast accurately future requirements and/or control the stock efficiently[3]). Some confusion, however, has been noticed in the literature regarding the definitions of different non-normal demand patterns, in that different

---

[1] The Economic Order Quantity (EOQ) formula is accredited to Harris F. W. (1915) Operations and Cost A. W. Shaw Company, Chicago, pp. 48-52.

[2] Exceptions to that rule could be, for example, an intermittent demand pattern where the inter-demand interval is known and constant or a demand pattern that can be represented by a Poisson arrival stream and unit-sized transactions.

[3] A limited number of papers appeared in the area of inventory control for non-normal demand items have explicitly addressed the issue of forecasting.

authors use different criteria in order to define a non-normal pattern. For example, low demand has been defined by the magnitude of demand over a lead time and by the magnitude of demand over a calendar year. In addition, the cut-off values are most often arbitrary in nature and therefore one could doubt the validity of a rule such as "*A slow moving item is an item whose demand is less than ten units during the lead time*", as is often used in practice. This definition begs the question: why ten and not nine or eleven units during the lead time?

Finally, it is important to note that the data, based on which alternative non-normal definitions are developed, usually refer to different demand contexts. Consequently, the definitions tend to be inconsistent with each other. For example, the slow moving items definition is unlikely be the same for a grocery wholesaler and a car parts dealer because the volume of demand, frequency etc. is not the same.

3.1.1 Chapter structure

In an inventory control context the objective in defining or categorising non-normal demand patterns is the proposal of the most appropriate forecasting and inventory control method for the situation under concern. It may, therefore, be beneficial to compare alternative methods for the purpose of establishing regions of superior performance and then to categorise the demand patterns based on the results. This approach appears in Johnston and Boylan (1996) and it is the approach followed in this thesis as well.

In this chapter we approach the problem of categorising non-normal demand patterns from a theoretical perspective. Specific rules that enable such a categorisation will be developed in chapter 6. The rules developed in chapter 6 will be based on approximate cut-off values and the accuracy of those approximations will be tested in chapter 7, where a theoretically generated data simulation experiment is developed. Finally the rules will be further evaluated on real demand data in chapters 10 and 11.

This chapter is structured as follows:

In section 3.2 we discuss the alternative non-normal demand patterns presented in the literature and in section 3.3 we present a theoretical framework to facilitate the conceptual distinction between non-normal demand patterns. The issue of categorisation is then addressed in section 3.4 and two categorisation schemes are discussed: one appeared in the academic literature and one is currently employed by a software package manufacturer. Finally the approach of Johnston and Boylan (1996) is further discussed in section 3.5 and the conclusions of the chapter are presented in section 3.6.

3.2 The definition of alternative non-normal demand patterns

3.2.1 Slow moving demand

Slow moving items are often infrequent demand items. When a demand occurs it is just for a single unit or very few units. The demand stream in the case of slow movement can be reasonably modelled as a Bernoulli process if time is treated as discrete and a Poisson process if time is treated as continuous.

A definition of slow moving items, discussed by Williams (1984), is the following: slow moving items are associated with low sporadicity and average demand during the lead time is less than ten units. In the same paper, Williams argued that there are a number of problems associated with the use of this definition in practice, since it does not comply with certain requirements that should always be considered when developing such rules (see sub-section 3.4.2). Nevertheless this definition was used by Sani (1995) in verifying that his real demand series did represent a slow moving demand pattern.

Gelders and Van Looy (1978) defined a slow moving item as an item whose demand in a year is at most two units, while the cut-off value implicitly assigned to the annual demand in Tavares and Almeida (1983) is one unit. Kwan (1991) commented on the lack of consistency in definitions that had appeared in the literature for slow moving items. For her own work, she used the British Steel definition, which was an annual demand rate of less than three units[4].

3.2.2 Intermittent demand

Silver et al (1998) defined intermittent demand as "*infrequent in the sense that the average time between consecutive transactions is considerably larger than the unit time period, the latter being the interval of forecast updating* (p. 127)". Intermittent demand is often referred to as sporadic demand. For intermittent demand items the observed demand during some periods is zero interspersed by periods with low or high, regular or irregular non-zero demand. Therefore intermittence (sporadicity) refers to the demand incidences and not to the demand size when demand occurs. Nevertheless sporadicity has often been associated with lumpiness in the academic literature (Ward, 1978; Schultz, 1987; Dunsmuir and Snyder, 1989).

Johnston and Boylan (1996) provided a decision rule for intermittent demand (inter-demand interval greater than 1.25 inventory review periods) specifying for the first time in this way, how sporadic the demand has to be in order to benefit from Croston's method more than from exponential smoothing. The originality of their approach lies in the fact that the particular cut-off value was the outcome of a formal theoretical comparison of the two estimation procedures in order to establish regions of relative performance. Their paper is further discussed in section 3.5.

Bartezzaghi et al (1996) argued that non-normal demand patterns in industrial markets emerge as the consequence of internal structural characteristics of the market. Intermittence or sporadicity in particular may emerge as the consequence of two market characteristics:

Numerousness (number: n) of potential customers.

Frequency (f) of customer requests. How often the customers place an order.

---

[4] The definition was communicated by British Steel to Kwan.

As n and f decrease, intermittence increases. As mentioned above, the researchers were focusing on an industrial context (i.e. business-to-business markets) where both n and f can be reasonably estimated by the marketing (or sales) department. In a more general context, the combined effect of both factors could be interpreted as the demand arrival process (for example a Bernoulli arrival of demand with a specified probability of demand occurrence or a Poisson arrival stream with a specified mean λ).

A special case of intermittent demand is "clumped" demand (Ritchie and Kingsman, 1985). In that case, demand occurs occasionally and is therefore sporadic, but when it occurs is of a constant size. Demand arrivals can be modelled as a Poisson stream and demand itself can be represented by the Poisson distribution, taking into consideration that transactions are not unit sized but rather they are for multiple items (for the same SKU) of a fixed "clump" size. The clump sizes may be determined by pack sizes (Package Poisson distribution, Vereecke and Verstraeten, 1994) or because demand naturally occurs in clumps (Clumped Poisson distribution, Ritchie and Kingsman, 1985). Another possibility that has appeared in the literature is to find the mode of the transaction sizes (still assuming a Poisson arrival of demands) and to consider that as a hypothetical SKU (h-SKU) such that the demand, when recalculated in units of the h-SKU, is Poisson (Williams, 1984).

3.2.3 Irregular/Erratic demand

Demand irregularity does not refer to the demand pattern (how often demand occurs) but rather to the size of demand when demand occurs. Irregular demand is associated with a high variability of the size of demand, when demand occurs. Silver (1970) defines an erratic item as "*one having primarily small demand transactions with occasional very large transactions* (p. 87)", and he specifically excludes the possibility of zero demand time periods. According to Silver et al (1998) an item is said to have an erratic demand pattern if the variability is large relative to the mean. Brown (1977) had operationalised this by saying that demand is erratic when the standard deviation of demand per unit time period is greater than the level of demand. Under this definition erratic demand may also be intermittent.

According to Bartezzaghi et al (1996) the main drivers of demand irregularity are the following:

Heterogeneity (h) of customers' requests. Heterogeneous requests occur when the potential market consists of customers with considerably different sizes, e.g. few large customers coexist with a large number of small customers.

Variety (v) of customers' requests. Variety occurs when a single customer's requests vary in size.

As mentioned in the previous sub-section, this particular research was referring to a business-to-business environment where there is the potential to reasonably estimate these factors. In a more general context the combined effect of h and v can be reflected in the coefficient of variation of the demand sizes.

3.2.4 Lumpy demand

For lumpy demand items the observed demand during some periods is zero, interspersed with irregular non-zero demand. This situation is graphically presented in Watson (1987). It is not only the variability of the demand size (irregularity) but also the variability of the demand pattern (intermittence) that make lumpy demand items so difficult to forecast. As mentioned in section 3.2.2 lumpy demand has often been associated with sporadic demand. But clearly lumpy demand is always sporadic while the opposite is not necessarily true. Nevertheless in Vereecke and Verstraeten (1994) lumpy demand items are defined as "*items whose demand frequency is less than 4 times a year* (p. 379)".

It is generally recognised in the Operational Research literature (see for example Silver, 1970; Bartezzaghi et al, 1996; Silver et al, 1998) that lumpiness is often generated in the higher levels of a multi-echelon inventory system because of specific replenishment decisions, lot-sizing rules or even over-reactions of judgement on the part of the inventory controller at the lower levels in the Bill of Materials (BoMs). In any event the result is that a non-lumpy demand pattern at the ultimate consumer level can be turned into a very lumpy pattern at a higher level in the production or distribution network. As discussed in chapter 2, unless demand at any level of this network is viewed as independent it is outside the scope of this research.

In other cases lumpiness is also related to some exogenous variables, e.g. atmospheric conditions or even fashion that induces similar behaviour of customers (see for example Fisher and Raman, 1996). Those cases are also not examined by this research, since forecasting their demand translates to a different forecasting problem, namely estimating the effect of the exogenous variable.

The degree of lumpiness for a SKU will depend on both the degree of intermittence and that of irregularity. It is therefore both the demand arrival pattern (or inter-demand interval) and the coefficient of variation of demand sizes that need to be considered in order to define an item as lumpy.

3.3 A framework for categorising non-normal demand patterns

In this section a theoretical framework is presented that aims at facilitating a conceptual categorisation of the non-normal demand patterns and overcoming the confusion noticed in the literature regarding the non-normal demand definitions. The theoretical framework (figure 3.1) reflects the synthesis of the arguments that have been appeared in the literature and discussed thus far in this chapter.

From the arguments discussed in sub-section 3.2.2 it becomes obvious that as demand arrives more intermittently, sporadicity (intermittence) increases. Therefore intermittence is related only to the demand arrival stream. It follows that a precise definition of intermittent demand should bear no relationship to any properties of the demand transaction sizes.

As the coefficient of variation of demand sizes increases, irregularity (demand erraticness) increases as well (see sub-section 3.2.3). Therefore, erraticness is associated only with the distribution of the demand sizes, when demand occurs.

Irregular demand can be either intermittent or it may appear at every single period.

An intermittent demand item is not necessarily a low demand item. Slow movement is usually characterised, in practice, by the (low) volume of demand per unit time period. Low demand can be the effect of intermittence and/or small transaction sizes. Consequently, a low demand item is not necessarily intermittent.

Lumpy demand is both intermittent and irregular (sub-section 3.2.4). It follows that a definition of lumpiness should refer to both demand incidence and transaction size.

Finally a clumped demand item is an intermittent demand item which is also characterised by a constant or approximately constant transaction size (sub-section 3.2.2).

An intermittent demand item is not necessarily a lumpy or clumped demand item. The opposite is always true.

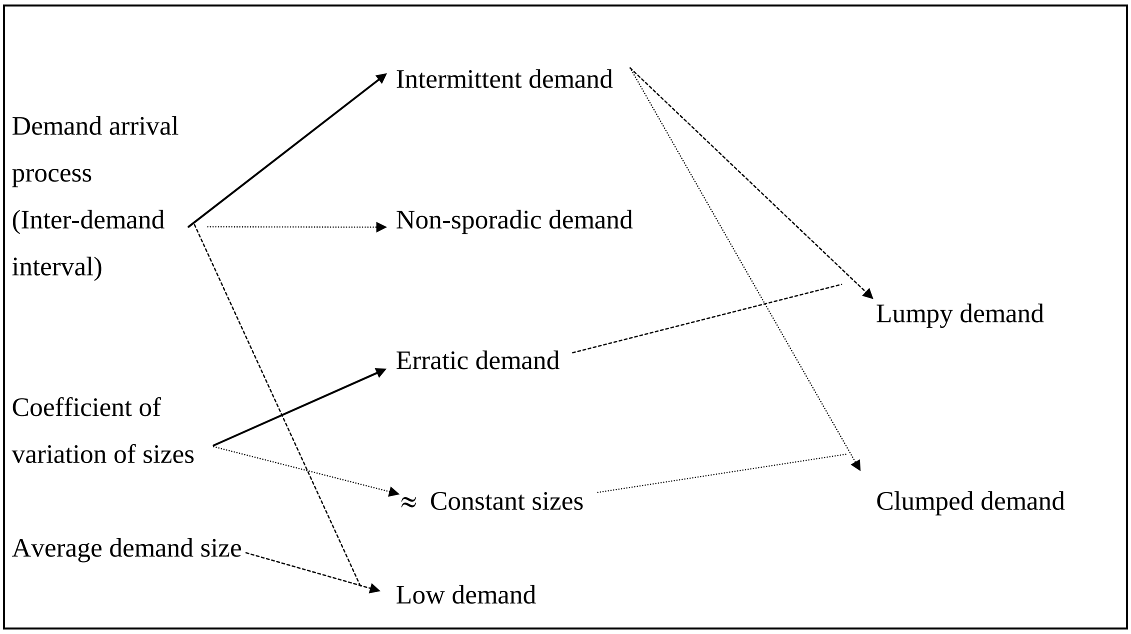

Figure 3.1. Categorisation of "non-normal" demand patterns

Considering the above we can formally define the alternative non-normal demand patterns.

- An intermittent demand item is an item whose demand is zero in some time periods.
- An erratic demand item is an item whose demand size is (highly) variable.
- A lumpy demand item is an item whose demand is zero in some time periods. Moreover demand, when it occurs, is (highly) variable.
- A slow moving item is an item whose average demand per period is low. This may be due to infrequent demand arrivals, low average demand sizes or both.

3.4 Demand categorisation schemes

3.4.1 A software package categorisation scheme

In this section the categorisation scheme employed by Unicorn Systems (UK) Ltd. is briefly discussed. The criteria used in this categorisation scheme reflect, according to the manufacturer, a typical approach taken in the inventory control software industry to the categorisation problem. That is, even though the exact cut-off values used in their particular scheme may differ from the values employed in other software packages, the main principles employed can be viewed as typical within the industry[5].

The purpose of the categorisation scheme is to indicate which estimation procedure and inventory control method should be used.

The demand per unit time period and the number of zero demand time periods are compared against the corresponding cut-off values to define lumpiness and slow movement of an SKU. Therefore, it is both the number of non demand occurring periods and the size of demand that are taken into consideration in order to classify demand as lumpy or slow. If demand has not been classified as lumpy or slow then it falls into the "normal" category. In this last case, if the variability of the demand sizes is greater than the specified cut-off value, the SKU is defined as erratic. Therefore, it is mainly the variability of the demand sizes that is considered in order to define an SKU as erratic.

[5] The software package manufacturer does not claim that the categorisation scheme employed in their software is more sophisticated than the industry norm. Rather, the core categorisations are present within differing inventory software.

Outliers can be explicitly handled for the lumpy and normal demand categories and as such may or may not be considered for categorisation purposes depending on the user.

The criteria used for categorisation purposes are dimensionless and the importance of this piece of information is further discussed in the following sub-section.

To apply the categorisation rules there is a minimum requirement for 4 demand periods one of which has to be a demand occurring period.

The major problem associated with the categorisation scheme under concern is that of a slow mover being classified as a lumpy SKU. In a slow moving demand pattern one outlier can cause the demand per unit time period to raise above the specified cut-off value and subsequently be classified as lumpy.

Finally, it is important to note that none of the cut-off values assigned to the categorisation criteria is the product of a “scientific” approach but rather they have all been subjectively estimated.

3.4.2 Williams’ categorisation scheme

Williams (1984) analysed the demand categorisation system employed by a public utility. At the time, products were classified into three categories, using an “ad hoc” method:

High sporadicity – one demand at least ten times the average weekly demand
Low sporadicity – average demand during the lead time less than ten
No sporadicity – neither of the above.

Williams (op. cit., p. 940) identified some limitations of the above discussed categorisation scheme which are quoted below:

“*1.* *The definitions do not suggest obviously in what different ways to treat the different categories.*

2. *Neither condition is dimensionless; thus the “low” condition depends on the product’s unit of measurement, and the “high” condition uses demand per week (ignoring, say, lead time).*

3. *A few outliers can cause a non-sporadic product to fall into the “highly sporadic” category.*

4. *75% of the products had insufficient data to reliably classify them by this method – obviously a major problem.*

5. *A very slow-moving item could be classified as “highly sporadic”(e.g. a product with demand of one unit every 11 weeks).*”

The limitations identified above reveal some theoretical and practical concerns regarding the validity of the particular categorisation scheme discussed in Williams (1984). Nevertheless, these theoretical and practical requirements should always be taken into account when developing rules for the purpose of distinguishing between alternative demand patterns. Considering the categorisation scheme presented in the previous sub-section, it becomes apparent that some of the limitations discussed by Williams (op. cit.) are still present in real world applications, 17 years after the publication of his paper.

Subsequently, we suitably modify Williams’ theoretical and practical considerations into a generic set of requirements which is as follows:

1. The categorisation scheme should suggest in what different ways to treat the resulting categories. The objective in categorising demand patterns is the identification of the most appropriate forecasting and inventory control methods to be applied to the different demand categories. As such, categorisation schemes should explicitly suggest which methods should be used under which circumstances.
2. The criteria considered in developing the rules should be dimensionless so that categorisation decisions regarding a SKU are independent of the product’s unit of measurement or of demand over any time period other than the lead time or the review period.

3. Sensitivity to outliers should be taken into account. The categorisation scheme should not allow products to move from one category to the other when few extreme observations are recorded.
4. The amount of data required to reliably classify demand patterns should also be considered. The decision rules should take into account the limited number of demand occurrences that characterise any intermittent demand pattern.
5. Slow moving demand patterns should not be "allowed" to be classified as lumpy.

The last requirement can be further amended to cover more general cases:

5. Logical inconsistencies should not allow demand for a SKU to be classified in an unintended category.

Williams (op. cit.) proposed a method of categorisation of demand patterns based on an idea that is called variance partition (we split the variance of the demand during lead time into its constituent parts). The purpose of categorisation was the identification of the most appropriate forecasting and inventory control methods for the resulting categories.

Using Williams' notation we denote by

n: the number of orders arriving in successive units of time with mean $\bar{n}$ and variance $Var(n)$ (the number of orders arriving in successive units of time are independent and identically distributed random variables, IIDRVs)
x: the size of these orders (transaction size) with mean $\bar{x}$ and variance $Var(x)$ (the sizes of the orders are IIDRVs) and
L: the lead time duration with mean $\bar{L}$ and variance $Var(\bar{L})$: (the lead times are IIDRVs).

Assuming that the three sets of random variables are independent of each other, we then have:

$$Var(Y_t) = \bar{n}\ Var(x) + \bar{x}^2\ Var(n) \tag{3.1}$$

$$\bar{Y}_t = \bar{x}\,\bar{n} \tag{3.2}$$

where $Var(Y_t)$ and $E(Y_t)$ are the variance of demand in a time period and the expected demand in a unit time period respectively.

The variance of demand over a lead time is given by (3.3) and the expected lead time demand by (3.4).

$$Var(Y_L) = \bar{L}\,\bar{n}\,Var(x) + \bar{L}\;\bar{x}^2\;Var(n) + \;\bar{n}^2\;\bar{x}^2\;Var(\bar{L}) \tag{3.3}$$

$$\bar{Y}_L = \bar{L}\;\bar{Y}_t = \bar{L}\,\bar{x}\,\bar{n} \tag{3.4}$$

while the squared Coefficient of Variation (CV^2) of the lead time demand is calculated as follows:

$$CV^2(Y_L) = \frac{CV^2(n)}{\bar{L}} + \frac{CV^2(x)}{\bar{n}\bar{L}} + CV^2(L) \tag{3.5}$$

If demand arrives as a Poisson stream with a mean λ, and also assuming constant lead times, (3.5) becomes:

$$CV^2(Y_L) = \frac{1}{\lambda\,\bar{L}} + \frac{CV^2(x)}{\lambda\,\bar{L}} \tag{3.6}$$

$\frac{1}{\lambda\,\bar{L}}$ indicates the number of lead times between successive demands (how often demand occurs or how intermittent demand is). The higher the ratio the more intermittent demand is.

$\frac{CV^2(x)}{\lambda\,\bar{L}}$ indicates the lumpiness of demand. Lumpiness in this context depends on both the intermittence and the variability of the demand size, when demand occurs. The higher the ratio the more lumpy demand is.

Categorisation of the items takes place in accordance with the following matrix:

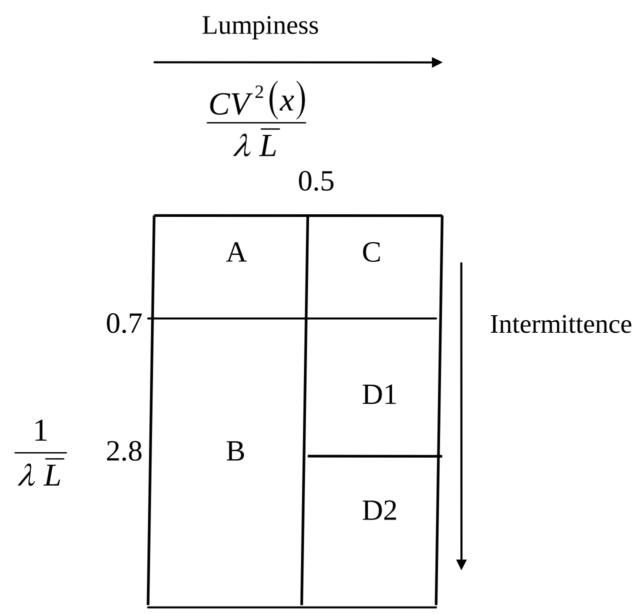

Figure 3.2. Williams' categorisation scheme

Category D2: Highly sporadic (lumpy). In that case we have few very irregular transactions.

Category B: Slow moving

Others: Smooth. According to Williams (1984), "*The assertion is that products in category A – and possibly categories C and D1 – have an essentially smooth demand pattern, and that continuous-demand stock control techniques can be used* (p. 942)".

3.4.3 Discussion of Williams' categorisation scheme

Consequently we assess Williams' categorisation scheme against the practical and theoretical requirements proposed in sub-section 3.4.2.

Williams verified that the categories resulting from his categorisation scheme have distinct demand patterns:

Category D2: no more than one demand occurrence during the lead time, Gamma distributed demand sizes (Williams, 1982; 1983)
Category B: Poisson demand distribution
Others: Gamma distributed demand per unit time period

and he proposed appropriate EWMA forecasting updating procedures and inventory control methods in order to deal with their specific requirements.

Both criteria employed for categorisation purposes are dimensionless:

$$\lambda \bar{L}: \ \frac{Incidences}{time\ period}\ no.\ of\ time\ periods$$

$$CV^2(x): \ \frac{units^2}{units^2}$$

“Buffer zones” were determined so that “borderline” products would not switch categories as the parameters vary from one side of the border to the other or categorisation would not be that easily affected by outliers. A buffer zone of ± 0.05 was suggested.

Initial estimates of the parameters required in the analysis can be calculated if just two demands have been recorded. As such Williams has taken into account the scarcity of observations that characterises categories B, D1 and D2.

Finally the categorisation scheme proposed does not allow SKUs to be classified in a category other than the intended one. Williams attributed this to the dimensionless nature of the criteria and his implicit argument was that requirements 2 and 5 (see previous sub-section) are closely related to one another. At this stage though we need to add that the consistency associated with his categorisation scheme cannot be attributed only to the dimensionless nature of the criteria, as applied on their own, but also to the double rather than single dimensional approach to the categorisation problem and consequently to the fact that both irregularity and intermittence are considered.

Williams' categorisation scheme meets all the theoretical and practical requirements that have been proposed in his paper. Nevertheless it is important to note that the cut-off values assigned to the criteria have been arbitrarily chosen so that they make sense only for the particular situation that was analysed in the 1984 paper. This creates certain doubts about the potential applicability of the proposed categorisation scheme to any different possible context.

The requirements proposed by Williams (1984) ensure the practical validity of a categorisation scheme for the particular situation under analysis. Furthermore a sixth requirement can be proposed that enables the applicability of a categorisation scheme to a wide variety of business contexts:

6. Determination of the cut-off values should be non-arbitrary thereby enabling the general applicability of the categorisation scheme.

The issue of meeting this requirement is explored in section 3.5 and in greater detail in chapter 6.

3.5 A different approach to the demand categorisation problem

Johnston and Boylan (1996) presented the Size-Interval method[6] (SI) for forecasting intermittent demand requirements. The method was based on Croston's concept of building demand estimates from constituent elements. The demand arrival process though was assumed to be Poisson rather than Bernoulli and consequently the inter-demand intervals were taken as exponentially rather than geometrically distributed. Their method was compared with EWMA on theoretically generated demand data over a wide range of possible conditions.

Many different average inter-demand intervals (negative exponential distribution), smoothing constant values, lead times and distributions of the size of demand

[6] The method's derivation is discussed in chapter 4.

(negative exponential, Erlang and rectangular[7]), were considered. The comparison exercise was extended to cover not only Poisson but also Erlang demand processes.

The results were reported in the form of the ratio of the MSE of one method to that of the other. For the different factor combinations examined in this simulation experiment the SI method was superior to EWMA for inter-demand intervals greater than 1.25 forecast revision periods. This result provided an answer to the question raised by Segerstedt (1994): "*When is it best to separate the forecasts like Croston suggests and when is it best with traditional treatment* (i.e. simple exponential smoothing)*? (p. 371)*".

The average inter-demand interval used as a decision criterion can be assessed against Williams' requirements that have been discussed in sections 3.4.2 and 3.4.3.

The decision rule clearly indicates which estimation procedure should be used for each category. However, inventory control issues were not addressed.

The criterion is clearly dimensionless:

$$\frac{\textit{no. of review periods}}{\textit{review period}} > 1.25$$

The rule is not affected by outliers since it deals only with intermittence (i.e. we refer only to the demand pattern, inter-demand intervals, rather than demand pattern and size of demand).

As in the case of Williams' categorisation scheme there is a minimum requirement of two demand incidences before the rule can be used.

Finally the rule deals only with speed of movement and in that sense there are only two resulting categories and no logical inconsistencies can occur.

---

[7] By considering these distributions a wide range of alternative shapes could be taken into account: From monotonically decreasing functions to unimodal positively skewed distributions to more normal type curves and finally to uniform functions.

Definition of intermittent demand in Johnston and Boylan (1996) results from a direct comparison of possible estimation procedures so that regions of relative performance can be identified. It seems more logical, indeed, working in the following way:

1. Compare alternative estimation procedures
2. Identify the regions of superior performance for each one of them
3. Categorise the demand patterns based on the method's comparative performance

rather than arbitrarily classifying demand patterns and then testing which estimation procedure performs best on each particular demand category.

The approach discussed above is the one adopted in this thesis. Because of its mathematically tractable nature, MSE is chosen for performing direct comparisons between existing and newly developed estimation procedures in chapter 6. The results, presented in the form of cut-off values assigned to the inter-demand interval and the squared coefficient of variation, enable us to specify regions of superior performance for each one of the methods considered. Non-normal demand patterns can then be defined based on the results. The combined use of both decision criteria will be assessed against Williams' theoretical and practical requirements in chapter 6.

3.6 Conclusions

Unless demand for any item can be reasonably represented by the normal distribution, significant problems should be expected in the area of forecasting and inventory control. Although it is not difficult to think of some exceptions to the rule, demand sporadicity and/or irregularity most often result in demand patterns that are difficult to forecast and consequently manage.

A certain confusion has been noticed in the academic literature as far as the definitions of the alternative non-normal demand patterns are concerned. Different authors use different criteria in order to define a specific demand pattern. Those criteria are hardly ever assessed against theoretical and practical considerations that should not be ignored if meaningful decision rules are to be constructed. Moreover, arbitrary cut-off values are, in the majority of cases, assigned to those criteria making their application to a more general context problematic.

The sources of non-normality were also discussed in this chapter for all possible non-normal demand patterns. Understanding the sources of non-normality is essential because the management may try to act on the sources in order to reduce the level of non-normality.

Consequently a theoretical framework was developed for the purpose of facilitating a conceptual distinction between non-normal demand patterns and intermittent, slow moving, erratic and lumpy demand items have been formally defined.

Williams (1984) expressed some theoretical and practical concerns regarding the development of categorisation rules for the purpose of distinguishing between alternative demand patterns. Williams' concerns are first drawn together to a generic set of requirements and, subsequently, they are amended to ensure that:

1. Items are always classified in the intended categories
2. The categorisation schemes developed are generally applicable.

The choice of the most appropriate forecasting (and inventory control) methods is the purpose of conducting any definition/categorisation exercise. Therefore it seems more logical to first compare alternative estimation procedures for the purpose of identifying their regions of superior performance and then, based on the results, categorise the demand patterns, rather than working the other way around. Such comparisons between existing and newly developed estimation procedures are performed in chapter 6. The decisions about when each method performs best will be based on the inter-demand interval and the squared coefficient of variation of the demand sizes. Both criteria will be shown to meet Williams' modified set of theoretical and practical requirements.

CHAPTER 4[1]

The Bias of Intermittent Demand Estimates

4.1 Introduction

Intermittent demand appears at random, with some periods having no demand (see section 3.3). Moreover when a demand occurs the request is very often for more than a single unit. As such, intermittent demand creates significant problems in the manufacturing and supply environment as far as forecasting and inventory control are concerned. It is not only the variability of the demand pattern but also, in many cases, the variability of the demand size that make intermittent demand so difficult to forecast.

In practice exponential smoothing is often used when dealing with intermittent demand. Exponential smoothing though places more weight on the most recent data, resulting, in the case of intermittence, in a series of estimates that are highest just after a demand occurrence and lowest just before demand occurs again.

Croston (1972) proposed a method that builds demand estimates taking into account both demand size and the interval between demand incidences. Despite the theoretical superiority of such an estimation procedure, empirical evidence suggests modest gains in performance when compared with simpler forecasting techniques; some evidence even suggests losses in performance.

In an effort to identify the causes of this forecast inaccuracy, as a first step towards improving Croston's method, a mistake was found in Croston's mathematical derivation of the expected estimate of demand. That mistake contributes towards the unexpectedly modest benefits of the method when applied in practice. Subsequently two new estimation procedures were developed that, theoretically, eliminate the forecast bias.

---

[1] This chapter is based on Syntetos and Boylan (2001).

4.2 Chapter overview

The estimation procedures that deal with intermittence can be divided into two categories: those that estimate the mean demand level directly (e.g. EWMA, Moving Average) and those that build mean demand level estimates from constituent elements (e.g. Croston's method, Croston, 1972). In both cases the ability of the alternative estimation procedures to deal with intermittence is judged on the accuracy of the estimates produced by these procedures. Therefore we are interested in the expected estimate of demand per unit time period or lead time demand (i.e. is the method biased or not?) and the variability associated with the estimates of the mean level of demand (i.e. the sampling error of the mean). Moreover, depending on what type of stock control system is utilised, not necessarily all estimates produced by the forecasting methods under concern are of interest to us. That is, if a continuous (re-order level) stock replenishment system is in place we are interested only in the estimates produced just after a demand occurrence (issue point) since only those estimates will be considered for replenishment purposes. On the other hand if a periodic (re-order interval or product group review) system is employed, all demand estimates are viewed as important. Therefore intermittent demand forecasting methods should be evaluated with respect to the accuracy of their estimates of the mean demand level for all points in time and for issue points only.

At the time of structuring this thesis it was decided that a separate discussion of the bias and variance issues was the best option. The analysis associated with the bias and variance of intermittent demand forecasting has led to a large number of findings. In order to do justice to the outcomes, bias and variance have been presented separately. In this chapter we focus on the issue of bias in intermittent demand forecasting and in chapter 5 the issue of variance of intermittent demand estimates will be discussed.

As discussed in chapter 2, the methodological approach taken in this thesis involves a mathematical analysis of the problem in hand, a theoretically generated data simulation, in order to check the accuracy of any approximated theoretical results and an empirical evaluation of the findings, to be conducted on real intermittent demand data series. The approximated results derived in this chapter are purely theoretical. The simulation experiment to be used for testing these results will be presented in chapter 7. In the experiment, theoretical demand data will be generated based upon the assumptions considered in this chapter. Finally the empirical evaluation of our findings will be discussed in chapters 10 and 11.

This chapter is structured as follows:

We first refer to the use of exponential smoothing for forecasting intermittent demand. Croston's method is presented followed by a discussion of the assumptions made in developing his model and the Size-Interval method developed by Johnston and Boylan (1996).

Empirical evidence for Croston's method is presented followed by a mathematical explanation of one factor contributing to the unexpectedly poor results found in practice. Croston's method is found to be biased and the bias is approximated for all possible smoothing constant values. Finally two methods: the λ Approximation method and the Approximation method are developed that produce approximately unbiased estimates of the mean demand level.

4.3 Theoretical background

4.3.1 Croston's critique of exponential smoothing

Croston (1972), as corrected by Rao (1973), proved the inappropriateness of exponential smoothing as a forecasting method when dealing with intermittent demands and he expressed in a quantitative form the bias associated with the use of this method when demand appears at random with some time periods showing no demand at all.

He first assumes deterministic demands of magnitude μ occurring every p review intervals. Subsequently the demand Y_t is represented by:

$$Y_t = \begin{cases} \mu, & t = np+1 \\ 0, & otherwise \end{cases} \tag{4.1}$$

where $n = 0,1,2...$ and $p \geq 1$.

Conventional exponential smoothing updates estimates every inventory review period whether or not demand occurs during this period. If we are forecasting one period ahead, Y'_t, the forecast of demand made in period t, is given by:

$$Y'_t = Y'_{t-1} + \alpha e_t = \alpha Y_t + (1-\alpha) Y'_{t-1} \quad (4.2)$$

where α is the smoothing constant value used, $0 \leq \alpha \leq 1$, and e_t is the forecast error in period t.

Under these assumptions, if we consider the demand estimates after demand occurs the expected estimate of demand per period is not μ / p, i.e. the population expected value, but rather:

$$\mathrm{E}(Y'_t) = \frac{\mu}{p} \frac{p\alpha}{1-(1-\alpha)^p} = \frac{\mu\alpha}{1-\beta^p} \quad (4.3)$$

where $\beta = 1 - \alpha$.

Croston then refers to a stochastic model of arrival and size of demand, assuming that demand sizes, z_t, are normally distributed, $N(\mu, \sigma^2)$, and that demand is random and has a Bernoulli probability $1/p$ of occurring in every review period (subsequently the inter demand intervals, p_t, follow the geometric distribution with a mean p). Under these conditions the expected demand per unit time period is:

$$\mathrm{E}(Y_t) = \frac{\mu}{p} \quad (4.4)$$

If we isolate the estimates that are made after a demand occurs, Croston showed that these estimates have the biased expected value:

$$\mathrm{E}(Y'_t) = \mu(\alpha + \beta / p) \quad (4.5)$$

The error, expressed as a percentage of the average demand, is shown to be $100\alpha(p-1)$ and reveals an increase in estimation error produced by the Bernoulli arrival of demands as compared with constant inter-arrival intervals.

4.3.2 Croston's method

Croston, assuming the above stochastic model of arrival and size of demand, introduced a new method for characterising the demand per period by modelling demand from constituent elements. According to his method, separate exponential smoothing estimates of the average size of the demand and the average interval between demand incidences are made after demand occurs. If no demand occurs, the estimates remain the same. If we let:

p'_t = the exponentially smoothed inter-demand interval, updated only if demand occurs in period t so that $\mathrm{E}(p'_t) = \mathrm{E}(p_t) = p$, and

z'_t = the exponentially smoothed size of demand, updated only if demand occurs in period t so that $\mathrm{E}(z'_t) = \mathrm{E}(z_t) = z$

then following Croston's estimation procedure, the forecast, $Y_t^{/}$ for the next time period is given by:

$$Y_t^{/} = \frac{z_t^{/}}{p_t^{/}} \tag{4.6}$$

and, according to Croston, the expected estimate of demand per period in that case would be:

$$\mathrm{E}(Y'_t) = \mathrm{E}\left(\frac{z'_t}{p'_t}\right) = \frac{\mathrm{E}(z'_t)}{\mathrm{E}\,(p'_t)} = \frac{\mu}{p} \tag{4.7}$$

(i.e. the method is unbiased.)

Now more accurate estimates can be obtained and an advantage of the method is that when demand occurs every period the method is identical to exponential smoothing.

Thus, it can be used not only for the intermittent demand items but for the rest of the Stock Keeping Units (SKUs) as well.

Croston (1972) claimed that the variance of the demand estimates per time period is given by:

$$Var\left(\frac{z'_t}{p'_t}\right) = \frac{\alpha}{2-\alpha}\left[\frac{(p-1)^2}{p^4}\mu^2 + \frac{\sigma^2}{p^2}\right] \tag{4.8}$$

Rao (1973) pointed out that the right hand side of equation (4.8) is only an approximation to the variance. In the following chapter we show that Croston's equation (4.8) is not only inexact but also incorrect.

Lead-time replenishment decisions take place only in the time periods following demand occurrence and are based on the equation:

$$R_t = z'_t + K\, m_t \tag{4.9}$$

where

R_t is the replenishment level to which the stock is raised,
m_t is the estimated mean absolute deviation of the demand size forecast errors and
K is a safety factor.

Schultz (1987) proposed a slight modification to Croston's method, suggesting that a different smoothing constant value should be used in order to update the inter-demand interval and the size of demand, when demand occurs. This modification to Croston's method has not been widely adopted, however, and it is not discussed further in this thesis.

4.3.3 Assumptions of Croston's model

Croston advocated separating the demand into two components, the inter-demand time and the size of demand, and analysing each component separately. He assumed a stationary mean model for representing the underlying demand pattern, normal distribution for the size of demand and a Bernoulli demand generation process, resulting in geometrically distributed inter-demand intervals.

No theoretical arguments have been identified in the literature in support of a model, other than the stationary mean, for describing the underlying intermittent demand pattern. As discussed in detail in chapter 2, the Steady State Model (SSM), which is viewed as a more realistic model than the stationary mean, cannot be justified from an empirical perspective in an intermittent demand context because of the scarcity of demand occurrences. That is, we cannot check whether the mean of demand per time period varies stochastically over time or not because we can hardly estimate the mean accurately. A stationary mean model will be assumed for all the theoretical analysis conducted in this thesis.

Three more assumptions implicitly made by Croston in developing his model are the following: independence between demand sizes and inter-demand intervals, independence of successive demand sizes and independence of successive inter-demand intervals. As far as the last assumption is concerned it is important to note that the geometric distribution is characterised by a "memory less" process: the probability of a demand occurring is independent of the time since the last demand occurrence, so that this particular assumption can be theoretically justified. With the only exception of Willemain et al (1994) (this paper is discussed in section 4.4) no other academic research has been identified in the intermittent demand literature that is not implicitly or explicitly based on the assumptions of Croston's model.

4.3.4 The Size-Interval (SI) method

If there is a random arrival of independent demands, the arrival process can be modelled as a Poisson stream. This idea was explored by Johnston and Boylan (1996). Their analysis was as follows:

If we set,

W : the demand per unit time with mean W_1 and variance W_2

S : the order size with mean S_1 and variance S_2

I : the inter-demand interval with mean I_1 and variance I_2

N : the number of orders per unit time with mean N_1 and variance N_2

then the demand in any period is the sum of the orders in that period and both the individual orders and the number of them in a given period are stochastic variables:

$$W = \sum_{i=1}^{N} S_i \tag{4.10}$$

Under the assumption that the order arrival process can be modelled as a Poisson stream and combining Clark's calculated mean and variance of the distribution of the summation of a number of stochastic random variables (1957):

$$W_1 = N_1 S_1 \tag{4.11}$$

$$W_2 = N_1 S_2 + N_2 (S_1)^2 \tag{4.12}$$

Using Cox's asymptotic equations (1962) for relating the number of orders (N) to the more easily measurable inter-demand interval (I) counting from a random point in time rather than an initial event (i.e. demand occurrence):

$$N_1 = \frac{1}{I_1} \tag{4.13}$$

$$N_2 \approx \frac{I_2}{(I_1)^3} + \frac{1}{6} + \frac{(I_2)^2}{2(I_1)^4} - \frac{I_3}{3(I_1)^3} \tag{4.14}$$

where I_3 is the third moment about the mean for the inter-order interval.

The authors proposed the following method (Size-Interval) for obtaining accurate (intermittent) demand per period estimates:

$$W_1 = \frac{S_1}{I_1} \tag{4.15}$$

$$W_2 = \frac{S_2}{I_1} + \frac{(S_1)^2}{I_1} \tag{4.16}$$

Thus, the forecasts can be generated from estimates of the mean and variance of the order size and the average inter-demand interval.

The SI method was compared with EWMA on theoretically generated demand data over a wide range of possible conditions. Many different average inter-demand intervals (negative exponential distribution), smoothing constant values, lead times and distributions of the size of demand (negative exponential, Erlang and rectangular), were considered. The comparison exercise was extended to cover not only Poisson but also Erlang demand processes. The results were reported in the form of the ratio of the MSE of one method to that of another. For the different factor combinations tried in this simulation experiment the SI method was superior to EWMA for inter-demand intervals greater than 1.25 review periods and in that way the authors showed how intermittent demand needs to be in order to benefit from the SI method (based on Croston's concept) more than the EWMA.

At this stage it is important to note that the estimate of mean demand is identical between Croston's method and SI method. Thus, later comments on bias of the $\frac{z_t^{/}}{p_t^{/}}$ (or $\frac{S_1}{I_1}$) estimator hold for both methods.

4.4 Performance of Croston's method

Croston's model is based on assumptions of independence (successive intervals are independent, successive demand sizes are independent and intervals and sizes are mutually independent) and normality of the demand size. The second assumption is important for inventory control purposes only (i.e. equation (4.9)). Willemain et al (1994) found correlations and distributions in real world data that violated Croston's assumptions. So they conducted a comparative evaluation of exponential smoothing and Croston's method under less idealised conditions using:

(a) Monte Carlo Simulation. Theoretical demand data were generated for different scenarios (lognormal distribution of demand size[2], cross-correlation between sizes and intervals, autocorrelated sizes and autocorrelated intervals) that violated Croston's assumptions. The comparison with exponential smoothing was mainly based on the Mean Absolute Percentage Error (MAPE).

(b) Industrial data, focusing on the MAPE for one step ahead forecasts.

The researchers concluded that Croston's method is robustly superior to exponential smoothing and can provide tangible benefits to manufacturers forecasting intermittent demand. A very important feature of their research, though, was the fact that industrial results showed very modest benefits as compared with the simulation results.

Sani and Kingsman (1997) conducted, with the use of simulation, a comparison between alternative forecasting methods evaluating them with respect to the cost and service level resulting from their implementation. The analysis was carried out on real and simulated low demand data. The forecasting methods compared were: a modification of Croston's method (Sani, 1995) as far as the variance, utilised for the replenishment levels calculation, is concerned; exponential smoothing updating every inventory review period and every 9 inventory review periods; an empirical

[2] The demand size lognormal distribution seems to have a considerable appeal to practitioners even though no theoretical arguments in its support have been presented in the academic literature. Forecast Pro XE versions 3 and 4 deals with intermittence by using Croston's method under the assumption that demand sizes are lognormally distributed.

forecasting method developed by one of the dealers who provided some of the real demand data that was used and the 1 year (26 periods) moving average updating every period (2 weeks). Sani's modification to Croston's method is based on the following calculation of demand variance:

$$Var(Y_t) \approx \max\left\{\frac{Var\, z'_t}{p'_t}, \frac{1.1 z'_t}{p'_t}\right\} \tag{4.17}$$

The proposed method of calculating the variance was motivated by the excessive replenishment stocking resulting from equation (4.9) and is based on the reasonable assumption (Kwan, 1991) that demand follows the negative binomial distribution and consequently the variance has to be greater than the mean[3].

The forecasting methods were compared across ten periodic inventory control methods (five empirically developed "simple" rules and five methods proposed in the academic literature).

The results showed that the best forecasting methods in terms of cost (ordering, holding and shortage costs are considered) were the empirical forecasting method and the moving average followed by Croston's method. When the service level was used as the performance criterion then the exponential smoothing updating every review period was the best method followed by the Croston forecast and the moving average. Overall the best forecasting method taking into account both cost and service level was concluded to be the 52 week moving average followed by Croston's method.

We may deduce from the work of Willemain et al and Sani and Kingsman that Croston's method may outperform other intermittent demand estimators when results are generated on simulated data. When real data is used, there is some evidence that less sophisticated, in fact very simple, forecasting methods seem to provide more accurate results and lead to more effective inventory control.

---

[3] The negative binomial distribution is a compound Poisson distribution and it is further discussed in chapter 5.

4.5 Expectation of the demand per time period

4.5.1 Expected estimate of demand – Croston's method

The empirical evidence suggests that the theoretical superiority of Croston's method is not reflected in the forecasting accuracy associated with the use of this method. The unexpectedly modest performance has not been explained by the method's sensitivity to the assumptions made by Croston in developing his model (Willemain et al, 1994). The normality assumption that could strongly be rejected in a practical situation does not affect the accuracy of forecasts given by Croston's method since any other distribution can be specified in order to estimate the mean size of demand when demand occurs.

Subsequently, in an attempt to identify the causes of this unexpected forecasting performance, a mistake was found in Croston's mathematical derivation of the expected estimate of demand per time period.

We know (assuming that order sizes and intervals are independent) that

$$\mathrm{E}\left(\frac{z'_t}{p'_t}\right) = \mathrm{E}(z'_t)\mathrm{E}\left(\frac{1}{p'_t}\right) \tag{4.18}$$

but

$$\mathrm{E}\left(\frac{1}{p'_t}\right) \neq \frac{1}{\mathrm{E}(p'_t)} \tag{4.19}$$

We denote by p_t the inter demand interval that follows the geometric distribution including the first success (i.e. demand occurring period) and by $\frac{1}{p_t}$ the probability of demand occurrence at period t. Now the case of $\alpha = 1$ is analysed since it is mathematically tractable; more realistic α values will be considered in the next sub-section. Assuming that $\alpha = 1$, so that $p'_t = p_t$ we then have:

$$\mathrm{E}\left(\frac{1}{p_t}\right) = \sum_{x=1}^{\infty} \frac{1}{x}\frac{1}{p}\left(1-\frac{1}{p}\right)^{x-1}$$

$$= \frac{1}{p}\sum_{x=1}^{\infty} \frac{1}{x}\left(\frac{p-1}{p}\right)^{x-1}$$

for $p > 1$ (i.e. demand does not occur in every single time period)

$$= \frac{1}{p}\sum_{x=1}^{\infty} \frac{1}{x}\frac{\left(\frac{p-1}{p}\right)^{x}}{\left(\frac{p-1}{p}\right)^{1}} = \frac{1}{p}\frac{1}{\frac{p-1}{p}}\sum_{x=1}^{\infty} \frac{1}{x}\left(\frac{p-1}{p}\right)^{x}$$

$$= \frac{1}{p-1}\left[\frac{p-1}{p} + \frac{1}{2}\left(\frac{p-1}{p}\right)^{2} + \frac{1}{3}\left(\frac{p-1}{p}\right)^{3} + \text{..........}\right]$$

$$= -\frac{1}{p-1}\log\left(\frac{1}{p}\right)$$

Therefore:

$$\mathrm{E}\left(\frac{z'_t}{p'_t}\right) = \mathrm{E}(z'_t)\mathrm{E}\left(\frac{1}{p'_t}\right) = \mu\left[-\frac{1}{p-1}\log\left(\frac{1}{p}\right)\right] \tag{4.20}$$

So if, for example, the average size of demand when it occurs is $\mu = 6$, and the average inter-demand interval is $p = 3$, the average estimated demand per time period using Croston's method (for $\alpha = 1$) is not $\frac{\mu}{p} = \frac{6}{3} = 2$ but it is $6 * 0.549 = 3.295$ (i.e. 64.75% bias implicitly incorporated in Croston's estimate).

The maximum bias over all possible smoothing parameters is given by:

$$Maximum\ bias = \mu\left[-\frac{1}{p-1}\log\left(\frac{1}{p}\right)\right]-\frac{\mu}{p} \tag{4.21}$$

This is attained at α = 1. For realistic α values, the magnitude of the error is smaller and it is quantified in the next sub-section.

4.5.2 An approximation of the bias implicitly incorporated in Croston's estimates

For α values less than 1 the magnitude of the error obviously depends on the smoothing constant value being used. We show, in this section, that the bias associated with Croston's method in practice can be approximated, for all smoothing constant values, by: $\frac{\alpha}{2-\alpha}\mu\frac{(p-1)}{p^2}$

The bias can be conveniently expressed as a percentage of the average demand and it is easily shown to be: $100\frac{\alpha}{2-\alpha}\left(1-\frac{1}{p}\right)$

The above approximation is proven as follows:

We apply Taylor's theorem to a function of two variables, $g(x)$

where:

x is the vector: $x=(x_1, x_2)$ and $g(x) = g(x_1, x_2) = \frac{x_1}{x_2}$ (with $x_1 = z'_t$ and $x_2 = p'_t$)

$\mathrm{E}(x_1)=\theta_1$, $\mathrm{E}(x_2)=\theta_2$ and

θ is the vector: $\theta=(\theta_1,\theta_2)$ with $g(\theta)=g(\theta_1,\theta_2)=\frac{\theta_1}{\theta_2}$

This is the case for the problem under concern, with $\theta_1=\mu$ and $\theta_2=p$.

$$g(x) = g(\theta) + \left[\frac{\partial g}{\partial \theta_1}(x_1 - \theta_1) + \frac{\partial g}{\partial \theta_2}(x_2 - \theta_2)\right]$$

$$+\frac{1}{2}\left[\frac{\partial^2 g}{\partial \theta_1^{\ 2}}(x_1 - \theta_1)^2 + 2\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1 - \theta_1)(x_2 - \theta_2) + \frac{\partial^2 g}{\partial \theta_2^{\ 2}}(x_2 - \theta_2)^2\right] + \ldots\ldots.. \quad (4.22)$$

$$\mathrm{E}[g(x)] =$$

$$\mathrm{E}[g(\theta)] + \mathrm{E}\left[\frac{\partial g}{\partial \theta_1}(x_1 - \theta_1) + \frac{\partial g}{\partial \theta_2}(x_2 - \theta_2)\right]$$

$$+\frac{1}{2}\,\mathrm{E}\left[\frac{\partial^2 g}{\partial \theta_1^{\ 2}}(x_1 - \theta_1)^2 + 2\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1 - \theta_1)(x_2 - \theta_2) + \frac{\partial^2 g}{\partial \theta_2^{\ 2}}(x_2 - \theta_2)^2\right] + \ldots\ldots\ldots\ldots \quad (4.23)$$

$$\frac{\partial g}{\partial \theta_1} = \frac{1}{\theta_2} \quad (4.24)$$

$$\frac{\partial g}{\partial \theta_2} = -\frac{\theta_1}{\theta_2^{\ 2}} \quad (4.25)$$

$$\frac{\partial^2 g}{\partial \theta_1^{\ 2}} = 0 \quad (4.26)$$

$$\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2} = -\frac{1}{\theta_2^{\ 2}} \quad (4.27)$$

$$\frac{\partial^2 g}{\partial \theta_2^{\ 2}} = -\theta_1\left(-\frac{2}{\theta_2^{\ 3}}\right) = \frac{2\theta_1}{\theta_2^{\ 3}} \quad (4.28)$$

Considering the assumption about independence between demand sizes and inter-demand intervals, equation (4.26) and the fact that the first moment about the mean is always zero, equation (4.23) becomes:

$$\mathrm{E}[g(x)] = \mathrm{E}[g(\theta)] + \frac{1}{2}\frac{\partial^2 g}{\partial \theta_2{}^2}\mathrm{E}\,(x_2 - \theta_2)^2 + ...$$

Therefore:

$$\mathrm{E}\left(\frac{x_1}{x_2}\right) = \frac{\theta_1}{\theta_2} + \frac{1}{2}\frac{\partial^2 g}{\partial \theta_2{}^2} Var(x_2) + ... \tag{4.29}$$

Assuming that the inter-demand interval series is not auto-correlated and that the inter-demand intervals (p_t) are geometrically distributed with a mean of p and homogeneous variance[4] of $p(p-1)$, it follows that:

$$Var(x_2) = Var(p'_t) = \frac{\alpha}{2-\alpha} Var(p_t) = \frac{\alpha}{2-\alpha}\, p(p-1)$$

Assuming that demand sizes (z_t) are distributed with a mean, μ, equation (4.29) becomes:

$$\mathrm{E}\left(\frac{x_1}{x_2}\right) \approx \frac{\theta_1}{\theta_2} + \frac{1}{2}\frac{\alpha}{2-\alpha}\frac{2\theta_1}{\theta_2{}^3}\, p(p-1) \tag{4.30}$$

$$\mathrm{E}\left(\frac{z'_t}{p'_t}\right) \approx \frac{\mu}{p} + \frac{\alpha}{2-\alpha}\,\mu\,\frac{(p-1)}{p^2} \tag{4.31}$$

Subsequently, the bias implicitly incorporated in Croston's estimates is approximated by (4.32):

$$Bias_{Croston} \approx \frac{\alpha}{2-\alpha}\,\mu\,\frac{(p-1)}{p^2} \tag{4.32}$$

In chapter 7 we show by means of experimentation on theoretically generated data that, for $\alpha \leq 0.2$, the difference between the theoretical bias given by (4.32) and the

[4]The issue of variance in the geometric distribution is discussed in chapter 5.

simulated bias lies within a 99% confidence interval of $\pm 0.2\,\%$ of the mean simulated demand.

4.6. The λ Approximation method

Since Croston's method is biased we consider applying a factor to the estimates produced by his method so that the second order bias term is directly eliminated.

We try to estimate the value of a parameter λ so that:

$$\mathrm{E}(Y'_t) = \mathrm{E}\left(\lambda \frac{z'_t}{p'_t}\right) = \frac{\mu}{p} \tag{4.33}$$

By applying a factor λ to Croston's updating procedure of sizes and intervals and considering approximation (4.31) we then have:

$$\mathrm{E}\left(\lambda \frac{z'_t}{p'_t}\right) \approx \lambda \frac{\mu}{p} + \lambda \frac{\alpha}{2-\alpha} \mu \frac{(p-1)}{p^2}$$

We can then set an approximation to the bias equal to zero in order to specify the value of parameter λ:

$$Bias \approx (1-\lambda)\frac{\mu}{p} - \lambda \frac{\alpha}{2-\alpha} \mu \frac{(p-1)}{p^2} = 0$$

$$1-\lambda = \frac{\lambda \frac{\alpha}{2-\alpha} \mu \frac{(p-1)}{p^2}}{\frac{\mu}{p}}$$

$$1 = \lambda\left(1 + \frac{\alpha}{2-\alpha} \frac{p-1}{p}\right)$$

$$\lambda = \frac{1}{1 + \frac{\alpha}{2-\alpha}\frac{p-1}{p}} = \frac{1}{\frac{2p - \alpha p + \alpha p - \alpha}{(2-\alpha)p}} = \frac{(2-\alpha)p}{2p-\alpha}$$

$$\lambda = \frac{1 - \frac{\alpha}{2}}{1 - \frac{\alpha}{2p}} \qquad (4.34)$$

Therefore we propose the following updating procedure for obtaining approximately unbiased intermittent demand estimates:

$$Y'_t = \frac{\left(1 - \frac{\alpha}{2}\right) z'_t}{\left(1 - \frac{\alpha}{2\,p'_t}\right) p'_t} = \frac{\left(1 - \frac{\alpha}{2}\right) z'_t}{p'_t - \frac{\alpha}{2}} \qquad (4.35)$$

We call this method, for the purpose of this research, the λ Approximation method. The expected estimate of mean demand per period for the λ Approximation method is given by equation (4.36).

$$\mathrm{E}(Y'_t) = \mathrm{E}\left[\frac{\left(1 - \frac{\alpha}{2}\right) z'_t}{p'_t - \frac{\alpha}{2}}\right] \approx \frac{\mu}{p} \qquad (4.36)$$

This approximation is not necessarily accurate when higher order terms are taken into account.

4.7 The Approximation method

From equation (4.34) we have:

$$\lambda = \frac{1 - \frac{\alpha}{2}}{1 - \frac{\alpha}{2p}}$$

But, as $p \to \infty$, $\lambda \to 1 - \frac{\alpha}{2}$

Therefore a possible estimation procedure, for intermittent demand data series with a large inter-demand interval, is the following:

$$Y'_t = \left(1 - \frac{\alpha}{2}\right)\frac{z'_t}{p'_t} \quad (4.37)$$

As in the case of the λ approximation method, the smoothing constant value is considered for generating demand estimates. The heuristic proposed seems to provide a reasonable approximation of the actual demand per period especially for the cases of very low α values and large p inter-demand intervals. We call this method, for the purpose of this research, the Approximation method. The expected estimate of mean demand per period for the Approximation method is given by equation (4.38).

$$\mathrm{E}(Y'_t) = \mathrm{E}\left(\left(1 - \frac{\alpha}{2}\right)\frac{z'_t}{p'_t}\right) \approx \frac{\mu}{p} - \frac{\alpha}{2}\frac{\mu}{p^2} \quad (4.38)$$

This approximation is not necessarily accurate when higher order terms are taken into account.

For the detailed derivation of (4.38) see *Appendix 4.A*.

4.8 Conclusions

Croston’s concept of building intermittent demand estimates from constituent elements, has been claimed to be of great value for organisations that deal with intermittent demand. Nevertheless when the method is tested on real demand data it shows very modest benefits. Moreover inferior performance has often been reported in the academic literature when the method is compared with less sophisticated methods such as exponential smoothing or the simple moving average. An attempt to explain one cause of this unexpected poor performance has been made in this chapter.

We first show that bias should always be expected when Croston's method is applied in practice. We have quantified the maximum bias that can be implicitly incorporated in Croston's estimates under the well-stated assumptions of his method. For $\alpha = 1$ the bias is given by (4.21). For the rest of the possible smoothing constant values the expected estimate of mean demand per period using Croston's method can be approximated by (4.31).

Moreover two new methods have been developed in this chapter based on Croston's concept of an explicit consideration of the demand size, when demand occurs, and the inter-demand interval. The first method is developed by estimating the value of a parameter which is directly applied to Croston's estimates in order to eliminate the bias. We call this theoretically unbiased method the λ Approximation method and its updating equation is given by (4.35). Subsequently, a heuristic is proposed based on the λ Approximation method. The heuristic is expected to work well for large inter-demand intervals and/or low smoothing constant values. We call this method the Approximation method and its updating equation is given by (4.37). All our derivations/approximations are accurate to the second order term in Taylor series.

CHAPTER 5

The Variance of Intermittent Demand Estimates

5.1 Introduction

In order to judge the ability of alternative estimation procedures to deal with intermittence we are interested in deriving the expected estimate of demand per unit time period or per lead time and the variance of the estimates produced by the forecasting methods under concern. In an inventory control context the replenishment quantities are determined by the forecasts of mean demand and the variance of the forecast errors. The latter consists of components relating to the variance of demand and the variance of the estimates of the mean demand level (i.e. the sampling error of the mean).

The issue of bias in intermittent demand forecasting was discussed in the previous chapter, where two new, approximately unbiased, estimation procedures were developed. In this chapter we focus on the issue of variance of the demand estimates produced by the following methods: EWMA, Croston's, λ Approximation and Approximation method.

The results (exact or approximate) derived in this chapter are purely theoretical. A theoretically generated data simulation experiment will be developed in chapter 7 in order to assess the accuracy of all the approximations derived in this chapter. Finally the empirical evaluation of our findings will be discussed in chapters 10 and 11.

This chapter is structured as follows:

We first refer to the variability associated with the estimates produced by the exponential smoothing method. In section 5.3 the issue of variance in Croston's method estimates is discussed. Croston, in developing his model, assumed that the inter-demand intervals follow the geometric distribution. However, by not correctly estimating the variance of inter-demand intervals, Croston failed to produce an accurate approximation of the variance of the demand estimates produced by his method. Moreover it has been shown, in chapter 4, that Croston's method is biased and in this chapter it is argued that the particular Stuart and Ord (1994) expression, used to derive the variance of Croston's estimates in his original paper, cannot be applied directly in order to calculate variance results when bias exists. Subsequently, we produce a correct approximation to the variance of Croston's estimates by considering the first three terms in a Taylor series.

In sections 5.4 and 5.5 we approximate the variance of intermittent demand estimates generated by the λ Approximation and Approximation method respectively. Finally, the conclusions of this chapter are presented in section 5.6.

5.2 The variance of EWMA estimates

Croston (1972) referred to a stochastic model of arrival and size of demand, assuming that demand sizes are normally distributed, $N(\mu, \sigma^2)$, and that demand is random and has a Bernoulli probability $1/p$ of occurring in every review period (consequently the inter demand intervals follow the geometric distribution with a mean, p). Under these conditions the expected demand per unit time period is:

$$\mathrm{E}(Y_t) = \frac{\mu}{p} \tag{5.1}$$

In this case Croston claimed that when demand estimates are updated every period using exponential smoothing, the expected estimate of demand per period is:

$$\mathrm{E}(Y'_t) = \frac{\mu}{p} \tag{5.2}$$

i.e. the method is unbiased.

The variance of demand per unit time period was given by Croston as:

$$Var(Y_t) = \frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p} \tag{5.3}$$

and the variance of the estimates as:

$$Var(Y'_t) = \frac{\alpha}{2-\alpha} Var(Y_t) = \frac{\alpha}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right] \tag{5.4}$$

(where α is the smoothing constant)

assuming a stationary mean model and homogeneous variance of demand per unit time period.

If we isolate the estimates that are made just after an issue (which are those that will be used for replenishment purposes by a continuous review stock control system) Croston showed that these estimates have the biased expected value:

$$\mathrm{E}(Y'_t) = \mu\left(\alpha + \frac{\beta}{p}\right) \tag{5.5}$$

and variance (as corrected by Rao, 1973):

$$Var(Y'_t) = \alpha^2\sigma^2 + \frac{\alpha\beta^2}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right] \tag{5.6}$$

where $\beta = 1 - \alpha$.

5.3 The variance of Croston's estimates

Croston suggested estimating the average interval between issues and the average size of an issue when it occurs and to combine those statistics to give an unbiased estimate of the underlying mean demand.

If we let:

p_t = the inter-demand interval that follows the geometric distribution with: $\mathrm{E}(p_t) = p$ and, according to Croston,

$$Var(p_t) = (p-1)^2 \tag{5.7}$$

p'_t = the exponentially smoothed inter-demand interval, updated only after demand occurs

z_t = the demand size, when demand occurs, that follows the normal distribution, $N(\mu, \sigma^2)$, and

z'_t = the exponentially smoothed size of demand, updated only after demand occurs

we then have:

$$\mathrm{E}(z'_t) = \mathrm{E}(z_t) = \mu \tag{5.8}$$

$$\mathrm{E}(p'_t) = \mathrm{E}(p_t) = p \tag{5.9}$$

$$Var(z'_t) = \frac{\alpha}{2-\alpha} Var(z_t) = \frac{\alpha}{2-\alpha} \sigma^2 \tag{5.10}$$

$$Var(p'_t) = \frac{\alpha}{2-\alpha} Var(p_t) = \frac{\alpha}{2-\alpha} (p-1)^2 \tag{5.11}$$

The variance of the ratio of two independent random variables x_1, x_2 is given in Stuart and Ord (1994) as follows:

$$Var\left(\frac{x_1}{x_2}\right) = \left(\frac{\mathrm{E}(x_1)}{\mathrm{E}(x_2)}\right)^2 \left[\frac{Var(x_1)}{(\mathrm{E}(x_1))^2} + \frac{Var(x_2)}{(\mathrm{E}(x_2))^2}\right] \tag{5.12}$$

For $x_1 = z'_t$ and $x_2 = p'_t$, considering equations (5.8), (5.9), (5.10) and (5.11), the variance of the estimates produced by using Croston's method is calculated by (5.13)

$$Var(Y'_t) = Var\left(\frac{z'_t}{p'_t}\right) = \frac{a}{2-a}\left[\frac{(p-1)^2}{p^4}\mu^2 + \frac{\sigma^2}{p^2}\right] \tag{5.13}$$

assuming that the same smoothing constant value is used for updating demand sizes and inter-demand intervals and that both demand size and inter-demand interval series are not auto-correlated and have homogeneous variances.

Rao (1973) pointed out that the right hand side of equation (5.13) is only an approximation to the variance. This follows since (5.12) is, in fact, an approximation.

5.3.1 The variance of inter-demand intervals

The number of independent Bernoulli trials (with a specified probability of success) before the first success is representing by the geometric distribution. An alternative form of the geometric distribution involves the number of trials up to and including the first success (demand occurring period). Considering the notation used in this chapter the variability of the geometrically distributed inter-demand intervals is $p(p-1)$, irrespectively of which form of the geometric distribution is utilised. Consequently (5.7) should be replaced by (5.14).

$$Var(p_t) = p(p-1) \tag{5.14}$$

5.3.2 The corrected variance of Croston's method estimates

By taking (5.14) into consideration, the variance of the demand per period estimates, using Croston's method, would become:

$$Var\left(\frac{z'_t}{p'_t}\right) \approx \frac{a}{2-a}\left[\frac{p-1}{p^3}\mu^2 + \frac{\sigma^2}{p^2}\right] \tag{5.15}$$

indicating that the approximated variance of the estimates produced by Croston's method is in fact greater than that calculated by Croston himself, equation (5.13)[1].

[1] Equation (10) in the original paper.

Nevertheless, approximation (5.15) is still not correct. In fact there is a fundamental problem in directly applying Stuart and Ord's result, given by (5.12), for the purpose of deriving the variance of the forecasts produced by a biased estimator.

This is proven as follows:

We apply Taylor's theorem to a function of two variables, $g(x)$

where

x is the vector: $x = (x_1, x_2)$ and $g(x) = g(x_1, x_2) = \frac{x_1}{x_2}$

with $\mathrm{E}(x_1) = \theta_1$ and $\mathrm{E}(x_2) = \theta_2$.

The vector θ is defined as: $\theta = (\theta_1, \theta_2)$, with $g(\theta) = g(\theta_1, \theta_2) = \frac{\theta_1}{\theta_2}$

$$g(x) = g(\theta) + \left[\frac{\partial g}{\partial \theta_1}(x_1 - \theta_1) + \frac{\partial g}{\partial \theta_2}(x_2 - \theta_2)\right] + ... \quad (5.16)$$

where $g(\theta) = \frac{\theta_1}{\theta_2}$ is just the first term in the Taylor series and not necessarily the population expected value.

For:

$$\mathrm{E}[g(x)] = g(\theta) + \varepsilon \quad (5.17)$$

where ε is an error term, which according to Croston, can be neglected, we then have:

$$Var[g(x)] = \mathrm{E}\left\{g(x) - \mathrm{E}[g(x)]\right\}^2 = \mathrm{E}\left[g(x) - g(\theta)\right]^2 \approx$$

$$\mathrm{E}\left[\frac{\partial g}{\partial \theta_1}(x_1-\theta_1)+\frac{\partial g}{\partial \theta_2}(x_2-\theta_2)\right]^2=\left(\frac{\mathrm{E}(x_1)}{\mathrm{E}(x_2)}\right)^2\left[\frac{Var(x_1)}{(\mathrm{E}(x_1))^2}+\frac{Var(x_2)}{(\mathrm{E}(x_2))^2}\right] \tag{5.18}$$

If we set:

$x_1 = z'_t$, the estimate of demand size, with $\mathrm{E}(z'_t)=\mu$
and $x_2 = p'_t$, the estimate of the inter-demand interval, with $\mathrm{E}(p'_t)=p$

so that $g(x)=Y'_t$

it has been proven, in chapter 4, that:

$\mathrm{E}(Y'_t)\neq\frac{\mu}{p}$ or $\mathrm{E}[g(x)]\neq g(\theta)$

Based on that, we argue that the error term in equation (5.17) cannot be neglected and therefore approximation (5.18) cannot be used to represent the problem in hand.

Our argument is discussed in greater detail in *Appendix 5.A*, where we also derive a correct approximation (to the second order term) of the variance of Croston's estimates. That variance expression is given by (5.19).

$$Var\left(\frac{z'_t}{p'_t}\right)\approx\frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]+\frac{\alpha^4}{1-(1-\alpha)^4}\frac{\mu^2}{p^4}\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2+1\right] \tag{5.19}$$

In chapter 7 we show, by means of simulation, that, across the whole range of control parameters to be considered, approximation (5.19) does not increase the accuracy of the calculated variance more than by only considering the first term of this approximation. In fact for certain regions of particular importance to us, approximation (5.19) performs worse. Taking that into account and in order also to simplify all the comparisons based on the MSE, to take place in chapter 6, we finally approximate the variance of Croston's estimates by (5.20).

$$Var\left(\frac{z'_t}{p'_t}\right) \approx \frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] \tag{5.20}$$

Simplification of the variance calculation will enable us, in chapter 6, to derive more tractable, and as it will be proven in chapter 7, more accurate decision rules based on which Croston's method performance can be assessed in detail.

The improvement in accuracy of approximation (5.20) over that given by (5.19) is assessed in detail in chapter 7.

5.4 The variance of the λ Approximation method estimates

The estimation equation for the λ Approximation method presented in the previous chapter is given by (5.21)

$$Y'_t = \frac{\left(1-\frac{\alpha}{2}\right)z'_t}{\left(1-\frac{\alpha}{2\,p'_t}\right)p'_t} = \frac{\left(1-\frac{\alpha}{2}\right)z'_t}{p'_t - \frac{\alpha}{2}} \tag{5.21}$$

and the expected estimate produced by this method was shown in chapter 4 to be as follows:

$$\mathrm{E}(Y'_t) = \mathrm{E}\left[\frac{\left(1-\frac{\alpha}{2}\right)z'_t}{p'_t - \frac{\alpha}{2}}\right] \approx \frac{\mu}{p} \tag{5.22}$$

In *Appendix 5.C* we perform a series of calculations in order to find the variance of the estimates of mean demand produced by the λ Approximation method. The variance is approximated by equation (5.23).

$$Var\left[\frac{\left(1-\frac{\alpha}{2}\right)z'_t}{p'_t-\frac{\alpha}{2}}\right] \approx \frac{\alpha(2-\alpha)}{4}\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2+p(p-1)\mu^2+\frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4}$$

$$+\frac{\alpha^4}{1-(1-\alpha)^4}\frac{\left(1-\frac{\alpha}{2}\right)^2\mu^2}{\left(p-\frac{\alpha}{2}\right)^6}\;p^2\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2+1\right] \tag{5.23}$$

In chapter 7 we show, by means of simulation (as in the case of Croston's method), that consideration of both terms of approximation (5.23) does not provide overall a more reliable estimate of the calculated variance than when only the first term of this approximation is considered. Exclusion of the fourth power term in approximation (5.23) is also expected to facilitate the MSE comparisons to be undertaken in chapter 6. By simplifying approximation (5.23) we will be able to derive meaningful decision rules that can be even further evaluated in order to assess the λ Approximation method's performance. The variance of the λ Approximation method is finally approximated by (5.24) and the issue of simplifying the variance calculation is further discussed in chapter 7.

$$Var\left[\frac{\left(1-\frac{\alpha}{2}\right)z'_t}{p'_t-\frac{\alpha}{2}}\right] \approx \frac{\alpha(2-\alpha)}{4}\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2+p(p-1)\mu^2+\frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4} \tag{5.24}$$

5.5 The variance of the Approximation method estimates

The estimation procedure for the Approximation method introduced in the previous chapter is:

$$Y'_t=\left(1-\frac{\alpha}{2}\right)\frac{z'_t}{p'_t} \tag{5.25}$$

with

$$\mathrm{E}(Y'_t) = \mathrm{E}\left(\left(1-\frac{\alpha}{2}\right)\frac{z'_t}{p'_t}\right) \approx \frac{\mu}{p} - \frac{\alpha}{2}\frac{\mu}{p^2} \tag{5.26}$$

The variance of the estimates produced by the Approximation method is calculated as:

$$Var(Y'_t) = Var\left(\left(1-\frac{\alpha}{2}\right)\frac{z'_t}{p'_t}\right) = \left(1-\frac{\alpha}{2}\right)^2 Var\left(\frac{z'_t}{p'_t}\right) \tag{5.27}$$

Considering approximation (5.20) we finally have:

$$Var\left(\left(1-\frac{\alpha}{2}\right)\frac{z'_t}{p'_t}\right) \approx \frac{\alpha(2-\alpha)}{4}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] \tag{5.28}$$

The effect of choosing equation (5.19) or (5.20) in order to approximate the variance of the Approximation method is discussed in chapter 7, where the choice of equation (5.20) is justified.

5.6 Conclusions

In this chapter the issue of variance of the estimates produced by alternative intermittent demand forecasting methods has been considered.

Croston assumed that the inter-demand intervals follow the geometric distribution including the first success (i.e. demand occurring period). We show, that by not correctly estimating the variance of inter-demand intervals and by assuming no bias in the estimates produced by his method, Croston fails to produce an accurate expression for the variance of those estimates. Subsequently a correct approximation to the variance is derived by applying Taylor’s theorem. This approximated variance is given by (5.20).

In chapter 4, two new estimation procedures were developed for forecasting intermittent demand. In this chapter the variance associated with their estimates is derived. The approximated variances of λ Approximation method and Approximation method are given by (5.24) and (5.28) respectively. Both approximations are accurate to the second order term in a Taylor series.

CHAPTER 6

The Mean Squared Error of Intermittent Demand Estimates[1]

6.1 Introduction

This chapter is based on Syntetos, Boylan and Croston (2000). The equation (6.15) in this chapter has also appeared in Strijbosch et al (2000). These two bodies of research, for all their similarities, have been conducted independently of each other.

In this chapter, exponential smoothing, Croston's method (see chapters 4 and 5), λ Approximation and the Approximation method will be compared (pair-wise comparisons) based on the theoretical Mean Squared Error (MSE) associated with their application over a fixed lead time of length L. The expressions for the bias and variance of Croston's method will be those derived in chapters 4 and 5. The pair-wise comparisons will result in the derivation of theoretical rules that indicate under what conditions one method is theoretically expected to perform better than another. These theoretical rules will be based on the squared coefficient of variation and the average inter-demand interval of the intermittent demand series. The cut-off values to be assigned to both criteria will be the outcome of a numerical analysis to be conducted on the theoretical results. Having obtained the cut-off values, we can then specify regions of superior performance of one method over another. In chapter 7 a theoretically generated data simulation experiment will be developed for the purpose of checking the accuracy of the decision rules developed in this chapter as well as all the other approximated results derived in the thesis.

[1] The author is indebted to Mr. J. D. Croston for his comments and constructive criticism regarding the material presented in this chapter.

The pair-wise comparisons to be conducted in this chapter will refer to issue points and all points in time as well. Croston (1972) has explored the theoretical properties of exponential smoothing, when dealing with intermittence. Croston showed that the performance of exponential smoothing is dependent upon which points in time are considered for generating forecasting accuracy results. When all points in time are considered the method is unbiased whereas a certain bias should be theoretically expected when issue points only (i.e. the estimates after demand occurrence) are taken into account. The performance of all other methods is not affected by which points in time we refer to. Hence, the results of any comparison between their MSEs will reflect performance differences in either context of application. Each method, though, will be separately compared against exponential smoothing for all and issue points in time only.

Why is MSE the accuracy measure chosen for the theoretical comparison of the alternative methods considered in this chapter? The reason is that MSE is a mathematically tractable accuracy measure. MSE is similar to the statistical measure of the variance of forecast errors (which consists of the variance of the estimates produced by the forecasting method under concern and the variance of the actual demand) but not quite the same since bias can also be explicitly considered.

The issue of bias in intermittent demand forecasting has been discussed in chapter 4, where we have also derived the expected estimate of demand per unit time period for all the estimation procedures developed (or corrected) during this research. Moreover, the variance of those estimates was derived in chapter 5. Being able to estimate the variance of the actual demand, we can then quantify the theoretical one step ahead MSE for all the forecasting methods, assuming that the demand population parameters are known. The relationship between the one step ahead MSE and the MSE over a lead time of fixed length L is derived in this chapter assuming auto-correlated forecast errors over the lead time but not auto-correlated demand.

6.1.1 The MSE comparisons on real data

Despite its attractive theoretical properties, MSE has some significant disadvantages when applied to real data, the most important being its scale dependent nature. The scale of the data often varies considerably among series. Then series with large numbers dominate the comparisons especially if quadratic loss functions, such as the MSE, are used to report error statistics. Moreover, MSE results are severely affected by the presence of outliers.

The theoretical and practical considerations in choosing an accuracy measure for the purpose of conducting a large-scale empirical accuracy comparison exercise will be discussed in detail in chapter 8. In that chapter we recognise that MSE is not the most appropriate measure for an accuracy comparison exercise. Therefore, other accuracy measures will be selected for the purpose of comparing the alternative intermittent demand forecasting methods on real data series. MSE will be used for empirically testing the theoretical results derived in this chapter, rather than determining overall accuracy performance differences.

6.1.2 Structure of the chapter

This chapter is structured as follows:

In section 6.2 the notation that will be used for deriving our results is presented. In section 6.3 the relationship between the variance of the forecast errors over a fixed lead time and the MSE is discussed for one step ahead forecasts. The standard results (e.g. Gilchrist, 1976) indicate that MSE is preferable to the variance of the forecast errors when bias is theoretically expected. In section 6.4 the variance of the forecast errors over a fixed lead time is derived, for biased and unbiased estimation procedures, assuming that, in both cases, the forecast errors are auto-correlated.

Consequently in section 6.5 the lead time MSE is derived as a function of bias, variance of the one step ahead estimates and variance of demand itself. In section 6.6 the results of section 6.5 are used to develop an expression for the MSE of the estimation procedures discussed in this thesis. In section 6.7, some important issues related to the theoretical accuracy comparison exercise are discussed. The detailed MSE comparisons are conducted in *Appendices 6.A – 6I* of the thesis. The theoretical rules derived from the pair-wise MSE comparisons are presented and analysed in section 6.8. Summary results are presented in section 6.9 where certain categorisation related issues discussed in chapter 3 are also revisited. Finally the conclusions of the chapter are presented in section 6.10.

6.2 Notation

We set as:

Y'_t: The estimate (made at the end of period t) of demand in any period $t+\kappa$ (assuming stationary mean model), obtained by any of the estimation procedures discussed in this thesis,

where $1 \leq \kappa \leq L$ and L is the forecast lead time

Note: Y'_t does not change over the forecast horizon

$\overline{Y'}$: The expected estimate of demand in any time period

Y_{t+k}: The actual demand in period $t+k$

$\overline{Y}$: The expected actual demand in any time period

e_{t+k}: The forecast error in period $t+k$

6.3 The relationship between MSE and the variance of the forecast errors

The following series of equations describe the relationship between MSE and variance of the forecast errors for one step ahead forecasts (i.e. $\kappa = 1$)

$$MSE = \mathrm{E}\left(Y'_t - Y_{t+1}\right)^2$$

$$= \mathrm{E}\left\{\left(Y'_t - \overline{Y'}\right) + \left(\overline{Y'} - \overline{Y}\right) + \left(\overline{Y} - Y_{t+1}\right)\right\}^2$$

(assuming stationary mean model and therefore independence of terms)

$$= \mathrm{E}\left(Y'_t - \overline{Y'}\right)^2 + \left(\overline{Y'} - \overline{Y}\right)^2 + \mathrm{E}\left(\overline{Y} - Y_{t+1}\right)^2$$

$$= Var(Estimates) + Bias^2 + Var(Actual\ Demand) \quad (6.1)$$

The variance of the forecast errors can be derived as follows:

$$Var(Forecast\ Error) = Var\left(Y'_t - Y_{t+1}\right)$$

$$= \mathrm{E}\left\{\left(Y'_t - Y_{t+1}\right) - \left(\overline{Y'} - \overline{Y}\right)\right\}^2$$

if Y'_t is an unbiased estimator:

$$= \mathrm{E}\left(Y'_t - Y_{t+1}\right)^2 = MSE \quad (6.2)$$

and if Y'_t is a biased estimator:

$$= \mathrm{E}\left\{\left(Y'_t - \overline{Y'}\right) - \left(Y_{t+1} - \overline{Y}\right)\right\}^2$$

$$= \mathrm{E}\left(Y'_t - \overline{Y'}\right)^2 - 2E\left[\left(Y'_t - \overline{Y'}\right)\left(Y_{t+1} - \overline{Y}\right)\right] + \mathrm{E}\left(Y_{t+1} - \overline{Y}\right)^2$$

(assuming stationary mean model and therefore independence of terms)

$$= \mathrm{E}\left(Y'_t - \overline{Y'}\right)^2 + \mathrm{E}\left(Y_{t+1} - \overline{Y}\right)^2$$

$$= Var(Estimates) + Var(Actual\ Demand) = MSE - Bias^2 \qquad (6.3)$$

6.4 The variance of the lead time forecast error

6.4.1 Theoretical arguments

In many short-term forecasting systems, the variance of the cumulative lead time forecast error is taken as the sum of the error variances of the individual forecast intervals.

$$Var\left(\sum_{\kappa=1}^{L} e_{t+\kappa}\right) = \sum_{\kappa=1}^{L} Var\left(e_{t+\kappa}\right) \qquad (6.4)$$

Assuming that the forecast errors are distributed with a constant variance V, equation (6.4) becomes:

$$Var\left(\sum_{\kappa=1}^{L} e_{t+\kappa}\right) = L\ V \qquad (6.5)$$

Equations (6.4) and (6.5) are consistent with serially uncorrelated forecast errors. Nevertheless, it has been argued (Fildes and Beard, 1991) that for lead times greater than one period "*the errors will typically be auto-correlated and this issue has received very limited attention* (p. 13)". Silver et al (1998) noted that the exact relationship between the variability of the forecast error during the lead time and that during the forecast interval "*depends in a complicated fashion on the specific underlying demand model, the forecast updating procedure and the values of the smoothing constant used* (p.114)". Subsequently they argued that the relationship under concern should be established empirically.

Under the Steady State Model assumption, it has been shown (Johnston and Harrison, 1986) that equation (6.5) neglects any correlation between the estimates of demand. This correlation exists, at least in part, because of the uncertainty in the estimate of the true underlying level of demand that is carried from one period to the other. This uncertainty does not consist only of the sampling error variance of the level (i.e. variance of the estimates) but of the variance of the mean demand (level) as well.

If the underlying mean level does not change (as assumed in this research) then the stationary mean model is satisfactory. In this case, the variance associated with the level of demand becomes zero (Johnston and Boylan, 1994) and the uncertainty reduces to the sampling error of the mean. If bias exists, the estimates of demand should also be auto-correlated although now it is not only the sampling error of the mean but also the bias that is carried forward from one period to the other. Whether bias exists or not, it is unreasonable to suppose that the forecast errors over a lead time are uncorrelated, assuming a stationary mean model.

By ignoring the auto-correlation term, we are most probably overstating the performance of the estimation procedure under concern since auto-correlations induced by bias or lumpiness are generally positive.

The variance of the lead time forecast error, taking auto-correlation into account, can be calculated as follows:

(We assume for simplicity that $\mathrm{L}=2$)

$$Var\left(\sum_{\kappa=1}^{2} e_{t+\kappa}\right) = Var\left(e_{t+1}+e_{t+2}\right)$$

$$= Var\left(e_{t+1}\right)+Var\left(e_{t+2}\right)+2Cov\left(e_{t+1}, e_{t+2}\right)$$

assuming constant variance of the unit time period forecast errors, V

$$= 2V + 2Cov\left(e_{t+1}, e_{t+2}\right) \quad (6.6)$$

6.4.2 The covariance of forecast errors

Subsequently we perform a series of calculations in order to quantify the covariance of the forecast errors.

$$Cov(e_{t+1}, e_{t+2}) = \mathrm{E}(e_{t+1}\ e_{t+2}) - \mathrm{E}(e_{t+1})\mathrm{E}(e_{t+2})$$

If Y'_t is a biased estimator (considering that the expected forecast error is the bias):

$$Cov(e_{t+1}, e_{t+2}) = \mathrm{E}(e_{t+1}\ e_{t+2}) - Bias^2 \tag{6.7}$$

The expectation of the product of the two forecast errors is calculated as follows:

$$\mathrm{E}(e_{t+1}\ e_{t+2}) = \mathrm{E}[(Y_{t+1} - Y'_t)\ (Y_{t+2} - Y'_t)]$$

$$= \mathrm{E}(Y_{t+1} Y_{t+2}) - \mathrm{E}(Y_{t+1} Y'_t) - \mathrm{E}(Y_{t+2} Y'_t) + \mathrm{E}(Y'_t Y'_t)$$

We assume:

- no auto-correlation in the actual demand data series
- no cross-correlation between estimates of demand (Y'_t) and demand itself ($Y_{t+\mathrm{K}}$). This assumption is not valid if we refer to Y'_t and Y_t (i.e. the demand in a particular time period and the subsequent forecast based on that demand figure are correlated).

Therefore:

$$\mathrm{E}(e_{t+1}\ e_{t+2}) = \overline{Y}^2 - 2\overline{Y}\ \overline{Y'} + \mathrm{E}\left({Y'_t}^2\right) \tag{6.8}$$

Two biased estimation procedures are discussed in this thesis: Croston’s method and EWMA (when issue points only are considered). The Approximation method (the derivation of its estimation procedure has been based on the λ Approximation method, chapter 4) is approximately unbiased for large p, inter-demand interval values. Nevertheless when the method is applied on series with very low p values (and large α values) a certain bias is theoretically expected.

We continue the derivation of the variance of lead time forecast error, for biased estimation procedures.

$\mathrm{E}\left({Y'_t}^2\right)$ can be calculated as follows:

$$\mathrm{E}\left({Y'_t}^2\right) = \overline{Y'}^2 + Var\left(Y'_t\right) \tag{6.9}$$

Considering (6.9), (6.8) becomes:

$$\mathrm{E}\left(e_{t+1}\ e_{t+2}\right) = \overline{Y}^2 - 2\overline{Y}\ \overline{Y'} + \overline{Y'}^2 + Var\left(Y'_t\right) = Bias^2 + Var\left(Estimates\right) \tag{6.10}$$

Considering (6.10), equation (6.7) becomes:

$$Cov\left(e_{t+1}, e_{t+2}\right) = Var\left(Y'_t\right) = Var\left(Estimates\right) \tag{6.11}$$

For the unbiased estimation procedures considered in this thesis (EWMA, all points in time; λ Approximation method) we have:

$$\mathrm{E}\left(e_{t+1}\ e_{t+2}\right) = Var\left(Y'_t\right)$$

and consequently the covariance of the forecast errors is given by (6.11).

6.4.3 The variance of the lead time forecast error (assuming error auto-correlation)

Having calculated the covariance of the forecast errors, we now proceed to find the variance of the lead time demand forecast errors.

Taking into account equation (6.11), equation (6.6) becomes:

$$Var\left(\sum_{\kappa=1}^{2} e_{t+\kappa}\right) = 2V + 2Var(Y'_t) \tag{6.12}$$

If $L = 3$, then:

$$Var\left(\sum_{\kappa=1}^{3} e_{t+\kappa}\right) = Var(e_{t+1} + e_{t+2} + e_{t+3})$$

$$= Var(e_{t+1}) + Var(e_{t+2}) + Var(e_{t+3}) + 2Cov(e_{t+1}, e_{t+2}) + 2Cov(e_{t+1}, e_{t+3}) + 2Cov(e_{t+2}, e_{t+3})$$

$$= 3V + 6\,Var(Y'_t)$$

and if $L = 4$, then:

$$Var\left(\sum_{\kappa=1}^{4} e_{t+\kappa}\right) = 4V + 12\,Var(Y'_t)$$

In fact equation (6.12) can be generalised to cover all possible lead times:

$$Var\left(\sum_{\kappa=1}^{L} e_{t+\kappa}\right) = LV + 2\,C(L,2)Var(Y'_t) \tag{6.13}$$

where

$$C(L,2) = \frac{L!}{2!(L-2)!} \tag{6.14}$$

$C(L,2)$ denotes the number of combinations of 2 out of L time period forecast errors.

Finally, based on (6.14), equation (6.13) can be written as:

$$Var\left(\sum_{\kappa=1}^{L} e_{t+\kappa}\right) = LV + 2\frac{L!}{2!(L-2)!}Var(Y'_t) = LV + L(L-1)Var(Y'_t)$$

$$= LVar(Y_t) + L^2 Var(Y'_t)$$

$$= LVar(ActualDemand) + L^2 Var(Estimates) \qquad (6.15)$$

6.5 The lead time MSE

The MSE over a lead time of duration L can be calculated as follows:

$$MSE_{L.T.} = \mathrm{E}\left(\sum_{k=1}^{L} e_{t+k}\right)^2$$

$$= \sum_{k=1}^{L} \mathrm{E}\left(e_{t+k}^2\right) + \sum_{i \neq j}\sum \mathrm{E}\left(e_{t+i}\ e_{t+j}\right) \qquad (6.16)$$

$$\mathrm{E}\left(e_{t+i}\ e_{t+j}\right) = \mathrm{E}\left(e_{t+1}\ e_{t+2}\right)$$

for all $i \neq j$ and $1 \leq i < L$ and $1 < j \leq L$

Therefore:

$$MSE_{L.T.} = L\ \mathrm{MSE} + L(L-1)\mathrm{E}\left(e_{t+1}\ e_{t+2}\right) \qquad (6.17)$$

where MSE is the Mean Squared Error of one step ahead forecasts.

For the biased intermittent demand estimation procedures discussed in this thesis (considering equations (6.1) and (6.10)), equation (6.17) becomes:

$$MSE_{L.T.} = L\{Var(Y'_t) + Bias^2 + Var(Y_t)\} + L(L-1)\{Var(Y'_t) + Bias^2\}$$

$$= L^2 Var(Y'_t) + L^2 Bias^2 + \mathrm{L} Var(Y_t)$$

$$= L\{L\mathrm{Var}(Estimates) + L\, Bias^2 + Var(Actual\ \mathrm{Demand})\} \tag{6.18}$$

Consequently the lead time MSE for the unbiased forecasting methods becomes:

$$MSE_{L.T.} = L\{L\mathrm{Var}(Estimates) + Var(Actual\ \mathrm{Demand})\} \tag{6.19}$$

6.6 The MSE of intermittent demand estimation procedures

In order to derive the MSE of the alternative estimation procedures discussed in this thesis, we have already quantified the expected one step ahead forecast error, for the biased estimation procedures (chapter 4) and the variance of the one step ahead estimates produced by all forecasting methods (chapter 5). The bias and the variance results are also summarised in this chapter, in sub-sections 6.6.2 and 6.6.3 respectively. The variance of the actual demand is presented in sub-section 6.6.1.

6.6.1 The variance of the actual demand

Under the assumptions considered in this thesis, intermittent demand data series are regarded as stationary following a compound binomial[2] distribution. The demand occurs as a Bernoulli process, with probability of demand occurrence $1/p$. In this case the inter-demand intervals, p_t, are geometrically distributed with a mean of p. The variance of the inter-demand intervals in this case (see also chapter 5) is $Var(p_t) = p(p-1)$.

The size of demand, z_t, is assumed to be arbitrarily distributed with a mean μ and variance σ^2. Under these conditions the demand pattern can be represented (see also chapter 4) by (6.20)

$$Y_t = \begin{cases} 0, & 1 - \frac{1}{p} \text{ probability of demand occurence} \\ z_t, & \frac{1}{p} \text{ probability of demand occurence} \end{cases} \tag{6.20}$$

with:

$$\mathrm{E}(Y_t) = 0\,\frac{p-1}{p} + \mathrm{E}(z_t)\frac{1}{p} = \frac{\mu}{p} \tag{6.21}$$

Consequently the variance of demand per unit time period is calculated as follows (see Croston, 1972, equation (6)):

$$Var(Y_t) = \frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p} \tag{6.22}$$

6.6.2 The bias of intermittent demand estimation procedures

The exponential smoothing method, when all points in time are considered, is an unbiased intermittent demand estimation procedure. The λ Approximation method developed in chapter 4 is an approximately unbiased forecasting method and therefore the bias associated with its application in practice is theoretically expected to be almost zero.

The bias implicitly incorporated in Croston's estimates was approximated in chapter 4 by applying Taylor's theorem. The bias is given by (6.23).

$$Bias_{CROSTON} \approx \frac{\alpha}{2-\alpha}\mu\frac{(p-1)}{p^2} \tag{6.23}$$

The Approximation method is approximately unbiased for large inter-demand intervals and low smoothing constant values. When those conditions are not satisfied, a small bias is expected. That bias is approximated by (6.24).

[2] Demand occurs as a Bernoulli process with the sizes of demand, when demand occurs, following an arbitrary distribution.

$$Bias_{APPROXIMATION} \approx -\frac{\alpha}{2}\frac{\mu}{p^2} \tag{6.24}$$

The negative sign in approximation (6.24) indicates that bias is in the opposite direction to that associated with Croston's method. Croston's method overestimates the mean demand level, whereas the Approximation method underestimates it.

When issue points in time are considered (consequently we refer only to the estimates following a demand occurrence) the expected estimate produced by the exponential smoothing method does not equal the population expected value, μ / p, but rather (Croston, 1972):

$$\mathrm{E}(Y'_t) = \mu\left(\alpha + \frac{\beta}{p}\right) \tag{6.25}$$

where $\beta = 1 - \alpha$.

The bias expected by applying exponential smoothing to a stationary intermittent demand series that follows a compound binomial distribution, considering only the issue points, is given by (6.26).

$$Bias_{EWMA} = \mu\left(\alpha + \frac{\beta}{p}\right) - \frac{\mu}{p} \tag{6.26}$$

6.6.3 The variance of intermittent demand estimation procedures

The variance of the one step ahead exponentially smoothed estimates, in case of intermittence and considering all points in time, is given by Croston (1972):

$$Var_{EWMA} = \frac{\alpha}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right] \text{, for all points in time} \tag{6.27}$$

When issue points only are considered, the variance of the one step ahead estimates for exponential smoothing is calculated as follows (Croston, 1972, as corrected by Rao, 1973):

$$Var_{EWMA} = \alpha^2 \sigma^2 + \frac{\alpha \beta^2}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right], \text{ for issue points only} \tag{6.28}$$

After correcting the theoretical expectation of the demand estimates produced by Croston's method in chapter 4, the variance of the one step ahead estimates was derived in chapter 5 by applying Taylor's theorem to a function of two variables. That variance is given by (6.29):

$$Var_{CROSTON} \approx \frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] \tag{6.29}$$

Moreover in chapter 5 the variance of one step ahead forecasts was also derived for the λ Approximation method and the Approximation method. Those variances are given by (6.30) and (6.31) respectively.

$$Var_{\lambda APPROXIMATION} \approx \frac{\alpha(2-\alpha)}{4} \frac{\left[\left(p-\frac{\alpha}{2}\right)^2 \sigma^2 + p(p-1)\mu^2 + \frac{\alpha}{2-\alpha} p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4} \tag{6.30}$$

$$Var_{APPROXIMATION} \approx \frac{\alpha(2-\alpha)}{4}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] \tag{6.31}$$

6.6.4 Estimation procedures for intermittent demand

Considering the derivations of section 6.5 and the results presented in sub-sections 6.6.1, 6.6.2 and 6.6.3, the exact or approximated MSE (over a fixed lead time of length L) associated with alternative intermittent demand estimation procedures is calculated as follows:

$$MSE_{EWMA} = L\left\{L\alpha^2\sigma^2 + L\frac{\alpha\beta^2}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right] + L\left[\mu\left(\alpha + \frac{\beta}{p}\right) - \frac{\mu}{p}\right]^2 + \left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right]\right\},$$

when issue points only are considered.

$$MSE_{EWMA} = L\left\{L\frac{\alpha}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right] + \left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right]\right\}$$, for all points in time

$$MSE_{CROSTON} \approx L\left\{L\frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] + L\left[\frac{\alpha}{2-\alpha}\mu\frac{(p-1)}{p^2}\right]^2 + \left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right]\right\}$$

$$MSE_{APPROXIMATION} \approx L\left\{L\frac{\alpha(2-\alpha)}{4}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] + L\left[\frac{\alpha}{2}\frac{\mu}{p^2}\right]^2 + \left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right]\right\}$$

$MSE_{\lambda APPROXIMATION} \approx$

$$L\left\{L\frac{\alpha(2-\alpha)}{4}\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2 + p(p-1)\mu^2 + \frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4} + \left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right]\right\}$$

6.7 Theoretical Comparison of the MSEs

We start the analysis by assuming that the approximated or exact MSE of one method is greater than the approximated or exact MSE of one other method, and we try to specify under what conditions the inequality under concern is valid. The comparisons for all methods, with the only exception of EWMA, refer to all and issue points in time as well. The EWMA estimation procedure is compared with the other methods separately for all and issues points in time as the particular method's performance depends on which points in time are considered.

It is easy to show that the comparison between any two estimation procedures is only in terms of the bias and the variance of the one step ahead estimates associated with their application. That is, the length of the lead time and the variance of demand itself do not affect the final results. Suppose for example that methods A and B are compared:

$$MSE_{MethodA} > MSE_{MethodB} \Leftrightarrow$$

$$L\left\{L\mathrm{Var}(EstimatesA) + L\ BiasA^2 + Var(Actual\ \mathrm{Demand})\right\} >$$

$$L\left\{L\mathrm{Var}(EstimatesB) + L\ BiasB^2 + Var(Actual\ \mathrm{Demand})\right\} \Leftrightarrow$$

$$L\mathrm{Var}(EstimatesA) + L\ BiasA^2 + Var(Actual\ \mathrm{Demand}) >$$

$$L\mathrm{Var}(EstimatesB) + L\ BiasB^2 + Var(Actual\ \mathrm{Demand}) \Leftrightarrow$$

$$\mathrm{Var}(EstimatesA) + BiasA^2 > \mathrm{Var}(EstimatesB) + BiasB^2$$

In order now to demonstrate our approach in detail, let us consider an example. We compare the MSE of Croston's method with that of the Approximation method over a lead time of length L ($L \geq 1$):

$MSE_{CROSTON} > MSE_{APPROXIMATION} \Leftrightarrow$

(see *Appendix 6.A* of the thesis)

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[\frac{4(p-1)}{2-\alpha} - \frac{2-\alpha}{p-1} - p(\alpha-4)\right]}{\frac{p(\alpha-4)(2p-\alpha)}{2-\alpha}} \quad (6.32)$$

for $p > 1$, $0 \le \alpha \le 1$

The implications of (6.32) are discussed in sub-section 6.8.1.

The theoretical rule developed above is expressed in terms of the squared coefficient of variation (CV^2) and the average inter-demand interval. In fact all the theoretical rules that will be developed in this chapter will have the same form. The theoretical rules can then be further analysed (considering different possible values of the control parameters: α, μ, p and σ^2) so that cut-off values can be determined. This numerical exercise will be conducted using *Microsoft Excel 97* (Windows 95).

All detailed pair-wise comparisons are conducted in *Appendices 6.A – 6.I.* of the thesis and only the decision rules (with similar form to that of inequality (6.32)) will be given in this chapter.

Since most of the theoretical MSEs given in sub-section 6.6.4 are approximate rather than exact results, it is important to note at this stage that the theoretical rules presented in this chapter are also approximate.

Moreover we need once more to consider that the rules reflect accuracy performance differences under the assumption of a Bernoulli demand generation process. Modelling the demand generation process as a Poisson or a condensed Poisson stream, the latter being a "censored" Poisson process in which every nth event is "marked" (Chatfield and Goodhardt, 1973), the inter-demand demand intervals will follow the exponential and Erlang (n) distribution respectively. In those cases the theoretical conclusions on the alternative estimation procedures' performance would not necessarily be the same.

Finally, it is important to note that all the pair-wise comparison results are generated assuming that the same smoothing constant value is employed by all of the estimation procedures under concern. Summary results, to be presented in sub-section 6.9, and the whole analysis on theoretically generated data as well as the greater part of the analysis on real demand data will also be based on the same assumption. We recognise that the use of the same smoothing constant may put one or more methods at a relative advantage/disadvantage but the issue of sensitivity of the comparative performance results to the application of the same α value has not been further explored, from a theoretical perspective. In the empirical part of the thesis some results will be generated with respect to the best α value performance for each of the forecasting methods.

6.8 Comparison results

The purpose of this section is to identify the most accurate estimation procedure(s) for different possible values of the control parameters α, μ, p and σ^2. The specific pair-wise comparison results will be studied in order to gain insight into the behaviour of alternative estimators operating in different conditions. Conclusive results regarding all methods, rather than two at the time, can then be generated and those results will be summarised in the following section.

6.8.1 MSE Croston's method – MSE Approximation method

As discussed in the previous section:

$MSE_{CROSTON} > MSE_{APPROXIMATION}$ if and only if:

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[\frac{4(p-1)}{2-\alpha} - \frac{2-\alpha}{p-1} - p(\alpha-4)\right]}{\frac{p(\alpha-4)(2p-\alpha)}{2-\alpha}} \text{ for } p > 1,\ 0 \le \alpha \le 1.$$

The denominator of the right hand side of the inequality is always negative and it can be shown, numerically, that for $p > 1.32$ the numerator of the right hand side of the inequality is always positive and therefore the inequality holds (superior performance is theoretically expected by the Approximation method).

For $p \leq 1.32$ the numerator becomes negative and the right hand side of the inequality becomes positive. In that case the right hand side of the inequality cannot take any values higher than 0.48.

- If $CV^2 > 0.48$ then $MSE_{CROSTON} > MSE_{APPROXIMATION}$
- If $CV^2 \leq 0.48$ then there is a p cut-off value ($1 < p \leq 1.32$) below which Croston's method performs better (i.e. the inequality is not valid). For example if $CV^2 = 0.15$, then the cut-off value is $p = 1.20$. For $1 < p \leq 1.20$ Croston's method performs better and for $p > 1.20$ the Approximation method performs better.

  As the ratio CV^2 decreases, the p cut-off value increases, and for $CV^2 = 0.001$ the cut-off value is $p = 1.32$ therefore Croston's method performs better for $1 < p \leq 1.32$.

The above results are valid for $\alpha = \frac{1}{7} \approx 0.145$ (see chapter 2) and approximately true for other realistic α values (refer to table 6.1). At this stage, it is important to note that the present analysis does not indicate the magnitude of the MSE differences. That is, we establish that one method performs better than another under a specific set of conditions but we do not indicate by how much. Performance differences will be explored in the following chapter along with the validity of all rules as well as their sensitivity to different control parameter values.

From the above analysis it is clear that there are four decision areas in which one method can be theoretically shown to perform better than the other. These areas are presented schematically in figure 6.1.

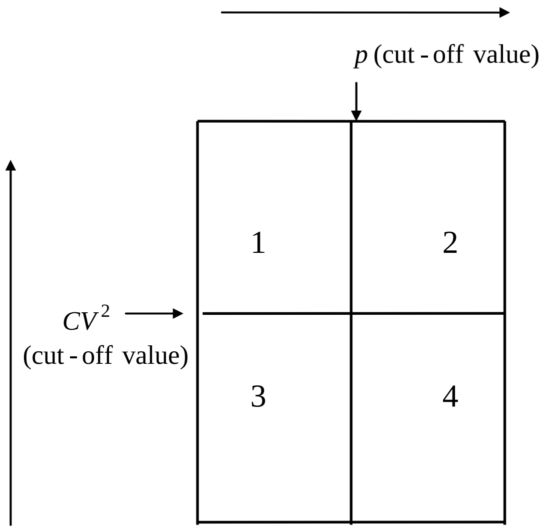

Figure 6.1. Decision areas

All the pair-wise comparisons to be conducted in this section will result in a categorisation matrix of the above form. The exact cut-off values for the categorisation criteria will be different for each comparison but the following comments apply to all of them.

For average inter-demand intervals and squared coefficients of variation above their corresponding cut-off values (area 1) we know with certainty which method is theoretically expected to perform better. The same happens when either of the criteria takes a value above its cut-off value, while the other takes a value below its cut-off value (areas 2 and 4). The only area that requires further examination is the one formed when both criteria take a value below their corresponding cut-off values (area 3).

In table 6.1 cut-off values that indicate regions of relative performance for Croston's and the Approximation method are presented, for smoothing constant values in the realistic range $0.05 - 0.2$ (refer to chapter 2).

α *smoothing constant value*	p *cut-off value*	CV^2 *cut-off value*
0.05	1.32	0.49
0.10	1.32	0.49
0.15	1.32	0.48
0.20	1.31	0.47

Table 6.1. MSE Croston's method – MSE Approximation

In the first column of table 6.1 we give the α smoothing constant values considered and in the second column the corresponding approximate p cut-off value above which the Approximation method always performs better. For average inter-demand intervals lower than the p cut-off value, the CV^2 cut-off values are presented in the third column, above which the Approximation method performs better.

From table 6.1 we can conclude that for all smoothing constant values that are likely to be applied in practice the Approximation method should always perform better than Croston's method for any $p > 1.32$ unit time periods and/or $CV^2 > 0.49$. Therefore figure 6.1 can now take the following form.

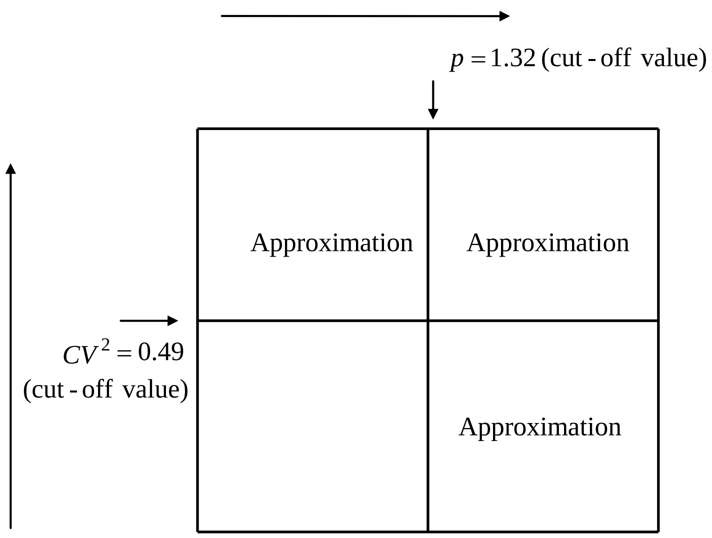

Figure 6.2. Decision areas (Croston's method – Approximation method)

At this stage the categorisation scheme is not complete. The area that corresponds to $p \le 1.32$ and $CV^2 \le 0.49$ has been left blank since neither method can be theoretically shown to perform better in all cases. Only when the values of both criteria are known, can we decide on which method should be chosen. The same will be true for all other pair-wise comparisons to be conducted in this section. Nevertheless, a numerical analysis of the comparison results (see also following chapter) shows that in this area of indecision the method that performs worse in decision areas 1, 2 and 4 (Croston's method in this particular comparison) turns out to outperform the other method (Approximation) for the majority of possible combinations of the control parameter values. When the opposite is the case (i.e. in the context of the current analysis, in the part of area 3 where the Approximation method still performs better) the MSE differences are so small that we are almost indifferent as to which method will be used.

We conclude that for the pair-wise comparison under analysis it is reasonable to assign the area of indecision (area 3 in figure 6.1) to Croston's method. Similarly, for all other comparisons between any two estimation procedures, the decision area 3 will be assigned to the method that performs worse in decision areas 1, 2 and 4. Clearly the derivation of a function, based on which the more accurate estimator can be identified, would be very welcome, but such an exercise is beyond the scope of this research.

Finally figure 6.2 can take the form which is indicated below (see figure 6.3).

Two important comments relating to the pair-wise comparison undertaken in this sub-section are the following:

- The bias of both methods does not differ significantly from one set of conditions to another while the variance of the estimates produced by the Approximation method is consistently lower than that of Croston's method.
- The pair-wise comparison results are insensitive to the smoothing constant value used.

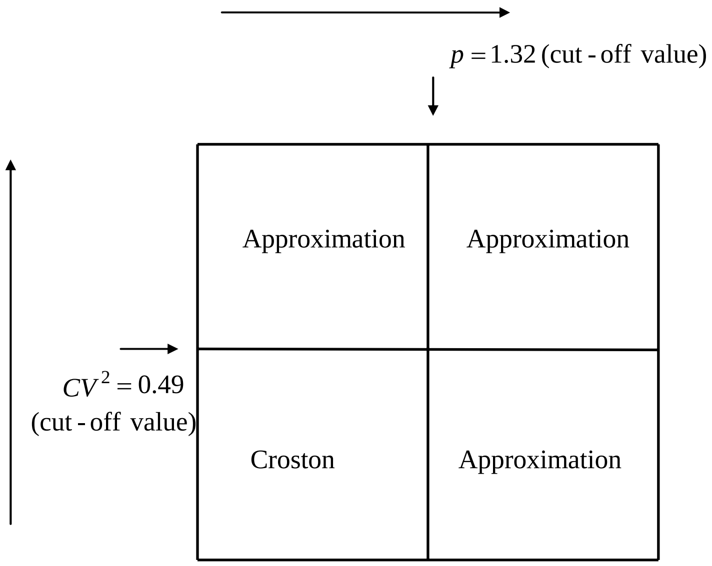

Figure 6.3. Cut-off values (Croston's method – Approximation method)

6.8.2 MSE EWMA – MSE Approximation method, issue points

$MSE_{EWMA} > MSE_{APPROXIMATION}$ if and only if (see *Appendix 6.B*):

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[\frac{(1-\alpha)^2}{2-\alpha} + \alpha(p-1) - \frac{\alpha}{4(p-1)p^2} - \frac{(2-\alpha)}{4p}\right]}{\frac{\alpha(p-1)}{4p} + \frac{2-\alpha}{4} - \alpha\, p^2 - \frac{\alpha(1-\alpha)^2 p}{2-\alpha}} \tag{6.33}$$

for $p > 1$, $0 \le \alpha \le 1$.

For $p > 1.29$ the numerator of the right hand side of (6.33) can be shown, numerically, to be always positive. Moreover we show in *Appendix 6.B* that the denominator is always negative. Under those conditions the inequality holds and superior performance is theoretically expected by the Approximation method.

For $p \le 1.29$ the numerator becomes negative and all right hand side of the inequality becomes positive (as the denominator is still negative). In that case the right hand side of the inequality cannot take any values higher than 0.48 and comments made to the previous sub-section apply to this comparison as well.

The above results are valid for $\alpha \approx 0.145$. The analysis is now extended to cover more α smoothing constant values that could be potentially used by practitioners. A table similar to that developed in sub-section 6.8.1 is presented in order to identify the regions of superior performance for the Approximation method.

α *smoothing constant value*	p *cut-off value*	CV^2 *cut-off value*
0.05	1.16	0.49
0.10	1.24	0.49
0.15	1.29	0.48
0.20	1.33	0.47

Table 6.2. MSE EWMA – MSE Approximation, issue points

From table 6.2 it is clear that the p cut-off value is more sensitive to the smoothing constant value used, in comparison with the results presented in the previous sub-section. It is important to note that sensitivity of the criteria cut-off values is noticed for all the pair-wise comparisons that involve the EWMA method.

For all smoothing constant values that are likely to be applied in practice the Approximation method should always perform better than the EWMA method for any $p > 1.33$ unit time periods and/or $CV^2 > 0.49$.

Even more favourable results for the Approximation method will be obtained, when the two methods' performance is compared on all rather than issue estimates only (see sub-section 6.8.4).

The Approximation method emerges as a promising method to deal with intermittence. The bias implicitly incorporated in the estimates produced by this method is very small and so is the variance of those estimates. For large p values and/or very low α values the method is approximately unbiased. It becomes evident though that even when those conditions cannot be justified the method's performance is not severely affected.

6.8.3 MSE EWMA – MSE Croston's method, issue points

$MSE_{EWMA} > MSE_{CROSTON}$ if and only if (see *Appendix 6.C*):

$$\frac{\sigma^2}{\mu^2} > \frac{\alpha(2-\alpha)\,p^3 + \left(1-4\alpha+2\alpha^2\right)p^2 - \frac{2}{2-\alpha}\,p + \frac{\alpha}{2-\alpha}}{\frac{\alpha}{2-\alpha}\,p - p^2 - \alpha(2-\alpha)\,p^3} \tag{6.34}$$

for $p > 1$, $0 \le \alpha \le 1$.

It can be shown numerically that for $\alpha \approx 0.145$ and $p > 1.26$ the numerator of the right hand side of (6.34) is always positive and as the denominator is always negative, superior performance is theoretically expected by Croston's method.

This is a very interesting result if we consider the fact that Johnston and Boylan (1996) found that Croston's method performs better than EWMA for average inter-order intervals greater than 1.25 review periods.

In that paper Johnston and Boylan compared Croston's method with EWMA on theoretically generated data under a wide range of simulated conditions (α value, average inter-order interval, lead time length, demand sizes distribution, all and issue points in time). The inter-order intervals were assumed to follow the negative exponential or Erlang (n) distribution. Based on MSE results the researchers concluded that Croston's method performs better than EWMA when the average inter-order interval is in excess of 1.25 forecast review periods.

This evidence shows that, when issue points in time are considered, the p cut-off value is not sensitive to the probability distribution of inter-order intervals.

For $p \le 1.26$ the numerator becomes negative and the right hand side of the inequality becomes positive. In that case the right hand side of the inequality cannot take any values higher than 0.21 which means that for $\alpha \approx 0.145$, if $CV^2 > 0.21$, Croston's method still performs better than EWMA.

A table similar to that presented in previous sub-sections is developed here in order to indicate the average inter-demand interval and squared coefficient of variation cut-off values for possible smoothing constant values in the realistic range 0.05 – 0.2.

α smoothing constant value	p cut-off value	CV^2 cut-off value
0.05	1.10	0.08
0.10	1.19	0.17
0.15	1.27	0.23
0.20	1.34	0.28

Table 6.3. MSE EWMA – MSE Croston, issue points

For issue points in time, Croston's method is expected to perform better than EWMA for $p > 1.34$ and/or $CV^2 > 0.28$. The validity of this rule is further discussed in the following chapter.

6.8.4 MSE EWMA – MSE Approximation method, all points in time

$MSE_{EWMA} > MSE_{APPROXIMATION}$ if and only if (see *Appendix 6.D*):

$$\frac{\sigma^2}{\mu^2} > \frac{p(p-1)\left[4p-(2-\alpha)^2\right]-\alpha(2-\alpha)}{(2-\alpha)\left(2p^2-\alpha p\right)-4p^3} \tag{6.35}$$

for $p > 1$, $0 \le \alpha \le 1$.

For $p > 1.17$ the numerator is always positive and as the denominator of the right hand side of the inequality is always negative (see *Appendix 6.D*), the inequality holds (superior performance is theoretically expected by the Approximation method) for all p values greater than 1.17.

For $p \le 1.17$ the numerator becomes negative and all the second part of the inequality becomes positive (as the denominator is still negative). In that case the second part of the inequality cannot take any values higher than 0.48 and all the aforementioned comments apply in this section as well. Those results are valid for $\alpha \approx 0.145$.

α *smoothing constant value*	p *cut-off value*	CV^2 *cut-off value*
0.05	1.12	0.49
0.10	1.15	0.48
0.15	1.18	0.48
0.20	1.20	0.47

Table 6.4. MSE EWMA – MSE Approximation, all points in time

From table 6.4 we can conclude that for smoothing constant values that are commonly applied in practice the Approximation method should always perform better than the EWMA method for any $p > 1.20$ unit time periods and/or $CV^2 > 0.49$.

For issue points only, the Approximation method was found to perform better than EWMA for any $p > 1.33$ and $CV^2 > 0.49$. This is an interesting result if we take into account the fact that EWMA is theoretically unbiased when all points in time are considered and therefore one may expect an improved performance of this estimation procedure in that context of application. The sampling error of the mean though, for issue exponentially smoothed estimates, is always lower than the error produced when all points in time are considered. The relationship between that difference and the bias of the method depends on all the control parameters.

6.8.5 MSE EWMA – MSE Croston' s method, all points in time

$MSE_{EWMA} > MSE_{CROSTON}$ if and only if (see *Appendix 6.E*):

$$\frac{\sigma^2}{\mu^2} > \frac{1-p}{p} \tag{6.36}$$

But: $\frac{\sigma^2}{\mu^2} > \frac{1-p}{p}$ for $p > 1$, $0 \le \alpha \le 1$.

When all points in time are considered, Croston's method is always expected to perform more accurately than EWMA. If we take into account the results of sub-section 6.8.3 it becomes obvious that there is a comparative improvement in Croston's method performance when all points in time are considered. The theoretical explanation of this result is the deterioration of EWMA's performance rather than a true improvement in Croston's method performance. This result though is not consistent with the findings of Johnston and Boylan (1996). They concluded that independently of which points in time are considered, Croston's method shows a superior performance to EWMA for inter-order intervals greater than 1.25 unit time periods. This is a matter that requires further examination but is beyond the scope of this thesis.

This results of the last two pair-wise comparisons indicate that if a re-order interval inventory control system is utilised then one of the estimation procedures specifically designed to deal with intermittence could be utilised for all SKUs. With the results obtained so far that could be either Croston's method or the Approximation method. For fast demand items the Approximation method's performance is in general though inferior to that of Croston's method. The opposite is the case when more intermittent and/or more irregular demand patterns are considered. This issue is further discussed in the following chapter.

6.8.6 MSE EWMA – MSE λ Approximation method, issue points

We conjecture that $MSE_{EWMA} > MSE_{\lambda APPROXIMATION}$ if and only if (see *Appendix 6.F*):

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[\frac{1}{2-\alpha}+\alpha(p-2)-\frac{(2-\alpha)p^3}{4\left(p-\frac{\alpha}{2}\right)^4}\right]}{p\left[\frac{(2-\alpha)p}{4\left(p-\frac{\alpha}{2}\right)^2}+\frac{\alpha}{4}\frac{p^2(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-\alpha p-\frac{\beta^2}{2-\alpha}\right]} \tag{6.37}$$

for $1 < p \leq 10$, $0 \leq \alpha \leq 1$.

This result has been illustrated graphically but not proven mathematically. However, since no counter-examples have been found in the stated ranges of p and α, we will proceed on the assumption that the inequality is correct.

It can be illustrated numerically that for $0.05 \le \alpha \le 0.2$ and $p > 1.40$ the numerator of the right hand side of (6.37) is always positive. We also illustrate in *Appendix 6.F* that the denominator is always negative for $1 < p \le 10$, $0 \le \alpha \le 1$. Under these conditions the inequality holds and therefore superior performance is theoretically expected by the λ Approximation method for $1.40 < p \le 10$.

The CV^2 cut-off value (for $0.05 \le \alpha \le 0.2$) is 0.40. Similar results were obtained when Croston's method and the Approximation method were compared to EWMA for issue points only.

6.8.7 MSE EWMA – MSE λ Approximation method, all points in time

We conjecture that $MSE_{EWMA} > MSE_{\lambda APPROXIMATION}$ if and only if (see *Appendix 6.G*):

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[\dfrac{1}{p^2} - \dfrac{\left(1-\dfrac{\alpha}{2}\right)^2 p}{\left(p-\dfrac{\alpha}{2}\right)^4}\right]}{\dfrac{\left(1-\dfrac{\alpha}{2}\right)^2}{\left(p-\dfrac{\alpha}{2}\right)^2}\left[1+\dfrac{1}{\left(p-\dfrac{\alpha}{2}\right)^2}\dfrac{\alpha}{2-\alpha}\,p(p-1)\right]-\dfrac{1}{p}} \tag{6.38}$$

for $1 < p \le 10$, $0 \le \alpha \le 1$.

This result has been illustrated graphically but not proven mathematically. However, since no counter-examples have been found in the stated ranges of p and α, we will proceed on the assumption that the inequality is correct.

In *Appendix 6.G* we illustrate that the denominator of the right hand side of the inequality is always negative for $1 < p \le 10$, $0 \le \alpha \le 1$. For $0.05 \le \alpha \le 0.2$ the p and CV^2 cut-off values for this pair-wise comparison are 1.17 and 0.22 respectively. When all points in time are considered, EWMA may perform better than the λ Approximation method only when both decision criteria take a value below their corresponding cut-off point.

In the previous sub-section, where the λ Approximation method was compared with EWMA for issue points only, the average inter-demand interval and squared coefficient of variation cut-off values were found to be 1.40 and 0.40 respectively. Taking into account these results it is important to note that the comparative improvement of the λ Approximation is not as marked as it was for Croston's method but is more substantial than that observed in the Approximation method.

6.8.8 MSE Croston's method – MSE λ Approximation method

We conjecture that $MSE_{CROSTON} > MSE_{\lambda APPROXIMATION}$ if and only if (see *Appendix 6.H*):

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[p + \frac{\alpha}{2-\alpha}(p-1) - \frac{\left(1-\frac{\alpha}{2}\right)^2 p^5}{\left(p-\frac{\alpha}{2}\right)^4}\right]}{\frac{\left(1-\frac{\alpha}{2}\right)^2 p^4}{\left(p-\frac{\alpha}{2}\right)^2} + \frac{\alpha}{2-\alpha} p(p-1)\left[\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^4} p^2 - 1\right] - p^2} \tag{6.39}$$

for $1 < p \le 10$, $0 < \alpha \le 1$ (for $\alpha = 0$ the denominator of the right hand side of the inequality is always zero).

This result has been illustrated graphically but not proven mathematically. However, since no counter-examples have been found in the stated ranges of p and α, we will proceed on the assumption that the inequality is correct.

In *Appendix 6.H* we illustrate that the denominator of the right hand side of (6.39) is negative for $1 < p \le 10$ and $0 < \alpha \le 1$. Moreover it can be illustrated, numerically, that for $0.05 \le \alpha \le 0.2$ and $p > 1.65$ the numerator of the right hand side of (6.39) is always positive. Therefore the λ Approximation method is theoretically expected to perform better than Croston's method for any $1.65 < p \le 10$.

For $p \le 1.65$ the squared coefficient of variation cut-off value is 1.17.

A consistently higher variance of the Croston's method estimates coupled with the bias incorporated in those estimates result in the superior performance of the λ Approximation method.

Nevertheless, note the very high cut-off value of both decision criteria. So far the squared coefficient of variation and average inter-demand interval cut-off values have not exceeded 0.49 and 1.40 respectively in any of the pair-wise comparisons. In the next sub-section we show that the Approximation method compares against Croston's method much more favourably than the λ Approximation method does.

6.8.9 MSE Approximation method – MSE λ Approximation method

$MSE_{APPROXIMATION} > MSE_{\lambda APPROXIMATION}$ if and only if (see *Appendix 6.I*):

$$\frac{\sigma^2}{\mu^2} < \frac{\dfrac{\alpha^2}{4} - \dfrac{\alpha(2-\alpha)}{4} p(p-1)\left(\dfrac{p^4}{\left(p-\dfrac{\alpha}{2}\right)^4} - 1\right)}{\dfrac{\alpha(2-\alpha)}{4}\left\{\dfrac{p^4}{\left(p-\dfrac{\alpha}{2}\right)^2} + p(p-1)\dfrac{\alpha}{2-\alpha}\left[\dfrac{p^4}{\left(p-\dfrac{\alpha}{2}\right)^4} - 1\right] - p^2\right\}} \tag{6.40}$$

for $p > 1$, $0 \le \alpha \le 1$.

In *Appendix 6.I* we show that the denominator of the right hand side of (6.40) is positive for $p > 1$, $0 \le \alpha \le 1$. It can also be shown numerically that for $0.05 \le \alpha \le 0.2$ and $p > 1.25$ the numerator of the right hand side of the inequality becomes negative and as the denominator is always positive the inequality does not hold i.e. superior performance is theoretically expected by the Approximation method for all p values greater than 1.25.

For $p \le 1.25$ the right hand side of the inequality cannot take any values greater than 0.48. Therefore if $CV^2 > 0.48$ then the inequality is not valid and superior performance is still theoretically expected by the Approximation method.

The Approximation method is by definition approximately unbiased only for large average inter-demand interval values. If this is not the case, the bias associated with that method's application gives an advantage to the λ Approximation method even though the variance of the Approximation method's estimates is comparatively lower.

For inter-demand intervals greater than 1.25 the bias associated with the Approximation method estimates is almost negligible. Moreover, as has been noted in earlier comparisons, the variance of these estimates is very low. The variance of the Approximation method estimates is so low that the method turns out to perform even better than the λ Approximation method, from which it was derived.

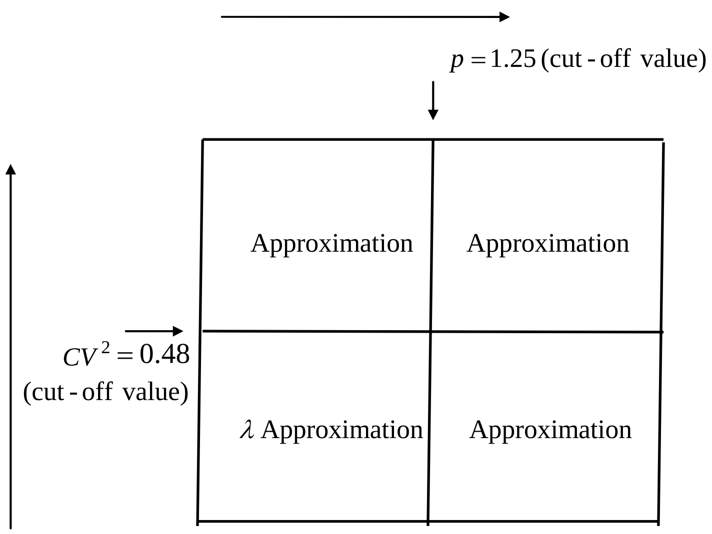

Figure 6.4. Decision areas (λ Approximation – Approximation method)

6.9 Summary results

In chapter 3 a theoretical framework was presented in order to facilitate a conceptual categorisation of alternative non-normal demand patterns and subsequently enable their formal definition. The definitions developed in chapter 3 were as follows:

- *Intermittent demand* appears at random with some time periods having no demand
- *Erratic demand* is (highly) variable. Erraticness relates to the demand size rather than demand per unit time period
- *Lumpy demand* appears at random with some time periods having no demand. Moreover demand, when it occurs, is (highly) variable.
- *Slow demand* is characterised by the low volume of demand per unit time period. This may be due to infrequent demand arrivals, low average demand sizes or both.

If we denote by x the average inter-demand interval cut-off value obtained in any of the pair-wise comparisons and by y the corresponding squared coefficient of variation cut-off value, each of the categorisation schemes developed in this chapter has the following characteristics:

The "$p \leq x$, $CV^2 \leq y$" condition tests for SKUs which are not very intermittent and erratic (i.e. faster moving products or products whose demand pattern does not raise any significant forecasting and inventory control difficulties).

The "$p > x$, $CV^2 \leq y$" condition tests for low demand items or intermittent demand patterns with constant, or more generally, no highly variable demand sizes (i.e. not very erratic).

The "$p > x$, $CV^2 > y$" condition tests for lumpy demand items, and finally

the "$p \leq x$, $CV^2 > y$" condition tests for erratic (irregular) demand items with rather frequent demand occurrences (i.e. not very intermittent).

The four resulting demand categories are graphically presented in figure 6.5.

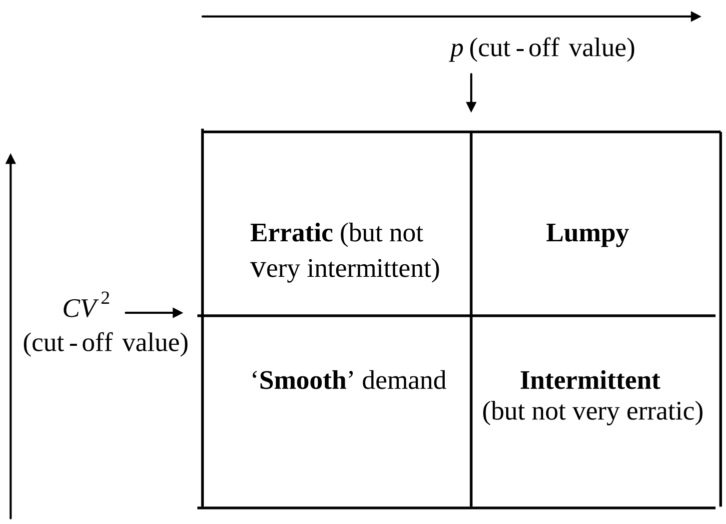

Figure 6.5. Categorisation of demand patterns

So far we have generated accuracy comparison results that indicate which estimation procedure performs better for each of the four demand categories. The results though have been developed by considering only two methods at a time. Therefore, it is necessary to extend the analysis conducted so far in order to propose rules which are valid across all the methods considered in this chapter.

The average inter-demand interval and squared coefficient of variation cut-off values that have been obtained in all the pair-wise comparisons are indicated below. The method that appears in italics is the one that theoretically performs better only in decision area 3.

Pair-wise comparison	p *cut-off value*	CV^2 *cut-off value*
Approximation-*Croston*	1.32	0.49
Approximation- λ *Approx.*	1.25	0.48
λ Approx.-*Croston*	1.65	1.17
ISSUE POINTS		
Approximation-*EWMA*	1.33	0.49
Croston-*EWMA*	1.34	0.28
λ Approx.-*EWMA*	1.40	0.40
ALL POINTS IN TIME		
Approximation-*EWMA*	1.20	0.49
Croston-*EWMA*	Croston always performs better than EWMA	
λ Approx.-*EWMA*	1.17	0.22

Table 6.5. Cut-off values ($0.05 \le \alpha \le 0.2$**)**

When issue points in time only are considered (i.e. in the context of a re-order level inventory control system) the Approximation method performs better than all the other methods for $p > 1.33$ unit time periods and/or $CV^2 > 0.49$.

For $p \le 1.33$ and $CV^2 \le 0.49$ EWMA and Croston's method are theoretically expected to perform best. In particular:

EWMA performs better than all the other methods considered, for average inter-demand intervals less than or equal to 1.33 and a squared coefficient of variation taking values in the range [0 – 0.28]

Croston's method is expected to show superior performance for low average inter-demand intervals (less than or equal to 1.33) and a moderate squared coefficient of variation (0.28 – 0.49].

Note that the λ Approximation method does not perform better than all the other methods in any of the decision areas.

The categorisation of demand patterns, in case that issue point estimates only are considered, takes the final form, which is indicated below:

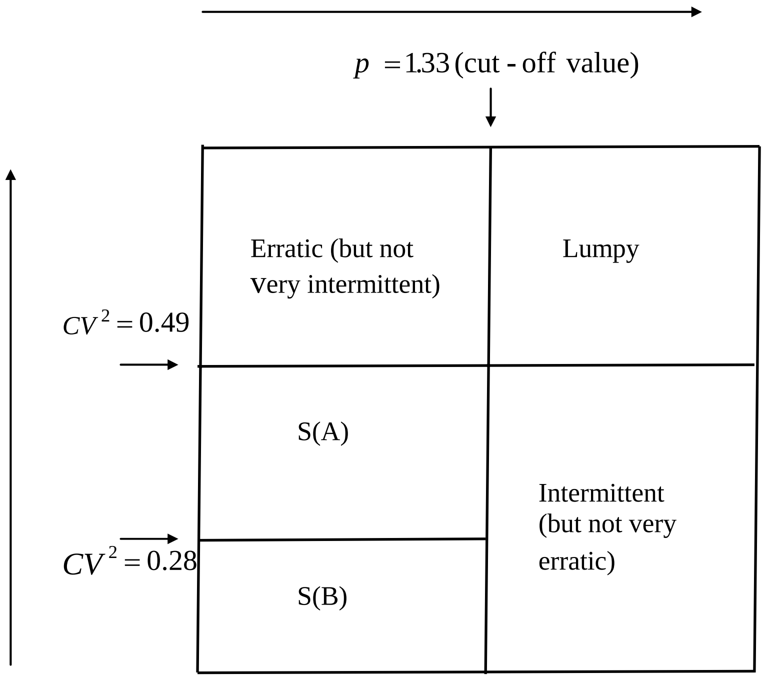

Figure 6.6. Categorisation of demand patterns (re-order level systems)

The ‘Smooth’ demand decision area has now been divided into two sub-areas: Smooth A, S(A) and Smooth B, S(B), in order to account for the theoretical differences in the forecasting accuracy performance observed during our analysis.

The recommended estimation procedures are as follows:

Erratic:	Approximation method
Lumpy:	Approximation method
Intermittent:	Approximation method
S(A):	Croston’s method
S(B):	EWMA

When all points in time only are considered (i.e. in the context of a re-order interval inventory control system) the Approximation method performs better than all the other methods for $p > 1.32$ unit time periods and/or $CV^2 > 0.49$.

For $p \le 1.32$ and $CV^2 \le 0.49$, Croston's method is theoretically expected to perform better than all the other methods. This result is, at least intuitively, not what one might expect. Based on the analysis conducted so far, this result is attributed to the large variance associated with the estimates produced by EWMA. The discussion on the poor EWMA performance, when all points in time are considered, is continued in the following chapter.

The categorisation of demand patterns, in the case that all point estimates are considered, takes the final form indicated below. Note that once more the λ Approximation method does not perform better than all the other methods in any of the decision areas.

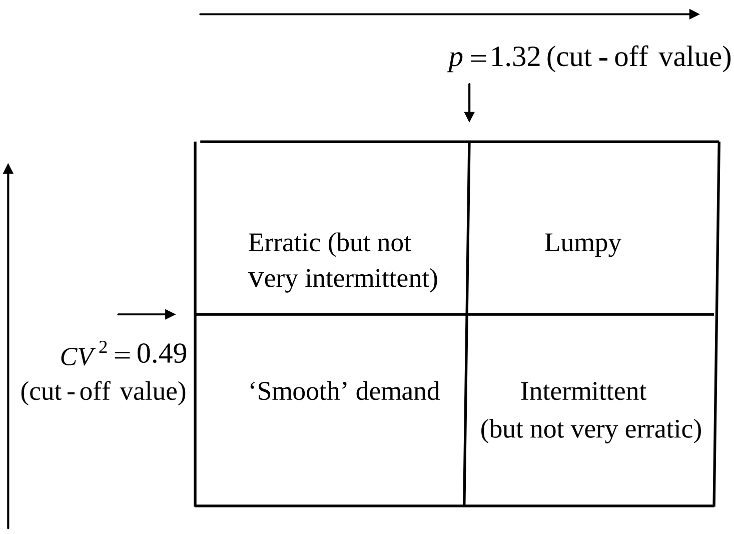

Figure 6.7. Categorisation of demand patterns (re-order interval systems)

The recommended estimation procedures are as follows:

Erratic:	Approximation method
Lumpy:	Approximation method
Intermittent:	Approximation method
Smooth:	Croston's method

One of the most interesting results of the analysis conducted so far is the poor performance of the λ Approximation method, even when it is compared against the Approximation method, considering that the latter estimation procedure was derived in chapter 4 only as a special case of the former. The λ Approximation is theoretically unbiased but the relatively large sampling error variance of the mean does not allow this estimation procedure to perform as well as expected. In fact the variance of the estimates appears, in this chapter, to play a more important role than the bias in determining overall accuracy performance results.

The λ Approximation method is outperformed by the Approximation method even for average inter-demand intervals as low as 1.25 review periods. For high average inter-demand intervals both methods are approximately unbiased. As the value of p reduces, the bias associated with the Approximation method increases but the variability of the estimates remains very low. On the contrary the very small (or zero) bias associated with the λ Approximation method remains the same but so does the large variance of the estimates. It is only for $p \leq 1.25$ review periods that the λ Approximation method may perform more accurately than the Approximation method. But in that region Croston's method and/or EWMA are expected to show superior performance. As such the λ Approximation method is not theoretically expected to perform better than all the other methods in any of the decision areas that can be formed in the categorisation matrices 6.6 and 6.7.

It is also important to note that when the λ Approximation method is compared with Croston's method, the resulting cut-off points are very high (the λ Approximation method is theoretically expected to show superior performance for $p > 1.65$ and/or squared coefficient of variation > 1.17). The categorisation rule in that case reduces effectively to $p > 1.65$ since squared coefficient of variation values greater than 1.17 can be expected (and they have been reported in the literature, Willemain et al, 1994) only for lumpy rather than just erratic demand patterns. But in that particular area ($p > 1.65$) the overall superiority of the Approximation method has been established.

When the λ Approximation method was developed, we aimed at eliminating the bias contained in Croston's estimates without considering the effect that such a bias elimination would have on the forecast variance. Unfortunately, the results presented thus far in this chapter indicate that the variance of the method estimates is relatively large, resulting in the relatively poor forecasting accuracy performance of this estimation procedure.

The purpose of this section is to identify the most accurate estimation procedure(s) for different combinations of the control parameter values. One of the main objectives of this Ph.D. as a whole is to introduce estimation procedures which, at least theoretically, outperform Croston's method. As mentioned above, the λ Approximation method is not expected to perform better than all the other methods in any of the possible factor combinations. The λ Approximation method will be considered in the following chapter (chapter 7) when its application on theoretical data will be simulated, for the purpose of verifying its poor forecasting accuracy performance. Consequently, upon verification of the poor performance of the λ Approximation method, and in order also to simplify the real data simulation experiment, to take place in chapters 10 and 11, this estimation procedure will be disregarded from chapter 8 onwards.

6.9.1 Modified Williams' criteria

The two categorisation schemes developed in this section (figures 6.6 and 6.7) are now assessed against the theoretical and practical requirements proposed by Williams (1984) as modified in sub-sections 3.4.2 and 3.4.3. To recap, the requirements are as follows:

1. The categorisation scheme should suggest in what different ways to treat the resulting categories. The objective in categorising demand patterns is the identification of the most appropriate forecasting and inventory control methods to be applied to the different demand categories. As such, categorisation schemes should explicitly suggest which methods should be used under which circumstances.
2. The criteria considered in developing the rules should be dimensionless so that categorisation decisions regarding a SKU are independent of the product's unit of measurement or of demand over any time period other than the lead time or the review period.

3. Sensitivity to outliers should be taken into account. The categorisation scheme should not allow products to move from one category to another when few extreme observations are recorded.
4. The amount of data required to reliably classify demand patterns should also be considered. That is, the decision rules should take into account the limited number of demand occurrences that characterise any intermittent demand pattern.
5. Logical inconsistencies should not allow demand for a SKU to be classified in an unintended category.
6. Determination of the cut-off values should be non-arbitrary thereby enabling the general applicability of the categorisation scheme.

The categorisation schemes clearly suggest which estimation procedure should be used. In fact, as it has already been shown, categorisation of the demand patterns results from a structured forecasting accuracy comparison between alternative estimation procedures. The categorisation rules do not indicate what type of inventory control approach should be taken but rather a set of categorisation rules can be chosen depending on whether the inventory system is re-order level or re-order interval. In this research we are exploring the possibility of improving intermittent demand forecasting practices assuming that an appropriate inventory control system is already in place. This is based on the assumption that there are no significant interactions between estimation procedures and inventory control models (so that the combination of estimator A and inventory model *x* performs better than the combination of estimator B and the same model *x*, even though the estimator A is in fact the same or less accurate than estimator B). Sani (1995) found no evidence of such interaction.

With respect to the second requirement developed by Williams (1984) both decision criteria are clearly dimensionless:

$$p: \textit{no. of review periods}$$

$$\frac{\sigma^2}{\mu^2}: \frac{units^2}{units^2}$$

Even though the categorisation rules have been derived based on a theoretical analysis of the MSE, their robustness on other accuracy measures will also be tested, in chapter 10. The MSE is known to be "unduly" influenced by outliers whereas outliers do not affect other error measures that will be used in our real data simulation experiment for generating results. The sensitivity of the categorisation rules to extreme observations is further discussed in chapters 10 and 11. It is important to note that Williams (1984) proposed using "buffer zones" so that "borderline" SKUs would not switch categories as the parameters vary from one side of the border to the other or categorisation would not be that easily affected by outliers. Nevertheless an "ongoing" application of the categorisation scheme (at the end of every review period re-categorisation occurs and the appropriate estimation procedure is chosen) will not be considered in chapters 10 and 11. Rather the properties of each demand data series will be used in order to define demand for the corresponding SKU, the best estimation procedure (considering forecasting and/or inventory control performance) will be recorded and the results will be checked for consistency against the categorisation schemes developed in this chapter.

As far as Williams' fourth criterion is concerned, there is an absolute minimum requirement of two demand occurrences before the categorisation rules can be applied in practice assuming that the first recorded zero or non-zero demand follows an issue point. If we start from a random point in time, then the minimum requirement increases to three demand occurring periods so that the average inter-demand interval can be determined. At this stage it is worth noting that even though two (or three) demand occurrences allow the application of the categorisation scheme, the sampling error of the demand size variance estimates will be very large if there are few observations.

The categorisation schemes developed in this section explicitly consider both the demand pattern and size of demand, when demand occurs. That characteristic eliminates any eventual logical inconsistencies and does not allow SKUs to be classified in a category other than the intended one. Finally the cut-off values assigned to the decision criteria are the product of structured MSE pair-wise comparisons and as such the categorisation schemes proposed are generally applicable.

6.10 Conclusions

In this chapter the corrected Croston's method, EWMA and the two estimation procedures that have been developed during this research are compared based on their lead time MSE performance. The MSE is similar to the statistical measure of variance of the forecast errors but not quite the same since bias is also taken into account. MSE is a mathematically tractable accuracy measure and allows significant theoretical insight to be gained.

Traditionally, the cumulative lead time MSE is taken as the sum of the MSEs of the individual forecast intervals. By estimating the demand variance in that way, we neglect any correlation between the estimates of demand. This correlation is generated by the sampling error of the mean and the bias (if any) that is carried forward from one period to the next. The correct expression of the MSE, over a fixed lead time, for biased and unbiased estimation procedures, has been derived in this chapter under the stationary mean model assumption.

Consequently, using results generated in chapters 4 and 5 and derivations given by Croston (1972) we quantify the MSE of all the estimation procedures discussed in this thesis. Pair-wise lead time MSE comparisons are then conducted for the purpose of deriving theoretical rules that indicate under what conditions one method is theoretically expected to perform better than another. These theoretical rules are based on the squared coefficient of variation and the average inter-demand interval of the intermittent demand series. The cut-off values assigned to both criteria are the outcome of a numerical analysis conducted on the theoretical results. Having obtained the cut-off values we then specify regions of superior performance of one method over the other. When all pair-wise comparison results have been collected we then extend the analysis in order to indicate overall superiority, specifying under what conditions one method is theoretically expected to perform more accurately than all the other methods.

Finally two categorisation schemes are developed: one referring to a re-order level and the other to a re-order interval situation. The results indicate that the Approximation method performs better than all the other estimation procedures for non-smooth demand patterns. For the smooth demand category, EWMA and Croston's method are found to perform best when issue points in time only are considered and Croston's method is the only recommended procedure when all point estimates are taken into account. The λ Approximation method does not outperform all the other estimation procedures in any of the decision areas.

CHAPTER 7

A Simulation Experiment Using Theoretically Generated Data

7.1 Introduction

In this chapter a simulation experiment using theoretically generated data will be developed for the purpose of assessing the accuracy of all the approximated results derived during this research.

The use of simulation in extending Operational Research and Management Science theory has been widely recognised in the academic literature (e.g. Meier et al, 1969; Ignall and Kolesar, 1979; Law and Kelton, 2000). Moreover issues related to the symbiosis of simulation and mathematical modelling have also attracted the attention of many researchers (e.g. Adam and Dogramaci, 1979; Santhikumar and Sargent, 1983).

Even though the exact use of a simulation model developed along with an analytic model[1] depends on the context of application, a simulation model often aims at representing elements and relationships of the system under study that make the analytic model mathematically intractable. In the case of this research, although we have only dealt with mathematically tractable problems, the analysis has resulted in solutions of an approximate nature. Therefore, the simulation model developed in this chapter is intended to assess the accuracy of those approximate mathematical results rather than representing a part of the system that cannot be modelled otherwise.

[1] The terms "analytic" and "mathematical" are used interchangeably for the purpose of this research in contrast with other academic work in which both analytic and simulation models/results are described as mathematical models/results (e.g. Bakir, 1994).

In chapter 4 the issue of bias in intermittent demand forecasting was analysed while in chapter 5 we derived the sampling error variance of the estimation procedures considered in this thesis. In both cases the final results were approximations[2].

Subsequently, in chapter 6, the MSE associated with alternative forecasting methods was obtained. The MSE results were used for establishing regions of superior performance of every estimator over one other and over all other estimators. The MSE results are of an approximate nature and, as such, the decision rules developed are also approximate. Moreover certain simplifying assumptions were made in chapter 6 in order to enable the development of meaningful categorisation schemes. Therefore, it is important to know the extent to which those assumptions can affect the validity of the rules proposed.

By simulating the performance of the alternative estimation procedures on theoretical data, which has been generated upon the assumptions considered in the mathematical analysis, we obtain bias, variance and MSE results which are close to the population expected values but not quite the same since sampling variation exists. This sampling variation can be reduced if we consider a sufficiently large number of simulated demand time periods (say 20,000). Further reduction of the sampling variation can be achieved if the simulation model runs, for each specific set of conditions, more than once (say 5 times) so that average results can be obtained[3]. Those results can then be compared against the mathematical results and conclusions about the accuracy of the approximations can be drawn.

The simulation model will cover a wide range of possible situations. It will be based on the assumption that demand occurs as a Bernoulli process and therefore the geometric distribution represents the inter-demand intervals. Many distributions have been proposed in the academic literature for representing the demand sizes, when demand occurs. Computational considerations dictate that only one such distribution can be selected for the purpose of our simulation experiment.

---

[2] The only exact results that we have discussed so far refer to EWMA, in which case the results were derived by Croston (1972). The only EWMA result derived in this thesis has been the lead time MSE (for all and issue points in time) which was an exact result as well.

[3] Equivalently, the model could also run just once considering, say, 100,000 demand time periods. Nevertheless, this approach cannot be further considered because of the current Excel limitations.

Consequently, issues related to the sensitivity of the results to the assumed demand size distribution will not be addressed in this thesis. In this chapter we will argue that the lognormal distribution is the most appropriate for the purpose of our simulation exercise because, amongst other reasons, it allows the coefficient of variation (or squared coefficient of variation) to take values above unity.

The control parameters of the simulation experiment will be: the average inter-demand interval, the mean and variance of the size of demand, the smoothing constant value and the length of the lead time. Results will be generated for both all and issue points in time only.

7.1.1 The mathematical-simulation model relationship

Before using the simulation results for any possible application, the simulation model needs first to be verified and validated. Verification is to be distinguished from validation since the former refers to internal consistency or correct logic of the model while the latter to the degree that the model corresponds to the real system. Exact analytic results can be used in the verification of simulation models. In that case the control parameters of the simulation model can be set to correspond with the analytic model so that we can judge the correctness of the simulation's model logic by comparing the output of the simulation model to the exact theoretical results of the analytic model. Therefore the analytic results are used in order to check the correctness of the simulation results. Once the correct structure of the simulation experiment has been confirmed, the simulation results (other than those that have been verified by referring to the analytic model) can be used to assess the accuracy of any approximate analytic results. This is precisely the way that we are going to work in this chapter.

The simulation model will be developed based on the main assumptions that have been considered in this thesis so far. Consequently exact theoretical results (and in particular Croston's derivations regarding the EWMA bias) will be used in order to verify the simulation model. Once the model has been verified, the simulation results can then be used to assess the accuracy of the corresponding approximate analytic results. If the simulation model does not exhibit the properties that hold for the analytic model we have reasons to question the accuracy of the approximation under concern.

Fortunately, the simulation results indicate good accuracy of our theoretical approximations. The empirical validity and utility of our theoretical results will be tested in chapters 10 and 11.

7.1.2 Chapter structure

A simulation model is a dynamic or an operating model of a system or problem entity that "mimics" the operating behaviour of the system or problem entity and contains its functional relationships (Santhikumar and Sargent, 1983).

A number of logical systematic processes have been designed to support the development of a simulation model (e.g. Watson and Blackstone, 1989; Law and Kelton, 2000). One such process was developed by Hoover and Perry (1989) and will guide the structure of this chapter:

Formulating the problem and planning the study. Before beginning to develop a simulation model, it is very important to have well defined objectives and to decide how the model will be used in the decision making process. As far as our simulation experiment is concerned, both issues have already been discussed in the beginning of this chapter's introductory section.

Collecting the data. To build the conceptual model the researcher should have sufficient data in order to develop the mathematical and logical relationships in the model so that it adequately represents the problem entity for its intended use. If for any possible reason real data is not used, theoretical data has to be generated. A literature overview on theoretically generated data simulation experiments is presented in section 7.2. For the simulation experiment described in this chapter demand will be built from constituent elements (demand sizes and inter-demand intervals). The assumed incidence and demand size distributions are discussed in section 7.3 while the actual generation process is described in *Appendix 7.B*.

Validation. Since any model is only an abstraction of the real system under study, the analyst should always retain a healthy scepticism about how the model represents the system. Validation is a confirmation that a model is a credible representation of the system. All the assumptions made during our mathematical analysis are still employed for the purpose of this simulation exercise and the validity associated with them has already been discussed in the previous chapters. In this chapter we will argue that demand sizes should be represented by the lognormal distribution and the validity of this new assumption will be explored in section 7.3.

Construction of the conceptual model. Technical details about our simulation model are given in section 7.4.

Verification. Verification is the process of determining whether the computer coding of the model corresponds to the model logic or not. The issue of verification is discussed in sub-section 7.6.1.

Determining the control parameters of the simulation. In section 7.5 the control parameters and the values assigned to them are presented. The selection of the particular values is explained in *Appendix 7.C*.

Analysing the output data. The results of our simulation experiment are analysed in sections 7.6, 7.7 and 7.8 where we directly compare them with our theoretical approximations. Finally the conclusions of the chapter are presented in section 7.9.

7.2 Simulation using theoretical data

7.2.1 Literature overview

Croston (1972) used theoretically generated data in order to demonstrate the improved forecasting accuracy and consequent inventory control savings of employing his method over traditional EWMA when dealing with intermittence. Croston's assumptions used for generating demand data were the same as those employed in developing his method. The control parameters used in his simulation experiment were the following:

- The mean of the geometrically distributed inter-demand intervals
- The mean and variance of the normally distributed demand sizes
- The smoothing constant value.

Results were generated for one step ahead forecasts considering the estimates made immediately after a demand occurrence.

Bretschneider (1986) compared the MAD and MSE smoothing approaches in estimating the variance of the forecast error for the simple exponential smoothing forecasting method. The simulation was conducted on theoretically generated demand data using three types of underlying structure for both demand mean and variance: constant, discrete step changes and ramp changes in level. The control parameters were the mean and standard deviation of the demand distribution.

Willemain et al (1994) found correlations and distributions in real world data that violated Croston's assumptions. Therefore, they compared EWMA and Croston's method under less idealised conditions. Theoretical demand data was generated for scenarios that violated Croston's assumptions. The control parameters used in the demand generation process were the following:

- The mean and standard deviation of the demand size distribution. Demand sizes were assumed to follow the lognormal distribution
- The average inter-demand interval. Inter-demand intervals were assumed to follow the geometric distribution
- Demand size and inter-demand interval auto-correlation coefficients
- Size-interval cross-correlation factors.

In addition industrial data was considered in order to test the validity of the results. The smoothing constant value was introduced as a control parameter for both sets of data and results were generated for the one step ahead estimates produced by the two methods. (Following Croston's scenario, the issue point estimates were considered only for the theoretical set of data. For the real demand data series, the accuracy comparison results were based on all point estimates.)

Croston's method was found to perform better than EWMA and the researchers concluded that this method can indeed provide tangible benefits to manufacturers forecasting intermittent demand.

Sani (1995) conducted research with the purpose of identifying the most appropriate inventory control policy and forecasting method when low demand items are considered. Selected inventory control methods and demand estimation procedures, presented in the academic literature or used by practitioners, were compared on both theoretically generated and real demand data. The real data files represented the demand history of 30 car spare parts and 54 agricultural machinery spare parts. The parameter settings used in order to derive the theoretical data was only a part of the factorial grid (288 parameter settings) used in deriving the Power Approximation method (Ehrhardt, 1979). In particular the control parameters used in the demand generation process were the following:

- Distribution of demand per period. Demand was not generated from constituent elements. It was rather assumed that demand per period can be reasonably represented by the Negative Binomial or Poisson distribution
- Variance to mean ratio
- Mean demand per period.

The lead time was also introduced as a control parameter. The smoothing constant value used for EWMA and Croston's method was set to 0.15 and all the points in time were considered for generating results.

Additional control parameters that were introduced in order to obtain inventory control results were:

- The ordering cost
- The unit shortage cost
- The unit holding cost.

Sani argued that even though simulation based on real data increases the credibility of the final results, real data may not cover all the situations that can possibly exist.

Moreover real world data may not be always available or, in case of availability, may not be sufficient for carrying out the particular simulation exercise.

Nevertheless, although theoretically generated data may cover a wide range of possible situations, they give equal weight to all the situations explored. In addition, theoretical data take no account of other situations that may actually occur (e.g. there may be SKUs that do not follow the hypothesised distribution(s)). Therefore it is desirable, in order to reach any definitive conclusions, to develop a simulation model and run it on both theoretically generated and real demand data.

Johnston and Boylan (1996) compared Croston's method with EWMA assuming that demand occurs as a Poisson process. The possibility of a condensed Poisson demand generation process was also taken into account. The demand sizes were assumed to follow the negative exponential, Erlang or rectangular distribution. Apart from the inter-demand interval distribution (and the corresponding average values selected) and the distribution of demand sizes (with specified mean and variance values) other control parameters taken into account were the smoothing constant value, the length of the lead time and the points in time considered for generating results.

7.2.2 Design of simulation

The control parameters introduced in a theoretically generated data simulation experiment can be divided in two categories. The first set of control parameters refers to the demand generation process. The selection of appropriate control parameters at this stage depends on the assumptions upon which demand is modelled. In the case of our simulation experiment we want to model demand from constituent elements assuming stationarity of demand sizes and inter-demand intervals and size-interval independence. Specific distributions will be assumed for representing demand sizes and inter-demand intervals and meaningful control parameters to be introduced are the following:

1. Mean and variance of the demand sizes
2. Mean of the inter-demand intervals.

Considering the assumptions based on which the simulation model will be developed, control parameters that explore different auto-correlation and cross-correlation structures (as per Willemain et al, 1994) are not applicable in this simulation study. Similarly the generation of the demand data from constituent elements does not allow the experimentation with control parameters that consider the aggregate demand in a period and model how this moves or develops through time (as per Sani, 1995).

The second set of parameters refers to the process of generating results and depends on the objectives of the simulation experiment. In our case we want to generate only forecasting accuracy results (rather than forecasting accuracy and inventory control results). The objective of the simulation experiment is to check the accuracy of our analytical results in different contexts of application (re-order level or re-order interval systems), assessing the effect of the lead time and the sensitivity of our analytical results to different smoothing constant values. Therefore the additional control parameters are required to be the following:

3. Lead time
4. Smoothing constant value
5. Points in time considered.

7.3 The distribution of demand incidence and demand size

7.3.1 The demand incidence distribution

In an intermittent demand context, three demand generation processes have been presented in the literature. In table 7.1 we present the three demand processes along with the corresponding inter-demand interval distributions.

	Demand generation process	*Inter-demand interval distribution*
Discrete	Bernoulli	Geometric
Continuous	Poisson	Negative exponential
	Condensed Poisson	Erlang-2

Table 7.1. Demand generation processes and the corresponding distributions of inter-demand intervals

The Erlang-2 distribution may be derived by considering a "censored" Poisson process in which only every second event is recorded. Chatfield and Goodhardt (1973) named this distribution "condensed Poisson" because its variance is less than its mean.

In the past the exponential, geometric and Erlang-2 distributions have been extensively considered for representing the inter-purchase time distribution for individual consumers or households. The Inverse Gaussian distribution can also be considered. Theoretical arguments and empirical evidence in support of those distributions for representing the time between two consecutive purchasing occasions for individual customers or households are summarised in Boylan (1997).

Unfortunately, although there is quite extensive theory in support of the exponential or the geometric distribution for representing the time interval between total demand generated by all customers, the empirical evidence in support of these distributions is very limited.

Kwan (1991) conducted research with the purpose of identifying the theoretical distributions that best fit the empirical distributions of demand sizes, inter-demand intervals and demand per unit time period for low demand items. Statistical tests were performed on a number of distributions, using two sets of real world data which contained the demand history of 85 spare parts.

The results, reported at a significance level of 5%, showed that the Erlang-2 inter-demand distribution did not fit the demand histories of any of the 85 spare parts. The geometric distribution fitted 20% of her sample of SKUs. This compared unfavourably to the negative exponential distribution, which fitted 42% of the SKUs.

The geometric distribution was found to be a reasonable approximation to the distribution of inter-demand intervals, for real demand data, in Dunsmuir and Snyder (1989) and Willemain et al (1994).

Janssen (1998) developed a (T, s, Q) periodic review model (where T is the review period, s the re-order point and Q the order quantity) for the control of intermittent demand items assuming that demand per unit time period can be reasonably represented by a compound Bernoulli distribution. The Bernoulli demand generation process was tested on a set of empirical data obtained from a Dutch wholesaler of fasteners. In particular the sample variance of the inter-arrival times: $\hat{\sigma}_p^2$ was compared against the theoretical variance for the geometric distribution: $p(p-1)$ (where p, the average inter-demand interval, was estimated from the sample average: $\bar{p}$). This was done for three classes of items:

- C-class: one or more customer orders in two weeks
- D-class: one or more customer orders in one month
- E-class: one or more customer orders in one quarter

For each class a regression analysis, without intercept in the model, was performed, which resulted in:

- C-class: $\bar{p}(\bar{p}-1) = 0.934\,\hat{\sigma}_p^2$, adjusted $R^2 = 0.97$
- D-class: $\bar{p}(\bar{p}-1) = 0.941\,\hat{\sigma}_p^2$, adjusted $R^2 = 0.95$
- E-class: $\bar{p}(\bar{p}-1) = 0.966\,\hat{\sigma}_p^2$, adjusted $R^2 = 0.96$

The results indicated that the Bernoulli demand generation process indeed might be a reasonable approximation for intermittent demand processes.

The empirical evidence, such as it is, does not support the Erlang-2 distribution for representing the inter-demand intervals. There is some evidence to support both geometric and negative exponential distributions. In this thesis, the geometric distribution will be assumed, since this is consistent with Croston's model (discussed earlier in chapters 4 and 5).

7.3.2 The demand size distribution

Preliminary analysis conducted on the real data files, to be used in chapter 10 for the purpose of this research, indicated that not only the variance to mean ratio but also the demand size coefficient of variation may take values higher than one. The same was noted with the industrial data used in Willemain et al (1994), where the coefficient of variation for the demand sizes was reported to be in the range: 0.34 –1.84.

Therefore a realistic distribution for representing the demand sizes would be one that allows the coefficient of variation to vary considerably. This is not only a theoretical but a practical requirement also, since the squared coefficient of variation cut-off values developed in the previous chapter were not in all cases less than one. A part of the simulation experiment will be devoted to exploring the accuracy of all the decision rules derived in chapter 6 as well as the categorisation schemes developed in that chapter regarding all estimation procedures. Hence, a distribution that allows a wide range of the coefficient of variation values is an important practical consideration.

Croston (1972) suggested the use of the normal distribution for representing the demand sizes when demand occurs. No empirical evidence has been presented in the literature to support the use of the normal distribution for this purpose. Moreover, the generation of demand sizes based on the normal distribution necessitates the use of a very small variance as compared to the mean size of demand so that negative demand values are avoided. Therefore the normal distribution cannot be used for representing the demand sizes. In fact the normality assumption for intermittent demand sizes has never appeared in the academic literature since 1972 with the only exception being Croston (1974) where the issue of whether or not a low demand item should be kept in stock is discussed.

Since the normality assumption does not appear to be realistic, another distribution must be chosen for generating the demand sizes in our simulation experiment.

7.3.2.1 Three continuous demand size distributions

Johnston and Boylan (1996) simulated the demand sizes as following the negative exponential, Erlang or continuous rectangular distribution. By doing so, a wide range of possible alternative distribution shapes were taken into account, from monotonically decreasing functions to unimodal positively skewed distributions to more normal type curves and finally to uniform functions.

The three distributions are discussed in *Appendix 7.A* of the thesis. All three of them cover a wide range of variance to mean ratios but none of them can be further considered since they require the standard deviation to be equal or less than the mean.

7.3.2.2 Two discrete demand size distributions

Kwan (1991) argued that Croston’s normality assumption cannot be valid in the case of a low demand item. When demand is very low, the sizes of demand, when demand occurs, will be low integer values. According to Kwan, the assumption that those values fit a continuous distribution, normal, is not reasonable. Consequently the distributions that were evaluated for the purpose of her research were: truncated Poisson, truncated Pascal and constant demand sizes of one unit. The truncated Pascal distribution was found to be the best, fitting 58% of the SKUs considered in her sample. The truncated Poisson distribution fitted only 27% of those SKUs.

The truncated Poisson distribution requires the variance to be always equal to the mean. The truncated Pascal distribution requires the variance to exceed the mean. Both distributions do not allow the standard deviation to take any value above the mean (see *Appendix 7.A*).

7.3.2.3 The lognormal distribution

Willemain et al (1994) found evidence in real world data that suggested the use of the lognormal distribution for representing the demand sizes. The lognormal distribution is also used by Forecast Pro, XE, versions 3 and 4 (Business Forecast Systems, Inc., Users Manual, 1997) when Croston's method is applied to intermittent demand data.

Boylan (1997) argued that a demand distribution should be assessed based on three criteria:

- A priori grounds for modelling demand
- The flexibility of the distribution to represent different types of demand
- Empirical evidence.

To satisfy the first criterion, a distribution must be explainable in terms of an underlying mechanism. We clearly cannot provide an a priori justification for using the lognormal distribution and we also recognise that for discrete demand, this distribution is only an approximation to the distribution of demand. Nevertheless, considering the flexibility of the lognormal distribution (the squared coefficient of variation for the lognormal distribution is derived at the end of this sub-sub-section) and the empirical evidence that exists in its support we finally decide to use that distribution for representing the demand sizes.

It is important to note at this stage that the Gamma (rather than Erlang) and truncated Negative Binomial (rather than truncated Pascal) distributions are as flexible as the lognormal distribution. No evidence have been put forward in the academic literature in support of the truncated NBD for representing the size of demand. In contrast, the Gamma distribution has been highlighted in the academic literature in regard to the representation of demand sizes. For example, Williams (1982) developed approximations for the re-order point s in (s,Q) policies (where Q is the order quantity) assuming that gamma-distributed sized orders arrive stochastically with no more than one during the lead time. Dunsmuir and Snyder (1989) and Strijbosch et al (2000) developed intermittent demand inventory control models in which positive demand during the lead time was assumed to be gamma distributed. Nevertheless, although empirical evidence has been presented in support of the Gamma distribution for modelling demand per unit time period/lead time demand (e.g. Burgin and Wild, 1967), this is not so when demand sizes are considered.

The probability density function of the lognormal distribution is given by (7.1).

$$f(x) = \frac{1}{x\sigma_N (2\pi)^{1/2}} \exp\left\{ \frac{-[\log(x/m)]^2}{2\sigma_N^{\ 2}} \right\} = \frac{1}{x\sigma_N (2\pi)^{1/2}} \exp\left[\frac{-(\log x - \mu_N)^2}{2\sigma_N^{\ 2}} \right] \tag{7.1}$$

The range of the lognormal variate $L: m, \sigma_N$ or $L: \mu_N, \sigma_N$ is $[0,\infty)$.

The scale parameter $m > 0$ is the median. The alternative parameter μ_N is the mean (location parameter) of $\log L$.

m and μ_N are related by: $m = \exp(\mu_N)$ or $\mu_N = \log m$

The shape parameter $\sigma_N > 0$ is also the standard deviation (scale parameter) of $\log L$.

We denote by μ the mean in the lognormal distribution and by σ^2 the variance.

$$\mu = m \exp\left(\frac{1}{2} \sigma_N^{\ 2} \right) \tag{7.2}$$

$$\sigma^2 = m^2 \omega (\omega - 1) \tag{7.3}$$

where $\omega = \exp(\sigma_N^{\ 2})$

Consequently the squared coefficient of variation is calculated as follows:

$$CV^2 = \frac{m^2\omega(\omega-1)}{m^2\omega} = \omega - 1 \tag{7.4}$$

From (7.4) it follows that the CV^2 can take any value below or above unity.

We conclude that the lognormal distribution meets our requirements in terms of flexibility and empirical support and therefore it will be utilised for representing the demand sizes in our simulation experiment.

To summarise, in this section we discussed alternative distributions that potentially could be used for representing demand incidences and demand sizes. Considering the purpose of our simulation experiment and certain theoretical and practical requirements for selecting an appropriate distribution, demand will be modelled as a compound Bernoulli process, i.e. with a fixed probability that there is positive demand during a time unit, otherwise demand is zero. The sizes of demand, when demand occurs, will be assumed to be lognormally distributed. The algorithms that will be used for the purpose of generating the demand data series are presented in *Appendix 7.B.*

7.4 Technical details

The code needed for performing our simulation exercise was written in *Visual Basic* (embedded in *Microsoft Excel 97,* Windows 95). The control parameters and the specific values assigned to them are discussed in section 7.6. Willemain et al (1994) generated, for each of the scenarios (set of conditions) that they considered in their research, data series long enough to contain 1,000 pairs of demand sizes and intervals. They claimed that for series of that length, five replications would be sufficient in order to keep the sampling variation to an acceptable level.

In our simulation experiment, each simulation run will consist of 20,000 simulated demand time periods. As such we simulate 20,000/p actual demand occurrences, where p is the mean inter-arrival period. The largest value to be assigned to the p control parameter (see section 7.6) is 10. Therefore we generate series with at least 2,000 demand occurrences. Following Willemain et al (1994), for each specific set of conditions the simulation model will run 5 times. The values related to each of the simulation results of interest to us (e.g. the bias incorporated in Croston's estimates) will be recorded and consequently the average value will be obtained for each of those results. For example, for a specific set of conditions, the simulated Croston's bias is taken to be the average bias produced in the 5 simulation runs. This average value is then compared with the theoretically expected value in order to determine the accuracy of our approximation. Both the length of the simulation and the number of "replications" (simulation runs for a specific combination of the control parameter values) are expected to reduce the sampling variation considerably.

The initial estimate of all forecasting methods is taken to be the population expected value. The updating procedure for all the forecasting methods is given in chapter 4. The "transient interval" is set, in all simulation runs, to 100. That is, the first 100 simulated demand time periods are not considered for generating any results. As such we finally consider only $19{,}900\text{-}L+1$ simulated demand time periods (where L is the lead time considered in the particular simulation run).

7.5 The control parameters

As discussed in section 7.2 the control parameters to be used for our simulation exercise are the following:

- *Average inter-demand interval*. Assuming geometrically distributed inter-demand intervals, the expected inter-demand interval (p) is specified
- *Mean and variance of demand sizes.* Assuming lognormally distributed demand sizes the mean and variance are specified
- *Points in time considered.* Since EWMA's performance depends on which points in time are considered, we distinguish between all and issue points in time only
- *Lead time*
- *Smoothing constant value.*

The values assigned to all the control parameters are presented in the summary table 7.2. The selection of the particular values is explained in *Appendix 7.C.*

Factors	System parameters: Levels	Number of levels
Inter-demand interval distribution	Geometric (mean: p) 1.1, 1.3, 1.5, 1.7, 1.9, 2, 4, 6, 8, 10	10
Demand size distribution	Lognormal L (μ, σ) L (2, 3), L (10, 10), L (2, 1.5), L (10, 6), L (10, 3) L (2, 0.25), L (10, 0.25)	7
Points in time considered	All – Issue points in time	2
Lead time length	1, 3, 5, 12	4
α smoothing constant value	0.05, 0.1, 0.15, 0.2	4

Table 7.2. The simulated conditions

There are 2,240 possible combinations of the control parameter values but the model will run only for 1,120 sets of conditions since the issue points in time estimates do not need to be re-generated, i.e. when all points in time estimates have been produced the estimates made when demand occurs can be isolated without having to re-run the model. Five replications will be applied to each of the 1,120 unique combinations of the parameter values. Thus, the model will finally run 5,600 times.

7.6 Simulation results. The bias of intermittent demand estimates

In this section, the simulation results are analysed in terms of the bias associated with the application of intermittent demand estimators on theoretically generated data. No bias expressions for the EWMA method have been derived in this thesis. Nevertheless, the standard exact results (Croston, 1972), regarding this method's application, are also tested against the simulation results to verify our simulation model.

A possible way of checking the accuracy of our approximations is to consider the signed difference between the theoretically expected bias (EB) and the average bias obtained over five replications via simulation (SB), and express that as a percentage of the SB.

$$\%Error = \frac{EB - SB}{SB}100 \qquad (7.8)$$

However, the EB of some of the estimation procedures considered in the simulation experiment equals zero. Consequently the %Error results for these estimators will always equal -100%. Moreover, the calculation of the %Error results based on (7.9) is not possible since, for the theoretically unbiased estimators considered in this chapter, the ratio cannot be defined.

$$\%Error = \frac{SB - EB}{EB}100 \qquad (7.9)$$

Finally an additional problem arises when using either (7.8) or (7.9) for calculating accuracy performance results for the theoretically biased estimators: in many cases the EB or SB take values very close to zero. In those cases the magnitude of the %Error will not reflect the true accuracy of the approximations.

Hence, we consider finding a different way of assessing the accuracy of our bias approximations. The simulated bias has been recorded as: *Forecast minus (-) Demand* (a (+) sign indicates that the forecasting method under concern overestimates the mean demand level whereas a (–) sign that it underestimates it). For methods that theoretically overestimate the mean demand level, an appropriate accuracy measure is the following:

$$\%Error = \frac{EB - SB}{MSD}100 \qquad (7.10)$$

where EB is the positive theoretically expected bias, SB the mean simulated (+ or -) bias and MSD is the mean simulated demand per period as calculated in the simulation experiment, for a particular combination of the control parameters, over five simulation runs (replications).

For methods that underestimate the mean demand level, (7.10) will be adjusted to take into account the negative sign of the bias:

$$\%Error = \frac{|EB| + SB}{MSD}100 \tag{7.11}$$

In both (7.10) and (7.11) the error of our approximations is expressed as a percentage of the mean simulated demand. The accuracy measure is not dependent on the scale of the theoretically generated demand data series and it can be averaged across all series of interest to us to calculate summary results (Avg.%Error). For a biased estimation procedure, on a single series (combination of control parameters) evaluation, a %Error = +(-)1% tells us that our approximation over (under)-estimates the true relative bias by 1% of the actual demand. For unbiased estimators the %Error results should always appear with a negative sign indicating that we underestimate the true (positive or negative) bias. Nevertheless, a plus/minus notation is employed, utilising (7.11) with $|EB| = 0$, to indicate whether demand is over (+) or under (-) estimated.

To check the statistical significance of the results, the following condition will be tested (significance level = 0.01, t = 2.576):

$$\frac{-2.576\, s.e}{\overline{MSD}}100 < \frac{\overline{EB} - \overline{SB}}{\overline{MSD}}100 < \frac{2.576\, s.e}{\overline{MSD}}100 \tag{7.12}$$

where:

$\overline{SB}$ is the average simulated bias obtained across all control parameter combinations
$s.e.$ is the standard error of the mean ($\overline{SB}$)
$\overline{EB}$ is the average theoretically expected bias, and
$\overline{MSD}$ is the average simulated demand.

So if, for example, the 99% confidence interval for the ratio $\frac{\overline{EB} - \overline{SB}}{\overline{MSD}}100$ is given by the range -0.15% to $+0.15\%$ the approximation will be said to be accurate to $\pm 0.15\%$ of the mean simulated demand (significance level = 0.01)[4].

[4] The confidence limits will be reported to the second decimal place.

It is important to note that the above hypothetical result will be true only for the range of the control parameter values considered in the simulation experiment. If other control parameter values were taken into account the results would not necessarily be the same. Nevertheless, it is important to note that the parameter values have been selected as representative of many real world applications. As such, the results relate directly to a potential application of the estimators under concern in a real system.

7.6.1 Verification of the simulation model

As discussed in the introductory section of this chapter, exact theoretical results will be used in order to assess the correctness of the simulation model's logic. In particular, Croston's derivations regarding the EWMA bias on all and issue points in time will be checked against the corresponding simulation bias to allow for an assessment of the degree of accuracy of the simulation output. Once the accuracy of the results produced by the simulation has been established (i.e. the theoretically expected results lie within the specified confidence limits) we can then run the simulation model and use the simulation results of interest to us in order to assess the accuracy of our theoretical approximations.

Croston (1972) derived the exact theoretically expected bias, implicitly incorporated in EWMA estimates, when issue points in time are considered (equation (7) in the original paper). The estimator is expected to be exactly unbiased, when all points in time are taken into account. Both results have been generated upon the same assumptions that we have considered for building our simulation experiment. Demand sizes, in Croston's case, were assumed to be normally rather than lognormally distributed, but this assumption does not affect the method's expected performance.

For both all and issue points in time only, the difference between the theoretically expected bias and the simulated bias lies within the specified 99% confidence limits. It is important to note that when issue points only are considered (1,120 simulated scenarios) there is not even one case in which EWMA is found to underestimate the mean demand level, as theoretically expected. The accuracy of the theoretically expected bias result shows no sensitivity to changes of any of the control parameter values and this is true for both all and issue points in time only.

Considering the above results we may conclude that the simulation model's logic is correct, i.e. the model has been verified. The simulation results can now be used to assess the accuracy of the corresponding approximate analytic results. If the simulation model does not exhibit the properties that hold for the analytic model we have reasons to question the accuracy of the approximation under concern.

7.6.2 Croston's method

In sub-section 4.5.2, Taylor's theorem was applied to a function of two variables in order to derive the bias implicitly incorporated in Croston's estimates. The first three terms of the Taylor series were considered and the bias was approximated by (4.32).

There are only 65 occasions (out of the 2,240 tested) where Croston's method appears to underestimate the mean demand level. The difference between the theoretical (approximate) bias given by (4.32) and the simulated bias lies within a 99% confidence interval of $\pm 0.20\%$ of the mean simulated demand. This is true when all 2,240 control parameter combinations are taken into account.

When results are generated on all points in time (1,120 simulated conditions) the bias approximation has also been found to be accurate to within the specified 99% confidence limits ($\pm 0.27\%$ of the mean simulated demand). However this is not so when results are generated on issue points only. In this last case the difference between the theoretically estimated and simulated bias is found to be statistically significant at the 1% significance level (confidence limits: $\pm 0.26\%$ of the mean simulated demand). This issue is further discussed later in this sub-section.

The average percentage error results (Avg.%Error), across all combinations, for a specific p value, are indicated in table 7.3, for all and issue points in time only.

Value of p	All points in time	Issue points only
1.1	0.05	0.08
1.3	0.09	0.22
1.5	0.12	0.28
1.7	0.19	0.29
1.9	0.19	0.30
2	0.19	0.17
4	-0.19	1.31
6	-0.35	2.05
8	-0.38	2.83
10	-0.30	2.55

Table 7.3. Croston's method. Average percentage error results (p value)

The results indicate that for a specific combination (μ, σ, α, L, all points in time) for $p \leq 2$ we tend to slightly overestimate the true bias of Croston's method whereas for $p > 2$ we tend to slightly underestimate it. The differences are quite consistent across all possible factor combinations. Consequently, it may be reasonable to assume that for $p \leq 2$ the sum of the Taylor series' terms, that have not been considered in generating our theoretical approximation, is negative whereas for $p > 2$ it is positive.

For $p \leq 2$ the results are similar between all and issue points in time, the very small difference being attributed to the increased sampling variation associated with the results obtained for issue points only. For $p > 2$ the difference between all and issue point results increases.

When issue points in time are considered, the sampling variation increases with the average inter-demand interval value and as such the absolute Avg.%Error is also expected to increase. Nevertheless, we cannot claim that the observed differences between the theoretically expected and simulated bias for Croston's method (issue points only, $p > 2$) can be attributed only to the sampling variation. The results demonstrate that, in a re-order level context, as the average inter-demand interval value increases the accuracy of Croston's method bias approximation decreases.

When issue points only are considered and for $p \leq 2$ (672 simulated scenarios) the difference between the theoretical bias given by (4.32) and the simulated bias lies within a 99% confidence interval of $\pm 0.27\%$ of the mean simulated demand. When results are generated on issue points only and for $p > 2$ (448 simulated scenarios) the difference between the theoretical and simulated bias is found to be statistically significant (99% confidence limits: $\pm 0.88\%$ of the mean simulated demand). The above results indicate the reduced accuracy of (4.32) in a re-order level context and for average inter-demand intervals greater than 2 periods.

The average percentage error results across all combinations for a specific α value are indicated in table 7.4, for all and issue points in time only.

α value	All points in time	Issue points only
0.05	-0.01	1.06
0.1	-0.05	0.92
0.15	-0.03	0.95
0.2	-0.05	1.12

Table 7.4. Croston's method. Average percentage error results (α value)

The approximation of the bias, implicitly incorporated in Croston's estimates, is shown to be insensitive to the smoothing constant value used.

The Avg.%Error results for different lead time lengths are presented in table 7.5. When all points in time are considered, the results are similar for all L values. For issue points only the Avg.%Error of the approximation reduces with the lead time. The more forward demand data periods are considered for generating error results, the less the effect of the sampling variation will be.

L value	All points in time	Issue points only
1	-0.03	1.72
3	-0.04	1.14
5	-0.04	0.81
12	-0.04	0.37

Table 7.5. Croston's method. Average percentage error results (L value)

Finally, the Avg.%Error of the approximation appears to be insensitive to the squared coefficient of variation of the demand sizes.

CV^2 value	All points in time	Issue points only
0.000625	-0.07	1.00
0.015625	-0.05	1.16
0.09	-0.03	0.96
0.36	-0.04	1.05
0.5625	-0.03	1.18
1	-0.01	0.93
2.25	-0.04	0.79

Table 7.6. Croston's method. Average percentage error results (CV^2 value)

Overall, the simulation results indicate the:

- good accuracy of the approximation of Croston's bias, for all points in time
- good accuracy of the approximation of Croston's bias, for issue points only and small average inter-demand intervals ($p \leq 2$). The approximation is found to be less accurate for higher values of p.

At this stage it is important to note that the p cut-off value determination in chapter 6 should be relatively unaffected by the reduced accuracy of Croston's method bias approximation for $p > 2$. This is because all the p cut-off values determined in chapter 6 were less than 2. However, the forecasting accuracy of the λ Approximation and Approximation method needs to be checked carefully since their derivation in chapter 4 was based on our approximation to Croston's bias.

Finally, we should also note that considering further terms in the Taylor series could eventually increase the accuracy of Croston's bias approximation at the expense though of the mathematical tractability that is ensured when the first three terms only are considered.

7.6.3 λ Approximation method

The λ Approximation method was designed to eliminate the bias implicitly incorporated in Croston's estimates. In the previous sub-section we showed that the true (simulated) Croston's bias may be over or under-estimated by its theoretical approximation derived in chapter 4. When Croston's bias is overestimated

(+Avg.%Error results) we account for a bias greater than the true one and in consequence the λ Approximation method is intuitively expected to underestimate the mean demand level. The opposite should be the case when Croston's bias is underestimated. In table 7.7 the Avg.%Error results are presented for the λ Approximation method for all and issue points in time, for different average inter-demand intervals. The results, unless zero, indicate that the true absolute bias is underestimated. The sign indicates whether or not demand is over or under-estimated.

The results should be compared with those presented for Croston's method in table 7.3.

Value of p	All points in time	Issue points only
1.1	0.04	0
1.3	0.10	-0.02
1.5	0.12	-0.03
1.7	0.09	-0.01
1.9	0.09	0
2	0.10	0.13
4	0.43	-1.08
6	0.52	-1.87
8	0.53	-2.68
10	0.43	-2.41

Table 7.7. λ Approximation method. Average percentage error results (p value)

When all combinations are considered, the difference between the theoretically expected zero bias and the simulated bias lies within a 99% confidence interval of $\pm 0.02\%$ of the mean simulated demand. The same is the case when results are generated on all the simulated conditions that refer to a re-order interval context.

When results are generated on issue points only and for $p \leq 2$, the difference between the theoretical and simulated bias lies within the specified 99% confidence limits ($\pm 0.02\%$ of the mean simulated demand). However, this is not so for issue points only and $p > 2$ (99% confidence limits: $\pm 0.14\%$ of the mean simulated demand).

Overall, the results are very similar to those presented for Croston's method. In particular, the accuracy of the approximation is insensitive to the length of the lead time, the squared coefficient of variation and the smoothing constant value used. Moreover, the results indicate a connection between the accuracy of Croston's method bias approximation and the resulting bias of the λ Approximation method. That is, the growth in bias of the λ Approximation method for higher value of p in a re-order level context can be attributed to the corresponding decline in accuracy of Croston's bias approximation.

7.6.4 Approximation method

The difference between the theoretically expected bias for the Approximation method and the corresponding simulated bias lies within a 99% confidence interval of $\pm 0.40\%$ of the mean simulated demand. This is true when all the simulated scenarios are taken into account. There are only 31 cases where the Approximation method overestimates the mean demand level.

The simulation results for this method can largely be explained (as in the case of the λ Approximation method) in terms of the way the method was derived. When Croston's bias approximation overestimates the true bias we do account for a bias greater than the one given by simulation. Consequently the simulated bias for the Approximation method appears to be greater than the theoretically expected one (i.e. we underestimate the true bias given by this method). The opposite is the case when Croston's simulated bias is underestimated. In the following tables the Avg.%Error results are presented for different α, L, CV^2 and p values, for all and issue points in time only. Note the sign of the Avg.%Errors which is in all cases different from that given in Croston's results (sub-section 7.7.2).

α value	All points in time	Issue points only
0.05	0.03	-1.03
0.1	0.08	-0.88
0.15	0.09	-0.88
0.2	0.12	-1.03

Table 7.8. Approximation method. Average percentage error results (α value)

L value	All points in time	Issue points only
1	0.08	-1.66
3	0.08	-1.09
5	0.08	-0.76
12	0.08	-0.33

Table 7.9. Approximation method. Average percentage error results (L value)

CV^2 value	All points in time	Issue points only
0.000625	0.07	-1.00
0.015625	0.07	-1.14
0.09	0.06	-0.93
0.36	0.09	-0.99
0.5625	0.09	-1.11
1	0.08	-0.85
2.25	0.13	-0.70

Table 7.10. Approximation method.

Average percentage error results (CV^2 value)

Value of p	All points in time	Issue points only
1.1	0	-0.03
1.3	-0.03	-0.16
1.5	-0.06	-0.22
1.7	-0.12	-0.23
1.9	-0.12	-0.24
2	-0.13	-0.11
4	0.22	-1.28
6	0.36	-2.02
8	0.41	-2.79
10	0.33	-2.50

Table 7.11. Approximation method. Average percentage error results (p value)

The statistical significance of the results has been checked separately for all and issue points in time only. The 99% confidence limits are as follows:

1. all points time (1,120 scenarios): $\pm 0.60\%$ of the mean simulated demand
2. issue points only, $p \leq 2$ (672 scenarios): $\pm 0.64\%$ of the mean simulated demand
3. issue points only, $p > 2$ (448 scenarios): $\pm 0.32\%$ of the mean simulated demand

The mean difference between the simulated and expected bias lies within the confidence limits in the first two cases, but not in the third case.

The accuracy of the bias approximation is insensitive to all the control parameters introduced in our simulation experiment with the exception of the lead time length and the average inter-demand interval, when issue points only are considered. The sensitivity to the L value (for issue points only) has already been explained in terms of the sampling variation that increases as the L value decreases. Moreover, the growth in bias of the Approximation method for higher value of p can be attributed to the corresponding decline in accuracy of Croston's method bias approximation.

7.6.5 Conclusions

In this section the accuracy of the bias approximations that have been derived during this research, has been tested against simulation results. The approximations are found to be insensitive to the squared coefficient of variation of demand sizes, the α smoothing constant value used and the lead time length (when all points are taken into account). The accuracy of the approximations increases with the lead time length when issue points only are considered. This is because the effect of the sampling variation reduces as the L value increases. For the control parameters and the range of the simulated values considered in the experiment, the bias approximations developed in this thesis are found to be accurate to within the specified 99% confidence limits. This is true when results are generated on

- all simulated scenarios
- control parameter combinations referring to a re-order interval context
- simulated conditions referring to a re-order level context, average inter-demand interval less than two.

When issue points only are considered, the accuracy of our approximation to Croston's bias (4.32) deteriorates for average inter-demand intervals greater than 2 review periods. In fact the difference between the theoretical and simulated bias is found to be statistically significant (at the 1% level) in the corresponding simulated scenarios. The reasons of the decline in accuracy of our expression (4.32) are not yet clear. However, the simulation results have shown that this decline in accuracy affects the bias of the λ Approximation and Approximation method. As such, both methods may not work as well in re-order level as in re-order interval systems. The issue of bias will be further considered in chapter 10, where empirical results will be generated for the methods discussed in this chapter[5].

7.7 Simulation results. The variance of intermittent demand estimates

In this section we check the accuracy of all variance expressions derived in chapter 5 of the thesis. The error results, difference between the theoretically expected variance (*EV*) and the simulated variance (*SV*), are now expressed as a percentage of the SV:

$$\%Error = \frac{EV - SV}{SV} 100 \qquad (7.13)$$

The usual approach to demonstrate statistical significance of the variance results would be to consider the *SV* as χ^2 (chi-square) distributed and then to perform a test using:

$$\chi_\nu^2 = \frac{(n-1) * SV}{EV}$$

where:

n is the sample size (i.e. $n = 19{,}900$, for all points in time and $n = 19{,}900 / p$ for issue points only) and

χ_ν^2 is the chi-square variable with $\nu = n - 1$ degrees of freedom.

[5] Empirical results will not be generated for the λ Approximation method for reasons explained in the previous chapter.

The chi-square variable is additive[6] and as such statistical results could be generated for each particular combination of the control parameters, across all five replications, or overall accuracy results across all combinations of the control parameter values considered in this simulation experiment. Nevertheless, the chi-square variable tends to the normal for large sample sizes, which is the case in the current analysis. Therefore, the results can be tested for statistical significance using an approach similar to that considered in the previous section.

The statistical significance of the results is checked by testing the following condition (significance level = 0.01, t = 2.576):

$$\frac{-2.576\,s.e}{\overline{SV}}100 < \frac{\overline{EV}-\overline{SV}}{\overline{SV}}100 < \frac{2.576\,s.e}{\overline{SV}}100 \qquad (7.14)$$

where:

$\overline{SV}$ is the average simulated variance obtained across all control parameter combinations

$s.e.$ is the standard error of the mean ($\overline{SV}$) and

$\overline{EV}$ is the average theoretically expected variance

7.7.1 Croston's method

In section 5.3 the variance of Croston's estimates was approximated by:

$$Var\left(\frac{z'_t}{p'_t}\right) \approx \frac{a}{2-a}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]+\frac{\alpha^4}{1-(1-\alpha)^4}\frac{\mu^2}{p^4}\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2+1\right] \qquad (7.15)$$

considering the first three terms in a Taylor series.

[6] The sum of k independent chi-squared variates is also a chi-square variate: $\sum_{i=1}^{\kappa}(\chi^2:\nu_i) \sim \chi^2:\nu$, where $\nu=\sum_{i=1}^{\kappa}\nu_i$

In order to enable the derivation of meaningful categorisation rules, in chapter 6, we finally approximated the variance of Croston's estimates by:

$$Var\left(\frac{z'_t}{p'_t}\right) \approx \frac{a}{2-a}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] \tag{7.16}$$

promising at that stage to explore in detail the accuracy differences between (7.15) and (7.16) in this current chapter.

The first important finding is that the difference between the simulated variance and the variance given by either (7.15) or (7.16) lies within a 99% confidence interval of $\pm$17.3% of the simulated variance. On average, approximation (7.15) underestimates the true variance by 0.01% whereas approximation (7.16) by 3%. The above results are valid when all 2,240 combinations of the control parameters are taken into account. The Avg.%Error results for both approximations are presented in table 7.12, for different p values, for all and issue points in time.

Value of p	All points in time		Issue points only	
	approx. (7.15)	**approx. (7.16)**	**approx. (7.15)**	**approx. (7.16)**
1.1	6.14	5.31	6.15	5.32
1.3	4.44	2.81	4.35	2.73
1.5	3.34	1.11	3.56	1.33
1.7	2.22	-0.46	2.43	-0.25
1.9	2.96	-0.08	3.21	0.15
2	0.78	-2.42	1.07	-2.14
4	-4.07	-8.53	-3.89	-8.38
6	-7.46	-12.27	-6.94	-11.80
8	-6.09	-11.09	-5.38	-10.40
10	-3.72	-9.08	-3.38	-8.81

Table 7.12. Croston's method. Average percentage error results (p value)

The results generated in the previous section have indicated that the accuracy of Croston's bias approximation deteriorates for issue points only and $p \geq 2$. To derive a variance expression, the expected estimate of demand per period needs to be taken into account.

Consequently, the statistical accuracy of the variance expressions has been also checked, separately, for all and issue points in time only, distinguishing between average inter-demand intervals less than or equal to two review periods and greater than two review periods. In all cases, both (7.15) or (7.16) are found to be within the specified 99% confidence limits. These limits, expressed as a percentage of the simulated variance are as follows:

- all points in time, all p values: $\pm 24.46\%$
- all points in time, $p \leq 2$: $\pm 25.90\%$
- all points in time, $p > 2$: $\pm 33.47\%$
- issue points only, all p values: $\pm 24.49\%$
- issue points only, $p \leq 2$: $\pm 25.92\%$
- issue points only, $p > 2$: $\pm 33.58\%$

The fact that the difference between the simulated variance and expected variance, given by either (7.15) or (7.16), lies within the specified 99% confidence interval indicates that the accuracy of the variance approximations is not affected by the reduced accuracy of Croston's method bias approximation in a re-order level context and for $p > 2$.

For $p \leq 2$ approximation (7.15) overestimates the true variance of Croston's estimates whereas for $p > 2$ it underestimates it. The extra term that distinguishes (7.15) from (7.16) is always positive, being between 1.5%-2% of the true variance when $p \leq 2$, and between 4%-6% of the true variance when $p > 2$. By not considering this extra term (i.e. using approximation (7.16)) we obtain, for $p \leq 2$, an Avg.%Error of 1.1% whereas the Avg.%Error given by (7.15) is 3.4%. For $p > 2$ (7.15) underestimates the true variance and by not considering the extra positive term we underestimate the variance by even more. For $p > 2$ the Avg.%Error given by (7.15) and (7.16) is -5% and -10% respectively.

Overall, both approximations are reasonably accurate. The decision about which one should be chosen depends on both tractability and overall degree of accuracy required. In the case of this research, approximation (7.16) is viewed as the more appropriate one to use for two reasons:

- Simplification of the variance calculation enables us to derive tractable decision rules based on which Croston's method performance can be assessed in detail
- The theoretical MSE performance differences are particularly sensitive to the average inter-demand interval values. For $p > 2$ the MSE differences are very well marked whereas for $p \leq 2$ it is difficult to decide which method performs best. As such, and in order to assess the accuracy of the theoretical categorisation rules on the simulated data, increased accuracy is required for $p \leq 2$.

Nevertheless, it is important to note that (7.15) could be usefully employed in inventory systems requiring estimation of the forecast variability, particularly for higher values of p, because of its accuracy superiority over (7.16) in this region.

The selection of (7.16) is further justified in the following section where we show that all the MSE categorisation rules derived in chapter 6 are accurate.

In the following tables we present the Avg.%Error obtained by using (7.16) for different L, CV^2 and α values.

L value	All points in time	Issue points only
1	-3.45	-3.20
3	-3.46	-3.23
5	-3.47	-3.24
12	-3.50	-3.26

Table 7.13. Croston's method. Average percentage error results (L value)

CV^2 value	All points in time	Issue points only
0.000625	-3.49	-3.17
0.015625	-2.27	-1.68
0.09	-2.46	-2.05
0.36	-4.27	-4.12
0.5625	-3.23	-3.00
1	-4.79	-4.51
2.25	-3.79	-4.06

Table 7.14. Croston's method. Average percentage error results (CV^2 value)

α value	All points in time	Issue points only
0.05	-0.09	0.10
0.1	-3.30	-3.18
0.15	-4.75	-4.40
0.2	-5.75	-5.42

Table 7.15. Croston's method. Average percentage error results (α value)

The variance approximation appears to be insensitive to changes of L and CV^2. The Avg.%Error increases in absolute terms with the smoothing constant value. It would be reasonable to suppose that as the α value increases, the future Taylor terms that have not been considered in deriving the variance approximation become more substantial and, subsequently, the accuracy of the approximation deteriorates.

7.7.2 The λ Approximation method

The variance of the estimates produced by the λ Approximation method is approximated by (7.17).

$$Var\left[\frac{\left(1-\frac{\alpha}{2}\right)z'_t}{p'_t-\frac{\alpha}{2}}\right]\approx\frac{\alpha(2-\alpha)}{4}\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2+p(p-1)\mu^2+\frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4}$$

$$+\frac{\alpha^4}{1-(1-\alpha)^4}\frac{\left(1-\frac{\alpha}{2}\right)^2\mu^2}{\left(p-\frac{\alpha}{2}\right)^6}\,p^2\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2+1\right] \qquad (7.17)$$

Nevertheless, it was decided, in chapter 5, that the fourth power term of approximation (7.17) should be excluded. This enabled more meaningful comparison results to be obtained in chapter 6. Finally the variance of the estimates was approximated by (7.18).

$$Var\left[\frac{\left(1-\frac{\alpha}{2}\right)z'_t}{p'_t-\frac{\alpha}{2}}\right]\approx\frac{\alpha(2-\alpha)}{4}\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2+p(p-1)\mu^2+\frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4} \qquad (7.18)$$

The validity of our decision is assessed in detail in this sub-section.

The results regarding the performance of the two approximations are very similar to those already discussed in the previous sub-section. The difference between the simulated variance and the variance given by either (7.17) or (7.18) is found to lie within the specified 99% confidence limits for all the cases analysed. These limits, expressed as percentage of the simulated variance, are as follows:

- all simulated scenarios: $\pm 17.29\%$
- all points in time, all p values: $\pm 24.45\%$
- all points in time, $p \le 2$: $\pm 25.70\%$
- all points in time, $p > 2$: $\pm 33.27\%$
- issue points only, all p values: $\pm 24.47\%$
- issue points only, $p \le 2$: $\pm 25.72\%$
- issue points only, $p > 2$: $\pm 33.37\%$

On average approximation (7.17) underestimates the true variance by 0.17% whereas approximation (7.18) by 3.1%. This is true when all 2,240 combinations of the control parameters are taken into account. The Avg.%Error results for both approximations are presented in table 7.16, for different p values, for all and issue points in time.

p value	All points in time		Issue points only	
	approx. (7.17)	**approx. (7.18)**	**approx. (7.17)**	**approx. (7.18)**
1.1	6.78	5.76	6.79	5.77
1.3	4.67	2.75	4.58	2.66
1.5	3.33	0.78	3.56	1.00
1.7	2.00	-1.01	2.21	-0.80
1.9	2.62	-0.75	2.87	-0.51
2	0.39	-3.13	0.68	-2.85
4	-4.61	-9.25	-4.43	-9.09
6	-7.92	-12.83	-7.38	-12.35
8	-6.45	-11.53	-5.74	-10.84
10	-4.02	-9.46	-3.68	-9.18

Table 7.16. λ **Approximation method. Average percentage error results (** p **value)**

For $p \le 2$ approximation (7.17) overestimates the true variance of the λ Approximation method estimates whereas for $p > 2$ it underestimates it. The extra term that distinguishes (7.17) from (7.18) is always positive, being between 1%-2.5% of the true variance when $p \le 2$, and between 5%-6% of the true variance for $p > 2$. By not considering this extra term (i.e. using approximation (7.18)) we obtain, for $p \le 2$, an Avg.%Error of 0.8% whereas the Avg.%Error given by (7.17) is 3.4%. For $p > 2$ (7.17) underestimates the true variance and by not considering the extra positive term we underestimate the variance by even more. For $p > 2$ the Avg.%Error given by (7.17) and (7.18) is -5.8% and -10.5% respectively.

The selection of (7.18) instead of (7.17) for representing the variance of the estimates is justified in terms of the increased accuracy required for very low average inter-demand interval values. The Avg.%Error results obtained across the different control parameter values are very similar to those presented in the previous sub-section. The variance approximation is insensitive to the lead time length and squared coefficient of variation of the demand sizes. The accuracy of the approximation deteriorates as the smoothing constant value increases.

Before we close this sub-section it is important to note that (7.17) should be preferred to (7.18) for employment in stock control systems requiring estimation of the forecast variability, particularly for higher values of p, because of its accuracy superiority in this region.

7.7.3 Approximation method

The variance of the estimates produced by the Approximation method is given by (7.19)

$$Var(Y'_t) = Var\left(\left(1-\frac{\alpha}{2}\right)\frac{z'_t}{p'_t}\right) = \left(1-\frac{\alpha}{2}\right)^2 Var\left(\frac{z'_t}{p'_t}\right) \tag{7.19}$$

where $Var\left(\frac{z'_t}{p'_t}\right)$ is the variability of Croston's estimates and can be approximated by (7.15) or (7.16). The latter approximation has been chosen and our decision is now justified.

The results are very similar to those presented in sub-sections 7.7.1 and 7.7.2 and they are summarised in the following two tables.

p value	All points in time		Issue points only	
	using (7.15)	**using (7.16)**	**using (7.15)**	**using (7.16)**
1.1	6.14	5.31	6.15	5.33
1.3	4.44	2.81	4.36	2.74
1.5	3.34	1.11	3.56	1.33
1.7	2.22	-0.46	2.43	-0.26
1.9	2.97	-0.09	3.21	0.15
2	0.78	-2.42	1.07	-2.15
4	-4.07	-8.54	-3.90	-8.39
6	-7.46	-12.27	-6.94	-11.81
8	-6.10	-11.09	-5.38	-10.41
10	-3.72	-9.09	-3.38	-8.81

Table 7.17. Approximation method. Average percentage error results (p value)

	Avg.%Error	
	using (7.15)	**using (7.16)**
$p \leq 2$	3.39%	1.11%
$p > 2$	-5.11%	-10.05%
all p values	-0.01%	-3.35%

Table 7.18. Approximation method. Summary (%) results

Both approximations are, overall, accurate to $\pm 16.97\%$ of the simulated variance (see equation (7.14)). The approximations are also statistically accurate when results are generated, separately, for all and issue points in time only, distinguishing between $p \leq 2$ and $p > 2$. In all cases, the difference between the simulated variance and the theoretical variance obtained by using either (7.15) or (7.16) lies within the specified 99% confidence limits. These limits, expressed as a percentage of the simulated variance, are as follows:

- all points in time, all p values: $\pm 23.99\%$
- all points in time, $p \leq 2$: $\pm 25.37\%$
- all points in time, $p > 2$: $\pm 32.76\%$
- issue points only, all p values: $\pm 24.01\%$
- issue points only, $p \leq 2$: $\pm 25.39\%$
- issue points only, $p > 2$: $\pm 32.86\%$

The accuracy of the approximations is not sensitive to any of the control parameters with the exception of the smoothing constant value. The Avg.%Error results obtained across the different control parameter values are very similar to those presented in the sub-section 7.7.1.

Even though the use of (7.16) is more meaningful for categorisation purposes, (7.15) should offer better results in real applications when the forecast variability needs to be estimated.

7.7.4 Conclusions

In this section, the variance approximations derived thus far in this research have been tested against simulation results. When all the simulated scenarios are taken into account, the difference between the simulated variance and the corresponding theoretically expected variance(s) is found, for all methods, to lie within a 99% confidence interval of $\pm 17\%$ of the simulated variance. The results presented in the previous section indicated that the accuracy of the bias approximations derived in this thesis reduces considerably in a re-order level context when the average inter-demand interval exceeds two review periods. However, the above mentioned limitation has no effect on the accuracy of the approximate variance expressions.

For Croston's, λ Approximation and Approximation method the variance expressions were simplified in chapter 5 by not considering a part of the expressions. By doing so, meaningful categorisation rules were derived in chapter 6 and (as shown in this section) increased accuracy is achieved for $p \leq 2$ where the MSE differences between alternative estimators are not very well-marked. However, the full variance expressions could be usefully employed by stock control systems requiring an estimation of the forecast variability for higher inter-demand intervals, since in that region the simplified expressions do not perform so well.

Finally, it is important to note that all the approximations are very well behaved across all the control parameters considered in the simulation experiment. Some sensitivity though is indicated to changes in the smoothing constant value and this could be attributed to the further Taylor terms that have not been considered in deriving the variance approximations.

7.8 Simulation results. The MSE of intermittent demand estimates

In this section, the accuracy of the theoretically expected MSEs, for the methods considered in this research, is tested against simulation results. In addition, we check the accuracy of all the categorisation rules derived in chapter 6. The Avg.%Error is calculated as follows:

$$\%Error = \frac{EMSE - SMSE}{SMSE} 100 \tag{7.20}$$

where EMSE is the theoretically expected MSE (see chapter 6) for the estimation procedure under concern, and SMSE is the MSE obtained via simulation. The confidence limits are expressed as a percentage of the SMSE.

7.8.1 Accuracy of the approximations

The simulation results, regarding the accuracy of the approximations, are summarised in table 7.19. In the second column we indicate the Avg.%Error (percentage by which the theoretically expected MSE under (-) or over (+) estimates, on average, the simulated MSE, for different control parameter combinations). In all cases, all the approximations (or exact results) are found to be within the specified confidence limits. These limits, expressed as a percentage of the simulated MSE, are given in the third column of the table.

estimator	points in time	inter-demand interval	Avg.%Error	99% confidence limits
Croston	all simulated scenarios		-0.08%	± 10.77%
	all	all p values	-0.02%	± 15.22%
		$p \leq 2$	0.42%	± 17.74%
		$p > 2$	-0.68%	± 19.03%
	issue	all p values	-0.13%	± 15.23%
		$p \leq 2$	0.45%	± 17.74%
		$p > 2$	-1.01%	± 19.05%
EWMA	all	all p values	0.38%	± 15.095
		$p \leq 2$	0.25%	± 18.08%
		$p > 2$	0.58%	± 20.85%
	issue	all p values	0.16%	± 14.76%
		$p \leq 2$	0.24%	± 18.41%
		$p > 2$	0.03%	± 22.93%
λ Approximat.	all simulated scenarios		-0.11%	± 10.74%
	all	all p values	-0.03%	± 15.20%
		$p \leq 2$	0.37%	± 17.67%
		$p > 2$	-0.63%	± 18.90%
	issue	all p values	-0.19%	± 15.20%
		$p \leq 2$	0.39%	± 17.67%
		$p > 2$	-1.06%	± 18.92%
Approximation	all simulated scenarios		-0.14%	± 10.57%
	all	all p values	-0.04%	± 14.96%
		$p \leq 2$	0.29%	± 17.42%
		$p > 2$	-0.55%	± 18.83%
	issue	all p values	-0.23%	± 14.96%
		$p \leq 2$	0.29%	± 17.42%
		$p > 2$	-1.00%	± 18.85%

Table 7.19. MSE simulation results

The results indicate that the Avg.%Error, across all the simulated scenarios, is in the range -0.07%, +0.38%. Considering that the simulated MSE can be as high as 2,700 it is obvious that the approximations[7] perform well. When all control parameter combinations are taken into account, the difference between the simulated and theoretically expected MSE lies, for all methods, within a 99% confidence interval of $\pm 10\%$ of the simulated MSE.

The results presented in section 7.6 indicated that the accuracy of the bias approximations derived in this thesis deteriorates in a re-order level context for average inter-demand intervals greater than two review periods. The MSE expressions consist of three components: the variance of demand, variance of the estimates and a bias squared term. The results presented in table 7.19 indicate that the accuracy of the approximate MSE expressions derived in this thesis is not affected by the above discussed limitation.

The accuracy of the MSE expressions is not sensitive to any of the control parameters used, for EWMA. Some sensitivity to the smoothing constant value has been observed for the remaining three estimation procedures (the approximations deteriorate as the smoothing constant value increases).

7.8.2 Categorisation rules

In this sub-section we check the accuracy of the categorisation rules developed in chapter 6. We first consider the categorisation rules that resulted from the pair-wise comparisons and then we assess the accuracy of the categorisation rules regarding all methods' performance.

The categorisation rules developed in chapter 6 are summarised in the following table. The method that appears in italics is the one that theoretically performs better only when both criteria take a value below their corresponding cut-off point.

[7] The MSE expressions for the EWMA estimator (all and issue points in time) are exact rather than approximate results. Nevertheless, these exact results are also tested against the simulation results to assess the extent to which our findings are affected by sampling variation.

Pair-wise comparison	p *cut-off value*	CV^2 *cut-off value*
Approximation-*Croston*	1.32	0.49
Approximation- λ *Approx.*	1.25	0.48
λ Approx.-*Croston*	1.65	1.17
ISSUE POINTS		
Approximation-*EWMA*	1.33	0.49
Croston-*EWMA*	1.34	0.28
λ Approx.-*EWMA*	1.40	0.40
ALL POINTS IN TIME		
Approximation-*EWMA*	1.20	0.49
Croston-*EWMA*	Croston always performs better than EWMA	
λ Approx.-*EWMA*	1.17	0.22

Table 7.20. Cut-off values ($0.05 \le \alpha \le 0.2$**)**

For each pair-wise comparison the conditions under which one method is theoretically expected to perform better (or worse) than the other are identified and the simulated MSEs given by the two methods under concern are compared. We report whether or not the expected MSE performances are confirmed by the simulation results, and also the percentage increase (decrease) of the MSE when using one estimation procedure instead of the other.

In particular, for every pair-wise comparison the simulation results are divided into two sets: one referring to all the simulated conditions when $p >$ *corresponding cut-off value* and/or $CV^2 >$ *corresponding cut-off value* (set A, recommended method: say estimator x) and the other to all the remaining simulated conditions (set B, recommended method: say estimator y). The results will be reported in the form of the MSE average percentage increase (decrease) achieved by employing estimator x instead of estimator y in both sets.

$$Avg.\%Difference = 100 Avg.\left(\frac{MSE_{estimator\ x} - MSE_{estimator\ y}}{MSE_{estimator\ y}} \right) \qquad (7.21)$$

The Avg.%Difference is expected to be negative for data set A and positive for data set B. The results are summarised for all pair-wise comparisons in table 7.21.

Pair-wise comparison	**Set A** $p >$ cut-off value and/or $CV^2 >$ cut-off value	**Set B** $p \leq$ cut-off value and $CV^2 \leq$ cut-off value
Approximation-*Croston*	-1.48	2.82
Approximation- λ *Approx.*	-1.00	4.35
λ Approx.-*Croston*	-0.63	0.43
ISSUE POINTS		
Approximation-*EWMA*	-20.36	0.94
Croston-*EWMA*	-18.80	0.47
λ Approx.-*EWMA*	-19.60	-0.97
ALL POINTS IN TIME		
Approximation-*EWMA*	-13.07	1.58
Croston-*EWMA*	-11.50	
λ Approx.-*EWMA*	-12.10	-2.68

Table 7.21. Simulation (%) results (pair-wise comparisons)

The results for data set B indicate that the differences between the MSEs are small. The differences are very well marked for most comparisons for data set A. This appears not to be the case for the first three pair-wise comparisons. A further analysis of the results, though, indicates that:

- The Approximation method performs better than Croston's method in 1,984 out of the 1,984 cases covered in data set A
- The Approximation method performs better than the λ Approximation method in 1,964 out of the 1,984 combinations of the control parameters tested
- The λ Approximation method performs better than Croston's method in 1,585 out of the 1,664 cases explored.

The categorisation rules were developed in chapter 6 in such a way that one estimation procedure always (theoretically) performs better in, what we now call for the purpose of this chapter, data set A. Therefore, one should expect consistent differences in all the cases covered by data set A. The estimator selected for data set B was an approximate solution, since in the case that both criteria take a value below their corresponding cut-off value, no estimator can be shown, theoretically, to perform better in all cases. Moreover the categorisation schemes proposed in chapter 6 covered all possible smoothing constant values in the realistic range 0.05 – 0.2. Obviously that also introduces a certain degree of uncertainty in the proposed rules. Recall, from chapter 6, that this uncertainty will only be reflected in the area formed when both p and CV^2 take a value below the specified cut-off point. It may be for those reasons that the λ Approximation method is found to perform better than EWMA in data set B, when the opposite is theoretically the case.

Apart from the unexpectedly good performance of the λ Approximation method, when compared with the EWMA, all other results appear to be consistent with the theory. That provides further justification for the selection of approximations (7.16) and (7.18) in calculating the variance of the estimates given by Croston's Approximation and λ Approximation method. As discussed in the previous section, approximations (7.15) and (7.17) do not perform particularly well for low average inter-demand interval values, the area in which increased accuracy is required for categorisation purposes.

In chapter 6, two categorisation schemes were proposed covering all estimation procedures. Both schemes were the result of a synthesis of all the pair-wise decision rules. The validity of those schemes is now assessed on the simulation results. For each of the decision areas we first identify the corresponding conditions (combinations of the control parameters) that have been covered in this simulation experiment. Subsequently, the MSE given by the theoretically expected superior estimator is compared with that of all the other forecasting methods and the Avg.%Difference, across all the relevant combinations, is reported. Each difference is expressed as a percentage of the MSE given by the theoretically inferior estimator. Negative signs indicate that the simulation results are consistent with the theory. The results are summarised in the following table.

ALL POINTS IN TIME			
Superior estimator	*Croston*	*Lambda*	*EWMA*
Approximation	-1.52	-0.97	-13.53
	Approximation	*Lambda*	*EWMA*
Croston	-2.79	-0.97	-4.78
ISSUE POINTS ONLY			
	Croston	*Lambda*	*EWMA*
Approximation	-1.44	-0.95	-20.36
	Approximation	*Lambda*	*EWMA*
Croston	-0.09	-0.14	-3.24
	Croston	*Approximation*	*Lambda*
EWMA	-0.49	-2.60	-0.27

Table 7.22. MSE comparison results (%)

All the results are consistent with what is expected from the theory. The categorisation schemes proposed in chapter 6 are found to yield simulation results consistent with the theory despite the approximations that were introduced in specifying the criteria cut-off values. The validity of the categorisation schemes will also be assessed on empirical data in chapter 10.

The relatively poor performance of the λ Approximation method has been established theoretically in chapter 6. The simulation results verify this poor performance and provide further justification for the decision made in chapter 6 (see section 6.9), to disregard this method for the rest of the thesis.

7.9 Conclusions

In this chapter a simulation experiment using theoretically generated data was developed for the purpose of assessing the accuracy of all the approximated results derived during this thesis. Demand has been assumed to occur as a Bernoulli process and the demand sizes, when demand occurs, were assumed to be lognormally distributed. The control parameters used were: the average inter-demand interval, the coefficient of variation of the demand sizes, the smoothing constant value, the lead time length and the points in time considered for generating results (all, issue points only). The control parameter values used for running the model were selected as representative of a range of "real world" applications. The number of values assigned to the control parameters reflects the compromise that we tried to achieve between the detailed investigation of the problem in hand and the size of the simulation experiment.

The simulation results indicate that the bias approximations developed during this research are reasonably accurate especially when all points in time are considered. For the control parameters and the range of the simulated values considered in the experiment, the bias approximations developed in this thesis are found to be accurate to within the specified 99% confidence limits. This is true when results are generated on

- all simulated scenarios
- control parameter combinations referring to a re-order interval context
- simulated conditions referring to a re-order level context, average inter-demand interval less than two.

When issue points only are considered, the accuracy of our approximation to Croston's bias deteriorates for average inter-demand intervals greater than two review periods. The reasons of this decline in accuracy are not yet clear. However, the simulation results have shown that this decline in accuracy affects the bias of the λ Approximation and Approximation method. Overall, the results appear to be more "firm" when all points in time are considered.

The decline in accuracy of the bias approximations for issue points only and average inter-demand intervals that exceed two review periods does not affect the accuracy of the approximate variance and MSE expressions derived in the thesis. The difference between the simulated variance and the corresponding theoretically expected variance(s) is found, for all methods, to lie within a 99% confidence interval of $\pm 17\%$ of the simulated variance. For Croston's, λ Approximation and Approximation method the variance expressions were simplified in chapter 5 by not considering a part of the expressions. By doing so, meaningful categorisation rules were derived in chapter 6 and, as demonstrated in this chapter, increased accuracy is achieved for $p \leq 2$ where the MSE differences between alternative estimators are not very well-marked. The approximations are very well behaved across all the control parameters considered in the simulation experiment. Some sensitivity though is indicated to changes in the smoothing constant value and this is true for all three estimators.

The difference between the simulated and theoretically expected MSE is found, for all methods, to lie within a 99% confidence interval of $\pm 10\%$ of the simulated MSE. The pair-wise categorisation rules are found to be accurate. Few small discrepancies from what is theoretically expected are attributed to the approximations introduced during the theoretical development of the rules. The categorisation schemes that refer to all methods are also found to be accurate. Whatever the limitations, the simulation results indicate the improvement achieved when the theoretically recommended estimator is utilised. Finally the relatively poor performance of the λ Approximation method has been confirmed in this chapter and this provides further justification for a decision made earlier in the thesis, namely to disregard this method for all future chapters.

CHAPTER 8

Accuracy Measures for Intermittent Demand Estimates

8.1 Introduction

The objective of this chapter is to select specific accuracy measures to be used for comparing alternative intermittent demand estimation procedures. It is to consider the objectives of such a comparison exercise and to choose accuracy measures and testing procedures that contribute towards meeting those objectives. Certain forecasting methods will be considered in chapter 10 in order to generate one step and one lead time ahead forecasts on 3,000 real intermittent demand data series. Subsequently these methods will need to be evaluated with respect to forecasting accuracy. Clearly, accuracy measures need to be chosen for that purpose, to enable meaningful comparisons to be made.

The main objectives of the comparison exercise are to identify which method is generally the most accurate and to specify under what conditions one method performs better than another or than all the rest of the methods. Identification of specific rules for selecting a method is one of the main objectives of the thesis as a whole. Certain decision rules have been derived so far, based on a mathematical analysis of the MSE. The theoretical conclusions have also been confirmed by means of simulation on theoretical data. It is therefore our intention to check whether the results can be replicated in a real forecasting application or not. For this to be done there is no need to restrict our analysis only to the MSE. Other error measures can also be considered in order to identify the extent to which MSE results are reflected by other criteria.

Moreover approximated theoretical results on bias, generated in chapter 4, have also been tested by means of the experiment discussed in chapter 7. Therefore it is important, at this stage, to assess whether these theoretical results can be replicated in practice or not.

To summarise, the main objectives of the accuracy comparison exercise are:

1. To check in practice the validity of the theoretical results on bias generated in chapter 4
2. To check in practice the validity of the theoretical results on Mean Square Error generated in chapter 6
3. To generate results on the conditions under which one method is more accurate than others
4. To determine which is the most accurate forecasting method

Therefore this chapter is oriented towards selecting a set of accuracy measures and methods that, collectively, will capture the information required in order to meet the above objectives.

8.2 Experimental structure

The accuracy measures and methods selected in this chapter will be used to evaluate alternative forecasting methods' performance across 3,000 real intermittent demand data series. A simulation experiment, details of which are presented in chapter 9, will be used to assess, in a dynamic way, how existing and newly developed estimation procedures would have performed if they had been applied in practice. The experiment simulates each method's performance on each one of the SKUs for one step ahead and one lead time forecasts considering all points in time and issue points only. That is, each method will be applied to a particular demand series, the forecasts will be generated and accuracy results of a certain type, that indicate the particular method's performance, will be obtained across time for the demand series under concern (aggregation over time for a particular demand series). The same will be repeated for all demand data series. Summary accuracy results will then be obtained across all series (aggregation over series following the aggregation over time for a particular series) for all methods and the results will be checked for statistical significance.

In the past a slightly different experimental structure has been employed in most of the forecasting accuracy competitions. Performance results are generated by time period across all series (aggregation over series for a particular time period) and then they are summarised across all time periods (aggregation over time following the aggregation over series for a particular time period), i.e. we do not focus on a particular file, SKU in our case, but rather on different forecast horizon estimates across all series. Moreover it has been argued (Fildes, 1992) that alternative forecasting methods' performance should be evaluated by series (across all time periods) and by time period (across all series). The reasoning behind this argument is that summary error statistics calculated in both alternative ways can provide us with a better summary description of the underlying distribution of the error measures. Although this is true, clearly in our case we want to associate the alternative methods' forecasting accuracy directly to specific demand data series characteristics.

As stated in the previous section it is not our intention to conduct a pure forecasting accuracy comparison for the purpose of simply identifying which is the most accurate estimation procedure. We rather try to develop an experiment which will yield insight into the final results, whatever those results may be. That is, once the summary error statistics are available, an effort will be made to link the results directly to specific series characteristics. It is for this reason that we discard the forecasting evaluation by time period across all series and employ only the experimental structure proposed in the beginning of this section.

8.3 Research concerns

Although specifying the objectives of the forecasting accuracy comparison exercise and deciding on the specific structure of the simulation experiment is an important step towards deciding which accuracy measures should be used, there are two more important issues that need to be considered before reaching any definitive conclusions.

Despite the fact that all the real demand data series considered in this research can be broadly categorised as "intermittent", because of the very fact that they have been so treated by a commercial software package, there are still significant scale differences between them, particularly with respect to the size of demand when it occurs. Therefore the mean demand per unit time period may differ significantly from one series to another. Scale differences need to be considered when selecting an accuracy measure so that we do not end up concluding that a forecasting method is superior to all the others simply because it performed well on very few data series. In addition if data in the different series are not expressed in the same unit of measurement, caution should be taken in selecting unit free accuracy measures. Relative accuracy measures are often used in order to overcome these problems.

Selecting accuracy measures in an intermittent demand context poses extra difficulties to the researcher because of the very nature of intermittent demand. The fact that there are some zero demand time periods in all the series, does not allow one to consider some of the most commonly used accuracy measures. In addition many outliers appear in the more lumpy demand files, which makes the selection process even more difficult.

Taking into account the main objectives of the forecasting accuracy comparison, the structure of the simulation experiment and the number and theoretical properties of the demand data series considered in this research it will be argued, in this chapter, that the following accuracy measures should be used:

1. Mean Signed Error (ME)
2. Wilcoxon Rank Sum Statistic (RSS)
3. Mean Square Forecast Error (MSE)
4. Relative Geometric Root Mean Square Error (RGRMSE)
5. Percentage of times Better (PB)

The Percentage of times Better (PB) and the Wilcoxon Rank Sum Statistic (RSS) will be evaluated for statistical significance by employing non-parametric tests. Non-parametric tests require no specific population assumptions to be made but they always sacrifice power in terms of using all available information (in the sense that they consider “relationships” between the errors rather than the error sizes themselves) to reject a false null hypothesis.

The PB reports the proportion of times that one method performs better than one or all other methods. The RSS will not be used for a direct accuracy comparison but rather for indicating which methods consistently underestimate or overestimate the level of demand.

The rest of the accuracy measures consider the size of all errors in order to generate results, which can then be tested for statistical significance using parametric tests.

What follows is the theoretical background for selecting these five accuracy measures along with a detailed discussion of their advantages, disadvantages and application in the simulation experiment. All accuracy measures to which we will refer throughout this chapter are defined in *Appendix 8.A* at the end of the thesis unless they are discussed and mathematically defined in the body of the chapter.

8.4 Literature review

Many empirical and theoretical studies have been conducted in the past that aimed to develop an understanding of the advantages and disadvantages of various forecasting methods (for example: Slovic, 1972; Armstrong, 1978; Makridakis and Hibon, 1979). More recently, empirical and experimental studies have taken the form of a "forecasting competition" where expert participants analysed and forecasted many real life time series coming from entirely different populations (M-Competition: Makridakis et al, 1982; M2-Competition[1]: Makridakis et al, 1993; M3-Competition: Makridakis and Hibon, 2000).

In the latest M3-Competition fully automated software packages, with an "expert selection" facility, were also considered. The major findings are summarised below:

1. Statistically sophisticated or complex methods do not necessarily produce more accurate forecasts than simpler ones
2. The performance of the various methods is very much dependent upon the length of the forecasting horizon and the category of data (yearly, quarterly etc)

[1] The M2-Competition was designed and carried out in such a way as to resemble the actual procedure used in budget (aggregate) forecasting by business firms. That is, extra inside information was also available to improve the predictive accuracy of quantitative methods.

3. Combining the forecasts of a few methods improves overall forecasting accuracy over and above that of the individual forecasting methods used in the combination.

Fildes (1992) conducted empirical research on a large group of series (263) coming from the same company and referring to the same variable. The paper concludes that exponential smoothing and “naïve” models, previously thought to be robust estimation procedures (also confirmed by the M-Competition), forecast poorly for the particular set of time series under analysis. Therefore the author suggests that forecasters should carry out a detailed evaluation of the data series, rather than relying on results from any of the earlier competitions.

The M-Competitions have attracted widespread interest. Nevertheless, one could argue that they did not meet their original objective (Makridakis et al, 1984), namely to facilitate the process of selecting alternative forecasting methods when a great number of series is considered. This is because the results regarding the performance of alternative methods are not particularly conclusive.

It has been claimed (Chatfield, 1992), that the greatest benefit of the M-Competition has been not the results as such, but the “by-products” in making us think more clearly about such issues as error measures and replicability. In addition to answering the question “which is the best forecasting method?” for the situation under concern, the researcher or practitioner now also needs to ask “how should we compare alternative forecasting methods?” in a way that ensures the validity required for our results. Therefore the additional problem of selecting an accuracy measure for the purpose of conducting forecasting accuracy comparisons has also emerged.

8.4.1 Accuracy measures: theoretical concerns

There are many theoretical issues that should be considered in the choice of an error measure. In this sub-section these issues are discussed in detail.

Relativeness

The first important theoretical but also practical consideration is whether or not the accuracy measure under concern is relative in the sense that scale and unit of measurement differences are taken into account. The scale of the data often varies considerably among series. In that case, series with large numbers dominate the comparisons especially if quadratic loss functions (e.g. MSE) are used to report error statistics. Moreover scale dependence can be a severe weakness for an error measure applied to business problems in that the unit of measurement in which a series is recorded is often arbitrary. Unless accuracy measures are unit-free it is possible that we end up comparing “apples and oranges” (Chatfield, 1988) in a way that makes little sense.

In the case of this research all demand data series are expressed in the same unit of measurement. We do not face the problem of distinguishing between pack-sizes and “single” units (see also section 2.5). Therefore we are only concerned with scale differences among data series and from now on “relativeness” should be interpreted as “scale independence” rather than “scale and unit of measurement independence”.

Amount of change

In general, comparisons are more difficult across series where large changes occur over the forecast horizon. In that case one approach would be to employ an accuracy measure that discards any information about the amount of change (e.g. the Percent of times Better measure or the Mean Absolute Percentage Error) or introduce a relative error measure and compare the forecast error of one method against that from another standard method (most commonly the random walk/naïve 1 method). This issue of sensitivity to changes becomes less of a problem when we refer to one step ahead forecasts. Changes in the underlying demand pattern are much more likely to affect longer horizon forecasts.

Outliers

The accuracy measure must be robust from one situation to another but also must not be unduly influenced by outliers. When outliers are present in the data series under concern, care should be taken in order to avoid measures that give considerably more weight to larger errors than smaller ones (e.g. quadratic loss functions).

Outliers create particular problems when the objective is to select from among a set of forecasting methods. It is less of a problem when we try to calibrate a model. The effect of outliers can be reduced by "trimming", discarding very high and very low errors. Using median error measures is an extreme way of trimming as it removes all values higher and lower than the middle value. Another way of dealing with outliers is "winsoring", replacing extreme error values by certain limits (Armstrong and Collopy, 1992). This procedure though, as Makridakis (1993) pointed out, creates the serious problem of non-continuous scales and begs the question of how those limits will be selected (if not in an entirely arbitrary way).

Division by zero

Computational considerations dictate that, whatever measure is used, the possibility of division by zero must never exist. This is the main problem associated with percentage error measures that are relevant only for ratio-scaled data (i.e. data not including zeroes).

8.4.2 Accuracy measures: practical concerns

Practical concerns in selecting an accuracy measure include the following:

1. How widely the accuracy measure is understood. Not all the accuracy measures are straightforward in their interpretation
2. What is the particular accuracy measure's relationship to decision making?

Those issues will have a great effect on the analysis of the simulation results that takes place in chapter 10. Obviously the greater the interpretability of the accuracy measure(s) and the stronger the relationship with decision-making, the more insight we may gain into the accuracy differences between the alternative estimation procedures.

8.5 Accuracy measures for intermittent demand

It is the very nature of intermittent (demand) data, and in particular the existence of some zero demand time periods, that creates additional difficulties in selecting an appropriate accuracy measure. Nevertheless, those special properties of intermittent demand series seem to have been underestimated or in fact completely ignored in the past by both practitioners and academicians.

8.5.1 Commercial software packages

In the latest M3 competition (Makridakis and Hibon, 2000) fully automated software packages, with an "expert selection" facility, were considered along with specific methods/estimators. Forecast Pro from Business Forecast Systems, Inc. was found to perform well.

On a single demand series analysis, Forecast Pro (XE, Versions 3 and 4) provides the user with two sets of statistics. The first refers to "within sample" evaluations[2] and the second to "out-of-sample" rolling evaluation statistics[3].

For the first set of statistics the Mean Absolute Percentage Error (MAPE) is calculated among other accuracy measures. In the case of intermittence, computation of the MAPE is allowed by excluding all the APEs that are associated with zero demand time periods.

For the second set of statistics, the MAPE and Geometric Mean Relative Absolute Error (GMRAE) are computed among other error measures. The Relative Absolute

[2] Statistics are generated for the demand time periods used to initialise our forecasts, in case that the user selects the method to be used, or identify the most appropriate model, select a forecasting method and initialise the forecasts in case that the "expert selection" option is chosen.
[3] Summary statistics are generated for the remaining demand data periods for all possible alternative forecast horizons.

Error (RAE) compares the absolute error of a given method to that from the random walk or naïve 1 forecast, which is the demand in the previous time period. In case of intermittence of course this accuracy measure does not give any meaningful results if we consider the case of consecutive zero demand time periods. For “out-of-sample” evaluations, Forecast Pro adds a small amount to all zero demands and calculates MAPEs and GMRAEs which cannot be interpreted (both measures appear to be equal to an error message).

Based on the above we may conclude that the issue of assessing forecasting performance in an intermittent demand context has not been well addressed by a leading software vendor.

8.5.2 Academic literature

Willemain et al (1994) compared exponential smoothing and Croston’s method using:

- **(a)** Monte Carlo Simulation. Theoretical intermittent demand data were generated for different scenarios that violated Croston’s assumptions. The comparison with exponential smoothing was made with respect to MAPE, Median APE (MdAPE), Root Mean square Error (RMSE), and Mean Absolute Deviation (MAD) for issue points in time only.
- **(b)** Industrial intermittent data, focusing on the MAPE for one step ahead forecasts (for all points in time).

For the theoretically generated data the Mean and Median Absolute Percentage Error were computed by considering the mean demand per unit time period as was set in the demand generation process. No information though is revealed about how the researchers calculated the MAPE when the industrial data, containing some zeroes, was considered. The only comment made by the authors was the following: “…*We focused our attention on the MAPE for one-step-ahead forecasts, comparing forecasted values per period with actual values, both zero and non-zero* (p. 535)”.

One other paper identified in the academic literature that considers accuracy measures for the purpose of comparing alternative intermittent demand estimation procedures is that of Johnston and Boylan (1996). In that case the Relative Arithmetic Mean Square Error (RAMSE) was selected for the purpose of comparing the Size-Interval method and EWMA on theoretically generated demand data. Relative accuracy measures are discussed in section 8.9 whereas MSE itself is presented in section 8.8 of this chapter.

It is important to note that in both papers no justification is given by the authors for their choice of the accuracy measures that were finally used in their research.

8.6 Notation

L : The lead time, $L \geq 1$

$Y'_{t,L}$: The estimate (made in period t) of demand in period $t+L$, obtained by any of the forecasting methods considered.

Y_{t+L} : The actual demand in period $t+L$

e_{t+L} : The forecast error in period $t+L$

n : The number of demand time periods considered for the purpose of comparison

$n = \text{m - r - L}$

where

m is the total number of demand time periods contained in any of the series and
r is the number of periods that were used for initialisation purposes, i.e. not considered for generating results.

8.7 Accuracy measures for indicating bias

The first objective of the accuracy comparison exercise as defined in section 8.1 of this chapter, is to determine whether specific methods consistently underestimate or overestimate the level of demand, i.e. to indicate whether the estimation procedures under concern are biased, and if so in which direction. Two accuracy measures have been selected for reporting bias results: the Mean Signed Error (ME) and a modified version of the Wilcoxon Rank Sum Statistic (RSS). Their choice is justified in the sub-sections that follows.

8.7.1 The Mean signed Error (ME)

The ME is defined as:

$$ME = \frac{\Sigma\left(Y_{t+L} - Y'_{t,L}\right)}{n} = \frac{\Sigma\, e_{t+L}}{n}\,,\ \ t = 1, 2, \ldots\ldots, \mathrm{n} \tag{8.1}$$

For a particular demand series the ME associated with an estimation procedure is calculated by adding all the forecast errors generated by this method's application in practice, considering their sign, and dividing the sum of the errors by the number of forecasts made. The sign of the ME can be either plus (+) or minus (-) indicating that the forecast method underestimates or overestimates, respectively, the level of demand for the particular demand data series. Subsequently, the ME associated with the estimation procedure is calculated across time for all other demand data series. An arithmetic average is then used to calculate the average ME across all series. The same is repeated for all other estimation procedures.

Note that the ME, as all the other accuracy measures, is calculated in a dynamic way. That is, we assess what would have been the bias associated with a particular method if this method had been used in generating forecasts for the particular demand data series.

Having calculated the average ME associated with all estimation procedures, for a particular estimator we can then develop the following hypotheses:

H_0: The average ME is zero, i.e. the estimation procedure is unbiased

H_1: The average ME is not equal to zero, i.e. the estimation procedure is biased

We cannot assume that the population of the MEs is normal or that the population variance is known. However, because our sample size is very large (equal to the number of demand data series considered) the test statistic for testing whether the independently drawn sample of MEs comes from a population with a mean: $\mu = 0$, can be the *t*-test:

$$T = \frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{\kappa}}} \tag{8.2}$$

where:

$\bar{x}$ is the average ME obtained by a particular method across all series

$\hat{\sigma}$ is the standard deviation of the MEs

κ is the number of demand data series considered and

$\mu = 0$

The ME meets all the practical considerations in choosing an accuracy measure for indicating bias in intermittent demand estimates. In particular it is very easy to calculate, it has a straightforward interpretation and it can be used in a dynamic way in order to indicate which methods consistently underestimate or overestimate the mean demand level. From a theoretical point of view, the accuracy measure can be developed for all data rather than ratio-scaled data only and it is not unduly influenced by outliers as the original sign of the error is considered.

The main disadvantage of the ME is that it is not a relative error measure in the sense that when MEs are averaged across all series, the series with large numbers (which potentially will appear to have the largest errors) may, in theory, dominate the final result, i.e. the average ME for a particular method. In practice it is not expected that scale differences will have a great effect on the average ME calculated for each of the forecasting methods, because the sign of the MEs is considered when we arithmetically average them across series. Nevertheless the validity associated with our results can be increased by repeating our analysis for "scaled MEs". That is, the originally calculated MEs, per method, per series, can be divided by the average demand per unit time period (for the series under concern) so that scale dependencies are eliminated. The *t*-test, equation (8.2), can then be applied to the new, scale independent, sample of MEs, in order to check the originally obtained results.

In addition, in order to account for a potentially skewed distribution of the MEs, a non-parametric procedure is introduced in the following section, that considers only the relationship between the MEs, in order to generate results, without taking into account the actual size of them. Results in that case will be generated based on the median, rather than the arithmetic mean, as the most appropriate measure of location in the distribution of the MEs.

At this stage it is important to note that the ME can also be used in order to compare alternative intermittent demand estimation procedures (pair-wise comparisons) across series. In that case the null and the alternative hypotheses can be developed as follows:

H_0: The average ME given by method x_1 equals the average ME given by method x_2, i.e. methods x_1 and x_2 have the same degree of bias

H_1: The average ME given by method x_1 is greater (or less) than the average ME of method x_2, i.e. method x_1 is more (less) biased than method x_2.

The sample size allows us to test the difference between the two population means by using the *t*-test (one sided):

$$t = \frac{|\bar{x}_1| - |\bar{x}_2|}{\sqrt{\frac{1}{\kappa}\left(\hat{\sigma}_1^{\,2} + \hat{\sigma}_2^{\,2}\right)}} \tag{8.3}$$

where:

$|\overline{x}_i|$ is the average ME obtained by method i across all series expressed in an absolute form, i.e. ignoring its sign,

$\hat{\sigma}_i^2$ is the variance of the MEs generated by method i in all series and

κ is the number of demand data series considered.

A comparison of the average ME produced by alternative estimation procedures would have a very straightforward interpretation and it would contribute towards meeting the fourth objective of the accuracy comparison exercise (see also section 8.8). The difference between the absolute average ME given by any two methods indicates how much more (or less) biased is one method in comparison with another. Obviously the problem of scale dependence still needs to be considered and this is why the *t*-test will be also applied to scaled MEs. Moreover, in section 8.9 of this chapter another accuracy measure, the PB, is selected, so that the results given by the ME, when it is used for a direct comparison of alternative estimators, can be checked for consistency.

8.7.2 The Wilcoxon Rank Sum Test (RST)

Ranking is a non parametric procedure for conducting comparisons between two or more than two (forecasting) methods. The Sign Test, Wilcoxon Rank Sum Test (Wilcoxon, 1949), Mann-Whitney U-Test, Wald-Wolfowitz Runs Test, are examples of non-parametric ranking tests for conducting pair-wise comparisons (Harnett and Soni, 1991). The first two are equivalent to the *t*-test for matched-pairs whereas the latter two correspond to the *t*-test, difference between the means of two independently drawn samples[4]. The Kruskal-Wallis Test is often used to indicate if there is a significant difference between more than two methods.

As discussed in section 8.3 of this chapter, the main advantage of non-parametric tests is that no assumptions are required for the population probability distribution. Moreover the test statistic associated with most methods can be standardised, allowing the use of the more familiar normal distribution (Stekler, 1991). The main

[4] In Wilcoxon (1947) another version of the test appears equivalent to the t-test, difference between population means.

disadvantage of non-parametric tests is that they always sacrifice power in terms of using all available information to reject a false null hypothesis.

Although it will be argued later on in this chapter that ranking non parametric procedures cannot (or it is preferable not to) be used for direct pair-wise comparisons, it was found that a different interpretation of the Wilcoxon Rank Sum Test could be employed in order to determine which methods consistently under or overestimate the level of demand.

The null hypothesis tested by the Wilcoxon Test is that the median difference between two populations equals zero. The test considers the magnitude of the difference between each matched pair. These magnitudes are then ranked according to their absolute value. Then each of these ranks is given either a positive (+) sign or a negative (-) sign, depending on whether the error produced by method *A* was larger than the error produced by method *B* (the plus sign) or vice-versa (the minus sign).

Now, if the null hypothesis is true, we would expect the sum of the ranks with the plus signs to be approximately equal to the sum of the ranks with minus signs. The inference, if the two sums differ by very much, would be that the two populations (methods) are not identical.

If we let T_+ equal the sum of the positive ranks and T_- equal the sum of the absolute value of the negative ranks, the test statistic T is defined as the minimum of T_+ and T_-; that is $T = \min\{T_+, T_-\}$. An interpretation of the test, that matches our specific requirements for reporting bias results, would be the following:

- Calculate the Mean Signed Error (ME) for a particular method on each one of the time series considered in the same way as we did in sub-section 8.7.1.
- Rank the errors according to their absolute value. There is no need to assign a plus or minus sign since the original ME sign can be used.
- Calculate the sum of the ranks for the positive and negative MEs
- H_0: The median of the estimates produced by the particular method is zero

  H_1: The median of the estimates produced by the particular method is non-zero.

To test the null hypothesis T_+ or T_- can be compared with critical values (Wilcoxon, 1949). Alternatively if the sample size is at least 8 (Harnett and Soni, 1991), a normally distributed statistic, Z , can be formulated as:

$$Z = \frac{W_1 - \frac{1}{4}\kappa(\kappa+1)}{\sqrt{\frac{1}{24}\kappa(\kappa+1)(2\kappa+1)}} \tag{8.4}$$

where:

W_1 is the sum of the ranks of one set, and
κ is the total number of errors (data series) considered

The test does not consider the size of the forecast errors produced by the estimation procedure under concern but rather the sign of those errors. Nevertheless we can now check whether the results obtained by the ME itself, applied as a descriptive measure, can be replicated or not. Comparison of the results obtained by the two accuracy measures will eventually indicate any scale dependencies in calculating the average ME itself and will validate our conclusions at minimum computational effort.

8.8 The empirical validity of the categorisation rules

The second objective of the accuracy comparison exercise, as defined in section 8.1 of this chapter, is to check in practice the validity of the theoretical results generated in chapter 6. That is, we want to assess the extent to which the theoretical categorisation rules developed earlier in the thesis reflect real world scenarios. The rules were derived based on a mathematical analysis of the MSE because of the theoretical properties of this particular accuracy measure. As a consequence, the first obvious candidate accuracy measure for meeting our objective is the MSE itself. Other accuracy measure(s) though, that could not have been considered during the theoretical part of the thesis, can now also be employed, in order to establish the credibility of the categorisation rules/schemes proposed. In the following sub-sections the MSE and its alternatives are discussed.

8.8.1 Mean Square Forecast Error (MSE)

The MSE is defined as:

$$MSE = \frac{\Sigma\left(Y_{t+L} - Y'_{t,L}\right)^2}{n} = \frac{\Sigma e_{t+L}^2}{n}, \quad t = 1, 2, \ldots\ldots, n \tag{8.5}$$

As discussed in previous chapters, equation (8.5) is similar to the statistical measure of the variance of forecast errors but not quite the same since it does take bias into account.

Although the MSE has desirable statistical properties, it also has significant disadvantages. Absolute error measures, like the MSE, when they are summarised across series are not easily interpreted unless the data series can be thought of as a sample from a well-defined population of series. The MSE statistics are scale dependent and without the necessary homogeneity of the population the MSE is uninterpretable (Fildes, 1983; Gardner, 1983; Newbold, 1983).

Chatfield (1992) argued that once we have applied the same forecasting method to a group of series, it could be disastrous to average raw MSEs across series, as MSE is scale dependent. This was done with the M-Competition (Makridakis et al, 1982) results, and, according to Chatfield, the MSE results from this competition should be disregarded.

In the context of our real data simulation experiment MSE will not be used for generating summary results across all series. Our approach will rather be as follows: The MSE associated with each method will be calculated for each demand series. Demand series characteristics will also be calculated and reported as part of the simulation output along with each method's MSE performance. Therefore, for any specific series we will be able to link the relationship between different methods' accuracy to the mean and variance of the demand sizes (when demand occurs) and the mean inter-demand interval. Moreover, for every specific series the theoretically expected superior estimator can also be identified. A chi-square test can then be employed in order to compare expected and observed frequencies for each of the decision areas formed by the categorisation rules proposed in chapter 6.

The approach discussed above should generate reliable results about the extent to which our categorisation rules apply to a real situation. Nevertheless, there is no reason at this stage why we should restrict our analysis to the MSE results only. Other error measure(s) can also be employed, which are not so sensitive to the presence of extreme observations (outliers), and which can also be averaged across series in order to supply us with summary error statistics. More conclusive results can then be generated about the conditions under which one method performs better than one or all other methods.

For example, the ME could be used in order to generate accuracy results for all methods on each series. All the data series can then be divided in four or five sets, depending on the theoretical categorisation rule or scheme under consideration, and summary results can then be generated, and consequently tested for statistical significance, for each one of those sets of data. The statistically significant superiority of the theoretically expected best estimator, on each of the decision areas, will obviously increase significantly our trust in the categorisation rules.

8.9 The forecast error

The last two objectives of the accuracy comparison exercise, as defined in section 8.1 of this chapter, are to determine which is the most accurate forecasting method and to specify under what conditions one method performs better than one or all other methods. As discussed above, the latter objective is associated with testing whether the chapter 6 results can be replicated for other accuracy measures or not. For this to be done, certain accuracy measures need to be selected from the plethora that have been proposed in the academic literature.

The fourth objective of the comparison exercise (determine which is the most accurate forecasting method) will be met by generating results, across all series, using the accuracy measure(s) chosen. Conclusions will be reached only after testing the results for statistical significance. In order to meet the third objective (identify the conditions under which one method is more accurate than others), the analysis will be repeated for different sub-samples that correspond to theoretically designated decision areas.

Before considering accuracy measures for the purpose of comparing alternative methods in an intermittent demand context, it would be wise to determine which ones can be computed, taking into account that there are some zero demand time periods. As stated in sub-section 8.4.1, computational considerations dictate that, whatever measure is used, the possibility of division by zero must never exist.

Accuracy measures can broadly be categorised as:

- *Absolute*: the forecast error in a particular time period is expressed in an absolute or square form.

- *Relative to a Base*: the forecast error produced by the method under consideration in a particular time period is related to some benchmark, usually the forecast error produced by the naïve 1 method ($Y'_{t,L} = Y_{t-1}$) for the same time period.

- *Relative to another method*: the forecast error for a particular time period produced by one method is related to that produced by one other method for the same time period.

- *Relative to the series*: the forecast error produced for a particular time period is related to either the actual demand in the time period under concern, the forecast itself or an arithmetic (equal weight) average of both.

(This is an adaptation of a categorisation scheme originally proposed, in a significantly different form, by Makridakis and Hibon, 1995.)

Once the errors have been expressed in one of these forms, arithmetic means, geometric means or medians are used to summarise the errors for all the time periods in a particular time series.

In table 8.1 some of the accuracy measures that appear most often in literature and/or most commonly used in practice are categorised, based on the above discussed scheme, and their relevance to an intermittent demand context is also indicated. For the purpose of deciding whether an accuracy measure is relevant to an intermittent demand context or not, we consider the accuracy measures' suitability for non-ratio scaled data.

	Relevance to Intermittent Demand	
	Yes	No
Absolute	ME MAE MSE RMSE Ranks	
Relative to a Base (naïve 1)		U-Statistic Batting Average GMRAE MdRAE
Relative to another method	RGRMSE RARMSE % Better Ranks	
Relative to the series	MAPEFF $MAPE_{sym}$	MAPE MdAPE

Table 8.1. Accuracy measures

ME:	Mean Signed Error
MAE:	Mean Absolute Error^
MSE:	Mean Square Forecast Error
RMSE:	Root Mean Square Error
Ranks:	Ranking non-parametric test procedure
U-Statistic:	Theil's U-Statistic^

Batting Average	McLaughlin's Batting Average^
GMRAE:	Geometric Mean Relative Absolute Error^
MdRAE:	Median Relative Absolute Error^
RGRMSE:	Relative Geometric Root Mean Square Error
RARMSE:	Relative Arithmetic Root Mean Square Error
% Better:	Percentage of times Better
MAPE:	Mean Absolute Percentage Error
MdAPE:	Median Absolute Percentage Error^
MAPEFF:	Mean Absolute Percentage Error From Forecast
$MAPE_{sym}$:	Symmetric Mean Absolute Percentage Error

^For a definition see Appendix 8.A

From table 8.1 it is apparent that not all accuracy measures can be considered since some of them cannot even be defined in an intermittent demand context. All "relative to the series" accuracy measures, in which the errors are related to the actual demand, need to be excluded from our analysis, since demand will very often be zero. In the same way all "relative to a base" accuracy measures (Theil's U-Statistic, Batting Average, Geometric Mean Relative Absolute Error, Median Relative Absolute Error) need to be excluded since the forecast error produced by the naïve 1 method could very often be zero as well.

8.9.1 Absolute accuracy measures

The absolute error measures, although they can be computed, cannot be considered further, since when averaged across many time series they do not take into account the scale differences between them (see also sub-section 8.8.1).

When those measures are averaged across series they can very often lead the analyst to unreliable conclusions. Chatfield (1988), in a re-examination of the M-Competition data (Makridakis et al, 1982), showed that five of the 1001 series considered in the competition dominated the RMSE rankings. The remaining 996 series had little impact on the RMSE rankings of the forecasting methods.

The only exception to the exclusion of absolute measures may be the ME, the selection of which was discussed in sub-section 8.7.1. Based on the arguments presented earlier, it is decided to use the ME not only for reporting bias results but also for directly comparing alternative intermittent demand estimation procedures.

8.9.1.1 Ranks

Ranking itself as an accuracy measure is a non parametric procedure (see also sub-section 8.7.2). Rankings do not refer to a single method, as the rest of the absolute measures do, and in that sense sometimes appear attractive in generating conclusions about alternative methods' performance. In the case of this research all pair-wise comparison ranking tests could have been used in order to indicate if there is a significant accuracy difference between the two methods under concern.

The errors produced by the two forecasting methods expressed in an absolute form (Mann-Whitney U-Test, Wald-Wolfowitz Runs Test) or the absolute difference between the absolute errors produced by the two methods (Sign Test, Wilcoxon Rank Sum Test) would be ranked in ascending order and an appropriate test statistic would be employed in order to determine statistically significant difference. In the former case the non parametric tests could be characterised as "absolute" in the sense that they do not consider the relationship between the error produced by the forecasting methods under concern and therefore are not considered further for the accuracy comparison exercise. In the latter case the tests could be characterised as "relative to another method" since it is the absolute difference between the errors rather than the absolute errors themselves that are pooled and ranked. Nevertheless another non parametric procedure, the Percentage of times Better, is generally regarded as more intuitive and it is the non parametric procedure to be finally selected for reporting accuracy performance results.

8.9.2 Accuracy measures relative to the series

As stated in the beginning of section 8.9, accuracy measures "relative to the series" cannot be defined in the context of this research when demand appears in the denominator (MAPE, MdAPE) since demand will very often be zero. Nevertheless the MAPE is discussed in order to consider the evolution of two relatively new accuracy measures that appear in the literature, the $MAPE_{sym}$ and the MAPEFF.

8.9.2.1 The Mean Absolute Percentage Error (MAPE)

The MAPE is defined as:

$$MAPE = \frac{\Sigma\left|\frac{Y_{t+L}-Y'_{t,L}}{Y_{t+L}}\right|}{n}*100 = \frac{\Sigma\left|\frac{e_{t+L}}{Y_{t+L}}\right|}{n}*100\text{ , } t = 1, 2, ..., n \qquad (8.6)$$

MAPE is probably the most widely used unit and scale free method (Armstrong and Collopy, 1992). MAPE is a relative measure that takes into consideration the unit of measurement and the scale of the data and expresses forecast errors as a percentage of the actual data. This is its biggest advantage as it allows us to average all MAPEs associated with a particular method across many time series. In fact MAPE adds control for any scale dependencies to the MAE.

The biggest disadvantage of the MAPE is that is relevant only for ratio scaled data (series that do not contain zero demand time periods). Therefore in an intermittent demand context, we would not be able to calculate the MAPE associated with the use of a method in any of the series.

Another disadvantage is that the MAPE lacks a statistical theory similar to that, for example, associated with the MSE. MAPE has been rejected for comparison purposes by Fildes (1992) on statistical grounds. The sampling distribution for the MAPE measured across series is often badly positively skewed and the accuracy measure suffers from being particularly affected by observations close to zero (Gardner, 1983).

MAPE puts a heavier penalty on forecasts that exceed the actual demand rather than on those that are less than the actual demand (Makridakis, 1993). For example the MAPE is bounded on the low side by an error of 100% but there is no bound on the high side. This asymmetry can be expressed in a different way. MAPE gives smaller percentage errors when actual demand is larger than the forecast, than when actual demand is smaller than the forecast. If, for example, actual demand for period t is 6 units and the forecast equals 4 units then the Absolute Percentage Error (APE) is 33.3%. However when actual demand is 4 units and the forecast equals 6 units, the

APE equals 50%. This difference in absolute percentage errors can create problems when the actual demand is very small and the forecast is big because the APE can become extremely large.

8.9.2.2 The Symmetric Mean Absolute Percentage Error ($MAPE_{sym}$)

The problems of asymmetry can be corrected (Makridakis, 1993) by introducing the Symmetric Mean Absolute Percentage Error ($MAPE_{sym}$).

The $MAPE_{sym}$ is defined as:

$$MAPE_{sym} = \frac{\Sigma \left| \frac{Y_{t+L} - Y'_{t,L}}{\left(Y_{t+L} + Y'_{t,L}\right)/2} \right|}{n} * 100 = \frac{\Sigma \left| \frac{e_{t+L}}{\left(Y_{t+L} + Y'_{t,L}\right)/2} \right|}{n} * 100, \quad t = 1, 2, ..., n \qquad (8.7)$$

Expression (8.7) does not add to the statistical support of MAPEs. Nevertheless, even though it was not Makridakis' intention when he introduced the $MAPE_{sym}$, the corrected accuracy measure can be used for data containing zeroes.

The $MAPE_{sym}$ seems to have many desirable properties. It is relative, it accounts for any changes in the underlying demand pattern that occur over the forecast horizon, it is not unduly affected by outliers, it is easily computed and interpreted and, most importantly for the purpose of this research, is relevant not only for ratio-scaled data.

Nevertheless, some problems arise in applying the $MAPE_{sym}$ to compare alternative intermittent demand estimation procedures. Considering that the greatest percentage of the observations in any of the demand data series included in our sample is zero (i.e. no demand data periods) the APE_{sym} for those periods turns out to be 2 (or 200%) irrespective of which estimation procedure has been utilised. Therefore, when this accuracy measure is used, the comparison of the intermittent demand forecasting methods can refer to the non-zero demand time periods only. The alternative methods' forecasting performance in the zero demand time periods cannot be allowed to affect the final results, since all methods appear to perform equally.

Moreover, although the modified MAPE is symmetric when $Y'_{t,L}$ and Y_{t+L} are interchanged, the measure creates a new problem of asymmetry which is more likely to be of practical concern than the problem resulting from the interchange (Goodwin and Lawton, 1999). The MAPE does not treat single errors above the actual value any differently from those below it. So if actual demand is 4 units, errors of +2 and -2 units both result in an APE of 50%. The $MAPE_{sym}$ does treat them differently. So in the same example the errors of +2 and -2 units would result in APE_{sym} s of 40% and 66% respectively.

The important thing to note is that not only an error of -*x* units results in a different APE_{sym} from an error of +*x* units but the rate of increase in the APE_{sym}, as the APE increases, depends also on the sign of the original error. For positive actual and forecast values (as is always the case for a demand data series) the $MAPE_{sym}$ has an upper bound of 200%. However while this bound is reached when the forecast value equals zero for positive errors, for negative errors it is only approached as the forecast error tends to minus infinity (Goodwin and Lawton, 1999):

$$MAPE_{sym} = \frac{200}{\frac{2Y_{t+L}}{e_{t+L}} + 1}, \quad when \; e_{t+L} < 0, \qquad t = 1,2,......., n \tag{8.8}$$

Considering the disadvantages of the $MAPE_{sym}$ discussed above, it has been decided not to use this measure in the accuracy comparison exercise.

8.9.2.3 The Mean Absolute Percentage Error From Forecast (MAPEFF)

In an internal decision making context the forecasts obtained by any estimation procedure may be viewed as the potential plan the decision makers may choose to implement. In such an environment the out-of-sample Absolute Percentage Error should represent the average extent to which the plan will not be realised, and/or additional action by the company will be required. It has been argued (Pearson and Wallace, 1999) that for a business assessment of the forecast performance, the relevant benchmark is the plan (i.e. the forecast itself) rather than the actual demand.

Hence, the forecast error should be calculated as the percent variation from the forecast and not the actual demand or an average of both (as per Makridakis, 1993).

The MAPEFF is defined as:

$$MAPEFF = \frac{\Sigma\left|\frac{Y_{t+L} - Y'_{t,L}}{Y'_{t,L}}\right|}{n} * 100 = \frac{\Sigma\left|\frac{e_{t+L}}{Y'_{t,L}}\right|}{n} * 100\ ,\ \ t = 1, 2, \ldots\ldots., n \qquad (8.9)$$

According to this approach, the purpose of using an accuracy measure is not to see how close the forecasts are to the actual values but rather to assess how close the performance is to the plan that has been set.

Moreover Pearson and Wallace (1999) argued that many papers that recommend an accuracy measure "*tend to focus on the mechanisms of computation as if they were mutually exclusive from the context of application* (p. 2)". In fact, though, the context of application should be capable of leading the choices for computation, rather than permitting the mathematical consequences of the computation to lead the choice.

According to the authors, many companies have adopted the MAPEFF despite the fact that forecasting textbooks and all forecasting software packages examined by them use the traditional Percentage Error (PE) definition. Moreover the Kahn (1998-1999) survey of sales forecasting performance measures found that the MAPE is the most popular statistic for measuring error. However, of the 26 firms using MAPE as the error statistic, he found that 12 calculate MAPE with actual in the denominator, 10 with the forecast in the denominator, and 2 use MAPEs calculated in both ways.

The main criticism that the authors received at the 19th International Symposium of Forecasting was the lack of a common base for comparison purposes. The assertion was that, regardless of its shortcomings, the traditional MAPE gives a common base for evaluating alternative forecasting methods, whereas the base in MAPEFF will be different for each method used. Another argument was that the outside analyst does not know the alternative plans (sets of forecasts produced by alternative estimation procedures) considered by management during the selection process and therefore the focus is not on what was the best among the plans considered.

In addition four more problems can be identified in a potential application of the MAPEFF in this research:

1. One general problem related to the use of MAPEFF is that it puts a heavier penalty on forecasts that underestimate the actual demand (pessimistic plans) rather than on those that overestimate it (optimistic plans). The asymmetry can be expressed in a way parallel to that described in the case of the $MAPE_{sym}$. If for example our forecast (i.e. the short range plan) for period t is 6 units and the demand turns out to be 4 units then the APEFF is 33.3%. However when forecast is 2 units and the demand when it is realised equals 4 units, the APE equals 100%. Figure 8.1 illustrates the behaviour of the APE, APE_{sym} and APEFF for specific errors expressed in a signed form as a percentage of the actual demand.

2. MAPEFF would be particularly meaningful from a marketing and sales perspective where the forecast is in fact the target and marketing and sales activities are oriented towards influencing external demand and eventually meeting this target. Demand generated from customers can be influenced by promotional activities. Moreover in that case we most probably refer to an aggregate forecasting level where forecasts generated focus on estimating demand in the medium term for a group of products rather than in the short term for an individual SKU.

   In the intermittent demand context, in particular, one could argue that the SKUs under concern are amongst the slowest movers and are not generally the focus of marketing activity. Forecasts are routinely generated and will almost always be independent of any plans developed in the sales or marketing department.

3. The MAPEFF should be considered inappropriate for forecasting accuracy comparisons when biased estimation procedures enter the comparison, which is clearly the case in this research.

4. As in the case of the $MAPE_{sym}$, a potential application of the MAPEFF can refer to the non-zero demand time periods only. The APEFF would equal 1 (or 100%) for all zero demand time periods, irrespective of which estimation procedure had been utilised. Therefore the alternative method's performance in the zero demand time periods cannot be allowed to affect the final results since all methods would appear to perform equally.

Based on the above it has been decided not to use MAPEFF as an accuracy measure in the forecasting accuracy comparison exercise.

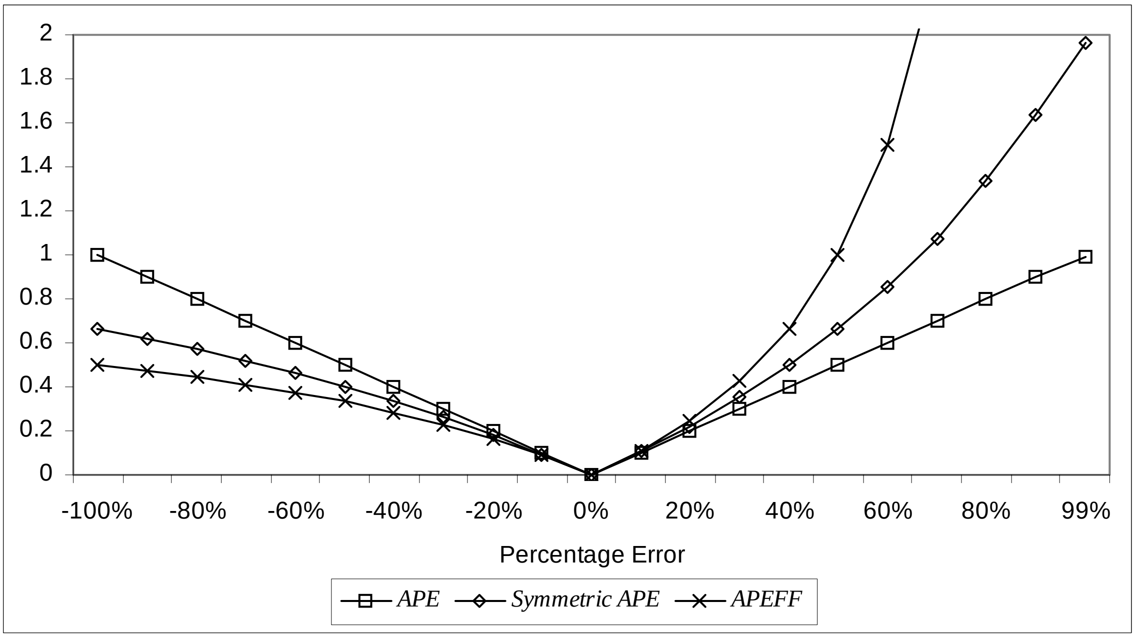

Figure 8.1. The asymmetry of accuracy measures relative to the series

8.9.3 Accuracy measures relative to another method

Accuracy measures relative to another method are widely used for pair-wise comparisons indicating how much better or how many times one method performs more accurately than another does. Descriptive error measures take into account the actual size of the forecast error in order to indicate how much better one method performs in comparison with one other. Non parametric tests would consider only the relationship between the absolute forecast errors produced by the two estimation procedures under concern in order to indicate whether the difference in the number of times that each method shows superior performance is statistically significant or not. For the purposes of this research, Ranks and "Percentage of times Better" could be used as the basis of non-parametric procedures. The Percentage of times Better (PB) is preferred as it can be more easily interpreted (Makridakis, 1993) and is generally regarded as a more intuitive non parametric procedure (Makridakis and Hibon, 1995).

The PB will tell us how many times a method performs better than another but not by how much. For this to be done a descriptive accuracy measure needs to be selected. The entire set of relative to a base, relative to the series and absolute accuracy measures (with the exception of ME) have been rejected for reasons given in the preceding sections. Therefore descriptive, relative to another method, accuracy measures will now be considered.

The relative to another method accuracy measures express the error produced by an estimation procedure as a fraction of the error generated by another procedure. In a particular time series, the error can be expressed in an absolute, signed, squared, or absolute percentage form and summarised by using arithmetic means, geometric means or, more rarely, medians. (In the case that a geometric summarisation occurs the signed errors cannot be considered.) The same is repeated for the other methods over that series and the errors are expressed (pair-wise comparisons) in a relative form, e.g. the Geometric MSE given by method A divided by that given by method B will form the Relative GMSE. In the following sub-section the RGRMSE is presented as a means of discussing the advantages and disadvantages associated with all relative to another method accuracy measures.

8.9.3.1 Relative Geometric Root Mean Square Error (RGRMSE)

The RGRMSE for methods *A* and *B* in a particular time series is defined as:

$$RGRMSE = \frac{\left(\prod_{t=1}^{n}\left(Y_{t+L} - Y'_{A,t,L}\right)^2\right)^{\frac{1}{2n}}}{\left(\prod_{t=1}^{n}\left(Y_{t+L} - Y'_{B,t,L}\right)^2\right)^{\frac{1}{2n}}} \qquad (8.10)$$

Fildes (1992) discussed extensively the theoretical properties of the RGRMSE. His analysis is as follows:

Assume that the squared errors produced by a particular method at different points in time (for a particular series) are of the form:

$$\left(Y_{t+L} - Y'_{M,t,L}\right)^2 = e^2_{M,t+L} = \varepsilon^2_{M,t+L}\ u_{t+L} \tag{8.11}$$

where:

u_{t+L} are assumed to be positive and can be thought of as errors due to the particular time period affecting all methods equally, while the $\varepsilon^2_{M,t+L}$ are the method's (*M*) specific errors.

According to Fildes, such a model corresponds to the case where the data and subsequently the errors are contaminated by occasional outliers. In that case MSE and RMSE could potentially be dominated by the errors due to the particular time period and subsequently, comparison of any two forecasting methods would also be unduly affected. Fildes shows that the use of a geometrically (rather than arithmetically) averaged RMSE (GRMSE) expressed in a relative way (RGRMSE of one method compared to another) is independent of the u_{t+L}.

Considering (8.10) and (8.11), the RGRMSE for methods *A* and *B* can also be defined as:

$$RGRMSE = \frac{\left(\prod_{t=1}^{n} e^2_{A,t+L}\right)^{\frac{1}{2n}}}{\left(\prod_{t=1}^{n} e^2_{B,t+L}\right)^{\frac{1}{2n}}} = \frac{\left(\prod_{t=1}^{n} \varepsilon^2_{A,t+L}\ u_{t+L}\right)^{\frac{1}{2n}}}{\left(\prod_{t=1}^{n} \varepsilon^2_{B,t+L}\ u_{t+L}\right)^{\frac{1}{2n}}} = \left(\prod_{t=1}^{n} \frac{\varepsilon_{A,t+L}}{\varepsilon_{B,t+L}}\right)^{\frac{1}{n}} \tag{8.12}$$

whereas the R(A)RMSE (Relative Arithmetic Root Mean Square Error) is given by:

$$R(A)RMSE = \frac{\left(\dfrac{\sum_{t=1}^{n}\left(Y_{t+L}-Y'_{A,t,L}\right)^2}{n}\right)^{\frac{1}{2}}}{\left(\dfrac{\sum_{t=1}^{n}\left(Y_{t+L}-Y'_{B,t,L}\right)^2}{n}\right)^{\frac{1}{2}}} = \frac{\left(\sum_{t=1}^{n} e_{A,t+L}^2\right)^{\frac{1}{2}}}{\left(\sum_{t=1}^{n} e_{B,t+L}^2\right)^{\frac{1}{2}}} = \left(\frac{\sum_{t=1}^{n} e_{A,t+L}^2}{\sum_{t=1}^{n} e_{B,t+L}^2}\right)^{\frac{1}{2}} \tag{8.13}$$

which is not independent of the u_{t+L}.

At this stage, we argue that it is the introduction of the Relative Geometric (rather than relative arithmetic) summarisation of the errors and not the error expression form, i.e. RMSE, that ensures independence of the u_{t+L}. In other words, the RGMSE and RGMAE would have exactly the same effect as RGRMSE, leaving us only with the errors produced by the methods under concern without any specific time period influences. This was not noticed by Fildes (1992).

In the past the RARMSE (Relative Arithmetic Root Mean Square Error) has been often used in order to indicate the improvement ratio of employing one method over one other, in an across series accuracy comparison exercise. Fildes, though, with a simple example, demonstrates that this accuracy measure does not take into consideration scale differences so that one method may appear to be the best simply because it performed well on a series with very large observations.

If a forecasting method has an error (expressed in any possible form) on one series 10% higher than an alternative and on another 10% lower, the GRMSEs are equal when they are averaged across these two series. If for example for series 1 method *A* has an error of 110 compared to method *B* at 100 while on series 2, method *A* has an error of 1000 compared to method *B* that has an error of 1100 then:

$$RGRMSE = \left\{\left(\frac{110}{100}\right)^2\left(\frac{1000}{1100}\right)^2\right\}^{1/4} = \left\{\frac{110}{100}\frac{1000}{1100}\right\}^{1/2} = 1$$

but

$$R(A)RMSE = \left(\frac{110^2 + 1000^2}{100^2 + 1100^2} \right)^{1/2} = 0.911$$

We conclude that a relative geometric, rather than relative arithmetic, summarisation of errors ensures minimising the effect of any outliers in a single series evaluation as well as any scale dependencies in an across series accuracy comparison exercise. Independently of how the errors are expressed (absolute or squared form), a relative arithmetic accuracy measure is always affected by extreme series observations and/or is scale dependent. The use of a relative geometric accuracy measure is to be preferred.

Moreover, we have argued that it is the relative geometric summarisation of the errors and not the error expression form that ensures certain desirable properties of an error measure. Therefore, any Geometric Relative measure could be employed for the purpose of this research. The RGRMSE though has been shown in Fildes (op. cit.), by using real data, to have a very well-behaved distribution across series. In addition this accuracy measure was found to be, as theoretically expected, not seriously affected by outliers in a single series analysis. Fildes claimed also that the GRMSE and RGRMSE per series can be modelled by the lognormal distribution, which is an intuitively appealing assumption, confirmed by real data analysis conducted in his research. Therefore considering its “independence” and its desirable statistical properties it has been decided to use the RGRMSE for reporting summary error results in this research.

$$RGRMSE = \frac{\left(\prod_{s=1}^{\kappa} \left(GRMSE_{A,s} \right)^2 \right)^{\frac{1}{2\kappa}}}{\left(\prod_{s=1}^{\kappa} \left(GRMSE_{B,s} \right)^2 \right)^{\frac{1}{2\kappa}}} = \left(\prod_{s=1}^{\kappa} \frac{GRMSE_{A,s}}{GRMSE_{B,s}} \right)^{\frac{1}{\kappa}} \qquad (8.14)$$

where the $GRMSE_{i,s}$ per series, for method i is calculated as:

$$GRMSE_{i,s} = \left(\prod_{t=1}^{n} \left(Y_{t+L} - Y'_{i,t,L} \right)^2 \right)^{\frac{1}{2n}} \tag{8.15}$$

For an across series evaluation where the $GRMSE_{i,s}$ s produced by a particular method in all demand data series are reported, the geometric mean of them could be also found by calculating the arithmetic mean of the natural logarithms of the $GRMSE_{i,s}$ s and then taking the antilogarithm of this arithmetic mean. Considering the large number of series, the arithmetic mean itself will be (from the central limit theorem) normally distributed. As such it is reasonable to assume that the geometric mean can be represented as lognormal. The validity of the latter assumption can be also verified as follows: If the $GRMSE_{i,s}$ s are lognormally distributed and independent (which is the case as it was shown in Fildes' paper), their geometric mean is also a lognormal variate.

The sample size allows us to use the *t*-test (sub-section 8.7.1), in order to assess whether or not the difference between the arithmetic means of the $\log GRMSE_{i,s}$ s produced by two methods differs significantly from zero. Testing whether or not that difference deviates significantly from zero is equivalent to testing whether or not the RGRMSE across series is significantly different from one.

8.9.3.2 The Percentage of times Better (PB)

PB is defined as the percentage of times (observations) that one method performs better than one other. For a particular time series the errors (expressed in any possible form) produced by two methods in each time period would be directly compared without considering the actual error difference in order to identify the percentage of times that one method performs better than the other. Alternatively, if the comparison exercise involves many time series, the summarised errors produced by the two methods in a series would be directly compared and the PB would be calculated across series.

The PB is an easy measure to calculate, with a simple interpretation. It will also be shown, later in this section, that the results given by the PB method can easily be checked for statistical significance. From a theoretical perspective, the method accounts for any changes in the underlying demand pattern, is relative, is not affected by outliers and can be defined for all data.

For intermittent demand data, PB is particularly meaningful since all series and all data periods within each series are considered in order to generate results. The accuracy measure provides precise information about the percentage of times that one method performs better than one other or, to be discussed later, than all the other methods, irrespectively of whether demand did or did not occur in the time period.

In the context of this research, we will apply this accuracy measure in order to generate results across all demand data series. For this to be done, we need to use one or more descriptive accuracy measures that will provide us with results about the alternative methods' performance in each one of the series. Accuracy measures that are not affected by outliers should be chosen for this purpose. So far we have argued for the robustness of the ME and the GRMSE generated per series. Consequently, these error measures will be used for generating PB results.

For the purpose of this comparison exercise the PB is applied as follows:

- Calculate the GRMSE (and ME) per series for methods *A* and *B*
- Compare the GRMSEs (and absolute MEs) for each one of the series across all series

  κ : total number of series (3,000)

  κ_1: number of series that method *A* performs better than method *B*

  κ_2: number of series that method *B* performs better than method *A*

  $p_1 = \frac{\kappa_1}{\kappa}$: sample proportion of series that method *A* performs better than method *B*

  $p_2 = \frac{\kappa_2}{\kappa}$: sample proportion of series that method *B* performs better than method *A*
- Formulate the hypotheses

  H_0: $\pi_1 = \pi_2$

  H_1: $\pi_1 > \pi_2$

  where π_1, π_2 are the corresponding population proportions

- Conduct a statistical test to compare the percentage better (PB) of one method against the null hypothesis that both methods have a 0.5 (50%) probability of performing better than the other method. The binomial distribution can be used to calculate the appropriate probabilities with:

  p (probability of success) $= 0.5$

  κ : number of trials

  κ_1 : number of successes

The binomial variate **B**: κ , p can be approximated by the normal variate with mean κp and standard deviation $\sqrt{\kappa p(1-p)}$ provided that $\kappa p(1-p) > 5$ and $0.1 \leq p \leq 0.9$ or if $\min\{\kappa p, \kappa(1-p)\} > 10$ (Evans et al, 1993) which is the case in our simulation experiment.

The test statistic, Z , can be formulated as follows:

$$Z = \frac{\kappa_1 - \kappa p}{\sqrt{\kappa p(1-p)}} = \frac{\kappa_1 - \frac{1}{2}\kappa}{\sqrt{\frac{1}{4}\kappa}} = \frac{\kappa_1 - 0.5\kappa}{\sqrt{0.25\kappa}} \tag{8.16}$$

which is equivalent to the population proportion test statistic using the normal distribution.

If we set $\kappa_1 = \kappa\, p_1$ we then have:

$$Z = \frac{\kappa_1 - \kappa p}{\sqrt{\kappa p(1-p)}} = \frac{\kappa p_1 - \kappa p}{\sqrt{\kappa p(1-p)}} = \frac{\kappa\left(p_1 - p\right)}{\sqrt{\kappa p(1-p)}} = \frac{p_1 - p}{\frac{\sqrt{\kappa p(1-p)}}{\kappa}} = \frac{p_1 - p}{\sqrt{\frac{p(1-p)}{\kappa}}} \tag{8.17}$$

The statistically significant results generated by the PB measure, applied on GRMSEs per series, will indicate whether any scale dependencies have affected the statistically significant results given by the RGRMSE. In a similar fashion, the statistically significant results generated by the PB measure, applied on MEs, will indicate whether any scale dependencies have affected the statistically significant results given by the ME accuracy measure (used for a direct accuracy comparison rather than for reporting bias).

The PB measure reports the proportion of times that one method performs better than one other method. When more than two estimators are involved in the accuracy comparison exercise, we may also report the proportion of times that one method performs better than all other methods, i.e. the proportion of times that each method performs best. In this case the measure is referred to as Percentage Best (PBt) rather than Percentage Better.

To test the statistical significance of the results, the Z test statistic (difference between population proportions) can be used referring to all methods rather than focusing on pair-wise comparisons. In this case, the GRMSE (or ME) per series can be calculated for all methods across all series and if we now set:

κ_1: number of series that method *A* performs better than all other methods (i.e. method *A* gives the lowest GRMSE/lowest absolute ME)
κ_2: number of series that method *B* performs better than all other methods

we then have:

$p_1 = \frac{\kappa_1}{\kappa}$, proportion of series that method *A* performs better than all other methods,

$p_2 = \frac{\kappa_2}{\kappa}$, proportion of series that method *B* performs better than all other methods

H_0: $\pi_1 = \pi_2$

H_1: $\pi_1 > \pi_2$

where π_1, π_2 are the corresponding population proportions.

The test statistic, Z, can then be formulated as follows:

$$Z = \frac{p_1 - p_2}{\sqrt{\frac{2}{\kappa} p(1-p)}} \quad (8.18)$$

where:

$$p = \frac{\kappa\, p_1 + \kappa\, p_2}{2\kappa} = \frac{p_1 + p_2}{2}$$

8.10 Conclusions

In this chapter we identified appropriate accuracy measures for the purpose of comparing alternative existing and new intermittent demand estimation procedures over a large number of demand data series. The objectives of the comparison exercise and the experimental structure of the simulation (to be described in detail in chapters 9 and 10) are clearly defined in order to enable the identification of specific accuracy measures and methods that collectively will capture the information required.

The Mean (Signed) Error (ME) and a modification of the Wilcoxon Rank Sum Statistic (RSS) are the accuracy measures chosen in order to meet the first objective of the accuracy comparison exercise, namely to determine whether the theoretical results obtained in chapter 4 concerning the bias of alternative estimation procedures, can be replicated in practice or not. The ME is a descriptive absolute accuracy measure and therefore suffers from scale dependency. Nevertheless it is expected to perform quite robustly since the sign of the errors is used in obtaining the final average. The Wilcoxon Rank Sum Test (RST) is a ranking non-parametric procedure and will be used in order to check the validity of the ME results.

The Mean Square Error (MSE) was the obvious accuracy measure to be chosen in order to meet the second main objective of the accuracy comparison exercise, namely to determine the validity of the theoretical rules proposed in chapter 6. In the context of our real data simulation experiment, the MSE will not be used for generating summary results across all series. It will rather be employed in order to determine the extent to which the theoretical categorisation rules reflect real world scenarios. For this to be done, a chi-square test will be used so that we can compare expected and observed frequencies of superior MSE performance across theoretically designated decision areas.

The difficulties associated with selecting accuracy measures, for the purpose of comparing alternative intermittent demand estimation procedures, have often been underestimated in the past by both practitioners and academicians. In order to facilitate the selection process for meeting objectives 3 and 4 (i.e. which is the most accurate forecasting method and under what circumstances one method performs better than another) the most common accuracy measures (descriptive and non parametric) are categorised based on the following scheme:

- Absolute
- Relative to a base (most commonly the forecast obtained by the naïve 1 method)
- Relative to another method
- Relative to the series (in which case the error could be expressed as a percentage of the actual demand, the forecast or an arithmetic, equally weighted, average of both).

The accuracy measures within the categories that are relevant to an intermittent demand context are further discussed and evaluated, and the Mean Error (ME), Relative Geometric Root Mean Square Error (RGRMSE) and Percentage Better (PB) accuracy measures are finally selected. The ME is now applied for a direct comparison between alternative estimation procedures rather than indicating bias results for one method at a time. The RGRMSE has desirable properties in terms of its scale independence and robustness to outliers. The PB is a non-parametric procedure and reports the proportion of series on which one method performs better than one or all other methods. In the latter case, the measure is referred to as Percentage Best (PBt). The comparisons on a single series in order to obtain the PB and PBt results are based on the ME and GRMSE given by the methods considered.

Finally in this chapter it is demonstrated that:

- The Mean Absolute Percentage Error From Forecast (MAPEFF) is not suitable for intermittent demand data

- Goodwin and Lawton's (1999) critique of the use of the Symmetric Mean Absolute Percentage Error ($MAPE_{sym}$) also applies to the MAPEFF
- Fildes' (1992) argument applies to all relative geometric measures rather than the Relative Geometric Root Mean Square Error (RGRMSE) only.

CHAPTER 9

Inventory Control for Intermittent Demand Items

9.1 Introduction

The purpose of this chapter is to propose an inventory control procedure suitable for the evaluation of forecasting methods for intermittent demand. This stock control method will be used in chapter 11 to generate service level and inventory holding results, regarding the performance of alternative estimation procedures on real demand data. To develop such an approach, the nature of our real demand data sample and the information available for each SKU need to be considered in detail. Moreover, existing stock control approaches for dealing with intermittence (drawn from the literature or used by practitioners) will be critically reviewed. This will allow a decision on their degree of relevance to this research but it will also assist in developing our understanding about issues of particular importance to managing the inventories of intermittent demand items.

A number of theory based conclusions have been derived thus far in this thesis with respect to forecasting intermittent demand. The practical validity of these conclusions will be assessed in the following chapter where forecasting performance will be simulated on real demand data. The results of this simulation exercise will indicate the extent to which our theoretical conclusions hold for real world cases. Expected forecasting accuracy improvements, though, are not always reflected in an inventory control situation. In addition, even if an improvement is achieved by employing the most accurate estimation procedure, the degree of improvement is unknown. It is essential to develop an appropriate inventory control simulation to allow for an evaluation of the empirical utility of our theoretical findings.

In this chapter, it will be argued that a periodic review system of the (T, S) form (every T periods enough is ordered to raise the inventory position up to S) is the most appropriate for the purpose of this research from a conceptual, computational and practical perspective also. The system will have the following characteristics:

demand not satisfied directly from the stock is fully backordered; the variability of demand over the lead time plus one review period is estimated by using the smoothed MSE approach; the distribution of demand over lead time plus review period is represented by the Negative Binomial Distribution (NBD). Moreover three managerial constraints are chosen for generating results: a specified customer service level, a specified shortage fraction per unit value short and a specified emergency delivery fraction per unit value short.

9.1.1 Data used

The intermittent demand data set available for simulation comes from the automotive industry and consists of 3,000 files. Each file consists of 24 demand time periods, demand being recorded monthly. The data has been provided by a forecasting and inventory control software package manufacturer (Unicorn Systems (UK) Ltd.). The exact nature (product description) of the SKUs has not been specified and no information has been revealed as to which part of the supply chain we refer to. The only other information accompanying the data sample is the following:

- An estimate of the annual inventory holding charge per unit value (25%)
- An estimate of the backorder cost per unit value short (28%)
- The lead time, which is approximately 3 months.

All three estimates refer to the whole data sample. No unit cost information is available for any of the SKUs.

In this chapter we are oriented towards selecting an inventory control method suitable for the evaluation of alternative estimation procedures on the particular demand data sample.

Our decision will be influenced by the following factors:

(a) Nature of the demand files

(b) Information for each file. The lack of unit cost information for our data sample is a major constraint and many intuitively appealing options will have to be disregarded.

(c) Objectives of this chapter. We endeavour to select a method that has a theoretical foundation but also corresponds to real world practices. The stock control method will be used for the purpose of comparison only and in that sense does not have to be theoretically optimal. That is, as long as the selected method enables meaningful comparisons to be made and as long as results can be generated for the empirical sample, the method can be further considered even though it may not be theoretically the best.

9.1.2 Notation

We denote by:

T: review period (1 month)

R: review period expressed as a fraction of the year (in our case $R = 1/12$).
It follows that the expected number of reviews in a year is $1/R$

D: average annual demand in units

A: ordering cost (£)

c: unit cost (£)

L: lead time (number of unit time periods)

$p(x)$: probability density function of demand x over $L+T$

P_2: fraction of demand satisfied directly from stock

B_1: shortage cost per occurrence of stockout (£)

B_2 : shortage cost per unit value short (%)

$B_2 c$: shortage cost per unit (£)

B_3 : overnight delivery premium charge per unit value (%)

$B_3 c$: extra acquiring cost per unit (£)

I : holding charge/year/item (%)

Ic : expected annual inventory holding cost per unit (£)

9.1.3 Chapter structure

This chapter is structured as follows:

First we discuss the main issues involved in the development of an inventory control system and we justify our decisions in the context of this research. For each important decision that needs to be made towards developing our inventory control system, the possible approaches are identified and one or more are selected considering: (a) the nature of our real demand data files, (b) additional information available for each one of those files and (c) the objectives of this chapter. In particular, in section 9.2, we review the literature on managing the stock of intermittent demand items and we specify the type of the system to be employed for the purpose of our research (periodic of the (T, S) form). In section 9.3 we discuss some issues related to setting the numerical values of the control parameters required in the inventory control system to decide when and how much to order. In section 9.4 we focus our attention on the calculation of the safety stock to be kept in the system. Finally a statistical distribution is selected, in section 9.5, in order to represent demand over lead time (plus one review period).

Having specified all the details of the system that will be employed we derive, in section 9.6, the conditions that allow the optimisation of the replenishment level under different managerial constraints imposed on the system. In section 9.7 technical details of the simulation experiment are discussed. In particular we refer to the initial conditions used (initial estimates, starting stock etc), the updating procedure of the control parameters, the content of the simulation output and the tests employed for demonstrating statistical significance. Finally the conclusions of the chapter are presented in section 9.8.

9.2 The inventory control process

The fundamental purpose of any inventory control system is to provide answers to the following four questions (Brown, 1967):

1. How should the stock status records be maintained?
2. How often should the test for re-ordering be tried?
3. When should a replenishment order be placed?
4. How large should the replenishment order be?

There are essentially only two ways of "posting" the stock status records. One is to add receipts and to subtract demand as they occur. In this case, each transaction triggers an immediate updating of the status and in consequence this type of control is known as "transactions reporting" (Silver et al, 1998). The second method of updating the stock status records is to do it periodically, that is, an update interval (T) elapses between two consecutive moments at which the stock level is known.

Once the stock status records have been updated, the inventory control system can then check the stock status against one or more control numbers so that a decision can be made about when and how much to order.

It is important to note at this stage that a continuous recording of each transaction does not necessarily imply a continuous review of the stock requirements. Porteus (1985) commented that: "*What really matters is not how the inventory levels are monitored but the relationship between recognising that an order should be placed, placing the order, and receipt of that order. Many, if not most, transactions reporting systems are equivalent to periodic review systems* (p. 145)". For example, if the inventory records are up to date on line continuously but the orders to a given supplier are issued at the end of the day (or theoretically at the end of any unit time period) then the system is one of periodic review with inventory levels being reviewed once a day.

9.2.1 Periodic versus Continuous review of the inventory level

Items may be produced on the same piece of equipment, purchased from the same supplier or shipped using the same transportation mode. In any of these situations, co-ordination of replenishments may be attractive (Silver et al, 1998). In such a case periodic review is appealing since all items in the co-ordinated group can be given the same review interval. Sani (1995) argued that re-order interval or product group review systems are the most commonly used in practice for intermittent demand items. Consequently, he explored various periodic inventory systems proposed in the academic literature and/or used by practitioners, to determine the most appropriate ones for the demand patterns under concern. The "simple" systems (i.e. lacking theoretical support) considered for his research were of the periodic (s, S) form (to be discussed later in this section) and employed by car and agricultural machinery spare parts dealers. These were the same dealers that provided the real demand data used in Sani's research.

The major advantage of continuous review is that, to provide the same level of customer service, it requires less safety stock (hence, lower inventory holding costs) than periodic review. This is because in a periodic review system, safety stock is used to compensate for any uncertainties regarding demand over the lead time plus one inventory review period. Under continuous review, the safety stock is calculated by considering lead time demand requirements only.

Moreover, for intermittent demand items very little costs are incurred by continuous review as updates are only made when a transaction occurs. The relationship between ordering cost and inventory holding charge could be further explored so as to decide on the appropriateness of each type of system. Nevertheless, in an intermittent demand context the inventory review period is relatively small in comparison to the average inter-order interval. Periodic review may be more effective than continuous because spoilage or pilferage can be more easily detected. In transactions recording no such automatic review takes place without a transaction occurring.

Although a periodic review system can be theoretically justified for application in an intermittent demand context, the decision about employing this type of system for the purpose of this research is also dictated by the type of data available for simulation. Our intermittent demand data series consist of monthly demand data, which cover the demand history of two years. The data is collected on a monthly basis so that one month can be viewed as the inventory review period ($T = 1$). At the end of every period the stock status will be reviewed and consequently compared against a control parameter so as to decide how much to order.

Considering the nature of the real demand data, there are two possible cases that could be analysed for the purposes of our research:

1. Consideration of forecasts on all points in time and periodic review of inventory
2. Consideration of forecasts on issue points only and periodic review of inventory

The former scenario is compatible and it will be simulated while the latter is not compatible and it will not be further considered for the purposes of our research. If the stock level is to be checked in every period there is little point in not considering the updated forecasts at the end of every period. The forecast review period will be taken to be one month so that the forecasts are updated at the end of each period by any of the estimation procedures considered. Forecasting accuracy results will be generated for both all and issue points in time only while the stock control performance of the estimators will be analysed only for all points in time.

9.2.2 Inventory control systems

The two most commonly encountered continuous review systems are of the (s, Q) or (s, S) form. After each transaction, the available stock (i.e. inventory position = stock on hand + on order – backorders) is compared with a control number, s, variously called an order point, a base stock or a minimum (Brown, 1959). If the inventory position is less than s (or in some cases at or below s) a replenishment order is released. The replenishment order can be for a standard order quantity Q or alternatively enough may be ordered to raise the inventory position to the value S, the replenishment level. If all demand transactions are unit-sized, the two systems are identical because the replenishment requisition will always be made when the inventory position is exactly at s (so that $S = s + Q$). If the demand sizes vary, then the replenishment quantity in the (s, S) system also varies. In this latter case optimisation of s and S occurs in parallel recognising that cost interactions exist between the two control parameters. Alternatively the parallel optimisation may be for s and Q (rather than s and S) in which case the (s, Q) and (s, S) systems are also equivalent in that the replenishment level can be determined as $S = s + Q$ (see for example Wagner, 1975).

In a periodic review context the inventory decision rules most usually take the form of a (T, S) or (T, s, S) system, the former being the one to be employed in this research. Under the regime of both policies, every T periods (constant inventory review interval) enough is ordered to raise the inventory position up to the replenishment level. The difference between the two systems is that the (T, s, S) system requires the inventory position to be less than or equal to s (or in certain cases strictly less than s) before an order is placed. Therefore the (T, S) system always results in higher ordering costs since even a unit-sized transaction during the review interval will trigger a replenishment requisition, whilst the (T, s, S) system will only place an order if the cumulative demand, over review period, exceeds some minimum level. The review interval, for both systems, can be optimised by classical economic lot size computations, meaning that the EOQ can also be expressed as a time supply T (see for example Brown, 1982). The (T, s, S) can be viewed as the periodic implementation of the (s, S) system, as the (T, s, S) reduces to (s, S) for $T = 0$.

In the optimisation of the control parameters process (see section 9.6) the results obtained from the (T, S) system can be easily transferred to a (s, Q) system by substituting s for S, L for $L + T$ and Q for DR (where D is the annual demand). The (T, Q) combination does not take into account the variability of demand and hence should not be applied in a probabilistic demand context.

In the sub-sub-sections that follow, the literature on inventory control for intermittent demand items is reviewed, with the purpose of specifying more precisely the periodic inventory control model that will be used in the simulation experiment. The main approaches to managing intermittent demand stocks will be discussed, mainly with respect to their relevance to our research and the objectives of this particular chapter.

9.2.2.1 The (T, s, S) policies

The (T, s, S) inventory control systems have been claimed, on the basis of theoretical arguments, to be the best for the management of low and intermittent demand items (Sani, 1995). The superiority of such systems has also been demonstrated by means of simulation on real demand data (Sani and Kingsman, 1997). From a practical perspective, though, the computational effort to find the best (s, S) pair in a (T, s, S) system is prohibitive unless we are dealing with an item where the potential savings in the Total Inventory Costs (TIC) are significant (Silver et al, 1998). This is not the case for the SKUs considered for the purpose of this research.

Many (T, s, S) policies have been developed in the academic literature, some giving optimal solutions (e.g. Veinott and Wagner, 1965) and some not (e.g. Wagner, 1975; Naddor, 1975; Ehrhardt, 1979; Ehrhardt and Mosier, 1984; Porteus, 1985). The problem in all cases is to specify the s, S values for a single item inventory in which unfilled demand is backlogged (see also sub-section 9.4.1). There is a constant review period T and a fixed lead time L. Demand during review periods is independent and identically distributed. At the end of each review period, costs I and B_2 (or B_3) are incurred for each unit on hand or backordered respectively. Most of the justification of the heuristics, and claims for their value, are based on the use of generated data with known properties. Nevertheless none of them is simple enough to be of potential benefit to practitioners. Many policies of this form are in fact used in practice (Sani, 1995). However, the values of the control parameters are often chosen arbitrarily.

Sani (1995) conducted research to identify the most appropriate periodic inventory control policies for intermittent demand items. The inventory control methods were evaluated on 84 real data files, considering alternative intermittent demand estimation procedures. The results (also summarised in Sani and Kingsman, 1997) showed that:

- For low demand items (below 20 units per year), Naddor's heuristic (Naddor, 1975) is the best, both on costs and service level
- For medium demand items (20-40 per year) the Power Approximation (Ehrhardt and Mosier, 1984) is the best, both on costs and service level
- For high demand items (above 40 per year) the Normal Approximation (Wagner, 1975) is best, both on costs and service.

The (T, S) system performed well only when the ordering costs were not considered for comparison purposes. The three heuristics require knowledge only of the mean and variance of the demand distribution. Ehrhardt (1979) noted that: "*The computation of an optimal policy* (but also of some heuristic procedures) *requires the complete specification of the demand distribution, and this level of demand information is particularly unrealistic in practical settings. Most managers would be very fortunate if they had accurate knowledge of only the first two moments of the demand distribution* (p.777)". Naddor's heuristic and the Normal Approximation have been developed based on normality assumptions and they require knowledge of the unit cost that is not available for the files considered in this research.

The Power Approximation was developed assuming that demand can be represented by the Poisson or Negative Binomial distribution. The heuristic does not require knowledge of the unit cost and it is clearly the least demanding (T, s, S) model from a computational perspective. Nevertheless, these advantages are outweighed by the fact that the heuristic has been developed considering a B_2 cost criterion (shortage cost per unit value short) and it does not allow comparisons to be made across other management specified system constraints (see section 9.4). Clearly the selected method needs to allow for a greater flexibility on the alternative conditions that will be simulated. Considering also that the (T, s, S) policies are treated in Silver et al (1998) under the heading: "Managing the most important (class A) inventories" the Power Approximation (and all the other order point – order up to level policies discussed in this section) will not be further considered.

9.2.2.2 The (*T*, *S*) policies

Hadley and Whitin (1963) noted the considerable differences, in the computational effort required, to generate results for the (T,S) system and the (T,s,S) or (T,s,Q) policies (the latter policy is discussed in the following sub-sub-sections). Subsequently they asked the question "*under what circumstances will the order up to* S *policy be essentially optimal, i.e. under what circumstances will the average annual cost differ so little from the average annual costs of the periodic* (s,S) *or* (s,Q) *policies that is not worthwhile to make the computations for the last two policies?* (p. 281)" The answer depends on the relative magnitudes of the review and ordering costs. When the review costs are high relative to ordering costs, an order up to S policy should be almost optimal. The authors noted that in many real world situations the review costs are considerably higher than the costs of placing an order. Hence, one would expect to find that in many practical situations one could use the (T,S) model without great deviations from optimality. When the opposite is the case (i.e. the ordering costs are high relative to the review costs) the authors suggested considering the possibility of switching to a transactions reporting system.

It is worthwhile mentioning that ordering costs are significantly lower in 2001 than they were in 1963 (when the particular book was published). The review costs have also declined, but is almost certain that review costs are still higher than ordering costs. Therefore, Hadley and Whitin's conclusions regarding the periodic order up to level systems are valid also in modern business contexts

Sani (1995) developed a stock control model that reflected the main characteristics of a real inventory system. The model is of the (T,S) form where an overnight emergency delivery is offered in the case of a stockout. The model was used to conduct a sensitivity analysis of the inventory costs and customer service levels achieved by employing the real system. In particular, using the model, an analysis was carried out in order to investigate:

1. The effect of the overnight emergency deliveries on cost and service level
2. The sensitivity of costs to changes in the replenishment level, S
3. The effect of the review period, T .

Demand was assumed to follow the negative binomial distribution and the mean and variance of demand were calculated from the data, rather than being estimated from a forecasting method. Sani argued that the (T,S) system represents many real world cases and is intuitively and computationally more appealing to practitioners than the (T,s,S) system. Sani also argued that investigation of sensitivity issues necessitates the use of an optimal system, one that has been explicitly developed based on minimising the total expected costs. Optimal systems enable us to see how much the optimal cost increases or decreases with any change to the control parameters. Optimal systems should be preferred to heuristics which "*are normally developed based on some inventory and mathematical techniques to determine the control parameters. The parameters are not determined based on the minimal cost* (p. 144)". Moreover, heuristics are always developed considering a specified managerial constraint, whereas optimal systems allow experimentation with different cost or service type restrictions imposed on the system by managers. Sensitivity results should be expressed as deviations from an optimal value, provided that this value can be obtained with "reasonable" computational effort so that practitioners can also see it as "optimal".

In the case of our research, we are also interested in assessing the sensitivity of inventory control results to changes in the control parameter values. These changes occur as a consequence of utilising different estimation procedures, whereas in Sani's research the changes were simply the result of experimentation with different possible inventory parameters. Nevertheless, an argument, similar to that made by Sani, can also be put forward in this research in order to justify the selection of an optimal system. In our case, all cost results (costs resulting from different estimation procedures) will be optimal if an optimal inventory system is utilised, and the lowest optimal cost can be regarded as a benchmark. Not all optimal systems, though, accurately reflect real world practices, either because of the restrictive assumptions upon which they are based or, as discussed above, because they are computationally prohibitive. In that respect, the standard (T,S) system seems to be the most appropriate one for conducting a sensitivity analysis in order to demonstrate potential gains or losses in a realistic situation.

9.2.2.3 The (*S-1*, *S*) policies

The (T, S) can be viewed as the periodic implementation of the (s, S) policy or a special case of the $(T,\ s,\ S)$ for $s = S - 1$. In this case the system is usually denoted by $(S - 1, S)$. Schultz (1987) discussed a periodic $(S - 1, S)$ policy for the control of intermittent demand items. At the end of every review period the inventory position (inventory on hand plus on order less the number of units backordered) is checked against the "base stock" level $S - 1$. If the inventory position is less than or equal to $S - 1$, an order is placed to bring the inventory position up to S; otherwise, no order is placed. This model covers the inventory system against the possibility that one demand will occur during the lead time plus review period. Necessary assumptions in this model's implementation are the following:

1. Lead times are small as compared to the average inter-demand interval
2. The cost of re-ordering is small relative to the cost of holding sufficient inventory to meet more than one order

The first assumption is clearly very restrictive. In the case of our research the lead time is 3 inventory review periods. Experimentation with other fictitious lead times will also be undertaken in the simulation experiment. However, since the forecast review period is the same as the inventory review period, the forecast lead time cannot be assumed to be less than one inventory review period. Hence, the $(S - 1, S)$ policies should not be used for generating results in this research.

9.2.2.4 Approaches to the re-order point calculation

Dunsmuir and Snyder (1989) developed a method for determining the re-order point (consistent with a specified customer service level) in an intermittent demand context. The model is of the (T, s, Q) form. The distinguishing feature of the method is the use of a probability distribution with a spike at zero to represent the relative frequency of periods with no transactions. Demand is assumed to occur as a Bernoulli process. The size of demand when demand occurs and the standard deviation of the demand sizes are estimated from the past data when a transaction occurs. The probability of demand occurrence, p_x, (frequency of occurrences per specified time period, e.g. a month) is also estimated based on the time elapsed between transactions. Consequently we can estimate the mean and variance of demand per unit time period as well as the mean (μ_y) and variance (σ_y^2) of demand over a lead time for both constant and variable lead times. The probability of demand occurring during the lead time, p_y, is calculated based on p_x (for the case of variable lead times, the gamma distribution is assumed in order to represent their length). Finally the mean (μ_{y+}) and variance (σ_{y+}^2) of the positive lead time demand can be estimated based on p_y, μ_y, σ_y^2.

The probability of positive demand during the lead time is approximated by a gamma density function. Based on this assumption and findings that appeared in Snyder (1984) the authors modified Brown's formula (1959, 1967):

$$Q(1-P_2) = \int_{x=s}^{\infty} (x-s)\,p(x)dx$$

in order to incorporate the effect of the joint probability that demand will occur during the lead time and it will be of a size, say ξ. Finally only μ_y, μ_{y+}, σ_{y+}^2 are required in order to calculate the re-order point.

The above described method is the first to appear in the literature that is explicitly developed based on Croston's approach to dealing with intermittence. The method is intuitively appealing and straightforward in its application. Nevertheless the proposed model can be used only in conjunction with a relevant estimation procedure, i.e. one that explicitly considers sizes and intervals. Methods that consider aggregate demand and model how this moves over time (e.g. EWMA) become redundant. The purpose of selecting an inventory control model in this chapter is to assess whether or not forecasting accuracy differences are reflected in an inventory control context. The theoretical properties of EWMA have already been discussed and we clearly want to generate inventory control results for this estimation procedure. Moreover we will argue, in a later section of this chapter, for the inclusion of a Moving Average method in our simulation experiment. Both methods could not be used if the above described approach was utilised. In consequence the Dunsmuir and Snyder approach cannot be further considered. Nevertheless, it is important to note that an interesting avenue of further research would be to compare the forecasting methods developed in this thesis, based on Croston's concept, with alternative inventory policies, including that of Dunsmuir and Snyder (1989).

One other, intuitively appealing, approach to the calculation of the re-order points, in an intermittent demand context, has appeared in Segerstedt (1994). In this case a model was developed based on an explicit consideration of the demand size, inter-demand interval and lead time. The three variables were assumed to be gamma distributed. Segerstedt's approach cannot be further considered for the reasons discussed earlier in this sub-sub-section. From a methodological perspective, the exclusion of this method can be justified also in terms of the demand distributional assumptions based on which it was developed (see also sub-section 9.5) and the assumed variability of the lead time. Our discussion on re-order point policies is continued in the following sub-sub-section.

9.2.2.5 Consideration of the deficit

The re-order point calculation in a lumpy demand context necessitates an explicit consideration of the distribution of the "deficit". The deficit is the amount by which the inventory position drops below the s value.

The exact distribution of the deficit is quite complex, depending in general on the difference between s and $s+Q$ (or S) and the probability distribution of transaction sizes (Silver et al, 1998). Different approaches to the calculation of the re-order point, when the deficit is taken into account, can be found in: Silver (1970), Ward (1978), Watson (1987), Janssen (1998), Janssen et al (1998), Strijbosch et al (2000).

Consideration of the deficit implies (Silver et al, 1998) that we deal with the most important (class A) inventories. In consequence all the relevant approaches that have been presented in the literature will be disregarded for the purposes of this research. At this point it is important to note that our decision reflects, to a certain extent, the Silver et al preference towards using simpler methods for class C (or even B) items. Moreover, we recognise that intermittent demand SKUs may be expensive and as such they may be classified as A items. Nevertheless, it is true that relatively simple methods are most commonly used in practice to deal with the categories C (and B) of inventories. In addition, it is also true that most of the intermittent demand items are usually classified as C (or more rarely as B) stock. Therefore, we may conclude that, from a practitioner's standpoint, simple inventory control models are required to manage intermittent demand stocks.

In the last two sub-sub sections, the literature on re-order point inventory control models for intermittent demand has been reviewed. The models were discussed with respect to their degree of relevance to our research. None of them was found to be consistent with the purposes of our simulation experiment and as such they will not be further considered in this chapter.

9.2.2.6 Conclusions

Having decided on the periodic nature of our inventory model (in sub-section 9.2.1), an attempt was made in this sub-section to specify the model more precisely.

Considering the information available for simulation purposes, the (T, S) model is the most appropriate one for meeting the objectives of this chapter. The model is simple, close to optimal and reflects to a great extent real world practices. Nevertheless it is important to note that the selection of the model did not result only from its comparative advantages over other models. It was also, in a sense, the natural consequence of the fact that other theoretically more advanced models had to be disregarded, owing to the objectives of this chapter and the information available for meeting those objectives.

In particular, the ($S-1, S$) models cannot be further considered because of the restrictive assumptions associated with their application. As far as the periodic order point - order up to level and order point inventory models are concerned, specific limitations have been identified but a general argument may also be put forward in order to justify their exclusion from the simulation experiment: both types of models are more suited, from a computational perspective, to class A of inventories[1]. The contribution of SKUs to sales can typically be described by the Pareto rule, meaning that approximately 80% of the SKUs contribute to approximately 20% of the turnover. Although some intermittent demand items may not fall within that 80% of the SKUs (because their price is very high) it is true to say that most of the sporadic demand items will eventually be classified as C (or even B) items. Relatively simple techniques are recommended in the literature and applied by practitioners to deal with the B and C items and in that respect the (T, S) model is the one to be preferred for our simulation.

9.3 The inventory management system

In a (T, S) inventory control system the objective is to optimise[2] the value of the control parameters T and S. Brown (1967) distinguishes between inventory control and inventory management systems. The former refers to a day-to-day physical operation of the approach chosen, whilst the latter to the process of setting the numerical values of the control parameters required in the inventory control system to decide when and how much to order. As stated in the previous section the value of T is optimised by converting the EOQ to a time supply. The S value is calculated as follows:

$$S = Y'_{T+L} + k\,\sigma_{\mathrm{T+L}} \tag{9.1}$$

where:

Y'_{T+L} is the forecast (estimate of the level) of demand over one inventory review period plus the fixed lead time L

$\sigma_{\mathrm{T+L}}$ is the standard deviation of the lead time forecast error, i.e. the estimate of the variability of demand over one review period plus lead time, and

k is a safety factor.

[1] The order point models necessitate the consideration of the deficit, which in turns often suggests that we deal with the class A stock.
[2] Strictly speaking, optimisation applies to the consideration of a cost criterion (see sub-section 9.4.2). The control parameters are “determined” rather than “optimised” when service criteria are used.

If we denote by Y'_t the estimate made by any forecasting method at the end of period $t-1$, of demand in period t, under the stationary mean model assumption:

$$Y'_{T+L} = (T+L)\, Y'_t \tag{9.2}$$

This is also true for the steady state model, although not assumed in this thesis.

Equation (9.2) will be used in order to obtain an estimate of demand over lead time plus review period for all methods that will be evaluated in our simulation experiment, including Croston's method. Croston suggested working in terms of guarding against one occurrence of a demand during the lead time (or lead time plus the review period). Thus he recommended using the latest estimate of the demand size (z'_t) as the estimate of demand over L (or $L+T$):

$$Y'_{T+L} = z'_t \tag{9.3}$$

Sani (1995) argued that when annual demand is very low, and the forecast review period is very short, as in the case of his research, there would be many periods with no demand at all. Croston's suggestion, in that case, would lead to significant extra stockholding. In addition one could also argue that if the average inter-demand interval is small in comparison with $L+T$ there could be more than one demand occurrences during that period. In that case Croston's suggestion would lead to a significant understocking.

The standard deviation of the lead time forecast error is traditionally calculated[3] as:

$$\sigma_{\mathrm{T+L}} = \sqrt{T+L}\, \sigma_t \tag{9.4}$$

where σ_t is the standard deviation of the one step ahead forecast error.

[3] This is true neither for the stationary mean model (see chapter 6) nor for the steady state model assumption (Johnston and Harrison, 1986).

The standard deviation of the one step ahead forecast error can be calculated by using either the MAD or MSE smoothing approach:

$$\sigma_t = \sqrt{MSE_t} \tag{9.5}$$

where $MSE_t = \alpha\left(Y_t - Y'_t\right)^2 + (1-\alpha)MSE_{t-1}$

or

$$\sigma_t \approx 1.25\, MAD_t \tag{9.6}$$

where $MAD_t = \alpha\left|Y_t - Y'_t\right| + (1-\alpha)MAD_{t-1}$.

In the above calculations, α is the smoothing constant value used and Y_t the actual demand in period t. Moreover, approximation (9.6) is based on the assumption that forecast errors are normally distributed. Nevertheless the ratio of the mean absolute deviation to the standard deviation is (within sampling error) almost constant for many distributions, being for example 0.75 for negative exponential and 0.85 for the rectangular probability density function (see for example Johnston, 1975).

The MSE smoothing approach to estimating the forecast variance was recommended by Brown (1982). Bretschneider (1986) showed that the MSE approach is more efficient than the MAD approach even in the presence of some outliers. In chapter 6 it was shown that MSE also takes into consideration the issue of bias, and as such, it is the accuracy measure chosen for estimating the demand variance in our simulation experiment.

Nevertheless in chapter 6 we showed that by estimating the variance of demand based on equations (9.4) and (9.5) the true variability is understated, since the auto-correlation of the forecast errors is not taken into account:

$MSE_{over\ L} = L\left\{Var\left(Y'_t\right) + Bias^2 + Var\left(Y_t\right)\right\}$,

ignoring the auto-correlation of the forecast error (9.7)

$MSE_{over\,L} = L\left\{LVar\left(Y'_t\right) + LBias^2 + Var\left(Y_t\right)\right\}$,
considering the auto-correlation of the forecast errors (chapter 6) (9.8)

Now, we also showed that, for an accuracy comparison exercise, consideration of either (9.7) or (9.8) does not result in different conclusions about the conditions under which one method performs better than one or all other methods. (That does not mean of course that we will not generate one lead time as well as one step ahead MSE results in our simulation experiment, in order to verify our theoretical conclusions.) Intuitively this is likely to be the case also in an inventory control context. Nevertheless we are still interested in analysing each individual method's stock control performance over lead time (plus review period). In addition, the explicit or implicit consideration of the auto-correlation terms is also required for ensuring the theoretical consistency of the thesis. Therefore, it has been decided to estimate the lead time (plus one review period) MSE directly rather than calculating the one step ahead MSE and then estimating the variability of the lead time (plus review period) demand.

There are two approaches in order to calculating the MSE over $\mathrm{L} + \mathrm{T}$. The first approach, which is the one adopted for the purpose of this research, requires the replacement of equations (9.4) and (9.5) by equations (9.9) and (9.10):

$$\sigma_{\mathrm{T+L}} = \sqrt{MSE_{t,T+L}} \tag{9.9}$$

where:

$$MSE_{t,T+L} = \alpha\left\{\sum_{i=t-T-L+1}^{t}\left(Y_i - Y'_i\right)\right\}^2 + \left(1-\alpha\right)MSE_{t-1,T+L} \tag{9.10}$$

That is, exponential smoothing will be used in order to update directly the lead time (plus review period) MSE at the end of every inventory review period.

The second approach is to directly estimate the MSE by considering the expressions derived in chapter 6 (sub-section 6.6.4). For example, the lead time MSE for EWMA (all points in time) is as follows:

$$MSE_{EWMA} = L\left\{L\frac{\alpha}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2+\frac{\sigma^2}{p}\right]+\left[\frac{p-1}{p^2}\mu^2+\frac{\sigma^2}{p}\right]\right\} \tag{9.11}$$

Based on (9.11) we could estimate the variability of demand over lead time plus one review period by utilising (9.12):

$$MSE_{t,T+L,EWMA} = (T+L)\left\{(T+L)\frac{\alpha}{2-\alpha}\left[\frac{p'_t-1}{{p'_t}^2}{z'_t}^2+\frac{{\sigma'_t}^2}{p'_t}\right]+\left[\frac{p'_t-1}{{p'_t}^2}{z'_t}^2+\frac{{\sigma'_t}^2}{p'_t}\right]\right\}$$

(9.12)

where p'_t, z'_t and σ'_t are the estimates of the inter-demand interval, demand size and variability of sizes respectively made at the end of period $t-1$. The variability of the demand sizes could be estimated based on either (9.5) or (9.6) considering the sizes of demand when demand occurs instead of demand per unit time period.

This approach, namely consideration of an analytic expression and substitution of the parts of the expression by their corresponding estimates, has appeared in Sani (1995) and Strijbosch et al (2000). However, such an approach can be used only for estimation procedures that explicitly consider the size of demand, when demand occurs, and the inter-demand interval. As such results cannot be generated for EWMA and subsequently analytic expressions, similar to that given in (9.12), will not be considered for generating MSE results in our simulation experiment.

Finally it is important to discuss Croston's approach to estimating the variability of lead time demand. Croston recommended using the variance of the demand transaction sizes for the variance of demand over L (or $L+T$). Following Croston, the variance of the demand transaction sizes, $Var(z_t)$, is estimated by using the MAD (Mean Absolute Deviation) of the forecast errors of the demand transaction sizes. Sani (1995) argued that when no demand occurs during the lead time Croston's approach would lead to significant extra stockholding. He consequently suggested the following method of estimating the variability of demand over $L+T$:

$$Var_{t,T+L} \approx (L+T)\max\left\{\frac{Var(z_t)}{p'_t}, \frac{1.1 z'_t}{p'_t}\right\}, \tag{9.13}$$

where $Var(z_t)$ is estimated as proposed by Croston.

This approximation is based on the assumption that lead time demand follows the negative binomial distribution (NBD) and, as such, is of direct relevance to our research since the NBD will also be assumed for the purpose of conducting the real data simulation experiment.

Nevertheless, by using approximation (9.13):

- The bias of Croston's method is not taken into consideration
- The variability of the inter-demand interval estimates is not taken into consideration
- The auto-correlation of the forecast errors over $L+T$ is not taken into consideration.

Therefore, approximation (9.13) will not be further considered for the purposes of this research.

In this sub-section some issues related to the optimisation of the replenishment level S in a periodic system of the (T, S) form have been discussed. To optimise the replenishment level, an estimate of demand and its variability, over one review period plus lead time, is required. Considering the assumptions made in this thesis, as well as the structure and objectives of our simulation experiment, demand over $L+T$ will be estimated based on (9.2) and demand variability based on (9.9) and (9.10).

9.4 The safety factor

To specify the safety factor k, decisions have to be made as to how we treat stock-outs, what criterion to use in our system and which theoretical distribution best fits the demand data. The first two issues are discussed in this section whereas issues related to the assumed distribution will be separately considered in the following section (9.5). Note that under a discrete demand distribution assumption the optimum control parameter value is, most commonly, specified directly (see for example section 9.6).

9.4.1 The assumption of backordering or lost sales

When a customer places an order, the demand is subtracted from the quantity on hand. If the quantity on hand is less than the customer wants, the negative difference is recorded, and the demand is either backordered (to be filled from the next scheduled receipt), or referred to another branch or plant, if the customer needs the SKU immediately. Alternatively, the demand is cancelled. In some cases it may be just the difference between demand and stock on hand that is cancelled, whilst it is also possible that the whole order placed for the particular SKU is cancelled. In either case, demand is still added to the current forecast time period demand, in order that a record is in place from which a forecast can be made. Given that the two cases (backordering and lost sales) represent the two extremes of a continuum, it is unlikely that companies would operate at one or other extreme point at all times. Nevertheless most models developed in the literature refer to one or other of the "absolute" cases. Silver et al (1998) noted that "*most of these models serve as reasonable approximations because the decisions that they yield tend to be relatively insensitive to the degree of backordering possible in a particular situation* (p. 234)". This is mainly because the high customer service levels, often specified, result in infrequent stockout occasions.

The average stock in the system depends on the safety stock, which is the expected stock just before a new replenishment arrives. The safety stock in turn depends on how unfilled demand is treated. Obviously if backorders are allowed, the net stock (= stock on hand – backorders) can take positive as well as negative values, whereas if sales are lost, the stock on hand is always greater than or equal to zero. Moreover, overnight deliveries at a premium charge to satisfy demand will have a different effect from satisfying demand from the next scheduled replenishment order. In the former case the average stock will be the same as in the lost sales situation, assuming that the units delivered overnight attract no inventory holding charges.

The data sample used in our research comes from the automotive industry. In this industry, exclusive dealerships (at the wholesale-retail link) or long-term supplier-(industrial) customer relationships and co-managed inventory schemes (at the manufacturing side of the supply chain) could justify the assumption of a complete backordering case. In general it is fair to claim that the very nature of intermittent demand necessitates the assumption of the complete backordering case. That is, from a business perspective, there is likely to be little, if any, competition for the supply of an intermittent SKU. Therefore the number of stockpoints for the SKUs under concern cannot be large (if it is not just one). Demand for intermittent demand SKUs is most often "captive" and as such the complete backordering situation will be assumed for the purpose of developing the (T, S) inventory control model in this chapter.

9.4.2 The policy variables

Different decision rules can be applied in deriving the safety stock, which is necessary for covering the system against the variability of demand. The safety stock is computed as the product of the standard deviation of the forecast error over review period plus lead time and a dimensionless safety factor. The decision rule for computing the safety factor contains a management policy variable. The value of the management policy variable controls the exchange of the capital investment in inventory for various measures of service.

The two main approaches in establishing safety stocks are the following:

1. Safety stocks based on the costing of shortages. This approach involves specifying a way of costing a shortage. Consequently the expected sum of ordering, holding and shortage cost is minimised. This approach is often referred to as a "cost perspective".

2. Safety stocks based on service considerations. Recognising the difficulties associated with costing shortages, an alternative approach is to introduce a service related control parameter. This approach is often referred to as a "service level perspective".

Safety stocks can be established through the use of a common safety factor, so that the safety factor (k) itself becomes the management policy variable. Moreover safety factors can also be calculated by considering a budget available for a collection of SKUs. The idea is to establish the safety stocks of individual items, so that for the budget given, the best possible aggregate service across the population of items is achieved. Equivalently, one selects the individual safety stocks to minimise a penalty function while meeting a desired aggregate service level. This thesis focuses on decision rules for individual SKUs rather than a collection of items and therefore aggregate rules will not be further considered.

The specific criteria to be used for the purposes of our simulation experiment, along with the justification of their selection, are discussed in the following two sub-sub sections.

9.4.2.1 Service measures

Similar to specifying a common safety factor is to determine a probability of no stockout per replenishment cycle. In both cases, fast and slower moving products are covered with the same chances of a shortage during a replenishment cycle, without regard to how many such opportunities to run short exist in a calendar year (see for example Brown, 1967).

Another popular service policy variable is a management specified Time Between Stockouts (TBS). In the decision rule resulting from using this measure, the frequency with which an order is placed is taken into account. It is easy to show that as the number of replenishment cycles and/or the TBS increase, so does the safety factor. Even though this particular decision rule is theoretically appealing, it is not particularly relevant to this research.

Slower moving items may have potentially severe shortage penalties when they are used to complete an order provided to an important customer. Shortages of items under concern can cause severe reductions in the usage of several faster moving items. Under those circumstances a safety factor is determined to provide a specified expected TBS. Large values of TBS can be selected (from 5 to 100 years, Silver et al, 1998) assuming that the inventory holding costs are very low. In the case of our research we are concerned with single item inventory decision rules. That is, we neglect the effect of our decisions, for a specific item, to the rest of the items in a group or order line. Moreover, the low inventory holding cost assumption is very restrictive. It is also important to note that even low TBS values should result, in the case of this research (few demand occurrences), in virtually no stock out occasions and in consequence it will be very difficult, if not impossible, to generate comparative results with respect to service criteria.

Another possible service criterion is the Ready-Rate (fraction of time during which the net stock = stock on hand – backorders, is positive), which serves mostly emergency purposes. This service criterion has had a wide application in the military context (see for example Silver, 1970) and it is not particularly relevant to the empirical data considered in this research. As such, the Ready-Rate criterion will not be used for generating simulation results.

The most commonly used service measure in an intermittent demand context (see for example Kwan, 1991; Janssen, 1998; Strijbosch et al, 2000) is a specified fraction of demand (P_2) to be satisfied directly from stock. The P_2 service measure is also called the fill-rate. Based on the decision rule resulting from this measure (to be discussed in greater detail in section 9.6) we can specify the safety factor by considering the variability of demand over lead time plus one review period. The P_2 service measure is the only criterion used by the software manufacturer, that provided our empirical data set, in order to specify the safety factor k. Due to the very few demand occurrences associated with our empirical data it will not be very easy to establish service level differences between the forecasting methods compared in the simulation experiment. Relatively low fill rate values (say 90% or 95%) may be proven useful in identifying performance differences between alternative forecasting methods with respect to a target service level. At this stage it is important to note that very high fill rate values are not commonly used in practice for intermittent demand items. That is, in an intermittent demand context the SKUs will most commonly be classified as C, or more rarely, as B items (see also sub-sub section 9.2.2.5). In that respect, it is generally reasonable to assume that the target service level in “real world” applications will not generally exceed 95%.

The P_2 service measure is the only service measure that will be employed for the purpose of this research due to its widespread usage and its relevance to our empirical data. Two values will be considered, in our simulation experiment, for this control parameter and those are 0.90 and 0.95.

9.4.2.2 Costing methods

Deciding on an appropriate shortage fraction involves developing and consequently minimising a Total Cost of Inventory (TCI) function. This function consists of the ordering, inventory holding and shortage costs. It is common to assume that the cost of placing an order is fixed, i.e. independent of the order size. The inventory holding cost is determined by considering the unit cost, the inventory holding charge per unit per year (which in turn is determined by the opportunity cost of capital tied up in inventory, the risk and costs of deterioration and obsolescence, the actual cost of storing this unit etc) and the expected number of units in stock at any point in time. The average stock in the system is affected by whether or not backorders are allowed. The shortage cost is derived after specifying the value of a chosen cost criterion:

1. Minimise backordering cost per unit of time/minimise the average backlog level
2. Minimise the cost/number of emergency overnight deliveries
3. Minimise number of replenishment lots needing expediting/minimise the number of stockout occasions per unit of time
4. Minimise number of backordered units

The first shortage costing method requires a premium amount to be charged for every unit short for every period of time. As such it has a natural interpretation in the manufacturing environment when a machine cannot be used because of a particular spare not being available when requested. This research is not restricted to that particular context and in consequence this criterion will not be further considered.

The second shortage costing method applies to serving very important customers and necessitates a certain degree of co-ordination so that demand can be satisfied overnight from another depot at a premium charge (say B_3) per unit value. In such a case highly efficient order-picking and transportation systems fill emergency orders using stock from the best location. This system of giving good service without holding inventory is known, within the automotive industry, as VOR, Vehicle Off the Road (see for example Sani, 1996). This criterion is particularly relevant to our data sample (that comes from the automotive industry) and as such it will be considered for generating stock control results.

The last two costing methods are the most commonly used ones. The first assumes that the only cost associated with a stockout is a fixed value B_1, independent of the magnitude or duration of the stockout. When the second one is used, we assume that a fraction B_2 of unit value is charged per unit short, independently of the duration of the stockout. The cost per unit short of an item i is $B_2 c_i$, where c_i is the unit variable cost. A situation where this type of costing would be appropriate is where units are made during overtime production (Silver et al, 1998). Nevertheless, this criterion is relevant to other environments as well, in that managerial concerns about customers' loss in goodwill and market share are often reflected in the B_2 measure (Kwan, 1991; Janssen, 1998).

The B_1, B_2 and B_3 criteria will be further considered in this chapter due to their widespread usage in practice and their relevance to our empirical data. The conditions that allow the optimisation of the replenishment level (for each one of the shortage criteria but for the P_2 service measure also) will be derived in section 9.6.

9.5 The demand distribution

9.5.1 Compound Poisson distributions

With Poisson arrivals of transactions and an arbitrary distribution of transaction sizes, the resulting distribution of total demand over a fixed lead time is compound Poisson. The stuttering Poisson distribution, which is a combination of a Poisson distribution for demand occurrence and a geometric distribution for demand size, has received the attention of many researchers (for example: Gallagher, 1969; Ward, 1978; Watson, 1987). Another possibility is the combination of a Poisson distribution for demand occurrence and a normal distribution for demand size (Vereecke and Verstraeten, 1994). Quenouille (1949) showed that a Poisson-Logarithmic process yields a negative binomial distribution (NBD). When order occasions are assumed to be Poisson distributed and the order size is not fixed but follows a logarithmic distribution, total demand is then negative binomially distributed over time.

Some work concerning variable lead times, intermittent demand and separate forecasts for the order size and the order intensity is also available in the literature. Demand has been modelled based on three components: order size, order intensity and lead time. For example, Nahmias and Demmy (1982) proposed the logarithmic-Poisson-gamma model, while Bagchi (1987) considered the geometric-Poisson-normal and the geometric-Poisson-gamma models.

9.5.2 The "package" Poisson distribution

Vereecke and Verstraeten (1994) presented an algorithm developed for the implementation of a computerised stock control system for spare parts in a chemical plant. Ninety per cent of the items were classified as lumpy with the remaining ten per cent consisting of slow or fast movers. The demand was assumed to occur as a Poisson process with a package of several pieces being asked for on each demand occurrence. The parameters of the distribution of the demand size could be calculated from the variance and the average of the demand history data of each item. The resulting distribution of demand per period was called "Package Poisson" distribution. The same distribution has appeared in the literature under the name "clumped

Poisson" distribution (Ritchie and Kingsman, 1985), for multiple item orders for the same SKU of a fixed "clump size", or "hypothetical" SKU (h-SKU) Poisson distribution where demand is treated as if it occurs as a multiple of some constant (m) (Williams, 1984). In an earlier work, Friend (1960) also discussed the use of a Poisson distribution for the demand occurrence, combined with demands of constant size.

9.5.3 Compound binomial distributions

With demand occurring as a Bernoulli process (i.e. the demand incidences follow the binomial distribution) and an arbitrary distribution of the demand sizes, the resulting distribution of total demand over a fixed lead time is compound binomial. When the order sizes are assumed to follow the Logarithmic-Poisson distribution (which is not the same as the Poisson-Logarithmic process that yields NBD demand) then the resulting distribution of total demand per period is the log-zero-Poisson (lzP, see for example Kwan, 1991). Alternatively the possibility of normally distributed demand sizes has appeared in the literature in Croston (1972, 1974).

9.5.4 Empirical evidence

Kwan (1991) conducted research with the purpose of identifying the probability distributions that best fit the distribution of demand sizes, inter-demand intervals and demand per unit time period when demand appears at random with some time periods showing no demand at all. Empirical tests were performed with the use of the chi-square test on a number of distributions, using two sets of real world data which contained the demand history of 85 spare parts.

Five distributions were considered for representing the demand per unit time period/lead time (or lead time plus review period) demand: Poisson, NBD, Hermite, log-zero-Poisson and Laplace.

For a Poisson distribution of the demand per period and Gamma distributed lead times the resulting distribution of lead time demand is NBD (Taylor, 1961). The obvious advantage of NBD over the Poisson distribution is that it allows a wide range of variance to mean ratios, since the variance is not required to be equal to the mean.

Another compound Poisson distribution for representing the lead time demand is the Hermite distribution (Kemp and Kemp, 1965; Bagchi et al, 1983). The Hermite distribution is an exact distribution of the lead time demand if the demand per unit time is Poisson distributed and the lead time is normally distributed.

The lzP distribution is, as discussed earlier, a compound binomial distribution. Kwan's main argument to justify the use of this distribution rests on its flexibility. The lzP distribution can be used to represent demand patterns where the variance is greater, equal or less than the mean.

Finally the Laplace or pseudoexponential distribution (Presutti and Trepp, 1970) was also considered by Kwan, based on a suggestion by Peterson and Silver (1979) suggestion that this distribution is the most appropriate for slow moving items with non-similar mean and variance.

In Kwan's research, the Negative Binomial distribution was found to be the best, fitting 90% of the SKUs, followed by the log zero Poisson.

Boylan (1997) tested the goodness-of-fit of four demand distributions (NBD, lzP, Condensed Negative Binomial Distribution (CNBD) and gamma distribution) on real demand data. The CNBD arises if we consider a condensed Poisson incidence distribution ("censored" Poisson process in which only every second event is recorded) assuming that the mean rate of demand incidence is not constant, but varies according to a gamma distribution. Although this distribution has been derived for demand incidence (Chatfield and Goodhardt, 1973) it may also be used for demand as an alternative to the NBD, on the grounds of "flexibility" since it has a wider range of variance to mean ratios than the NBD. In particular, the CNBD extends the lower end of the range of the variance to mean ratio to 0.5.

The empirical sample used for testing goodness-of-fit contained the six months histories of 230 SKUs, demand being recorded weekly.

The analysis showed strong support for the NBD. The results for the gamma distribution were also encouraging, although not as good for slow moving SKUs as the NBD.

9.5.5 The Negative Binomial Distribution (NBD)

Boylan (1997) proposed three criteria for assessing demand distributions (see also chapter 7):

- A priori grounds for modelling demand
- The flexibility of the distribution to represent different types of demand
- Empirical evidence

To satisfy the first criterion, a distribution must be explainable in terms of an underlying mechanism. Considering the assumptions made in this thesis a compound binomial distribution (lzP) could be utilised for modelling the demand over lead time plus review period. The lzP is also very flexible in the sense that the variance to mean ratio can take any possible value.

Kwan (1991) tested the lzP for the first time in an inventory control application. The results of her research showed that, despite its greater flexibility, the lzP does not perform better than the NBD. In fact the control parameters given by the lzP result, in many cases, in inventory costs higher than those associated with the use of the NBD.

The NBD is a compound Poisson distribution and therefore it can be theoretically justified for application in an intermittent demand context in terms of the underlying Poisson demand generation process.

Another possible distribution (not taken into account in Kwan's thesis) for representing demand is the gamma distribution. The NBD is the discrete analogue of the gamma. Boylan (1997) noted that "*although not having a priori support, the gamma is related to a distribution which has its own theoretical justification* (p. 168)". The gamma covers a wide range of distribution shapes, it is defined for non-negative values only and it is generally mathematically tractable in its inventory control applications (Burgin and Wild, 1967; Burgin, 1975; Johnston, 1980). Nevertheless if it is assumed that demand is discrete, then the gamma can be only an approximation to the distribution of demand. Moreover no empirical evidence exists in support of the gamma distribution for representing intermittent demand.

The NBD seems to be the best choice for representing demand in our simulation experiment. The theoretical results that have been developed so far in this thesis are based on the assumption that demand arrives as a Bernoulli process. In the simulation experiment all our derivations will be tested against real data forecasting accuracy results. The generation of the bias and MSE simulated results though requires no assumptions about the underlying demand distribution. The NBD assumption is necessary only for estimating the inventory cost and service level resulting from the application of an estimation procedure in real data.

The compound Bernoulli process can be seen as the discrete time variant of the compound Poisson process. As the probability of success (demand occurrence) tends to zero, the binomial variate tends to the Poisson variate. (Strictly speaking the sample size needs also to tend to infinity so that the Poisson variate becomes the limiting form of the binomial one.) Therefore for small time units the compound Bernoulli process is an approximation to a compound Poisson process (see for example Boylan, 1997; Janssen et al, 1998). Nevertheless, in the case of our research, demand has been recorded monthly and as such we cannot claim that the above discussed approximation is necessarily valid. In addition, in chapter 7 the lognormal distribution was assumed in order to represent the demand sizes, when demand occurs, whereas the use of NBD implies demand sizes that follow the logarithmic distribution.

Hence, we recognise that there is a theoretical inconsistency between what has been assumed so far in the thesis and the assumption that demand per unit time period/lead time demand follows the negative binomial distribution. Nevertheless, it is important to note that:

- Use of any non-compound distribution cannot, theoretically at least, be justified since such distributions bear no relationship to the observed frequency of periods without demands

- No other compound Poisson distribution, apart from NBD, is known to have been used in an inventory control context
- No other compound Binomial distribution, apart from lzP, is known to have been used in an inventory control context.

As such, and assuming that the lzP is unlikely to be considered by practitioners for inventory control purposes, the NBD still seems to be our best option. The next step is obviously the mathematical (or even intuitive) justification of another compound Bernoulli distribution. Alternatively, consistency between Croston's methodology and current theory on statistical distributions can be achieved only by modifying the estimation procedures, that have been either developed or corrected in this thesis, so that they reflect a Poisson rather than a Bernoulli demand generation process. Clearly both approaches are beyond the scope of this research.

Of course one could argue that explicit consideration of the sizes and intervals in an inventory control context would resolve all theoretical inconsistencies. Dunsmuir and Snyder (1989) proposed a method for determining re-order points consistent with a specified customer service level. The distinguishing feature of that method is the explicit consideration of the positive lead time demand and the probability that demand occurs in the lead time so that the spike at zero is accurately estimated. Nevertheless, their approach cannot be considered for the purpose of our research, for reasons stated in sub-sub-section 9.2.2.4.

Subsequently, the negative binomial distribution is adopted, to generate probabilities in our empirical data simulation experiment, in lieu of any better alternative.

9.6 The (*T*, *S*) system development

In this section the optimum replenishment level is derived for a (T, S) system for each of the four selected policy variables: B_1, B_2, B_3 and P_2, assuming that demand not satisfied directly from the stock is backordered. It is assumed that demand is satisfied from the next replenishment quantity or in the following day by placing an emergency order. Moreover demand over $L+T$ is discrete and represented by the NBD.

Additional necessary assumptions are the following:

1. Ordering cost is independent of the order size
2. Cost of carrying out a review is ignored
3. Unit cost c is a constant
4. L is constant
5. A unit acquired through emergency has no holding cost

Hadley and Whitin (1963) noted that the difference between the backordering and lost sales case is the on hand inventory, $f(x,S)$, just before the replenishment arrives.

For the backordering case:

$$f(x,S) = S - x \qquad \text{for } x \geq 0$$

whereas for the lost sales case:

$$f(x,S) = \begin{cases} S - x & if \quad x \leq S \\ 0 & if \quad x > S \end{cases}$$

The expected amount of on hand inventory just before a procurement arrives is:

$$Safety\ stock = \sum_{x=0}^{\infty} f(x,S)\, p(x)$$

The average replenishment quantity in the (T, S) system is:

$$Expected\ replen.quantity = DR$$

Therefore the stock drops from:

$$\sum_{x=0}^{\infty} f(x,S)\, p(x) + DR$$

after a replenishment quantity arrives, to

$$\sum_{x=0}^{\infty} f(x,S)\, p(x)$$

just before a replenishment quantity arrives again.

The average stock level is calculated as follows (assuming a linear consumption of the inventory):

$$Average\ stock\ level\ (backorders) = \sum_{x=0}^{\infty}(S-x)\, p(x) + \frac{DR}{2} = S - \mu_{L+T} + \frac{DR}{2}$$

$$Average\ stock\ level\ (lost\ sales) = \sum_{x=0}^{S}(S-x)\, p(x) + \frac{DR}{2}$$

and the expected annual inventory holding cost (IC)

$$IC\ (backorders) = Ic\left(S - \mu_{L+T} + \frac{DR}{2}\right) \tag{9.14}$$

or

$$IC\ (lost\ sales) = Ic\left(\sum_{x=0}^{S}(S-x)\, p(x) + \frac{DR}{2}\right) \tag{9.15}$$

The expected annual ordering cost (OC) is calculated as:

$$OC = \frac{A}{R} \tag{9.16}$$

9.6.1 Cost of shortage per unit value short (B_2)

The expected number of units short in a replenishment cycle is:

$$Expected\ number\ of\ units\ short\ = \sum_{x=S+1}^{\infty}(x-S)p(x) \tag{9.17}$$

and the expected annual shortage cost (SC) is given by (9.18):

$$SC = B_2 c \frac{1}{R} \sum_{x=S+1}^{\infty} (x - S) p(x) \tag{9.18}$$

The total annual inventory cost (TCI) is calculated as follows:

$$TCI = \frac{A}{R} + Ic\left(S - \mu_{L+T} + \frac{DR}{2} \right) + B_2 c \frac{1}{R} \sum_{x=S+1}^{\infty} (x - S) p(x) \tag{9.19}$$

Since R is a constant, considering only the part of equation (9.19) which is a function of S, we obtain the expected annual relevant cost of inventory $TCI(S)$:

$$TCI(S) = Ic(S - \mu_{L+T}) + B_2 c \frac{1}{R} \sum_{x=S+1}^{\infty} (x - S) p(x) \tag{9.20}$$

The optimisation condition under the B_2 criterion is the following (see, for example, Kwan, 1991):

$$\sum_{x=0}^{S} p(x) > 1 - \frac{IR}{B_2} \geq \sum_{x=O}^{S-1} p(x)$$

Theoretically $1 - \frac{IR}{B_2}$ may be negative in which case we have:

$$\sum_{x=O}^{S-1} p(x) \leq \max\left(1 - \frac{IR}{B_2}, 0 \right) < \sum_{x=0}^{S} p(x) \tag{9.21}$$

If $p(0) > \max\left(1 - \frac{IR}{B_2}, 0 \right)$

then the optimal replenishment level should be set to zero and no stock is carried for that particular item.

9.6.2 Penalty cost per stockout occasion (B_1)

The probability of a stockout in a replenishment cycle is:

$$Stockout\ \ probability = \sum_{x=S+1}^{\infty} p(x) \tag{9.22}$$

and the expected annual shortage cost (SC) is calculated as follows:

$$SC = B_1 \frac{1}{R} \sum_{x=S+1}^{\infty} p(x) \tag{9.23}$$

The expected annual relevant cost of inventory $TCI(S)$ in that case becomes:

$$TCI(S) = Ic(S - \mu_{L+T}) + B_1 \frac{1}{R} \sum_{x=S+1}^{\infty} p(x) \tag{9.24}$$

The optimisation condition under this cost criterion is the following (see, for example, Kwan, 1991):

$$p(S+1) < \frac{IcR}{B_1} \leq p(S) \tag{9.25}$$

Even though identifying the optimal re-order level under this shortage criterion requires very little computation time, its application on our data sample is not possible since the unit cost, c, is not given. Therefore, this criterion will not be used for generating results in our simulation experiment.

9.6.3 Emergency delivery cost per unit value short (B_3)

The shortage function in this case is the same as for the B_2 criterion while the expected annual inventory cost is the same as in the lost sales case (assuming that the units delivered to meet the demand backordered attract no inventory holding cost).

The $TCI(S)$ is calculated as follows:

$$TCI(S) = Ic\sum_{x=0}^{S}(S-x)\,p(x) + B_3 c\frac{1}{R}\sum_{x=S+1}^{\infty}(x-S)p(x) \qquad (9.26)$$

The optimisation condition under the B_3 criterion is as follows (see, for example, Sani, 1995):

$$\sum_{x=0}^{S-1} p(x) \leq \frac{B_3}{IR + B_3} < \sum_{x=0}^{S} p(x) \qquad (9.27)$$

If an item is of very low demand, so that $p(0)$ is so large that it is greater than $\frac{B_3}{IR + B_3}$, S should be set equal to zero and therefore no stock should be kept for that item. When a demand occurs, a VOR order should be placed.

9.6.4 Specified customer service level (P_2)

We define customer service level (P_2) as the fraction of demand satisfied directly from the shelf.

$$P_2 = \frac{Demand\ \ satisfied\ \ from\ shelf}{Total\ \ demand} \qquad (9.28)$$

If DR is the average size of a replenishment and Ψ is the expected number of units short during a replenishment cycle (equation (9.17)) we then have:

$$P_2 = \frac{DR - \Psi}{DR} \Rightarrow \Psi = DR(1 - P_2) \qquad (9.29)$$

It follows that the smallest S that can maintain the system with service level not less than P_2, is the optimal re-order level.

$$DR(1 - P_2) \geq \sum_{x=S+1}^{\infty}(x - S)p(x) \qquad (9.30)$$

9.7 The simulation model

The information available for the demand data sample that will be used for our research (see also sub-section 9.1.1) is the following:

$I = 25\%$, $B_2 = 28\%$, $L = 3$

Of course other values can also be considered for each one of those control parameters so that we account for a wider range of possible real world scenarios. Similarly, certain values can also be hypothesised for the B_3 and P_2 criteria for which an estimate is not available.

The lead time in particular can be set to: 1, 3, and 5 periods as in the theoretically generated data simulation experiment (the $L = 12$ case cannot be simulated due to the few demand data periods available for generating results, see also sub-section 9.7.3). Two values will be considered for the P_2 criterion and those are 0.90 and 0.95 (see also sub-sub-section 9.4.2.1). Regarding the I, B_2 and B_3 criteria, the values are specified "indirectly" as follows:

The optimisation condition for the B_2 criterion is given by (9.17). To simulate the situation we can either assign a value to I and B_2 or alternatively assign a value to $1 - \frac{IR}{B_2}$. The latter approach enables us to implicitly consider many possible combinations of the I / B_2 ratio and is the one to be considered for generating results. The $1 - \frac{IR}{B_2}$ is called, for the purpose of this research, the "target value". In the case of the B_3 criterion we work in a similar way. In that case, the target value equals $\frac{B_3}{IR + B_3}$ (see (9.19)). The control parameter values considered for both the B_2 and B_3 policies are given in table 9.1 along with the values assigned to the other control parameters introduced in our simulation experiment.

α	0.05 to 0.2 step 0.05
L	1 to 5 step 2
B_2 policy, target value = $1-\frac{IR}{B_2}$	0.93 to 0.96 step 0.03
B_3 policy, target value = $\frac{B_3}{IR+B_3}$	0.95 to 0.98 step 0.03
P_2	0.90 to 0.95 step 0.05
where I is the annual inventory holding charge.	
$1-\frac{IR}{B_2}=0.93 \Leftrightarrow \frac{I}{B_2}=0.84$, $1-\frac{IR}{B_2}=0.96 \Leftrightarrow \frac{I}{B_2}=0.48$ (for $R=1/12 \approx 0.08$)	
$\frac{B_3}{IR+B_3}=0.95 \Leftrightarrow \frac{I}{B_3}\approx 0.63$, $\frac{B_3}{IR+B_3}=0.98 \Leftrightarrow \frac{I}{B_3}\approx 0.24$ (for $R=1/12 \approx 0.08$)	

Table 9.1. The inventory control parameter values

The target values have been specified after consultation with Unicorn Systems (UK) Ltd. They are viewed, from a practitioner's perspective, as realistic, covering a wide range of real world inventory control systems.

In figures 9.1 and 9.2 we indicate the control parameter (I, B_2, B_3) values implicitly considered in our simulation experiment for the B_2 and B_3 policy respectively. Similar values have been considered, for conducting simulation on real (intermittent) data, in Kwan (1991) and Sani (1995).

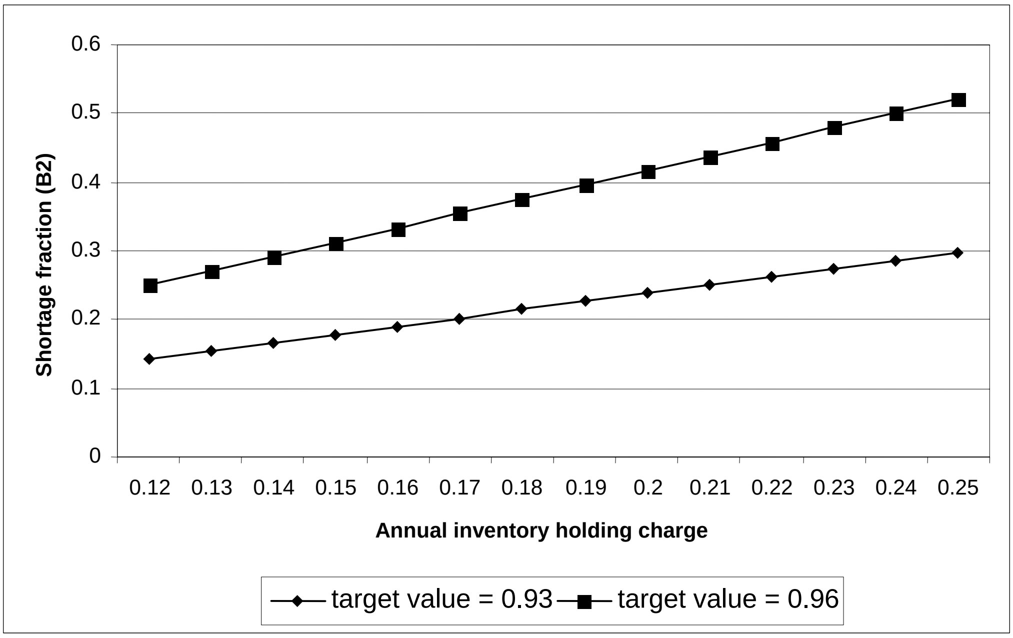

Figure 9.1. Control parameter values (B_2 policy)

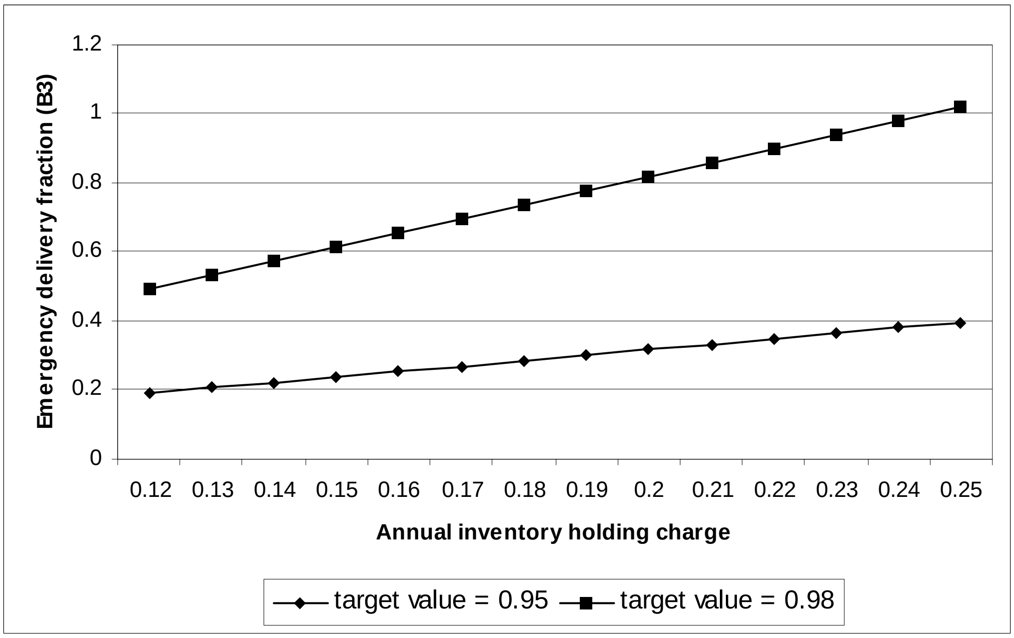

Figure 9.2. Control parameter values (B_3 policy)

9.7.1 The optimum replenishment levels

The (T, S) system will be separately simulated for the B_2, B_3 and P_2 criteria. In all cases the probability density function is necessary in order to apply the corresponding optimal solutions. Under the assumption that demand follows the negative binomial distribution, Prichard and Eagle (1965) assist the calculation of the probability that a demand of certain size will occur as follows:

$$p(0) = \left(\frac{1}{Z}\right)^{\frac{\mu_{L+T}}{Z-1}} \qquad (9.31)$$

and

$$p(x) = \frac{\left[\left(\frac{\mu_{L+T}}{Z-1}\right) + x - 1\right]}{x} \frac{Z-1}{Z} p(x-1) \qquad (9.32)$$

where:

$x = 1, 2, 3........$

$\mu_{\mathrm{T+L}}$ is the expected demand over $T + L$

$\sigma_{\mathrm{T+L}}$ is the variability of demand over $T + L$, and

$Z = \frac{\sigma^2_{L+T}}{\mu_{L+T}}$

The proof of this recursive formula is given in Kwan (1991).

The simulation will be performed in a dynamic way. We will explore what would have happened if any of the estimation procedures had been used in association with the (T, S) model in practice. Hence, μ_{L+T} and σ^2_{L+T} will be estimated based on the forecasts produced by the estimation procedures and the smoothed MSE over $L + T$ (see section 9.3) rather than the actual mean and variance of the intermittent demand series under concern:

$\mu_{L+T} = Y'_{T+L} = (T + L)\ Y'_t$ and

$\sigma^2_{T+L} = MSE_{t,T+L}$

where:

$$MSE_{t,T+L} = \alpha \left\{ \sum_{i=t-T-L+1}^{t} (Y_i - Y'_i) \right\}^2 + (1-\alpha) MSE_{t-1,T+L}$$

A significant restriction associated with the use of the NBD is the fact that σ^2_{L+T} must be greater than μ_{L+T}. That is, the estimate of the variance of demand over lead time plus review period must be greater than the estimate of the mean demand over the same period if the probabilities are to be calculated. No theoretically justified remedies to this problem have been identified and the author resorts to the method employed by Kwan (1991): if the variance is less than the mean, the variance is taken to be 1.05 times the mean.

9.7.2 A Simple Moving Average estimator (SMA)

The estimation procedure employed by Unicorn Systems (UK) Ltd., when dealing with intermittence, is a Simple Moving Average method (SMA). The method requires specification of only one control parameter, namely the length of the moving average (number of periods to be considered). It has been reported that for systems employing monthly forecast revisions, the length of the SMA is most commonly somewhere between 3 and 12 points, whereas for weekly revisions it is between 8 and 24 points (Johnston et al, 1999a). In the particular software package the number of periods is set to 13 (independently of the length of the review period). This value has been reported, by the software manufacturer, to provide the best results.

Sanders and Manrodt (1994) conducted a survey of forecasting practices in US corporations and found moving averages to be the most familiar and most used quantitative technique. Similar results are reported in Hughes (2001) with respect to the electronics and financial industries in Scotland and in Kogetsidis and Mathews

(1998) with respect to British companies in general. It is important to note that the theoretical properties of SMA under the stationary mean model assumption have been well known for decades (e.g. Gilchrist, 1976) whereas its theoretical properties for employment with a steady state model have just recently been explored (Johnston et al, 1999b). No theoretical results have been presented in the academic literature regarding the application of moving averages in an intermittent demand context.

Sani (1995) reported that moving averages are used in many real world cases in order to deal with intermittence. Sani and Kingsman (1997) found that the simple moving average performs significantly better than EWMA or even Croston's method in an intermittent demand context (see also chapter 4).

The SMA has not been discussed in this thesis and its theoretical properties have not been explored in the context of our research. Nevertheless, it can still be included in the simulation experiment since it reflects a popular industry approach to forecasting intermittent demand. The length of the Moving Average can be set to 13 (MA(13)) and the method can be viewed as a benchmark for the purpose of analysing the simulation results. That is, we will explore how much better or worse other methods perform in comparison with the estimation procedure currently employed by a commercial software package.

9.7.3 Initial conditions

In order to apply this method in practice the last 13 periods of demand data are necessary. The first estimate will be produced at the end of period 13, and this will be the estimate of demand in period 14. The exponential smoothing forecasting process can also be initiated by considering the average demand per unit time period (for the first 13 periods). The first exponentially smoothed estimate (of demand in period 14) is the average demand over the first 13 periods. In a similar way, the first exponentially smoothed estimate of demand size and inter-demand interval can be based on the average corresponding values over the first 13 periods.

In order to calculate the first estimate of the MSE at the end of period 13, we need to know the last $L+T$ estimates of demand. So if, for example, the lead time is 3 periods, and the review period is 1 period, the one step ahead estimates made at the end of months 9, 10, 11 and 12 are required. Those estimates will also be approximated by the average demand over the first 13 periods. Ideally, we would prefer to split all the demand data series into three parts. The first would be used to initialise the estimates of the level, the second to obtain an initial estimate of the MSE (using an out-of-sample rather than an in-sample estimator of the level) and the third to generate the out-of-sample results to be used for comparison purposes. Nevertheless, given the short data series available for simulation, the above discussed approximation is the most reasonable that we can make. Demand sizes and inter-demand intervals estimates will be approximated in the same way.

If no demand occurs in the first 13 periods, the initial EWMA and Moving Average estimates are set to zero and the inter-demand interval estimate to 13. As far as the demand size is concerned, it is more reasonable to assign an initial estimate of 1 rather than 0.

Iyer and Schrage (1992) discussed the application of a deterministic (T,s,S) model with constant lead times and a shortage cost per unit value per unit of time. The initial stock in that case was assumed to be equal to the first replenishment level S that was calculated. Despite the fact that we refer to a probabilistic demand context, this simplified assumption still seems reasonable. For the purpose of our research it will also be assumed that no orders are due to arrive at any point in time.

9.7.4 Updating of parameters

The estimation procedures that will be considered for the purpose of our simulation experiment are: Croston's method, Approximation method, EWMA, MA(13). For Croston's method and the Approximation method, the estimates are updated only at the end of the periods when demand occurs. If no demand occurs, the estimates remain the same. When EWMA and Moving Average are utilised, the estimates are updated at the end of every period, independently of whether demand has occurred or not during this particular period. Once the one step ahead estimates have been produced they are multiplied by $L+T$ (stationary mean model assumption) so that an estimate of demand during lead time plus one review period can be generated. The

MSE (over the lead time plus review period) associated with the application of an estimation procedure in practice is updated at the end of every period ($MSE_{t,T+L}$, equation (9.10)). The MSE updating procedure will be the same for all methods, including the moving average one.

The application of the moving average method is consistent with a different demand variability estimation procedure, namely one that averages the squared error over the last *N (13)* periods. Nevertheless, it is important to note that with this simulation exercise we wish to explore inventory control differences between alternative methods of estimating the mean demand level and not between alternative approaches of estimating the variability of demand. In that respect we want to introduce a standard way of updating the variability of the forecast error, so that all inventory control performance differences can be attributed to the forecasting accuracy of the estimators. As such the variability of the forecast error given by the moving average method will be updated based on (9.10) and consequently the inventory control results given by the MA method will vary with the smoothing constant value.

The estimated mean and variance of demand over $L+T$ are incorporated in the Prichard and Eagle formulae (equations (9.31) and (9.32)) so that the optimum replenishment level S can be derived. No safety factor needs to be calculated since the optimum S is derived directly by satisfying inequalities (9.21), (9.27) and (9.30) for the B_2 , B_3 and P_2 criterion respectively. Each time that the mean and variance of demand over $L+T$ are updated, a new replenishment level is also calculated.

The net stock (stock on hand – backorders) at the end of the previous period plus any receipts during this period are used to satisfy the demand occurring at any point during the current month/week. Therefore an implicit assumption that we make is that orders are received at the very beginning of the current month so that the current end of the period net stock = net stock at the end of the previous period + receipts during this period – demand during this period. If there are no backorders, the net stock equals the stock on hand. If there are backorders, the net stock is negative. Under the B_2 policy we proceed with the following period calculations using the negative net stock. If the B_3 criterion is used, the net stock at the end of the period is set to zero, since unfilled demand will be satisfied in the following day from an emergency order.

In the same way, the amount to be ordered (scheduled end of the review period order rather than emergency order) equals S - inventory position, where the inventory position is the net stock plus any order to arrive in the following $L+T$ periods. For the B_2 policy the net stock may be positive, negative or zero, whereas for the B_3 criterion the net stock cannot be less than zero.

Under the P_2 policy, the effect of satisfying unfilled demand from the next scheduled replenishment quantity will be evaluated.

9.7.5 Simulation output and statistical significance of the results

The unit cost information is not available for any of the SKUs included in our empirical data sample. Therefore, no inventory cost results can be generated in the simulation exercise. Volume differences can be considered instead, regarding the number of units kept in stock for the alternative estimators assessed in our experiment. Customer Service Level (CSL) results will also be generated and they can be related directly to performance differences as far as the number of units backordered is concerned.

The lack of availability of the unit cost information is viewed as a major restriction. This is because we cannot assess the cost consequences of employing one estimator instead of one or all others and as such, we cannot fully demonstrate the empirical utility of our theoretical findings. Nevertheless, the inventory holding and CSL results should allow us to comment on performance differences between the alternative estimators and derive conclusions that will be of a significant practical value.

We record the average monthly number of units in stock and the Customer Service Level (CSL) achieved (percentage of units satisfied directly from stock) by using the estimator under concern on each of the real demand data series, for all the control parameter combinations.

The stock on hand at the end of period t, which is the starting stock for the period $t+1$ (OH_t) reduces or increases to the stock on hand at the end of the period $t+1$

(OH_{t+1}). Therefore it is reasonable to assume that the average stock hold during period $t+1$ is:

$$Average\ Holding\ Stock_{t+1} = \frac{OH_t + OH_{t+1}}{2} \tag{9.33}$$

Subsequently, the average number of units in stock over any time period can be calculated as follows:

$$Average\ Holding\ Stock\ per\ period = \frac{OH_t + 2\sum_{i=t+1}^{t+n-1} OH_i + OH_{t+n}}{2n} \tag{9.34}$$

where $n = 11$.

The customer service level (CSL) achieved, by the estimator under concern, is calculated based on (9.35):

$$CSL = \frac{Total\ Demand\ \text{-}\ Backorders}{Total\ Dema\ nd} \tag{9.35}$$

where *Total Demand* and *Backorders* is the total number of units demanded and backordered respectively over all the time periods (n) that are taken into account in the simulation experiment.

Kwan (1991) and Sani (1995) argued (and in the case of the latter research demonstrated empirically) that the normality assumption cannot be met when stock control results are considered. Kwan (1991) tested the effect that different demand distributional assumptions have on the stock control of slow moving items. Sani (1995) compared alternative forecasting and inventory control methods in an intermittent demand context. In both cases it was suggested that non-parametric procedures should be used to test whether or not the stock control performance differences are statistically significant.

In particular, Kwan (1991) used the Sign test, while Sani (1995) tested the statistical significance of his results by employing the Percentage Best (PB) measure, Wilcoxon test (Wilcoxon, 1947) and the Kruskal-Wallis test. The Kruskal-Wallis test is a generalisation of the Wilcoxon rank sum test and is used in testing whether the means of more than two populations are the same or not. The test is an alternative non-parametric procedure to the *F* test for testing equality of means in one-way Analysis Of Variance (ANOVA). As such, the test cannot be used to test whether there is a significant difference between two estimators. The Sign test and the Wilcoxon rank sum test are non-parametric equivalents to the *t*-test for matched pairs and the *t*-test for independent samples respectively.

It has been argued, in the academic literature (see chapter 8), that the Percentage Better and Percentage Best (PBt) are more intuitive non-parametric measures and they are easier to interpret than ranking non-parametric procedures. Sani (1995) demonstrated that the Percentage Best measure gives similar results to the Wilcoxon test in an inventory control context. In his research, the PBt measure resulted in straightforward ordering as to which method performs best, second best etc. Finally, the PBt is likely to be of greater importance (than the PB), from a practitioner's perspective, since each method is tested against all other estimators. As such, we choose the PBt measure for generating pair-wise comparison results. We will test the hypothesis that the two methods under concern perform identically. The alternative hypothesis is that one estimator performs better than the other in terms of (a) percentage of times (series) that it results in the lowest number of units in stock, (b) percentage of times (series) that it results in the highest service level. The *Z*-test statistic (difference between population proportions, see sub-sub-section 8.9.3.2, equation (8.18)) will be used to test the statistical significance of the results.

The Percentage Best measure can provide us with valuable information about which method performs better/best but not by how much. Therefore, a relative measure needs to be introduced, to indicate the performance differences in descriptive terms. An appropriate measure (and particularly relevant to the purposes of our research) is the "Average Percentage Regret (APR)" introduced in Sani and Kingsman (1997). The APR measures the regret of using a particular method (estimator) compared to the best possible attainable performance over a number of series.

In the context of our research, and when results are generated with respect to the average number of units in stock, the APR of using estimator x, across all series, is given by (9.36):

$$StockAPR_x = \frac{\sum_{i=1}^{n} \frac{S_{x,i} - Mn_i}{Mn_i}}{n} \tag{9.36}$$

where:

i is the particular demand data series considered

$n = 3{,}000$

$S_{x,i}$ is the average number of units in stock resulted from the employment of estimator x on series i, and

Mn_i is the lowest average number of units in stock achieved (by one of the estimators considered) on the particular series.

When results are generated on CSL, the APR is the amount each estimator falls short of the maximum possible CSL across all series:

$$CSLAPR_x = \frac{\sum_{i=1}^{n} \frac{Mx_i - CSL_{x,i}}{Mx_i}}{n} \tag{9.37}$$

where:

$CSL_{x,i}$ is the CSL (%) resulted from the employment of estimator x on series i, and

Mx_i is the maximum CSL achieved (by one of the estimators considered) on the particular series.

The APR gives a simple and clear idea of the relative performance of each estimator and it was argued (Sani and Kingsman, 1997) to be particularly meaningful from a practitioner's perspective. As such, this measure will be used for generating comparative descriptive results in our simulation experiment. The interpretation of this measure is very straightforward. The measure indicates how much better each estimator could perform in comparison with the best possible attainable performance. The method with the least APR is obviously the one that should be preferred in practice.

9.8 Conclusions

A periodic inventory control model of the (T, S) form has been developed in this chapter. The model will serve the purpose of assessing the empirical utility of our theoretical findings. That is, the thesis has focused so far on the issue of forecasting intermittent demand and how the forecasting accuracy can be improved. However, we recognise that improvements in forecasting accuracy are not of great practical importance unless they are reflected in an inventory control situation.

The periodic nature of our model is dictated by the empirical data available for simulation. The model has the following characteristics: all demand not satisfied directly from stock is backordered and met from the next scheduled replenishment quantity or from an emergency delivery on the following day; the variability of demand over lead time plus review period is estimated by using the smoothed MSE approach; the demand over lead time plus one review period is approximated by the NBD; the managerial constraints imposed on the system are:

- a specified shortage fraction per unit value short, equation (9.21)
- a specified emergency delivery fraction per unit value short, equation (9.27)
- a specified customer service level, equation (9.30).

Comparative results will be generated with respect to volume differences, regarding the average number of units in stock for each of the estimators considered, and the Customer Service Level (CSL) achieved. No inventory cost results will be generated due to the limited information available for our empirical data sample. Two accuracy measures will be employed for comparison purposes: the Percentage Best (PBt) and the Average Percentage Regret (APR).

CHAPTER 10

Empirical Analysis - Forecasting

10.1 Introduction

The purpose of our empirical analysis is to assess the practical validity and utility of the main theoretical findings of this research. It is to discuss the extent to which the theory developed in chapters 3, 4, 5 and 6, but also elsewhere in the academic literature and in particular in Croston (1972), holds for a large number of real world cases. For this to be done a simulation experiment will be conducted on 3,000 real demand data series that come from the automotive industry. The simulation experiment consists of two main modules: forecasting and inventory control. The former module will assist our efforts to assess the empirical validity of our theoretical findings while the latter will be used to check the empirical utility of the theory developed in this research.

Using the first module the forecasting performance of Croston's method, EWMA, Approximation method and the 13 period Moving Average (MA) will be simulated on the real demand data. Subsequently, the performance of the estimators will be analysed with respect to their bias and their forecasting accuracy. The issue of categorisation of intermittent demand patterns (under what conditions one method performs better than one or all other methods?) will also be explored, by using the real data set available for this research. The accuracy measures to be used for those purposes and the tests to be employed for demonstrating statistical significance of the results have been discussed in detail in chapter 8.

The issue of the variability of the forecast errors is not explicitly addressed in our simulation experiment. By researching the validity of the theoretical rules developed in chapter 6 some conclusions can also be drawn regarding the variability of the forecast errors.

Using the second module, stock control results will be generated for all estimators. Some conclusions can then be drawn about the practical implications of the theory developed in this research. This thesis focuses on forecasting for periodic inventory control (see chapters 2 and 9). The model to be used for simulation purposes is of the periodic, order-up-to-level nature. The stock control model has been discussed in the previous chapter, where simulation details regarding both the inventory control and the forecasting module of our empirical data simulation experiment are also given (see section 9.7).

At the time of analysing the empirical results it was decided that a separate discussion of the forecasting and inventory control findings was conceptually the best option. As mentioned above, the forecasting results relate to the empirical validity of our theory while the inventory control results to the utility of that theory in “real world” applications. Distinguishing between forecasting and inventory control results is also important for presentation purposes, because of the considerable size of the simulation output. Finally, the empirical analysis has led to a large number of findings across both areas. In order to do justice to the outcomes, forecasting and inventory control results will be presented separately. In this chapter we focus on the empirical validity of our results (forecasting module) and in chapter 11 the empirical utility of our findings (inventory control module) will be discussed.

This chapter is structured as follows: in section 10.2, information regarding the real demand data series available for simulation is presented. In section 10.3 the simulation results are analysed with respect to the bias of intermittent demand estimates, while in section 10.4 the intermittent demand estimators are compared with respect to their forecasting performance. The issue of categorisation of “non-normal” demand patterns is explored in section 10.5 and, finally, the conclusions of this chapter are presented in section 10.6.

10.2 Real demand data series

The data sample available for this research consists of the demand histories for 3,000 SKUs. The data sample comes from the automotive industry and has been provided by Unicorn Systems (UK) Ltd. Some information regarding the company's activities has been presented in section 2.5. Neither product description nor cost information has been given to us for any of the files. Each series covers the monthly demand history of a SKU over a two-year period. All series have been treated as intermittent, by Unicorn Systems, in practice. The average inter-demand interval ranges from 1.04 to 2 months and the average demand per unit time period from 0.5 to 120 units. The average demand size, when demand occurs, is between 1 and 194 units and the variance of the demand sizes between 0 and 49,612 (the squared coefficient of variation ranges from 0 to 14). The statistics have been calculated considering all 24 demand data periods. The sample contains slow movers, erratic and lumpy demand items as well as intermittent demand series with a constant (or approximately constant) size of demand, when demand occurs (see figures 10.2 and 10.3). The distribution of the real demand data files, with respect to their average inter-demand interval and the squared coefficient of variation, is indicated in figure 10.1.

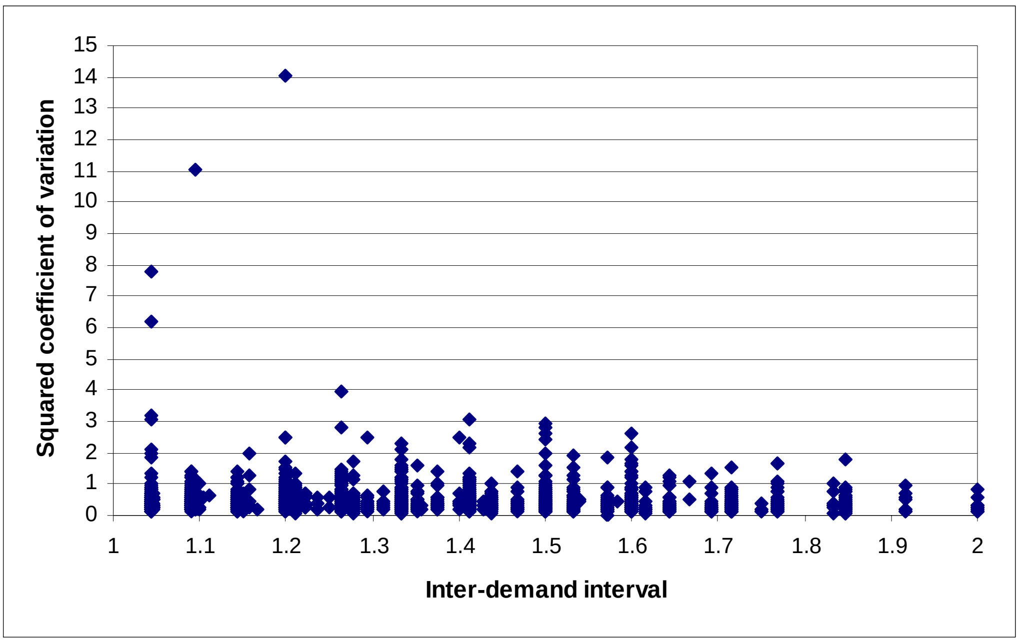

Figure 10.1. The real demand series characteristics

The range of the squared coefficient of variation values is very wide. The demand files are well-suited to the testing of the categorisation results developed in this thesis since, as shown below, each of the categories are well represented in the sample. However, the sample does not include highly intermittent demand items and therefore our results cannot be assessed for such data. Considering the demand categorisation schemes developed in chapter 6, we indicate, in figures 10.2 and 10.3, the number of files that fall within each of the designated categories.

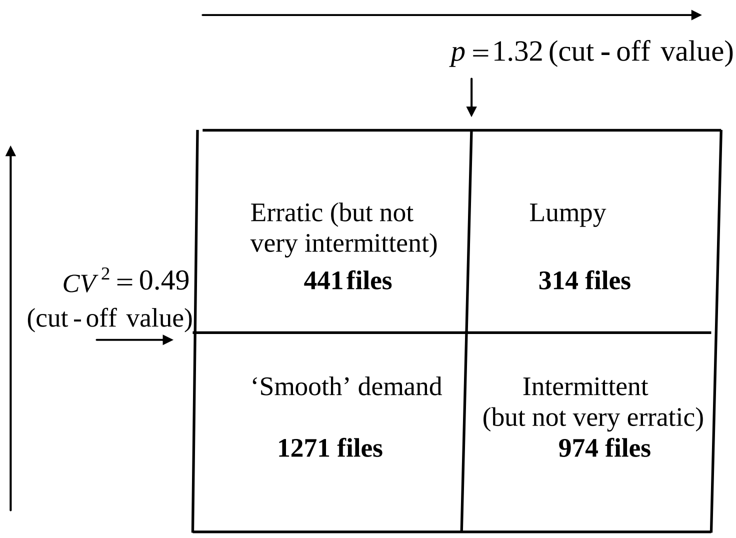

Figure 10.2. Categorisation of empirical data series (re-order interval systems)

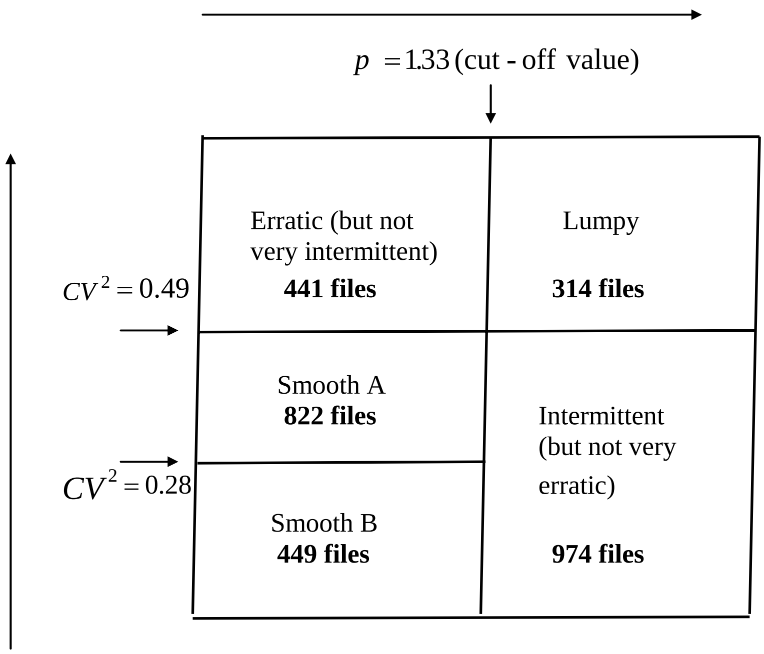

Figure 10.3. Categorisation of empirical data series (re-order level systems)

As mentioned above, the product description of the SKUs has not been specified and no information has been revealed as to the part of the supply chain in which they reside. The only other information accompanying the data sample is the following:

- An estimate of the annual inventory holding charge per unit value (25%)
- An estimate of the backorder cost per unit value short (28%)
- The lead time, which is approximately 3 months.

All three estimates refer to the whole data sample. No unit cost information is available for any of the SKUs.

Despite the fact that each series consists of 24 demand data periods, results will be generated only on the latest 11 periods. The MA method has not been discussed in this thesis and no theoretical derivations have been made regarding this method's performance on intermittent demand data. Nevertheless, a 13 period MA is currently used by Unicorn when dealing with intermittent demand. That is, this particular method has been used in practice to deal with the "non-normal" nature of the series available for this research. As such, the method's performance can be viewed as a benchmark against which the performance of the other estimators can be compared. The Moving Average method has been shown empirically to outperform EWMA and Croston's method in an inventory control context (Sani and Kingsman, 1997). The use of this method for simulation purposes necessitates the exclusion of the first 13 periods data for generating results. That is, to initialise the particular method's application, the first 13 period demand data need to be used. The other estimators can be initialised by withholding fewer demand data periods. Nevertheless, for consistency purposes it has been decided to initialise all methods' application by using the same number of periods. Therefore, the "out of sample" comparison results will refer to the latest 11 monthly demand data.

10.3 The bias of intermittent demand estimates

In this section the empirical results are analysed with respect to the biased or unbiased nature of the estimators considered in our simulation experiment. Results are first presented for the Mean signed Error (ME) accuracy measure by employing the *t*-test (for the population mean). To account for any potential scale dependencies and a potential asymmetric shape of the error distribution the original analysis is repeated for "scaled MEs" and a non-parametric test, the Wilcoxon test, is also employed. Statistically significant results at the 1% level are emboldened while significance at the 5% level is presented in italics.

10.3.1 ME results

The most obvious accuracy measure to be employed for testing the biased or unbiased nature of an estimator is the Mean signed Error (ME). The ME has been defined in sub-section 8.7.1, where details regarding the ME results generation process are also given. The ME is a scale dependent accuracy measure. The ME distribution though, across series, is expected to be fairly symmetric (or at least not as badly skewed as the distribution of other absolute error measures) since the original sign of the error is retained when summary results are generated. The mean and variance of the Mean Errors (MEs) across all 3,000 files for the 13 period Moving Average method (MA(13)) and all smoothing methods (for all α smoothing constant values and lead times considered in the simulation experiment) are presented in *Appendix 10.A* of the thesis. The test statistic, for testing whether the independently drawn sample of MEs comes from a population with a mean $\mu = 0$, is the *t* statistic (see equation (8.2)). The critical values of the *t* statistic for rejecting the null hypothesis, that the estimation procedure under concern is unbiased, are (+,-) 1.96 and (+,-) 2.57 for the 5% and 1% significance level respectively. The bias has been recorded as: Forecast minus Demand, so that a plus (+) sign indicates that we overestimate the demand level whereas a minus (-) sign shows that we underestimate it. The results obtained for each of the three smoothing estimation procedures are presented in the following tables for different α values. The results regarding the performance of the MA(13) are presented only in the first table.

α = 0.05	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	1.59	*1.99*	*2.25*	**2.59**	**4.02**	**4.74**
Croston	*2.38*	**2.66**	**2.91**	*2.44*	**3.89**	**4.66**
Approxim.	0.18	0.67	1.04	0.31	1.89	**2.72**
MA(13)	*2.42*	*2.44*	*2.20*	**3.72**	**4.81**	**5.09**

Table 10.1. *t*-test for the population mean, α = 0.05

α = 0.1	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	*2.19*	**2.66**	**2.93**	**4.13**	**5.56**	**6.20**
Croston	**3.24**	**3.42**	**3.62**	**3.00**	**4.53**	**5.29**
Approxim.	-1.82	-0.95	-0.32	-1.84	0.16	1.16

Table 10.2. *t*-test for the population mean, α = 0.1

α = 0.15	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**2.64**	**3.16**	**3.49**	**5.73**	**7.04**	**7.54**
Croston	**4.20**	**4.19**	**4.31**	**3.66**	**5.19**	**5.90**
Approxim.	**-4.28**	**-2.81**	-1.79	**-4.36**	-1.82	-0.52

Table 10.3. *t*-test for the population mean, α = 0.15

α = 0.2	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**2.96**	**3.53**	**3.84**	**7.34**	**8.44**	**8.74**
Croston	**5.22**	**4.93**	**4.96**	**4.36**	**5.84**	**6.46**
Approxim.	**-7.10**	**-4.87**	**-3.35**	**-7.13**	**-3.98**	*-2.29*

Table 10.4. *t*-test for the population mean, α = 0.2

The ME results, for α = 0.05, are consistent with the theory presented in this thesis. At the 1% significance level, the EWMA method is shown to be unbiased when all points in time are considered whereas the method is biased when forecasting performance is simulated on issue points only. Croston's method is biased (at the 5% significance level for $L=1$ and at the 1% significance level for $L>1$) independently of which points in time are taken into account for generating results. The Approximation method is unbiased, at the 5% level, in all but one of the simulated conditions. For issue points only, the method performs slightly worse than it does for all points in time and, when the lead time is set to 5 periods, the method is significantly biased at the 1% level. The MA(13) method is unbiased, at the 1% level, for all points in time. The method is shown to be biased, at the 1% level, when issue points only are considered.

For all points in time, and as the smoothing constant value increases, the performance of the EWMA method deteriorates. In fact for $\alpha \geq 0.1$, the method can now be shown to be unbiased (at the 1% significance level) only when one step ahead forecasts are generated and for $\alpha = 0.1$. For all other simulated scenarios the method is shown to be biased. This finding is not consistent with the theory developed by Croston (1972). The empirical evidence suggests that the EWMA systematically overestimates the mean demand level. Results to be presented in the following sub-section support this statement. The unexpectedly poor performance of EWMA, observed in this sub-section, is clearly an issue that requires further examination. We will not attempt to resolve this theoretical inconsistency in the rest of this thesis but in the following sections and sub-sections we will attempt to provide some empirical insight regarding the EWMA's performance that may be of interest to researchers who will take this issue forward.

For issue points only and for a specified lead time length, the bias of EWMA increases with the smoothing constant value and this is consistent with the theory. The same is shown to be the case for Croston's method, for both all and issue points in time only (see *Appendix 10.A*, tables *10.A.1 – 10.A.4*).

For $\alpha = 0.1$ the Approximation method is still unbiased at the 5% level and the sign of the ME indicates that, in the majority of cases, the method slightly underestimates the level of demand, as theoretically expected. As the smoothing constant value increases the Approximation method underestimates the mean demand level by even more. The

Approximation method is theoretically, approximately, unbiased for "low" smoothing constant values. The simulation results indicate that the deterioration in bias as α increases may be more rapid, in practice, than was anticipated by the theoretical results. The ME results also indicate that, for "higher" α values (0.15, 0.2), the method's performance improves according to the *t*-test as the lead time length increases. This is because the variability of the forecast errors produced by the Approximation method naturally increases significantly with the lead time but the average ME itself does not necessarily increase. In fact the absolute average ME may remain the same or even be reduced as the lead time increases.

Overall, we can conclude that Croston's method is biased, in all simulated conditions, as was expected in theory. EWMA is biased for issue points only (as expected) but, for $\alpha \geq 0.1$, it is also biased for all points in time. This result is inconsistent with theory and requires further research. The Approximation method is unbiased for low values of α ($\alpha \leq 0.1$) but shows a faster deterioration in bias for higher α values than was expected from our theoretical investigation. The MA(13) is unbiased, at the 1% significance level, for all points in time whereas the method is biased when issue points only are considered. No theoretical results have been developed in this thesis (or elsewhere in the academic literature) regarding the performance of MA(13) (or in general of the moving average estimator) in an intermittent demand context. Nevertheless, and if Croston's model is assumed, the MA(13) is theoretically expected to perform as the EWMA estimator, i.e. a certain bias should be expected when issue points only are considered whereas the estimator is theoretically unbiased in a re-order interval context. In that respect the ME results presented for the MA(13) are consistent with the theory according to the *t*-test, at the 1% significance level.

10.3.2 Scale independent ME results

As discussed in chapter 8, the main disadvantage of the ME is that it is not a relative error measure. Therefore, the analysis of the simulation output is repeated for "scaled MEs". The originally calculated MEs, per method, per series, are divided by the average demand per unit time period (for the series under concern) so that scale dependencies are eliminated. The *t*-test, equation (8.2), can then be applied to the new, scale independent, sample of MEs, in order to check the originally obtained results. In addition, in order to account for a potentially skewed distribution of the MEs, a non-parametric ranking test, the Wilcoxon test, is introduced, that considers only the relationship between the MEs, in order to generate results, without taking into account the actual size of them. The test has been discussed in detail in sub-section 8.7.2. The sample size (3,000 series) allows us to test the statistical significance of the results by using the normally distributed Z test statistic (see equation (8.4)). The critical values for rejecting the null hypothesis that the estimator under concern is unbiased are as discussed in the previous sub-section. The plus (+) sign indicates that the method under concern overestimates the demand level, whereas the minus (-) sign that it underestimates it. We first present the scaled ME results and then the Wilcoxon test results.

α = 0.05	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**4.48**	**5.20**	**5.56**	**7.55**	**10.33**	**11.12**
Croston	**6.84**	**7.36**	**7.63**	**7.51**	**10.31**	**11.27**
Approxim.	**2.93**	**3.72**	**4.23**	**4.03**	**6.75**	**7.91**
MA(13)	**6.79**	**6.88**	**6.26**	**10.30**	**13.00**	**12.62**

Table 10.5. *t*-test for the population mean (scaled MEs), α = 0.05

α = 0.1	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**5.29**	**6.23**	**6.56**	**10.61**	**13.82**	**14.07**
Croston	**8.30**	**8.68**	**8.78**	**8.26**	**11.40**	**12.15**
Approxim.	-0.79	0.53	1.48	0.38	**3.43**	**4.95**

Table 10.6. *t*-test for the population mean (scaled MEs), α = 0.1

α = 0.15	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**5.85**	**7.00**	**7.34**	**13.91**	**17.46**	**16.98**
Croston	**10.07**	**10.20**	**10.03**	**9.26**	**12.68**	**13.13**
Approxim.	**-5.35**	**-3.17**	-1.52	**-3.77**	-0.40	1.74

Table 10.7. *t*-test for the population mean (scaled MEs), α = 0.15

α = 0.2	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**6.14**	**7.54**	**7.89**	**17.35**	**21.12**	**19.75**
Croston	**12.01**	**11.81**	**11.30**	**10.43**	**14.07**	**14.15**
Approxim.	**-10.70**	**-7.38**	**-4.76**	**-8.37**	**-4.70**	-1.67

Table 10.8. *t*-test for the population mean (scaled MEs), α = 0.2

When the scaled MEs are considered, the EWMA, Croston's method and MA(13) are shown to be always biased independently of the simulated scenario. The original error distribution for these methods appears to be positively skewed. By reducing the scale dependencies the average ME decreases. At the same time, though, there is a considerable reduction of the variability of the MEs, across series. As such, it becomes easier to demonstrate the biased nature of these methods. Similar comments can be made for the Approximation method when α is set to 0.05. For higher smoothing constant values and in the simulated cases that the Approximation method has been already shown to be biased (negative bias), the original error distribution is negatively rather than positively skewed. By reducing the scale dependencies the average ME reduces in absolute terms but there is also a considerable reduction of the variability of the MEs across series.

The *t* test statistic value obtained on the scaled MEs is greater, for most methods, than the value obtained on the originally calculated MEs (or lower in case that the original value is negative). This is not true for the statistically significant results obtained in the previous sub-section for the Approximation method, when the average ME is negative. In those cases the performance of the Approximation method seems to improve when results are generated on the scaled MEs. The Approximation method performs particularly well for $\alpha = 0.1$ and 0.15. As mentioned above, for $\alpha = 0.05$ we greatly overestimate the mean demand level, the correction to Croston's estimates being insignificant in volume terms. For $\alpha = 0.2$ we greatly underestimate the mean demand level and, as discussed in the previous sub-section, the estimator is biased with the exception of the method's performance in a re-order level context when the lead time length equals 5 periods. The results presented in tables 10.5 – 10.8 indicate that the Approximation method is, overall, the least biased of the four methods discussed in this chapter and this matter will be further assessed later in this sub-section but also in section 10.4.

Apart from the discrepancies between the original ME and scaled ME results when $\alpha = 0.05$, the scale differences have not affected considerably the validity of the comments made in the previous sub-section. (Please note that the MA(13) which is now found to be biased was, in the previous sub-section, shown to be unbiased only at the 1% significance level.) This is because the sign of the MEs is taken into account when we arithmetically average them across series and as such the scale dependence does not, as expected, have a great effect on the average ME calculated for each of the estimators considered. Overall the original ME results are proven to be fairly accurate, without being particularly affected by scale differences among the real demand series. This is regarded as a significant finding since empirical evidence is now provided, in an intermittent demand context, to justify the fact that the ME, despite its absolute nature as an error measure, does not suffer from the serious scale dependence problem that other absolute measures like the Mean Squared Error (MSE) or Root Mean Squared Error (RMSE) do. Nevertheless, it is important to note that the results generated on the scaled MEs have given more confidence in our conclusions. The original results that have been affected by the scale dependence problem have been highlighted and we can now comment with more certainty on the biased (or unbiased) nature of alternative estimators.

We conclude that the results obtained by the ME are fairly reliable without being particularly affected by scale dependencies. Nevertheless, the confidence in our conclusions can be increased by generating scaled ME results. From a computational perspective, the only new input that is required is the average demand per period value, for each one of the series considered, and this piece of information will, most probably, be already available in any research exercise.

The final test to be employed for assessing the biased (or unbiased) nature of the estimation procedures considered in this chapter is the Wilcoxon test. The *Z* test statistic values are presented, for different smoothing constant values, in tables 10.9 – 10.12.

α = 0.05	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**3.37**	**4.78**	**5.97**	**6.05**	**8.80**	**10.12**
Croston	**5.22**	**6.40**	**7.61**	**6.01**	**8.82**	**10.18**
Approxim.	1.57	**3.14**	**4.49**	*2.32*	**5.57**	**7.22**
MA (13)	**5.48**	**6.49**	**6.65**	**8.57**	**11.12**	**11.31**

Table 10.9. *Wilcoxon test,* α = 0.05

α = 0.1	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**4.40**	**5.90**	**7.16**	**8.67**	**11.73**	**12.77**
Croston	**6.78**	**7.78**	**8.85**	**7.07**	**9.93**	**11.26**
Approxim.	-1.90	0.28	*2.14*	-1.08	**2.74**	**4.83**

Table 10.10. *Wilcoxon test,* α = 0.1

α = 0.15	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**4.67**	**5.65**	**6.01**	**9.68**	**12.63**	**12.31**
Croston	**6.47**	**6.61**	**7.40**	**7.17**	**8.69**	**9.68**
Approxim.	**-4.88**	**-2.48**	-0.47	**-3.16**	-0.02	1.80

Table 10.11. *Wilcoxon test,* α = 0.15

α = 0.2	*All Points in Time*			*Issue Points Only*		
	L.T.=1	L.T.=3	L.T.=5	L.T.=1	L.T.=3	L.T.=5
EWMA	**5.53**	**7.35**	**8.55**	**13.97**	**17.41**	**17.28**
Croston	**10.12**	**10.65**	**11.12**	**8.89**	**12.34**	**13.30**
Approxim.	**-11.75**	**-7.39**	**-3.70**	**-10.32**	**-5.07**	-1.19

Table 10.12. *Wilcoxon test,* α = 0.2

The Wilcoxon test results confirm the unbiased nature of the Approximation method for α = 0.1 and all points in time. For issue points only the estimator is unbiased when one step ahead forecasts are considered. The estimator is also unbiased for α = 0.05 and one step ahead forecasts, as well as for higher smoothing constant values when the lead time length is set to 3 or 5 periods. All other estimators are shown to be biased in all simulated cases.

The EWMA error distribution is always badly positively skewed and as discussed in the previous sub-section this finding is not consistent with the theory. Our discussion on the forecasting performance of EWMA is continued in the following sections.

The MA(13) is found to be biased at the 1% significance level and we may now comment on the biased nature of this particular estimator in both a re-order level and a re-order interval context. As discussed earlier in this section, if Croston's model is assumed, the MA(13) is theoretically expected to perform as the EWMA estimator, i.e. to be biased when issue points only are considered and unbiased when results are generated on all points in time. In that respect, the poor performance of the MA(13), in a re-order interval context, is an issue that requires further examination.

The Wilcoxon test results lead to exactly the same conclusions as the scaled ME results (with the exception of the unbiased nature of the Approximation method for α = 0.05 and lead time equals one). In all other cases the sign of the test statistic is the same and identical inferences can be made for the populations of the MEs obtained by alternative estimators across series. The Wilcoxon test results indicate that the Approximation method is, overall, the least biased of the four methods discussed in this chapter.

10.3.3 Conclusions

In this section the performance of four intermittent demand estimators was analysed with respect to their biased (or unbiased) nature. The natural accuracy measure to be used for that purpose is the Mean signed Error (ME). The ME though is theoretically a scale dependent error measure. Therefore results were also generated on the scaled MEs (the original ME is divided by the average demand per unit time period for the series under concern) and a non-parametric procedure was also introduced that does not take into account the actual size of the errors. In summary the results demonstrate that:

- Croston's method is biased and this is consistent with the theory developed in this thesis

- EWMA is biased, in a re-order level context, and this is consistent with the theory developed by Croston (1972)

- EWMA is also biased in a re-order interval context and this issue requires further examination

- The MA(13) method is biased in both a re-order level and a re-order interval context. The poor performance of MA(13), when all points in time are taken into account, is an issue that requires further research.

- The Approximation method can be claimed to be unbiased for α = 0.1. For α = 0.05 the method does not sufficiently correct for the bias implicit in Croston's estimates, unless one step ahead forecasts are considered. For α = 0.15, 0.2 the method performs well only in combination with longer lead times.

  The results discussed above will be re-considered in the following chapter where the performance of the Approximation method is analysed in terms of the service level resulting from its implementation in practice.

- Where bias had been theoretically anticipated, the sign of the bias is for all methods the theoretically expected one.

- The ME as an accuracy measure is not particularly affected by scale differences amongst the series considered for generating results. Nevertheless, in the next section it is shown that the choice of the scaled or non-scaled ME measure, for comparison purposes, affects the significance of the *t*-test results.

10.4 The accuracy of intermittent demand estimates

In this section the accuracy of the intermittent demand estimators considered in this chapter is tested, and comparative results are presented by using a range of accuracy measures. The Mean signed Error (ME) is first used to conduct pair-wise accuracy comparisons. Statistical significance is tested by using the *t*-test (difference between population means). The ME results have already been shown to be reliable without being particularly affected by the scale dependence problem. Nevertheless, to increase confidence in our conclusions the same exercise is repeated for scaled MEs.

A relative error measure (introduced in chapter 8) is also employed in this section, namely the Relative Geometric Root Mean Square Error (RGRMSE), which, theoretically and empirically (Fildes, 1992) does not suffer from the scale dependence problem (when results are generated across many series). In chapter 8 it was also argued that the RGRMSE, generated in a particular series, is not affected by extreme observations (outliers). Therefore, in this section we present results on forecasting accuracy starting with an analysis of a scale dependent error measure and then progressively eliminate the scale dependencies as well as the effect of the outliers on the error distributions.

In addition a non-parametric approach is also considered for generating results based on the Percentage Better and Percentage Best accuracy measures. Non-parametric tests sacrifice power in terms of using all available information to reject a false null hypothesis. This is because they consider relationships between the errors rather than the error sizes themselves. Nevertheless, they require no specific error population assumptions to be made and as such they may provide further insight as to how the alternative estimators behave.

At this point it is important to note that all smoothing methods are compared against MA(13) for different smoothing constant values. The performance of the moving average method is evaluated only for *N* (moving average length) = 13 and not for any other *N* value. In that respect, we put the class of simple moving average estimators at a relative disadvantage since a different (potentially more accurate) performance may be the case for the moving average, for a different moving average length. Nevertheless, the purpose of including the moving average method in this empirical experiment is to introduce a benchmark type of forecasting (and inventory control) performance against which the performance of other estimators can be assessed. All comments to be made in this section (but also in the following chapter) with respect to the performance of the MA(13) should not be misinterpreted as remarks regarding the performance of any other moving average length.

The statistically significant results at the 1% significance level to be presented in this section are emboldened while statistical significance at the 5% level only is indicated in italics.

10.4.1 Parametric tests

10.4.1.1 ME results

The average MEs across all series and the corresponding variances associated with the application of the estimators considered in this chapter on the real data are presented in *Appendix 10.A*. The *t*-test (difference between population means, equation (8.3)) has been used to test for the statistical significance of the pair-wise comparison results. We test the null hypothesis that both methods perform the same. The alternative hypothesis is that the first method performs less accurately that the second one (i.e. that the absolute average ME given by the first method is greater than the absolute average ME given by the second method). The critical values for rejecting the null hypothesis are 1.64 and 2.32 (one-sided test) for the 5% and 1% significance level respectively.

α = 0.05		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	0.60	0.97	1.60	0.40	0.25	1.43
	L.T.=3	0.52	0.91	1.42	0.20	0.34	1.14
	L.T.=5	0.50	0.85	1.33	-0.11	0.62	0.75
Issue Points Only	L.T.=1	-0.06	1.59	1.51	0.53	-0.59	*2.21*
	L.T.=3	-0.03	1.48	1.43	0.32	-0.36	*1.85*
	L.T.=5	0	1.41	1.40	0.01	-0.01	1.47

Table 10.13. *t*-test (difference between population means), α = 0.05

α = 0.1		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	0.86	0.18	1.01	0.21	0.64	0.38
	L.T.=3	0.63	1.16	*1.75*	-0.12	0.74	1.03
	L.T.=5	0.55	*1.82*	**2.34**	-0.52	1.06	1.31
Issue Points Only	L.T.=1	-0.70	1.55	0.83	-0.23	-0.46	1.30
	L.T.=3	-0.62	**3.78**	**3.11**	-0.47	-0.15	**3.28**
	L.T.=5	-0.56	**3.54**	**2.95**	-0.85	0.28	**2.71**

Table 10.14. *t*-test (difference between population means), α = 0.1

α = 0.15		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	1.32	-1.38	-0.06	0.14	1.08	-1.13
	L.T.=3	0.87	0.13	0.98	-0.32	1.15	-0.19
	L.T.=5	0.69	1.13	*1.80*	-0.83	1.51	0.29
Issue Points Only	L.T.=1	-1.32	0.80	-0.51	-1.01	-0.25	-0.24
	L.T.=3	-1.16	**3.62**	**2.41**	-1.24	0.11	**2.25**
	L.T.=5	-1.05	**4.95**	**3.84**	*-1.68*	0.62	**3.23**

Table 10.15. *t*-test (difference between population means), α = 0.15

α = 0.2		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	*1.92*	**-3.38**	-1.41	0.15	1.52	**-2.79**
	L.T.=3	1.20	-1.16	0.04	-0.44	1.55	-1.50
	L.T.=5	0.91	0.27	1.16	-1.06	*1.93*	-0.79
Issue Points Only	L.T.=1	*-1.97*	-0.14	*-2.05*	*-1.79*	-0.01	*-1.88*
	L.T.=3	*-1.68*	**3.03**	1.33	*-1.98*	0.38	0.89
	L.T.=5	-1.50	**4.55**	**3.00**	**-2.46**	0.96	*2.02*

Table 10.16. *t*-test (difference between population means), α = 0.2

No significant differences between the estimators are indicated for α = 0.05. As the smoothing constant value increases the superior performance of the Approximation method becomes apparent when issue points only are considered. Moreover it is important to note that as the α value increases the Approximation method performs significantly better than other estimators when the lead time is set to 3 or 5 periods (i.e. not for one step ahead forecasts). This is also the case when all points in time are taken into account.

For all points in time the Approximation method does not perform as well as it does when issue points only are considered, with the exception of the simulated scenarios referring to α = 0.1. For α = 0.05 the forecasting accuracy improvements achieved when the Approximation method is employed are not statistically significant. The results also indicate that α = 0.2 is a prohibitively large smoothing constant value for the Approximation method unless long forecasting lead times are the case.

The comparative results generated considering the bias of the alternative estimators indicate that the EWMA performs better than Croston's method, when all points in time are considered, and this is consistent with the theory. Even though the differences are not statistically significant (in all but one simulated scenarios), EWMA performs consistently better than Croston's method in a re-order interval context. The opposite is the case when issue points only are considered. The same comments can be made when we consider the comparison between Croston's method and MA(13). The Moving Average method seems to perform better than EWMA for

$\alpha \geq 0.1$, with statistically significant differences obtained when issue points only are taken into account.

Before we close this sub-sub-section, we view as important to conduct an additional ME comparison of the estimators discussed in this chapter considering their best α value performance. Our empirical analysis focuses on "out of sample" comparisons only, and the few (if any) demand occurrences during the "transient" interval have not allowed the optimisation of the α values used for the smoothing estimators (Approximation method, Croston's method, EWMA). By considering the same α value for comparison purposes, we may have put some smoothing estimators at a relative advantage/disadvantage. Therefore, in order to investigate more thoroughly the ME performance differences between all the estimators discussed in this chapter, pair-wise comparisons are conducted considering the best ME performance of each smoothing estimator with respect to the smoothing constant value. That is, for every lead time length (and for both all and issue points in time only) the best ME performance (lowest absolute ME) of each smoothing estimator, across the four simulated α values, is recorded and the *t*-test (difference between population means, equation (8.3)) is then used to test the statistical significance of the pair-wise comparison results.

The best performance is identified by considering all 3,000 series contained in our empirical sample since optimisation in a real world system will, most probably, occur across a number of files and not for individual SKUs. The MA(13) is still considered for comparison purposes even though the average length has not been "optimised".

The smoothing constant values for which each (smoothing) estimator performs best are indicated in the following table, for all the simulated scenarios (see also Appendix 10.A of the thesis).

		Best smoothing constant value		
		EWMA	Croston	Approximation
All points in time	L.T. = 1	0.05	0.05	0.05
	L.T. = 3	0.05	0.05	0.05
	L.T. = 5	0.05	0.05	0.1
Issue points only	L.T. = 1	0.05	0.05	0.05
	L.T. = 3	0.05	0.05	0.1
	L.T. = 5	0.05	0.05	0.15

Table 10.17. Best α value performance – lowest (absolute) ME

The bias of EWMA and Croston's method becomes minimum for α = 0.05 and this is true for all the scenarios considered in our experiment. As the lead time increases, the best α value for the Approximation method increases as well (see also sub-section 10.3.1). The *t*-test results of the best α value analysis are presented in table 10.18.

best α		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	0.60	0.97	1.60	0.40	0.25	1.43
	L.T.=3	0.52	0.91	1.42	0.20	0.34	1.14
	L.T.=5	0.50	1.39	*1.88*	-0.11	0.62	1.31
Issue Points Only	L.T.=1	-0.06	1.59	1.51	0.53	-0.59	*2.21*
	L.T.=3	-0.03	**2.83**	**2.76**	0.32	-0.36	**3.28**
	L.T.=5	0	**3.12**	**3.09**	0.01	-0.01	**3.23**

Table 10.18. *t*-test (difference between population means), best α value

The results presented in the above table indicate no significant bias differences, between the estimators considered in this chapter, in a re-order interval context, with the exception of the Approximation – Croston pair-wise comparison for L.T. = 5. For all points in time, the Approximation method performs better than any other estimator but not significantly so. When issue points only are taken into account the Approximation method performs best and the superior performance of the Approximation estimator is statistically significant.

The control parameter regarding the application of the Moving Average estimator (i.e. the moving average length) has not been "optimised" for the purposes of this comparison exercise. In that respect the MA(13) – Croston and MA(13) – EWMA pair-wise comparison results are rather surprising since no significant differences are indicated in any of the simulated scenarios. In fact, MA(13) performs better than EWMA for L.T. = 5 and all points in time but also better than Croston's method for all the scenarios referring to a re-order interval context.

10.4.1.2 Scaled ME results

Results have also been generated considering the scaled MEs and they are subsequently presented in tables 10.19 – 10.22.

α = 0.05		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	*1.91*	0.96	**2.80**	0.91	1.18	*1.93*
	L.T.=3	*1.72*	0.95	**2.62**	0.69	1.14	*1.67*
	L.T.=5	1.59	0.88	**2.44**	0.20	1.46	1.10
Issue Points Only	L.T.=1	0.15	**2.38**	**2.49**	1.17	-0.98	**3.67**
	L.T.=3	0.20	**2.41**	**2.57**	1.01	-0.78	**3.52**
	L.T.=5	0.24	**2.21**	**2.43**	0.50	-0.25	**2.78**

Table 10.19. *t*-test (scaled MEs, difference between population means), α = 0.05

α = 0.1		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**2.63**	**2.99**	**5.36**	1.09	1.59	**4.01**
	L.T.=3	*2.14*	**3.88**	**5.82**	0.54	1.60	**4.37**
	L.T.=5	1.81	**3.51**	**5.22**	-0.20	*2.00*	**3.31**
Issue Points Only	L.T.=1	-1.33	**7.08**	**5.61**	-0.15	-1.18	**6.90**
	L.T.=3	-1.31	**7.16**	**5.71**	-0.47	-0.84	**6.66**
	L.T.=5	-1.13	**6.39**	**5.19**	-1.02	-0.13	**5.38**

Table 10.20. *t*-test (scaled MEs, difference between population means), α = 0.1

α = 0.15		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**3.78**	-0.15	**3.43**	1.37	*2.22*	1.16
	L.T.=3	**2.86**	**2.42**	**5.07**	0.50	*2.23*	**2.76**
	L.T.=5	*2.24*	**4.01**	**6.11**	-0.49	**2.66**	**3.42**
Issue Points Only	L.T.=1	**-2.88**	**6.93**	**3.93**	*-1.68*	-1.10	**4.93**
	L.T.=3	**-2.87**	**11.88**	**8.77**	*-2.11*	-0.67	**9.22**
	L.T.=5	**-2.47**	**10.77**	**8.18**	**-2.62**	0.18	**7.91**

Table 10.21. *t*-test (scaled MEs, difference between population means), α = 0.15

α = 0.2		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**5.23**	**-4.12**	1.08	*1.71*	**2.97**	*-1.98*
	L.T.=3	**3.80**	-0.36	**3.31**	0.54	**2.93**	0.20
	L.T.=5	**2.83**	*2.08*	**4.80**	-0.69	**3.37**	1.30
Issue Points Only	L.T.=1	**-4.52**	**6.09**	1.52	**-3.36**	-0.84	**2.26**
	L.T.=3	**-4.47**	**11.44**	**6.78**	**-3.86**	-0.33	**6.71**
	L.T.=5	**-3.75**	**12.88**	**8.99**	**-4.26**	0.61	**8.15**

Table 10.22. *t*-test (scaled MEs, difference between population means), α = 0.2

The scaled ME results confirm the results obtained on the original MEs but more of the accuracy differences identified earlier in this section are now statistically significant. The superior performance of the Approximation method is clearly demonstrated when issue points only are considered. The same is the case when we refer to a re-order interval context. The estimator under concern performs significantly better than Croston's method in almost all simulated cases. The same is true for the Approximation – EWMA and the Approximation – MA(13) pair-wise comparisons. It is only for α = 0.2 and one step ahead forecasts that the MA(13) and EWMA perform significantly better than the Approximation method. Croston's method is shown to perform significantly worse than the EWMA method for all points in time, while the opposite is the case for issue points only, especially for higher smoothing constant values. Croston's method is also shown to perform significantly worse than the MA(13) in a re-order interval context while the superiority of Croston's method for issue points only is not statistically significant. The MA(13) method still performs better than EWMA for issue points only. No significant differences are indicated between them in a re-order interval context.

The results obtained on the scaled MEs demonstrate very clearly the differences in bias between the estimators considered in this chapter. Overall, there are not great discrepancies between the ME and scaled ME results in terms of which method performs better or worse. Statistical significance though is now more easily demonstrated and as such more conclusive results can be generated across all estimators as opposed to the pair-wise comparison level. Considering the scaled ME results, the estimators considered in this chapter can be ordered as follows in terms of their bias (the first method being the one with least bias):

	All points in time	**Issue points only**
1.	Approximation	Approximation
2.	EWMA	Croston's method
3.	MA(13)	MA(13)
4.	Croston's method	EWMA

(The overall measure of bias is calculated for each estimator as an equally weighted average of the absolute scaled MEs across all the control parameter combinations.)

At this point we view as important to assess the validity of the above proposed ordering scheme when the best smoothing constant values are considered for comparison purposes. As such, an additional scaled ME comparison exercise of the estimators discussed in this chapter is performed considering the best α value performances.

The best α values, with respect to the scaled ME (lowest scaled absolute ME) are indicated, for all the control parameter combinations and smoothing methods considered in our experiment, in table 10.23 (see also sub-section 10.3.2).

		Best smoothing constant value		
		EWMA	Croston	Approximation
All points in time	L.T. = 1	0.05	0.05	0.1
	L.T. = 3	0.05	0.05	0.1
	L.T. = 5	0.05	0.05	0.1
Issue points only	L.T. = 1	0.05	0.05	0.1
	L.T. = 3	0.05	0.05	0.15
	L.T. = 5	0.05	0.05	0.2

Table 10.23. Best α value performance – lowest (absolute) scaled ME

In all simulated cases the bias of EWMA and Croston's method becomes minimum for α = 0.05. When all points in time are taken into account, the scaled ME associated with the Approximation method becomes minimum for α = 0.1. In a re-order level context, as the lead time increases, the best α value for the Approximation method increases as well. The *t*-test results of the best α value analysis are presented in the following table.

best α		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	*1.91*	**2.78**	**4.68**	0.91	1.18	**4.01**
	L.T.=3	*1.72*	**3.45**	**5.15**	0.69	1.14	**4.37**
	L.T.=5	1.59	**3.00**	**4.59**	0.20	1.46	**3.31**
Issue Points Only	L.T.=1	0.15	**5.33**	**5.39**	1.17	-0.98	**6.90**
	L.T.=3	0.20	**7.65**	**7.74**	1.01	-0.78	**9.22**
	L.T.=5	0.24	**7.33**	**7.51**	0.50	-0.25	**8.15**

Table 10.24. *t*-test (scaled MEs, difference between population means), best α value

The scaled ME results confirm the results obtained on the original MEs (see table 10.18) but the superiority of the Approximation method (in both a re-order level and a re-order interval context) is now better marked. No significant differences are indicated in the MA(13) – EWMA and MA(13) – Croston pair-wise comparisons. EWMA performs better than Croston's method in all simulated cases but the performance differences are statistically significant only in the re-order interval context.

When all points in time are considered, the best smoothing constant value analysis (on MEs and scaled MEs) results in an ordering scheme that is the same as the one presented earlier in this sub-sub-section. When issue points only are taken into account the Approximation method still performs best, followed by Croston's method and EWMA. The latter estimators perform very similarly and their performance is slightly better than that of MA(13).

10.4.1.3 RGRMSE results

In this section accuracy results are generated by considering the RGRMSE measure. Results with a value less than 1 indicate that the second method performs better than the first (i.e. the GRMSE associated with the second method across series is lower than the GRMSE that corresponds to the first one). The error results per series (per method) are generated based on the GRMSE associated with the application of each one of the estimators on the series under concern. To test the statistical significance of the results the natural logarithm of the GRMSE per file is calculated for all methods and the *t*-test (difference between population means) is used on the new series of logGRMSEs. For more details on the validity of the assumptions behind such an approach as well as the results generation process please refer to sub-sub-section 8.9.3.1. In brackets we present the test statistic value.

α = 0.05		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	0.99 (0.45)	0.99 (0.34)	0.98 (0.79)	0.94 (**2.77**)	1.05 (-*2.30*)	0.93 (**3.11**)
	L.T.=3	0.99 (0.43)	0.99 (0.46)	0.98 (0.90)	0.97 (1.24)	1.02 (-0.79)	0.96 (*1.71*)
	L.T.=5	0.99 (0.44)	0.99 (0.36)	0.98 (0.80)	0.98 (0.83)	1.01 (-0.38)	0.97 (1.19)
Issue Points Only	L.T.=1	1.00 (0.06)	0.98 (0.65)	0.98 (0.71)	0.94 (**2.65**)	1.06 (-**2.59**)	0.93 (**3.31**)
	L.T.=3	1.00 (-0.12)	0.98 (0.91)	0.98 (0.78)	0.98 (1.02)	1.03 (-1.14)	0.95 (*1.94*)
	L.T.=5	1.00 (0.11)	0.98 (0.75)	0.98 (0.86)	0.98 (0.70)	1.02 (-0.59)	0.96 (1.46)

Table 10.25. *t*-test (GRMSE, difference between population means), α = 0.05

α = 0.1		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	0.99 (0.37)	0.98 (0.78)	0.97 (1.16)	0.93 (**3.03**)	1.06 (-**2.65**)	0.92 (**3.82**)
	L.T.=3	0.99 (0.57)	0.97 (1.08)	0.96 (*1.65*)	0.97 (1.22)	1.02 (-0.64)	0.95 (*2.31*)
	L.T.=5	1.00 (0.16)	0.97 (1.10)	0.97 (1.25)	0.98 (0.64)	1.01 (-0.46)	0.96 (*1.75*)
Issue Points Only	L.T.=1	1.01 (-0.25)	0.97 (1.31)	0.98 (1.06)	0.94 (**2.74**)	1.07 (-**3.00**)	0.91 (**4.08**)
	L.T.=3	1.00 (-0.17)	0.96 (*1.67*)	0.96 (1.49)	0.98 (0.87)	1.03 (-1.05)	0.94 (**2.56**)
	L.T.=5	1.00 (-0.14)	0.96 (1.61)	0.96 (1.47)	0.99 (0.54)	1.02 (-0.68)	0.94 (*2.18*)

Table 10.26. *t*-test (GRMSE, difference between population means), α = 0.1

α = 0.15		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
All Points in Time	L.T.=1	0.99 (0.31)	0.97 (1.40)	0.96 (*1.71*)	0.94 (**2.84**)	1.06 (**-2.51**)	0.91 (**4.26**)
	L.T.=3	0.99 (0.47)	0.97 (1.40)	0.96 (*1.86*)	0.98 (0.91)	1.01 (-0.43)	0.95 (**2.32**)
	L.T.=5	1.01 (-0.36)	0.95 (*1.91*)	0.96 (1.53)	1.00 (-0.05)	1.01 (-0.31)	0.95 (*1.87*)
Issue Points Only	L.T.=1	1.01 (-0.46)	0.95 (*2.06*)	0.96 (1.60)	0.95 (**2.39**)	1.07 (**-2.87**)	0.90 (**4.51**)
	L.T.=3	1.01 (-0.48)	0.95 (*2.20*)	0.96 (*1.72*)	0.99 (0.46)	1.02 (-0.96)	0.94 (**2.70**)
	L.T.=5	1.02 (-0.88)	0.93 (**2.57**)	0.96 (*1.68*)	1.00 (-0.14)	1.02 (-0.76)	0.94 (**2.47**)

Table 10.27. *t*-test (GRMSE, difference between population means), α = 0.15

α = 0.2		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
All Points in Time	L.T.=1	0.99 (0.59)	0.96 (*1.78*)	0.95 (**2.36**)	0.94 (**2.77**)	1.05 (*-2.16*)	0.90 (**4.59**)
	L.T.=3	0.99 (0.25)	0.95 (*2.21*)	0.94 (**2.44**)	1.00 (0.13)	1.00 (0.12)	0.94 (**2.36**)
	L.T.=5	1.01 (-0.56)	0.93 (**2.80**)	0.94 (*2.21*)	1.03 (-1.15)	0.98 (0.58)	0.96 (*1.66*)
Issue Points Only	L.T.=1	1.01 (-0.41)	0.94 (**2.60**)	0.95 (*2.19*)	0.95 (*2.16*)	1.06 (**-2.60**)	0.90 (**4.85**)
	L.T.=3	1.02 (-0.94)	0.92 (**3.19**)	0.95 (*2.25*)	1.01 (-0.44)	1.01 (-0.51)	0.93 (**2.79**)
	L.T.=5	1.04 (-1.30)	0.91 (**3.61**)	0.94 (*2.29*)	1.04 (-1.36)	1.00 (0.04)	0.94 (*2.28*)

Table 10.28. *t*-test (GRMSE, difference between population means), α = 0.2

The results indicate quite conclusively that the Approximation method performs significantly better than the MA(13). The Approximation method performs better than

EWMA and Croston's method in all simulated cases but the accuracy differences become significant only for higher smoothing constant values (0.15, 0.2). This result is not entirely consistent with the results presented in the previous sub-sections. Considering the bias results, we might have expected to see some greater advantage of using the Approximation method for α = 0.1 but we were not expecting to see a relative advantage for α = 0.2. No significant differences are indicated between Croston's method and EWMA, although the GRMSE (across series) associated with Croston's method is consistently lower than that of the EWMA for issue points only, while the opposite is the case for all points in time. Croston's method and EWMA perform significantly better than MA(13) and this result does not agree with the findings presented in the previous sub-sub-sections. When the effect of outliers is taken out, the MA(13) method does not outperform any of the other estimators considered in this chapter. The methods can now be ordered as follows in terms of their forecasting accuracy:

	All points in time	**Issue points only**
1.	Approximation	Approximation
2.	EWMA	Croston's method
3.	Croston's method	EWMA
4.	MA(13)	MA(13)

(The overall measure of accuracy is calculated for each estimator as an equally weighted average of the GRMSEs across all the control parameter combinations.)

The above ordering is accurate for all simulated scenarios, with the exception of α = 0.15 and 0.2 (L.T. = 3, 5) where no significant differences are indicated in the MA(13) - Croston and MA(13) - EWMA pair-wise comparisons. This issue is further discussed in the following sub-section.

The ordering of the methods discussed above is also valid when the best α value performance (lowest GRMSE across series) is considered for all the smoothing estimators. The best α values and the *t*-test results are presented in tables 10.29 and 10.30 respectively.

		Best smoothing constant value		
		EWMA	Croston	Approximation
All points in time	L.T. = 1	0.1	0.1	0.2
	L.T. = 3	0.05	0.05	0.2
	L.T. = 5	0.05	0.1	0.15
Issue points only	L.T. = 1	0.1	0.1	0.2
	L.T. = 3	0.05	0.05	0.2
	L.T. = 5	0.05	0.15	0.15

Table 10.29. Best α value performance – lowest GRMSE across series

best α		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	0.99 (0.37)	0.97 (1.51)	0.96 (*1.88*)	0.93 (**3.03**)	1.06 (**-2.65**)	0.90 (**4.59**)
	L.T.=3	0.99 (0.43)	0.97 (1.09)	0.96 (1.52)	0.97 (1.24)	1.02 (-0.79)	0.94 (**2.56**)
	L.T.=5	0.99 (0.35)	0.97 (1.02)	0.96 (1.38)	0.98 (0.83)	1.01 (-0.46)	0.95 (*1.87*)
Issue Points Only	L.T.=1	1.01 (-0.25)	0.96 (*1.72*)	0.97 (1.47)	0.94 (**2.74**)	1.07 (-3.00)	0.90 (**4.51**)
	L.T.=3	1.00 (-0.12)	0.96 (*1.65*)	0.96 (1.52)	0.98 (1.02)	1.03 (-1.14)	0.94 (**2.70**)
	L.T.=5	1.00 (-0.05)	0.95 (*1.73*)	0.96 (*1.68*)	0.98 (0.70)	1.02 (-0.59)	0.94 (**2.47**)

Table 10.30. *t*-test (GRMSE, difference between population means), best α value

10.4.1.4. Conclusions (parametric tests)

The results obtained on the ME, scaled ME and RGRMSE measure indicate that, in both a re-order interval and a re-order level context, the Approximation method is the most accurate estimator. When higher smoothing constant values are considered, the gain in accuracy is particularly marked for longer lead times.

When all points in time are considered, EWMA is shown to perform more accurately than Croston's method. Even though the differences are not statistically significant (and this is true only when the RGRMSE is considered), EWMA performs consistently better than Croston's estimator. This finding does not agree with the empirical results presented in Willemain et al (1994). In that paper EWMA and Croston's method were compared by means of simulation on 32 real demand data series, in a re-order interval context. The accuracy measures used, for comparison purposes, were the MAPE (Mean Absolute Percentage Error), MdAPE (Median Absolute Percentage Error), RMSE (Root Mean Squared Error) and MAD (Mean Absolute Deviation). Even though the accuracy differences were not as marked as when the methods were compared on theoretically generated data, Croston's method was shown to outperform EWMA.

The above discussed inconsistency clearly indicates the importance of the selection process of accuracy measures and provides some justification for devoting an entire chapter of this thesis to identifying the most appropriate error measures to be used for comparison purposes in an intermittent demand context.

An important finding is also considered to be the poor performance of the MA(13) estimator. Sani and Kingsman (1997) have given some empirical support for the simple moving average method. The MA(13) seems to be fairly robust to the presence of outliers. When the effect of the extreme observations though is not considered, this method is the least accurate one.

For issue points only, Croston's method performs better than EWMA and this is in accordance with the theory (at least as far as the bias measure is concerned). As discussed above, the MA(13) method is shown to be the least accurate estimator when the RGRMSE measure is considered. This is not true when bias (or scaled bias) results are generated. In that case MA(13) performs better than EWMA and this is particularly true for higher smoothing constant values. This difference can also be attributed to a significant improvement of the performance of EWMA when the effects of outliers are reduced and this is what intuitively one should expect, especially for issue points only.

In this sub-section results have also been generated on all three accuracy measures considered (ME, scaled ME and GRMSE) with respect to the best α value performance of the smoothing estimators. The results demonstrate the superiority of the Approximation method in both a re-order level and a re-order interval context. When all points in time are taken into account EWMA performs better than Croston's method. The opposite is the case when results are generated on issue points only. The MA(13) is always the least desirable estimator but this is what we intuitively expected since the average length has not been "optimised" on our empirical data sample.

10.4.2 Non-parametric tests

Percentage Better (PB) and Percentage Best (PBt) results have also been generated considering the absolute ME (i.e. ignoring the sign) and RGRMSE per series associated with each of the estimation procedures discussed in this chapter. The PB measure refers to pair-wise comparisons (proportion of times that one method performs better than the other, i.e. proportion of series that the ME or GRMSE of one method is lower than that of the other). The statistical significance of the results is checked by using equation (8.17). The PBt measure indicates the proportion of times that one method performs better than all other estimation procedures. Statistical significance is also checked on a pair-wise basis (see sub-sub-section 8.9.3.2, equation (8.18)).

10.4.2.1 Percentage Better results (ME, RGRMSE)

The Percentage Better results are presented, for different smoothing constant values, in *Appendix 10.B*. In tables *10.B.1 – 10.B.4* we present the *Z*-test statistic value (difference between population proportions) when results are generated on the ME. In brackets we present the number of files that the absolute ME of the first method is greater than that of the second method. Positive values for the test statistic indicate that the second method performs better than the first. The critical values are as discussed in the previous section. Some ties occur in the Croston – EWMA pair-wise comparison (i.e. $ME_{EWMA} = ME_{Croston}$). No ties occur in any of the other pair-wise comparisons. For the Croston – EWMA comparison, in brackets we present the number of times that EWMA performs better and the total number of series considered (i.e. 3,000 minus (-) no. of ties). The number of times (series) that ties occur changes with the lead time but not with the smoothing constant value.

In tables *10.B.5 – 10.B.8* results are presented for the case that the RGRMSE per series is considered for generating the PB results. The results can be generated either by considering the RGRMSE per series (values less than or greater than 1) or by directly comparing the GRMSE per series obtained for the two estimators under concern.

The PB results on MEs indicate that:

- The Approximation method is significantly better than Croston's method. This is not true for α = 0.2 and one step ahead forecasts in which case the latter estimator performs (not significantly) better than the former.

- The Approximation method performs better than EWMA only when results are generated on issue points

- The Approximation method performs significantly better than MA(13) for α = 0.15, 0.2 but the opposite is the case for α = 0.05 and α = 0.1.

- The MA(13) method performs better than Croston's method for α = 0.05 and α = 0.1. The opposite is the case for the remaining smoothing constant values

- The MA method performs better than EWMA only for α = 0.05

- EWMA performs better than Croston's method in all simulated cases.

The MA(13) performs very well when PB results are generated on the MEs. This is what one would expect from the theory for all points in time but not for issue points only (see also previous section).

The Percentage Better results on GRMSEs indicate that the Approximation method performs significantly better than all other three estimation procedures. In fact there are only two cases (in the Approximation – MA(13) comparison) where the Approximation method does not perform significantly better than one other method. EWMA is significantly better than Croston’s method in the majority of the simulated cases, for all points in time. When issue points only are considered Croston’s method is always better than EWMA. The MA(13) method does not perform particularly well. Both Croston’s method and EWMA outperform MA(13) in all cases for α = 0.05, 0.1 as well as for α = 0.15, 0.2 when one step ahead forecasts are considered.

We have attempted to synthesise the results discussed above to generate an ordering of the performance of all estimators. For each particular combination of the control parameter values, the estimators were ranked in terms of the number of pair-wise comparisons that outperform any of the other estimators considered (the first method being the best). The results are as follows:

<u>RGRMSE</u>

	All points in time	**Issue points only**
1.	Approximation	Approximation
	α = 0.05, 0.1, (0.15, 0.2, 1 step ahead forecasts)	
2.	EWMA	Croston
3.	Croston	EWMA
4.	MA(13)	MA(13)
	α = 0.15, 0.2 (L.T. = 3, 5)	
2.	MA(13)	MA(13)
3.	Croston	Croston
4.	EWMA	EWMA

(The above ordering is not valid for α = 0.1, L.T. = 5 (all points in time) where Croston’s method performs better than EWMA and α = 0.1, L.T. = 5 (issue points only) where MA(13) and EWMA perform identically.)

There is a remarkable consistency between the results generated based on the RGRMSE (across series) applied as a descriptive measure and the Percentage Better results generated on the GRMSE per series. In sub-sub-section 10.4.1.3 summary results were presented regarding the comparative forecasting accuracy performance of the alternative estimators, based on the RGRMSE, and few simulated scenarios were identified where the accuracy differences were not very well marked. The Percentage Better results confirm the validity of those conclusions and demonstrate considerable accuracy differences in the scenarios discussed above.

Mean signed Error

	All points in time		**Issue points only**	
	α = 0.05	α = 0.1	α = 0.05	α = 0.1
1.	MA(13)	EWMA	MA(13)	Approx.
2.	EWMA	MA(13)	Approx.	EWMA
3.	Approx.	Approx.	EWMA	MA(13)
4.	Croston	Croston	Croston	Croston
	α = 0.15, 0.2			
1.	EWMA		Approx.	
2.	Approx.		EWMA	
3.	Croston		Croston	
4.	MA(13)		MA(13)	

(The above ordering is not valid for α = 0.1, L.T. = 1 (issue points only) where the ordering is as follows: 1. EWMA 2. MA(13) 3. Approximation 4. Croston.)

The same is not true when we compare the results given by the ME applied as a descriptive measure across series and the Percentage Better results generated based on the ME given by alternative estimators in a single series. In fact the two sets of results are considerably different. It is reasonable to suppose that the results generated in a single series, considering the bias accuracy measure, may be particularly sensitive to extreme observations. In that respect, isolating the ME results, for each of the series considered, in order to compare alternative estimators on a pair-wise comparison level may not be very good practice. The above discussed inconsistency possibly also indicates the importance of the high bias SKUs in comparing the alternative estimators considered in this chapter. Similarly, it would be reasonable to attribute the consistency between the two sets of RGRMSE results to the “relative” nature of the accuracy measure under concern. Finally, the scale dependence problem may also help to explain why RGRMSE results are not completely consistent with the bias results. We have more to say on this issue in the following sub-sub-section.

10.4.2.2 Percentage Best results (ME, GRMSE)

The Percentage Best results (*Z*-test statistic value, difference between population proportions) on GRMSE are presented in the following tables, for different smoothing constant values.

α = 0.05		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**7.09**	**4.83**	**11.84**	**-7.68**	**14.62**	**-2.86**
	L.T.=3	**4.37**	**7.25**	**11.55**	**-11.02**	**15.26**	**-3.82**
	L.T.=5	**3.98**	**8.35**	**12.24**	**-12.72**	**16.54**	**-4.44**
Issue Points Only	L.T.=1	**3.77**	**13.71**	**17.31**	**-13.69**	**17.28**	0.03
	L.T.=3	**2.47**	**14.20**	**16.55**	**-15.89**	**18.22**	*-1.74*
	L.T.=5	**3.03**	**13.61**	**16.50**	**-16.64**	**19.48**	**-3.14**

Table 10.31. Percentage Best results (GRMSE), α = 0.05

α = 0.1		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**5.35**	**9.71**	**14.93**	-1.15	**6.49**	**8.57**
	L.T.=3	**6.11**	**10.27**	**16.20**	**-4.61**	**10.65**	**5.71**
	L.T.=5	**2.05**	**9.08**	**11.09**	**-8.35**	**10.37**	0.73
Issue Points Only	L.T.=1	**3.95**	**17.80**	**21.49**	**-6.47**	**10.37**	**11.52**
	L.T.=3	**3.50**	**17.97**	**21.22**	**-11.13**	**14.51**	**7.03**
	L.T.=5	*1.79*	**17.16**	**18.84**	**-11.96**	**13.68**	**5.37**

Table 10.32. Percentage Best results (GRMSE), α = 0.1

α = 0.15		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**5.36**	**10.35**	**15.57**	1.54	**3.83**	**11.86**
	L.T.=3	**6.67**	**7.81**	**14.35**	**-4.17**	**10.79**	**3.66**
	L.T.=5	**4.00**	**8.97**	**12.88**	**-7.50**	**11.43**	1.48
Issue Points Only	L.T.=1	**3.47**	**19.61**	**22.82**	**-6.54**	**9.96**	**13.31**
	L.T.=3	**2.73**	**17.84**	**20.38**	**-12.27**	**14.89**	**5.75**
	L.T.=5	**3.40**	**17.76**	**20.92**	**-12.62**	**15.87**	**5.31**

Table 10.33. Percentage Best results (GRMSE), α = 0.15

α = 0.2		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**7.58**	**10.35**	**17.71**	**3.81**	**3.79**	**14.08**
	L.T.=3	**5.05**	**9.97**	**14.88**	**-7.96**	**12.91**	*2.03*
	L.T.=5	*2.21*	**10.29**	**12.44**	**-11.84**	**13.98**	-1.58
Issue Points Only	L.T.=1	**5.35**	**20.16**	**25.06**	**-5.24**	**10.52**	**15.14**
	L.T.=3	**2.49**	**18.32**	**20.62**	**-14.32**	**16.68**	**4.15**
	L.T.=5	**2.58**	**18.63**	**21.00**	**-16.58**	**18.99**	*2.13*

Table 10.34. Percentage Best results (GRMSE), α = 0.2

The number of times that one method performs better that all other methods are indicated in *Appendix 10.C* (tables *10.C.1 – 10.C.4*). Ties occur when Croston's method and EWMA perform identically.

The Percentage Best results on the GRMSE indicate that the Approximation method performs significantly better than all other three estimators. (The Approximation method is outperformed only by MA(13), for α = 0.05.) The MA(13) performs better than Croston's method and EWMA and finally EWMA performs better than Croston's method. Those findings are valid for the majority (if not all) of the simulated scenarios.

Very similar conclusions are obtained when the Percentage Best results on bias are analysed. In fact, the only significant difference is that the Approximation method is outperformed by the MA(13) for both α = 0.05 and α = 0.1 (all points in time). The Percentage Best results on ME are presented in tables 10.35 – 10.38. The number of times that one method performs better that all other methods are indicated in *Appendix 10.C* (tables *10.C.5 – 10.C.8*). Ties occur when Croston's method and EWMA perform identically.

α = 0.05		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**4.02**	**5.64**	**9.58**	**-37.01**	**40.10**	**-32.27**
	L.T.=3	**4.31**	**5.63**	**9.87**	**-32.72**	**36.22**	**-27.88**
	L.T.=5	**3.08**	**6.71**	**9.74**	**-28.26**	**30.91**	**-22.13**
Issue Points Only	L.T.=1	**8.69**	**10.77**	**18.85**	**-35.31**	**41.77**	**-25.77**
	L.T.=3	**9.64**	**11.78**	**20.71**	**-30.50**	**38.11**	**-19.61**
	L.T.=5	**9.79**	**12.18**	**21.29**	**-25.82**	**33.99**	**-14.20**

Table 10.35. Percentage Best results (ME), α = 0.05

α = 0.1		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**10.79**	0.60	**11.38**	**-10.42**	**20.81**	**-9.83**
	L.T.=3	**9.95**	**2.93**	**12.81**	**-9.57**	**19.20**	**-6.67**
	L.T.=5	**8.45**	**3.74**	**12.11**	**-9.05**	**17.27**	**-5.35**
Issue Points Only	L.T.=1	**14.60**	**11.70**	**25.44**	**-9.10**	**23.06**	**2.64**
	L.T.=3	**14.66**	**12.52**	**26.26**	**-7.61**	**21.74**	**4.98**
	L.T.=5	**14.60**	**14.14**	**27.64**	**-8.16**	**22.17**	**6.08**

Table 10.36. Percentage Best results (ME), α = 0.1

α = 0.15		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**14.25**	*-2.32*	**11.99**	**8.59**	**5.79**	**6.29**
	L.T.=3	**11.48**	*2.17*	**13.59**	**4.47**	**7.08**	**6.63**
	L.T.=5	**9.43**	**4.26**	**13.59**	-0.21	**9.63**	**4.05**
Issue Points Only	L.T.=1	**18.94**	**12.04**	**29.92**	**7.13**	**12.20**	**18.97**
	L.T.=3	**17.27**	**15.32**	**31.27**	**4.22**	**13.29**	**19.38**
	L.T.=5	**15.26**	**16.69**	**30.61**	-1.52	**16.68**	**15.23**

Table 10.37. Percentage Best results (ME), α = 0.15

α = 0.2		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	12.15	*-1.77*	**10.42**	**14.75**	**-2.69**	**13.04**
	L.T.=3	10.03	**3.18**	**13.15**	**8.64**	1.42	**11.77**
	L.T.=5	7.60	**6.83**	**14.30**	0.51	**7.10**	**7.33**
Issue Points Only	L.T.=1	19.15	**11.51**	**29.71**	**12.41**	**7.17**	**23.48**
	L.T.=3	16.77	**16.26**	**31.74**	**8.68**	**8.42**	**24.48**
	L.T.=5	14.65	**18.01**	**31.29**	0.40	**14.27**	**18.39**

Table 10.38. Percentage Best results (ME), α = 0.2

Considering the Percentage Best results, on the RGRMSE, the estimators discussed in this chapter can be ordered based on the following scheme. For each particular combination of the control parameter values, the estimators were ranked in terms of the number of pair-wise comparisons that outperform any of the other estimators considered. The first method is the best.

	All points in time		**Issue points only**	
	α = 0.05	α = 0.1, 0.15, 0.2	α = 0.05	α = 0.1, 0.15, 0.2
1.	MA(13)	Approx.	MA(13)	Approx.
2.	Approx.	MA(13)	Approx.	MA(13)
3.	EWMA	EWMA	EWMA	EWMA
4.	Croston	Croston	Croston	Croston

The above ordering is also valid for the Percentage Best results on ME but with the following differences:

- The MA(13) performs better than the Approximation method for α = 0.1, all points in time
- The EWMA estimator performs better than the Approximation method for α = 0.15, 0.2, L.T. = 1, all points time
- The EWMA estimator performs better than MA(13) for α = 0.2, issue points only.

The Percentage Best results are probably more meaningful than the Percentage Better results since, ultimately, only the best method will be used following the categorisation approach developed in the thesis. What matters from a practitioner's perspective is the accuracy in determining the first place (best estimator) rather than the accuracy associated with the determination of the lower places. In addition, the Percentage Best measure has resulted in a straightforward ordering of the estimators considered in this chapter, with respect to their forecasting accuracy, while the same was not the case when the PB measure was used. Finally, it is important to note that the PBt results appear to be insensitive to the descriptive accuracy measure chosen for generating results in each of the series included in the empirical sample. In the previous sub-sub section we showed that the PB results on ME and RGRMSE were not consistent with each other.

10.4.2.3 Conclusions (non-parametric tests)

In this section a non-parametric approach was taken, to assess the forecasting accuracy of the alternative estimators discussed in this chapter. Percentage Better and Percentage Best results have been generated, across series, to evaluate the alternative methods' performance with respect to the number of times (series) that each estimator performs better than one other or all other methods.

To generate the results the ME and RGRMSE (per series) have been used, to indicate better or best performance in each single series. The PB results generated based on the RGRMSE per series are consistent with the RGRMSE results obtained across series. The same was not the case when the bias measure was used. Isolating the ME results for each of the series considered in order to compare alternative estimators on a pair-wise comparison level may not be very good practice since the results are particularly sensitive to the presence of extreme observations (see also sub-sub-section 10.4.2.1).

It has been argued, in the literature, that the Percentage Better is an easily interpreted and a very intuitive non-parametric approach to generate comparative forecasting accuracy results (see chapter 8). The Percentage Better accuracy measure has been recommended for use not only on a pair-wise comparison level but also for generating overall (across all estimators) accuracy results (Makridakis and Hibon, 1995): "*The percentage better measure requires* (as a minimum) *the use of two methods…..If more than two methods are to be compared the evaluation can be done for each pair of them* (p. 7)."

Nevertheless, the empirical evidence presented in this section demonstrates that the use of the Percentage Better measure for more than two estimators leads to rather complex ordering schemes. In addition, the PB results are heavily dependent upon which accuracy measure is used to obtain results in a single series. The Percentage Best measure is more meaningful, from a practitioner's perspective, as evidence has been provided for its insensitivity to the descriptive measure chosen. As such it is recommended for use in large-scale empirical exercises.

The results generated in this sub-section indicate that statistically significant accuracy differences obtained based on a descriptive accuracy measure across series are not necessarily reflected on the number of times (series) that one estimator performs better than one or all other estimators. Results should always be generated based on parametric and non-parametric tests and this should increase confidence in the conclusions.

The Approximation method is shown in this sub-section to be the most accurate estimator. The MA(13) does as well especially when Percentage Best results are generated. EWMA performs significantly better than Croston's method and this is not what one should expect from the theory especially for all points in time (see chapter 6). The MA(13) is fairly robust to the presence of outliers. When the effect of the extreme observation is "taken out" the estimator under concern does not perform so well. Overall, the most striking finding is considered to be the considerable improvement (as compared to the results given in the previous sub-section) in the performance of the MA(13).

10.4.3 Conclusions

In this section EWMA, Croston's method, MA(13) and the Approximation method were compared with respect to their forecasting accuracy. Results have been generated by considering parametric and non-parametric tests. In the former case the ME, scaled ME and RGRMSE descriptive measures were considered, while in the latter case Percentage Better and Percentage Best results have been generated based on the ME and RGRMSE per series. Parametric results have also been generated considering the best α value performance of the smoothing estimators, with respect to all the descriptive measures used for comparison purposes. The best smoothing constant value was "optimised" across all 3,000 series available for simulation and that corresponds to real world practices. The findings of this section can be summarised as follows:

- The Approximation method can be claimed to be the most accurate estimator

- EWMA performs better than Croston's method in a re-order interval context and this is true when both parametric and non-parametric procedures are used to generate results

- EWMA performs less accurately than Croston's estimator when issue points only are considered. This superiority of Croston's method is not reflected on the number of series that Croston's method performs better.

- The MA(13) method is robust to the presence of outliers. It compares very favourably with the smoothing methods for low α values. The opposite is true as the α value increases and this is, intuitively, what one should expect. The MA(13) does very well; in fact it can be claimed to be as accurate as the Approximation method, with respect to the number of times that it gives the lowest error.

- Different accuracy measures can lead to different conclusions in an intermittent demand context

- The ME results generated in a single series are particularly sensitive to the presence of extreme observations

- The Percentage Best (PBt) measure should be preferred to the Percentage Better (PB) measure because of its relevance to the choice of the best forecasting method and the fact that it gives more consistent results.

- The RGRMSE is a very well behaved accuracy measure.

10.5 The categorisation of "non-normal" demand patterns

A chi-square test was decided to be the most appropriate way of assessing the validity of the theoretical rules proposed in chapter 6. The average inter-demand interval and the squared coefficient of variation of the demand sizes were generated for all 3,000 files and the theoretical rules were used to indicate which method is theoretically expected to perform better (in the case of pair-wise comparisons) or best (when the overall categorisation rules are considered) in each one of those files. The rules were developed based on a theoretical analysis of the MSE, associated with each one of the estimators, and therefore MSE results are first generated for comparison purposes (other measures and tests will also be considered later in this section to assess the empirical validity of the theoretical rules). It is important to note that the MA(13) method is not considered at this stage of the empirical analysis, since no theoretical results have been derived regarding this method's performance[1]. The null hypothesis developed is that the performance of the methods is independent of what is expected from the theory. For the pair-wise comparisons (2 rows x 2 columns contingency table, see figure 10.4), the critical values for rejecting the null hypothesis (1 degree of freedom) are 3.84 and 6.63 for the 5% and 1% significance level respectively.

[1] The MA(13) will be considered later in this section when the empirical validity of the categorisation rules is assessed by means of parametric rather than non-parametric tests.

		Theory (chapter 6)		
		x method better than y method	y method better than x method	
Simulation	x method better than y method			row 1 sum
	y method better than x method			row 2 sum
		column 1 sum	column 2 sum	3,000 files

Figure 10.4. Chi-square test (pair-wise comparisons)

The categorisation rules that can be tested on the empirical data refer to the Croston-Approximation (see sub-section 6.8.1), EWMA-Approximation (see sub-sections 6.8.2 and 6.84) and Croston-EWMA (see sub-sections 6.8.3 and 6.8.5) pair-wise comparisons. The theoretical cut-off values are indicated in table 10.39.

Pair-wise comparison	p *cut-off value*	CV^2 *cut-off value*
Approximation-*Croston*	1.32	0.49
ISSUE POINTS		
Approximation-*EWMA*	1.33	0.49
Croston-*EWMA*	1.34	0.28
ALL POINTS IN TIME		
Approximation-*EWMA*	1.20	0.49
Croston-*EWMA*	Croston always performs better than EWMA	

Table 10.39. Cut-off values ($0.05 \le \alpha \le 0.2$**)**

The chi-square test cannot be utilised for the Croston-EWMA comparison when we refer to a re-order interval context. In that case Croston's method is always (on every demand data series) expected to perform better than EWMA, independently of the demand data series characteristics. As such, the *Z*-test statistic for the population proportion (see equation (8.17)) will be used to test whether or not the number of files on which Croston's method performs better (i.e. $MSE_{Croston} < MSE_{EWMA}$) is significantly greater from the number of files on which EWMA performs better (i.e. $MSE_{EWMA} < MSE_{Croston}$). The critical values for rejecting the null hypothesis are 1.64 and 2.32 for the 5% and 1% significance level respectively.

At this point it is important to note that in certain cases the MSE given by Croston's method equals the MSE given by the EWMA estimator. Obviously this equality does not raise any difficulties as to how the corresponding series will be treated when we consider the Approximation – EWMA and the Approximation – Croston pair-wise comparisons. The same is true when the EWMA – Croston comparison is considered in a re-order interval context since Croston's method is always, theoretically, expected to perform better than EWMA. For the EWMA – Croston pair-wise comparison (when issue points only are considered) the ties need to be taken into account and the contingency table takes the form indicated in figure 10.5. The critical values for rejecting the null hypothesis (2 degrees of freedom) are 5.99 and 9.21 for the 5% and 1% significance level respectively.

		Theory (chapter 6)		
		x method better than y method	y method better than x method	
Simulation	x method better than y method			row 1 sum
	y method better than x method			row 2 sum
	Ties			row 3 sum
		column 1 sum	column 2 sum	3,000 files

Figure 10.5. Chi-square test (EWMA – Croston, issue points only)

At this stage it is important to note that the number of ties is large enough to allow an expected frequency of at least 5 to appear in each "cell" (corresponding shaded demand categories). As such the chi-square test can be used for testing goodness-of-fit.

In testing the overall categorisation rules, ties occur when:

$$MSE_{Croston} = MSE_{EWMA} < MSE_{Approx.}$$

When the lead time equals one and for all points time ties never occur. For the rest of the scenarios covered in the simulation experiment there is always a certain number of ties. For testing the categorisation rules regarding all methods' performance in a re-order interval system (figure 10.2) a 3x2 (2 degrees of freedom) or 4x2 (3 degrees of freedom) table will be used depending on whether or not ties occur. The contingency table has the form indicated in figure 10.6. When we refer to a re-order level context (see figure 10.3) ties always occur and the contingency table (4x3, 6 degrees of freedom) is as indicated in figure 10.7.

		Theory (chapter 6)		
		Approx. better than Croston and EWMA	Croston better than Approx. and EWMA	
Simulation	Approx. best			row 1 sum
	Croston best			row 2 sum
	EWMA best			row 3 sum
	Ties			row 4 sum
		column 1 sum	column 2 sum	3,000 files

Figure 10.6. Chi-square test (re-order interval systems)

		Theory (chapter 6)			
		Approx. better than Croston and EWMA	Croston better than Approx. and EWMA	EWMA better than Approx. and Croston	
Simulation	Approx. best				row 1 sum
	Croston best				row 2 sum
	EWMA best				row 3 sum
	Ties				row 4 sum
		column 1 sum	column 2 sum	column 3 sum	3,000 files

Figure 10.7. Chi-square test (re-order level systems)

The critical values for rejecting the null hypothesis are as follows:

degrees of freedom	5% significance level	1% significance level
1	3.84	6.63
2	5.99	9.21
3	7.81	11.30
6	12.6	16.80

10.5.1 MSE rules

The χ^2 test statistic values are indicated, for different smoothing constant values, in tables 10.40 – 10.43. In brackets we present the number of files that the methods perform as expected. The shaded area refers to the Croston – EWMA pair-wise comparison, in a re-order interval context, and the values presented are the *Z*-test statistic values for the population proportion. Statistically significant results at the 1% level are emboldened while significance at the 5% level is presented in italics.

α = 0.05		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**Overall categorisation rules**
All Points in Time	L.T.=1	0.33 (1509)	0.84 (1633)	**16.45** (1634)	*9.19* (979)
	L.T.=3	**5.29** (1645)	1.10 (1691)	**15.77** (1631)	**12.80** (1111)
	L.T.=5	**4.35** (1619)	*4.41* (1726)	**13.18** (1624)	**20.47** (1160)
Issue Points Only	L.T.=1	1.53 (1628)	**33.60** (1713)	**19.32** (1644)	7.49 (1185)
	L.T.=3	3.24 (1660)	**33.53** (1717)	**24.85** (1669)	**19.46** (1239)
	L.T.=5	**16.13** (1568)	**23.44** (1688)	**19.08** (1655)	**25.58** (1221)

Table 10.40. chi-square test results, α = 0.05

α = 0.1		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**Overall categorisation rules**
All Points in Time	L.T.=1	**3.83** (1605)	**6.96** (1709)	**20.75** (1653)	1.60 (1081)
	L.T.=3	**10.11** (1777)	**17.98** (1827)	**17.22** (1640)	**11.94** (1238)
	L.T.=5	**9.02** (1747)	**19.02** (1826)	**13.02** (1625)	**17.43** (1271)
Issue Points Only	L.T.=1	2.05 (1745)	**62.73** (1777)	**28.78** (1675)	**20.27** (1305)
	L.T.=3	2.67 (1775)	**46.79** (1754)	**27.10** (1679)	**26.18** (1315)
	L.T.=5	**15.76** (1676)	**37.52** (1730)	**20.45** (1663)	**25.79** (1291)

Table 10.41. chi-square test results, α = 0.1

α = 0.15		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**Overall categorisation rules**
All Points in Time	L.T.=1	**6.83** (1687)	**17.34** (1753)	**30.56** (1688)	*6.20* (1160)
	L.T.=3	**13.04** (1857)	**53.58** (1925)	**28.27** (1678)	**13.05** (1339)
	L.T.=5	**13.04** (1857)	**44.66** (1913)	**16.79** (1644)	**22.20** (1364)
Issue Points Only	L.T.=1	2.03 (1844)	**84.28** (1813)	**39.15** (1703)	**38.55** (1381)
	L.T.=3	2.56 (1871)	**88.41** (1825)	**40.86** (1718)	**40.03** (1403)
	L.T.=5	**15.80** (1787)	**51.18** (1761)	**27.50** (1687)	**35.62** (1371)

Table 10.42. chi-square test results, α = 0.15

α = 0.2		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**Overall categorisation rules**
All Points in Time	L.T.=1	**8.47** (1732)	**28.99** (1802)	**53.58** (1744)	**16.22** (1230)
	L.T.=3	**14.83** (1906)	**84.10** (1991)	**31.96** (1694)	**20.86** (1405)
	L.T.=5	**15.12** (1914)	**60.77** (1949)	**18.70** (1654)	**21.37** (1418)
Issue Points Only	L.T.=1	2.95 (1931)	**114.68** (1854)	**50.44** (1734)	**53.29** (1406)
	L.T.=3	4.52 (1947)	**139.07** (1886)	**57.60** (1758)	**61.80** (1467)
	L.T.=5	**15.22** (1846)	**18.54** (1965)	**35.15** (1706)	**43.34** (1406)

Table 10.43. chi-square test results, α = 0.2

The results generated on MSEs indicate the practical validity of the rules proposed in chapter 6. The performance of the methods is clearly not independent of what the categorisation rules suggest. This is true at a pair-wise comparison level (with the only exception of the EWMA – Croston comparison for issue points only and when the lead time is 1 or 3 periods) and when the overall rules (regarding all methods' performance) are considered.

In particular, the results demonstrate the empirical validity of the Approximation – EWMA and Approximation – Croston pair-wise comparison rules in both a re-order interval and a re-order level context. The results regarding the EWMA - Croston comparison are also consistent with the theory when all points in time are considered. For issue points only the particular pair-wise categorisation rules have been empirically validated only for L.T = 5.

The MSE is similar to the statistical measure of the variance of the forecast errors but not identical since bias is also taken into account. The issue of the variability of the forecast errors is not explicitly addressed in this chapter. By researching the validity of the MSE theoretical rules developed in chapter 6 and by analysing the bias results generated in the previous section some conclusions can also be drawn regarding the variability of the forecast errors. The Percentage Better (PB) results on bias presented in sub-sub-section 10.4.2.1 can be used for comparison with the chi-square analysis results on the pair-wise categorisation rules. The Percentage Best (PBt) results on ME presented in sub-sub-section 10.4.2.2 could also be compared with the chi-square analysis results on the overall categorisation rules. Nevertheless, the PBt results involved also the MA(13) estimator which is not considered at this stage of the research (see also sub-section 10.5.2).

In order to check the validity of the Croston – EWMA categorisation rule in a re-order interval context, the *t*-test for the population proportion (PB measure) has been used instead of the chi-square test since Croston’s method is expected to perform better in all demand data series. The *t*-test results show significant differences in favour of Croston’s method. Nevertheless, the PB results on bias indicate that EWMA performs better than Croston’s method in all simulated cases. This inconsistency could be attributed to the increased variability of the forecast error associated with EWMA. When issue points only are considered, the pair-wise categorisation rule under concern has been validated only for L.T. = 5. For the rest of the lead time lengths considered in our experiment the rule has not been validated even though there is always a substantial number of series where the methods perform as expected. Again there is some evidence to support the assertion that the variability of the estimates produced by Croston’s method is lower than that of the EWMA estimator.

Regarding the EWMA – Approximation pair-wise comparison, the PB results on bias indicate that the former estimator performs better than the latter in all simulated cases, in a re-order interval context. The corresponding categorisation rule has been validated for all simulated scenarios and that could be attributed to the high sampling error of the mean of the EWMA estimator, when all points in time are taken into account. The results presented in this section for the Croston – Approximation pair-wise comparison are consistent with the PB results given in sub-sub-section 10.4.2.1 and as such no further comments can be made regarding the variability of the forecast error. That is, the reduction in MSE (per series) achieved by the Approximation method may be the result of the bias reduction only.

The comparison of the two sets of results discussed above indicates, implicitly, the high variability of the errors produced by the EWMA estimator. The decomposition of the empirical MSE into its constituent components (bias squared + variability of demand + variability of the estimates) would obviously enable the rigorous assessment of the contribution of these components to the empirical MSE. Nevertheless, no such results have been generated in our empirical study and this issue requires further simulations on real data.

Returning to our discussion on the validity of the categorisation rules, the overall rules have been empirically validated in this section in both a re-order interval and re-order level context. This is viewed as a significant finding since one could argue that, from a practitioner's perspective, what matters is the validity of the overall categorisation schemes rather than the accuracy of each one of the pair-wise categorisation rules. This issue will have certain implications in the tests performed in the following sub-sections.

10.5.2 Sensitivity of the categorisation rules

The categorisation rules tested in this chapter have been developed based on a theoretical analysis of the MSE, associated with alternative intermittent demand estimators. This particular accuracy measure was chosen (in chapter 6), for theoretical comparison purposes, because of its mathematically tractable nature. In the previous sub-section, the empirical MSEs were considered to assess the validity of the theoretical rules. Nevertheless, the MSE is known to suffer in practice from being particularly sensitive to extreme observations (outliers).

There is no reason at this stage why we should restrict our analysis to the MSE results only. Other error measure(s), that are not so sensitive to the presence of outliers, can also be used to test our theory. More conclusive results can then be generated about the conditions under which one method performs better than one or all other methods. The natural choice for that purpose is the RGRMSE (per series) measure since it is related by definition to the MSE measure and it has been, theoretically (Fildes, 1992) and empirically (section 10.4), shown to be relative.

Moreover, it is true to say that, from a practitioner's perspective, what matters is the validity of the overall categorisation schemes rather than how accurate is each one of the pair-wise categorisation rules. That is, in real world application we are interested in which method performs best (rather than better) and under what conditions. Consequently, it has been decided to test the overall rules. Finally it is important to note that the rules indicate when each estimator is expected to perform best but not by how much. Parametric tests will be performed later in this section to assess the validity of the categorisation rules in "descriptive" terms. Method *x* is said to perform better than methods *y* and *z* in a particular file if:

$$RGRMSE_{x,y} = \frac{GRMSE_x}{GRMSE_y} < 1 \text{ and } RGRMSE_{x,z} = \frac{GRMSE_x}{GRMSE_z} < 1 .$$

In the following table the chi-square test statistic value is indicated, for all simulated scenarios considered in the empirical experiment. In brackets we present the number of times (series) that the methods perform as expected.

RGRMSE		α = 0.05	α = 0.1	α = 0.15	α = 0.2
All Points in Time	L.T.=1	**18.03** (920)	*6.91* (1031)	**10.85** (1060)	2.88 (1072)
	L.T.=3	**35.43** (974)	**16.55** (1082)	**21.07** (1094)	**11.68** (1125)
	L.T.=5	**49.46** (976)	**28.93** (1074)	**24.33** (1082)	**25.56** (1129)
Issue Points Only	L.T.=1	*16.65* (1126)	**20.79** (1221)	**28.18** (1253)	**30.64** (1252)
	L.T.=3	**22.31** (1165)	**21.31** (1220)	**21.68** (1260)	**37.06** (1317)
	L.T.=5	**42.33** (1134)	**29.90** (1212)	**40.51** (1219)	**41.72** (1246)

Table 10.44. chi-square test results (RGRMSE)

All results are statistically significant at 5% level with the only exception of the simulated scenario: L.T. = 1, α = 0.2. The overall categorisation rules have been validated, using the RGRMSE measure, which is not sensitive to the presence of extreme observations (outliers). The MA(13) has not been included in the chi-square test because its theoretical properties were not investigated in the thesis. However, it is important to note that the MA(13) performs better than the Approximation method for α = 0.05 when Percentage Best (PBt) results are generated on the empirical data considering the RGRMSE per series (see Appendix 10.C).

10.5.3 Parametric tests

The theoretical rules indicate under what conditions one method performs better than one or all other methods but not by how much. Therefore, some parametric tests need to be performed. The descriptive accuracy measure chosen for that purpose is the RGRMSE (across series). The RGRMSE is scale independent and not sensitive to the presence of extreme observations.

As discussed in the previous sub-section, what matters from a practitioner's perspective is the validity of the overall rules rather than how accurate is each one of the pair-wise categorisation rules. As such, it has been decided to test only the overall rules (see figures 10.2, 10.3). The whole data set (3,000 series) has been divided in two (re-order interval systems) or three (re-order level systems) sub-sets depending on which method is theoretically expected to perform best. RGRMSE results have then been generated for each of the data sub-sets and all possible pair-wise comparisons[2]. Results with a value less than 1 indicate that the second method performs better than the first (i.e. the GRMSE associated with the second method across all the series contained in the particular sub-set is lower than the GRMSE that corresponds to the first one). The error results per series per method are generated based on the GRMSE associated with the application of each one of the estimators on the series under concern. To test the statistical significance of the results, the natural logarithm of the GRMSE per file is calculated for all methods and the *t*-test (difference between population means) is used on the logGRMSEs.

The *t*-test results for all points in time are presented in table 10.45[3]. The actual RGRMSE value for all the pair-wise comparisons (and simulated scenarios considered) are separately presented in *Appendix 10.D* of the thesis (table *10.D.1*). The Approximation method is theoretically expected to perform best when p (average inter-demand interval) > 1.32 and/or CV^2 (squared coefficient of variation of the demand sizes) > 0.49. The corresponding data sub-set is termed "non-smooth" for presentation purposes. Croston's method is expected to perform best in all other cases (smooth demand). Positive values for the test statistic indicate that the second method performs better than the first.

[2] Parametric results are generated on a pair-wise comparison level and as such there is no reason why we should not also consider the MA(13) method.
[3] Some areas in table 10.45 have been shaded in order to make the information displayed easier to read.

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
		α (smoothing constant value) = 0.05					
L.T.=1	non-smooth	0.54	0.38	0.92	**2.90**	**-2.34**	**3.28**
	smooth	0.32	0.28	0.59	**2.60**	*-2.28*	**2.89**
L.T.=3	non-smooth	0.38	0.55	0.92	1.31	-0.92	*1.87*
	smooth	0.52	0.34	0.86	1.12	-0.59	1.47
L.T.=5	non-smooth	0.67	0.19	0.86	0.98	-0.29	1.17
	smooth	0.07	0.65	0.71	0.59	-0.51	1.24
		α (smoothing constant value) = 0.1					
L.T.=1	non-smooth	0.46	1.04	1.50	**3.00**	**-2.51**	**4.05**
	smooth	0.23	0.37	0.60	**3.13**	**-2.90**	**3.51**
L.T.=3	non-smooth	0.65	1.15	*1.79*	1.43	-0.76	**2.59**
	smooth	0.44	0.99	1.43	0.89	-0.44	*1.89*
L.T.=5	non-smooth	0.34	1.24	1.56	0.85	-0.50	*2.10*
	smooth	-0.12	0.89	0.76	0.30	-0.41	1.19
		α (smoothing constant value) = 0.15					
L.T.=1	non-smooth	0.43	1.51	**1.93**	**2.99**	**-2.53**	**4.52**
	smooth	0.12	1.24	1.37	**2.62**	**-2.50**	**3.89**
L.T.=3	non-smooth	0.83	1.45	*2.27*	1.42	-0.58	**2.88**
	smooth	-0.10	1.33	1.23	0.09	-0.19	1.44
L.T.=5	non-smooth	-0.34	**2.34**	*1.97*	0.24	-0.57	**2.59**
	smooth	-0.38	1.23	0.83	-0.50	0.10	0.74
		α (smoothing constant value) = 0.2					
L.T.=1	non-smooth	0.63	*1.88*	**2.50**	**2.87**	**-2.20**	**4.79**
	smooth	0.52	*1.64*	**2.17**	**2.64**	**-2.12**	**4.32**
L.T.=3	non-smooth	0.41	**2.48**	**2.87**	0.70	-0.28	**3.20**
	smooth	-0.01	*1.79*	*1.78*	-0.78	0.77	1.03
L.T.=5	non-smooth	-0.63	**3.30**	**2.63**	-0.73	0.08	**2.59**
	smooth	-0.45	*2.01*	1.56	**-1.83**	1.38	0.18

Table 10.45. *t*-test (GRMSE, difference between population means), all points in time

The results presented above indicate that the Approximation method performs better than all other estimators, in all simulated cases, for the non-smooth demand patterns and this is in accordance with the theory. This is also true for the Approximation – MA(13) pair-wise comparison, although no theoretical results have been developed regarding the performance of the latter estimator in an intermittent demand context. Nevertheless, it is important to note that the performance differences are not significant in all cases.

For the non-smooth demand patterns, the performance of the Approximation method improves as the smoothing constant value increases. In particular, for $\alpha \geq 0.15$ the superiority of the Approximation method is significant in almost all simulated cases. The rules developed in chapter 6 covered all possible smoothing constant values in the realistic range 0.05 – 0.2. The empirical parametric results presented thus far in this sub-section suggest that the rules are valid for higher α values.

For the smooth demand category Croston's method is theoretically expected to perform best. The empirical results presented in table 10.45 do not support the theory. In fact Croston's method is shown to be outperformed by EWMA in the majority of the simulated cases and by the Approximation method in all simulated scenarios. The empirical non-parametric results generated in this section indicate that there is some merit in considering Croston's estimator for the smooth demand patterns. No such empirical evidence has been found in this sub-section

The pair-wise categorisation rules were developed in chapter 6 in a way that one estimation procedure always (theoretically) performs better in, what we now call for the purpose of this chapter, "non-smooth" data set[4]. Therefore, one should expect well-marked differences in all the cases covered by the data set under concern. The estimator selected for the "smooth" data set[5] was an approximate solution, since in the case that both criteria (p and CV^2) take a value below their corresponding cut-off value, no estimator can be shown, theoretically, to perform better in all cases.

The overall rules were constructed by synthesising the pair-wise rules. Theoretically there are no doubts as to which estimator performs best in the "non-smooth" demand category but there is still uncertainty governing the area formed when both p and CV^2 take a value below their specified cut-off points. It may be for this reason that

[4] This is the area that corresponds to quadrants 1, 2 and 4 in any of the categorisation schemes developed in chapter 6.
[5] This is the area that corresponds to quadrant 3 in any of the categorisation schemes developed in chapter 6.

Croston's method does not perform better than EWMA and the Approximation method in the smooth demand category, in a re-order interval context. The theoretically coherent delineation of the "smooth" demand quadrant in the categorisation schemes discussed in chapter 6 would be an interesting avenue for further research. A graphical illustration of alternative possibilities follows in figure 10.8.

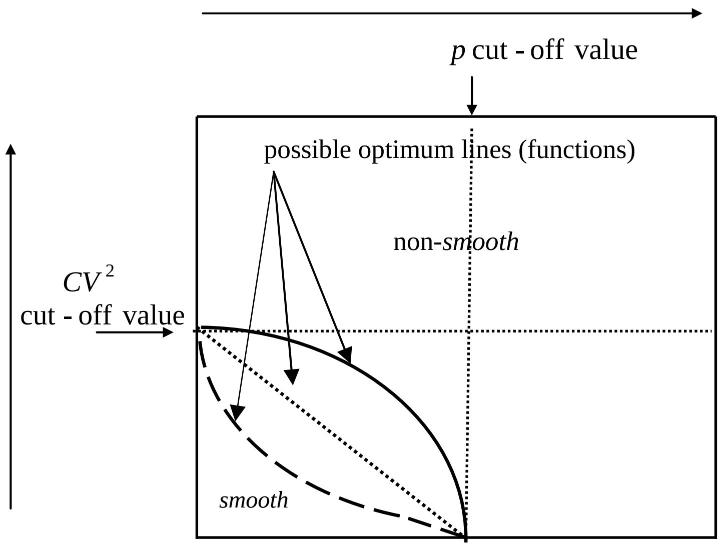

Figure 10.8. Delineation of the "smooth" demand quadrant

From a practitioner's perspective, the "smooth" demand category does not raise any significant difficulties as far as forecasting and inventory control are concerned. Moreover, this research is not concerned with improving the management of "smooth" demand items. What will be, most probably, required in a real world system is a rule according to which the "non-smooth" demand patterns can be identified and an estimator, other than the one already in place, can be recommended to deal with the "non-normal" nature of the corresponding demand data series. In that respect, the re-order interval overall categorisation rules have been empirically validated by means of parametric tests. Clearly, further research is required to "refine" the cut-off points, but at this stage we can claim that for p (average inter-demand interval) > 1.32 and/or CV^2 (squared coefficient of variation of the demand sizes) > 0.49 the Approximation method performs better than all the other estimators considered in this chapter. This is true theoretically, but it has also been validated by means of simulation on theoretically generated data (see chapter 7) and by means of simulation on real data by using both parametric and non-parametric tests.

In a re-order level context, the conditions under which each estimator is expected to perform best are as follows:

Approximation method: $p > 1.33$ and/or $CV^2 > 0.49$ (non-smooth demand)

Croston's method: $p \leq 1.33$ and $0.28 < CV^2 \leq 0.49$ (smooth A category)

EWMA: $p \leq 1.33$ and $CV^2 \leq 0.28$ (smooth B category).

The parametric results regarding the validity of the categorisation rules in a re-order level context are similar to the results presented above in this sub-section. In particular, the Approximation method is shown to perform best in the "non-smooth" demand category. The performance of the estimator under concern improves with the smoothing constant value but now statistically significant differences are indicated for $\alpha \geq 0.1$. The conclusions of the analysis are the same as those made before in this sub-section. Therefore, the *t*-test results (as well as RGRMSE results) are separately presented in *Appendix 10.D* of the thesis (tables *10.D.2 – 10.D.5*).

10.5.4 Conclusions

In this section the empirical validity of the categorisation rules proposed in chapter 6 was tested on the real data sample available for this research. The chi-square test was first used to test whether or not the methods' performance is independent of what is expected from the theory. Results on each file are generated by using the Mean Squared Error (MSE) since the rules were developed based on that accuracy measure. The MSE though is known to suffer from being particularly sensitive to the presence of outliers. Therefore, the analysis of the results is repeated by considering the RGRMSE per series. One can argue that, from a practitioner's perspective, what matters is the validity of the overall categorisation schemes and not the accuracy of each single pair-wise comparison rule. Consequently, and in order to simplify the results analysis process, the chi-square test results for the RGRMSE are developed only for the overall rules.

The rules indicate under what conditions one estimator is expected to perform better than all other methods but not by how much (i.e. error differences in volume terms are not addressed). As such, parametric tests were also used in this section to test the accuracy of the overall rules (across all series). The accuracy measure used for that purpose was the RGRMSE (across series) due to its scale independent nature and insensitivity to extreme observations (outliers).

Our conclusions can be summarised as follows:

- The pair-wise categorisation rules have been validated, with the only exception of the EWMA – Croston rule, when issue points only are considered and for L.T. = 1 or 3. The rules have been validated using non-parametric tests.

- The validation of the rules demonstrates, implicitly, the high variability of the EWMA forecast errors.

- In a re-order interval context the rule given in chapter 6 was that the Approximation method performs best for
  p (average inter-demand interval) > 1.32 and/or
  CV^2 (squared coefficient of variation of the demand sizes) > 0.49.
  This has been validated by using both parametric and non-parametric tests. In the remaining demand patterns, non-parametric tests show that there is some merit in adopting Croston's method but the parametric results do not support this finding.

- In a re-order level context the rule given in chapter 6 was that the Approximation method performs best for
  p > 1.33 and/or CV^2 > 0.49.
  This has been validated by using both parametric and non-parametric tests. In the remaining demand patterns, non-parametric tests show that there is some merit in adopting either Croston's method ($0.28 < CV^2 \leq 0.49$) or EWMA ($CV^2 \leq 0.28$) but no such empirical evidence has been found when parametric tests are used.

- Some further research is required in the area of "categorisation for non-normal demand patterns". In particular the theoretically coherent delineation of the "smooth" demand quadrant in the categorisation schemes discussed in chapter 6 remains unaddressed.

10.6 Conclusions

In this chapter the forecasting results of a real data simulation experiment have been presented. The purpose of the simulation exercise was to assess the empirical validity and utility of the main theoretical findings of this research. In this chapter we have discussed the empirical validity of our theoretical findings. Issues related to the utility of those findings are separately addressed in the following chapter.

In the preceding sections Croston's method, Approximation method, EWMA and a 13 period Moving Average (MA(13)) have been examined with respect to their biased (or unbiased) nature and subsequently they have been compared with respect to the forecasting accuracy resulting from their implementation in practice. The categorisation rules developed in chapter 6 have also been validated with this simulation experiment. To meet the objectives of this chapter, a wide range of accuracy measures has been employed. We have argued theoretically, and in this chapter we have demonstrated empirically the importance of the selection of appropriate accuracy measures. Our conclusions reflect the synthesis of our empirical findings with respect to all the accuracy measures used for generating results, and they can be summarised as follows:

- Croston's method is biased and this is consistent with the theory developed in this thesis. Croston's method has been shown to be biased mathematically (chapter 4), by means of simulation on theoretically generated data (chapter 7) and also using simulation on real data (in this current chapter).

- The Approximation method is approximately unbiased over the range of α values 0.05 to 0.2, showing slight bias (in opposite directions) at the extremes of this range. The Approximation method is the least biased of the four estimators examined in this chapter.

- EWMA and MA(13) are biased. The biased nature of both estimators in a re-order interval context is not what was theoretically expected and this issue requires further examination.

- Where bias had been theoretically anticipated, the sign of the bias is for all methods the theoretically expected one.

- All the pair-wise categorisation rules developed in chapter 6 have been validated, with the exception of the EWMA – Croston rule in a re-order level context and when the lead time is one or three periods. The rules have also been validated by means of simulation on theoretically generated data in chapter 7. The validation of the rules indicates, implicitly, the high variability of the EWMA forecast errors.

- The Approximation method performs best in the "non-smooth" demand category (p > 1.32 and/or CV^2 > 0.49 in a re-order interval context; p > 1.33 and/or CV^2 > 0.49 in a re-order level context). This has been shown to be correct mathematically (chapter 6), by means of simulation on theoretically generated data (chapter 7) and by means of simulation on empirical data in this chapter (using both parametric and non-parametric tests).

- Some further research is required in the area of "categorisation for non-normal demand patterns". In particular the theoretically coherent delineation of the "smooth" demand quadrant in the categorisation schemes discussed in chapter 6 remains unaddressed.

- The unbiased nature of the Approximation method is reflected in the superior forecasting accuracy of this estimator when compared with the other methods considered in this chapter.

- EWMA performs better than Croston's method in a re-order interval context. When issue points only are considered the comparison results are inconclusive.

- The MA(13) compares favourably with the smoothing methods for low (0.05) smoothing constant values. This estimator is also very robust to the presence of outliers.

Our findings regarding the performance of the accuracy measures used for generating results can be summarised as follows:

- The Mean Signed Error (ME), generated across series, does not particularly suffer from the scale dependence problem. The ME though, in a single series evaluation, it is particularly sensitive to the presence of extreme observations (outliers).

- Different accuracy measures can lead to different conclusions, in an intermittent demand context.

- The Percentage Best measure should be preferred to the Percentage Better for large scale comparison exercises.

- The Relative Geometric Root Mean Square Error (RGRMSE) is a very well behaved accuracy measure in an intermittent demand context.

CHAPTER 11

Empirical Analysis – Inventory Control

11.1 Introduction

The purpose of this chapter is to assess the empirical utility of the main theoretical findings of this research. For this to be done the inventory control performance of Croston's method, Approximation method, EWMA and a 13 period Moving Average (MA(13)) has been simulated on real demand data. The data comes from the automotive industry (3,000 SKUs) and details regarding our empirical data sample have been presented in the previous chapter. No theoretical results have been developed in this thesis regarding the application of MA(13) in an intermittent demand context. Nevertheless, the MA(13) is the estimator that has been used in practice to forecast demand for the SKUs covered in our empirical sample. As such the MA(13) performance can be viewed as the benchmark against which the performance of the other estimators can be compared.

Owing to the nature of the demand data available for this research, we focus on intermittent demand forecasting for re-order interval, rather than re-order level inventory control systems (see chapter 9). The inventory control model to be used for simulation purposes is of the periodic, order-up-to-level nature. The stock control model has been discussed in chapter 9, where details regarding our empirical data simulation experiment are also given (see section 9.7).

This chapter is structured as follows: in section 11.2 we recap the scenarios (control parameters and their corresponding values) considered for simulation purposes as well as the measures used for generating inventory control comparative results. In section 11.3 the inventory control results are presented, for a service driven system (specified Customer Service Level – CSL), and in section 11.4 the case of a cost driven system is empirically analysed (specified cost criteria). The "overall" re-order interval categorisation rule developed in chapter 6 is tested on the inventory control results in section 11.5.

Chapter 11 concludes the empirical part of this thesis and, consequently, some overall comments are made, at the end of the chapter, regarding our real data simulation experiment. In particular, certain problems may arise in practical applications that have not been accounted for in our empirical analysis. These issues are discussed in section 11.6 whereas in section 11.7, with hindsight, we discuss our reflections on the experimental design and the empirical data limitations. Finally, the conclusions of this chapter are presented in section 11.8.

11.2 Simulated scenarios

In this chapter comparative inventory control results will be presented for a wide range of real world scenarios. The inventory control system, that has been used for simulation purposes, is of the periodic order-up-to-level nature (T, S) system. At the end of every period T (one month) the inventory position is raised to the replenishment level S. Demand per unit time period/lead time demand is assumed to follow the Negative Binomial Distribution (NBD). The managerial constraints imposed on the system are:

a specified shortage fraction per unit value short (B_2)

optimisation condition: $\sum_{x=O}^{S-1} p(x) \leq \max\left(1 - \frac{IR}{B_2}, 0\right) < \sum_{x=0}^{S} p(x)$

a specified emergency delivery fraction per unit value short (B_3)

optimisation condition: $\sum_{x=0}^{S-1} p(x) \leq \frac{B_3}{IR + B_3} < \sum_{x=0}^{S} p(x)$

a specified customer service level (P_2)

“optimisation” condition: $DR(1 - P_2) \geq \sum_{x=S+1}^{\infty} (x - S) p(x)$

where:

R: review period expressed as a fraction of one year (in our case R = 1/12).

D: average annual demand in units

I : annual inventory holding charge

T : review period (1 month)

L : lead time (number of unit time periods)

$p(x)$: probability density function of demand x over $L+T$

The simulated control parameter values are as follows (see also section 9.7):

Control parameter	*Values assigned*
α	0.05 to 0.2 step 0.05
L	1 to 5 step 2
B_2 policy, target value = $1-\frac{IR}{B_2}$	0.93 to 0.96 step 0.03
B_3 policy, target value = $\frac{B_3}{IR+B_3}$	0.95 to 0.98 step 0.03
P_2	0.90 to 0.95 step 0.05

Table 11.1. The inventory control simulated conditions

Results are separately generated (and tested for statistical significance) for all three managerial policies. In total there are 72 simulated scenarios covered in our empirical experiment. Details regarding the results generation process can be found in chapter 9 of the thesis.

11.2.1 Performance measures

We record the average monthly number of units in stock and the Customer Service Level (CSL) achieved by using the estimator under concern on each of the real demand data series, for all the control parameter combinations. The CSL is defined as follows:

$$CSL = \frac{Total\ \ Demand\ - Backorders}{Total\ \ demand} \tag{11.1}$$

Since cost information is not available, no inventory cost results can be generated. Volume differences are considered instead, regarding the number of units kept in stock for the alternative estimators assessed in this chapter. In addition, CSL results are generated and they can be related directly to performance differences as far as the number of units backordered is concerned.

The Percentage Best (PBt) measure (see sub-section 9.7.5) will be used to generate pair-wise comparison results with respect to the inventory control performance of the estimators. We test the hypothesis that the two methods under concern perform identically. The alternative hypothesis is that the second method performs better than the first one in terms of (a) percentage of times that it results in the lowest number of units in stock or (b) percentage of times that it results in the highest service level. Positive values of the *Z*-test statistic (difference between population proportions) indicate statistical significance in favour of the second method. The critical values for rejecting the null hypothesis are 1.64 and 2.32 (one-sided test) for the 5% and 1% significance level respectively.

The Percentage Best measure can provide us with valuable information about which method performs best but not by how much. Therefore, a relative measure is also introduced, to indicate the performance differences in descriptive terms. The accuracy measure considered does not indicate performance differences at a pair-wise comparison level bur rather it indicates the cost and service level "regret" associated with using a particular estimator, considering the best possible attainable performance. In particular, we calculate the Average (per series) Percentage Regret (APR, Sani and Kingsman, 1997) of using any of the estimators considered in this chapter (see sub-section 9.7.5). The estimator with the least APR is the one that should be preferred for real world applications.

When results are generated with respect to the average number of units in stock, the APR of using estimator x, across all series, is given by (11.2):

$$StockAPR_x = \frac{\sum_{i=1}^{n} \frac{S_{x,i} - Mn_i}{Mn_i}}{n} \tag{11.2}$$

where:

i is the particular demand data series considered

$n = 3{,}000$

$S_{x,i}$ is the average number of units in stock resulted from the employment of estimator x on series i, and

Mn_i is the lowest average number of units in stock achieved (by any of the estimators considered) on the particular series.

When results are generated on CSL, the APR is the amount each estimator falls short of the maximum possible CSL across all series:

$$CSLAPR_x = \frac{\sum_{i=1}^{n} \frac{Mx_i - CSL_{x,i}}{Mx_i}}{n} \tag{11.3}$$

where:

$CSL_{x,i}$ is the CSL (%) resulted from the employment of estimator x on series i, and

Mx_i is the maximum CSL achieved (by any of the estimators considered) on the particular series.

11.3 Inventory control results - P_2 policy

In table 11.2 we present the average (across all files) customer service level (to the third decimal place) resulting from the implementation of alternative estimators on our empirical data sample, for both P_2 values considered in our experiment. Note that the MA(13) results change with the smoothing constant value. This is because the updating procedure of the variability of the MA(13) forecast error is the same as that used for all the other estimators (see also sub-section 9.7.4):

$$MSE_{t,T+L} = \alpha \left\{ \sum_{i=t-T-L+1}^{t} \left(Y_i - Y'_i\right) \right\}^2 + (1-\alpha) MSE_{t-1,T+L} \tag{11.4}$$

where:

Y_i and Y'_i are the actual demand and the forecast corresponding to period i respectively, and

t is the current time period.

		P_2 = 0.90				P_2 = 0.95			
		EWMA	Croston	Approx.	MA(13)	EWMA	Croston	Approx.	MA(13)
α =0.05	L.T.=1	0.932	0.915	0.913	0.913	0.954	0.944	0.942	0.942
	L.T.=3	0.935	0.920	0.915	0.917	0.953	0.942	0.940	0.940
	L.T.=5	0.942	0.932	0.928	0.931	0.957	0.950	0.946	0.948
α =0.1	L.T.=1	0.949	0.927	0.922	0.925	0.965	0.953	0.949	0.951
	L.T.=3	0.949	0.926	0.918	0.923	0.963	0.947	0.941	0.944
	L.T.=5	0.952	0.936	0.928	0.934	0.964	0.953	0.947	0.951
α =0.15	L.T.=1	0.958	0.936	0.929	0.934	0.972	0.959	0.954	0.957
	L.T.=3	0.956	0.931	0.920	0.928	0.968	0.951	0.943	0.948
	L.T.=5	0.958	0.939	0.927	0.936	0.969	0.955	0.946	0.952
α =0.2	L.T.=1	0.965	0.942	0.934	0.941	0.976	0.963	0.958	0.962
	L.T.=3	0.961	0.935	0.921	0.932	0.971	0.954	0.944	0.952
	L.T.=5	0.962	0.941	0.927	0.938	0.971	0.956	0.945	0.954
Average		**0.952**	**0.932**	**0.924**	**0.929**	**0.965**	**0.952**	**0.946**	**0.950**

Table 11.2. Customer Service Level results (P_2 policy)[1]

Before we discuss the results presented in table 11.2 it is important to comment on the considerable number of ties that occur with respect to the CSL achieved on every series by the alternative estimators. Ties are the natural consequence of the small number of demand occurring periods considered for conducting the "out of sample" comparisons between the alternative methods. In the majority, if not all cases, ties occur when:

$$CSL_{EWMA} = CSL_{Croston} = CSL_{Approx.} = CSL_{MA(13)}.$$

[1] **Explanatory note:** Some areas in this table, but also in most of the other tables to follow in this chapter, have been shaded in order to make the information displayed easier to read.

A high number of ties has also been observed when the stock holding results are generated (see the following tables). The number of files on which each method performs best and the number of ties occurring, for the P_2 = 0.90, 0.95 managerial constraints (and all the resulting possible combinations of the control parameter values) can be found in *Appendix 11.A* of the thesis. Similar numbers of ties have been observed for the rest of the managerial constraints considered in our experiment (B_2, B_3 - and their corresponding values).

The results presented in table 11.2 indicate that as the smoothing constant value increases the customer service level achieved by the alternative estimators (achieved CSL) increases as well. This is true for both theoretically specified CSL values (i.e. 0.90, 0.95). The results also indicate that EWMA gives the highest CSL in all simulated conditions.

For P_2 = 0.90, the CSL achieved by all estimators exceeds the target CSL for all the control parameter combinations. This is particularly true for the EWMA estimator in which case the percentage difference can be as high as 6%. Nevertheless, the fact that the EWMA offers the highest CSL does not imply that it is the best estimator. Under the P_2 managerial policy the objective is to meet a specified CSL. Achieved CSLs over and above the theoretically specified ones are obviously desirable, but they will, most probably, occur at the expense of a higher number of units in stock. Our discussion on that issue is continued later in this chapter. For P_2 = 0.90, the difference between achieved and desired CSL becomes minimum when the Approximation method is utilised. This result can be attributed to the approximately unbiased nature of the Approximation method which has been empirically demonstrated in section 10.3. In a similar manner the higher CSL achieved by Croston's method and EWMA is viewed as the natural consequence of the bias associated with their implementation in practice. This issue is further discussed in the following section. In the previous chapter, the MA(13) estimator was shown to perform comparatively well for low smoothing constant values when bias results were generated. That explains, partly at least, the very similar performance between the Approximation method and MA(13) estimator across both target service level values.

For P_2 = 0.95, the achieved CSL is in all cases relatively close (plus or minus) to the desired one. This is not true for the EWMA estimator in which case the achieved CSL always exceeds the desired one, the percentage difference being between 0.3% and 2.7%. When the target CSL is set to 0.95 the overall (across all α values) biased nature of Croston's method and MA(13) in not reflected on the achieved CSL.

The replenishment levels are calculated as a function of the estimate of demand over lead time plus one review period (corresponding level of demand + bias) and of the MSE (variability of the forecast errors + bias squared). The greater the bias associated with an estimator and the greater the variability of the corresponding forecast errors the higher the replenishment levels are going to be (i.e. actual replenishment level > theoretically correct replenishment level). As a result the positive discrepancy between the actual CSL and the theoretically desired one increases as well, most probably, at the expense of a very high number of units kept in stock. This issue is also further discussed in the following section. The average CSL (to the third decimal place) achieved by Croston's method, Approximation and MA(13) is, 0.952, 0.946 and 0.950 respectively. It is important to note that this average CSL reflects only the simulated conditions considered in our experiment, which nevertheless correspond to many real world applications.

The results presented for both P_2 values indicate that there is very little to choose between MA(13) and the Approximation method in terms of CSL.

Watson (1987) showed, by means of simulation on theoretically generated data, that demand-forecast fluctuations, in a lumpy demand context, can very often cause a significant (either positive or negative) shift in the CSL achieved as compared to the target CSL. The demand per unit time period was assumed to follow the stuttering Poisson distribution and the desired CSLs considered in that case were lower (0.7 – 0.9) than those considered in our simulation experiment. In addition, the mean demand level was calculated directly from the data rather than being estimated in a dynamic way as in the case of this experiment. Those may be the reasons that our empirical results do not fully support Watson's findings. The shifts do occur but (a) are not necessarily very large, at least in the case that P_2 = 0.95 and for all the estimators apart from EWMA, (b) the shifts tend to occur in the same direction and this is particularly true for P_2 = 0.90. The relationship between the desired and achieved CSL in an intermittent demand context, and for any of the estimators discussed in this chapter, is obviously of great practical importance and this may be an interesting avenue for further research.

The results presented thus far in this section indicate that all methods meet the target CSL. Strictly speaking, this is not true for the Approximation method and P_2 = 0.95 in which case the achieved CSL is slightly below the theoretically specified one. Under a different managerial constraint imposed on the system (other than the P_2 policy), detailed comparative results should be generated at this stage (with respect to the CSL achieved) to determine which estimator is the best, second best etc. In the context of this particular policy though, high CSLs, that exceed the theoretically specified ones, are not necessarily desirable, if we consider that high CSLs are achieved at the expense of a high number of units kept in stock. In fact the greater the discrepancy between achieved and specified CSL, the less our confidence will be in the particular estimator (since the system "returns" a CSL which is different from what has been theoretically specified). In that respect, the Approximation method and MA(13) perform best for P_2 = 0.90 and 0.95 respectively. At this stage we view as important to evaluate performance differences with respect to the number of units kept in stock to support the CSL achieved by the alternative estimators.

In tables 11.3 and 11.4 the non-parametric stock control results, with respect to the average number of units kept in stock, are presented for both P_2 values considered in our experiment. For each simulated scenario we present the *Z*-test statistic value (difference between population proportions). Positive values of the *Z*-test statistic indicate differences in favour of the second method. The critical values for rejecting the null hypothesis are 1.64 and 2.32 for the 5% and 1% significance level respectively (one-sided test). Statistically significant results at the 1% level are emboldened while significance at the 5% level is presented in italics.

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**7.77**	**8.46**	**15.90**	**-6.97**	**14.48**	1.51
	L.T.=3	**-12.03**	**28.49**	**18.09**	**-21.12**	**9.93**	**8.41**
	L.T.=5	**-19.99**	**37.77**	**21.66**	**-24.10**	**5.06**	**16.88**
α = 0.1	L.T.=1	**5.78**	**13.34**	**18.78**	**-4.64**	**10.34**	**8.81**
	L.T.=3	**-13.03**	**33.85**	**22.90**	**-16.84**	**4.12**	**19.06**
	L.T.=5	**-22.84**	**46.14**	**28.63**	**-19.55**	**-4.12**	**32.24**
α = 0.15	L.T.=1	0.19	**19.39**	**19.56**	**-7.74**	**7.92**	**11.98**
	L.T.=3	**-15.03**	**37.24**	**24.76**	**-14.90**	-0.13	**24.89**
	L.T.=5	**-22.52**	**47.95**	**31.11**	**-17.81**	**-5.86**	**36.08**
α = 0.2	L.T.=1	**-6.78**	**22.64**	**16.37**	**-13.24**	**6.61**	**9.93**
	L.T.=3	**-18.49**	**37.91**	**22.22**	**-17.45**	-1.16	**23.30**
	L.T.=5	**-22.49**	**47.29**	**30.43**	**-18.83**	**-4.58**	**34.37**

Table 11.3. PBt stock results, P_2 = 0.90

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**12.31**	**4.56**	**16.61**	**-3.03**	**15.17**	1.53
	L.T.=3	**-8.62**	**26.75**	**19.29**	**-20.38**	**12.46**	**7.12**
	L.T.=5	**-19.94**	**37.34**	**20.93**	**-24.33**	**5.30**	**15.89**
α = 0.1	L.T.=1	**9.59**	**9.03**	**18.28**	0.45	**9.15**	**9.47**
	L.T.=3	**-9.48**	**30.98**	**22.85**	**-15.13**	**5.92**	**17.28**
	L.T.=5	**-20.09**	**44.45**	**29.14**	**-18.26**	*-2.20*	**31.06**
α = 0.15	L.T.=1	**4.60**	**14.52**	**18.87**	**-3.51**	**8.07**	**11.10**
	L.T.=3	**-11.59**	**34.16**	**24.31**	**-13.81**	**2.34**	**22.14**
	L.T.=5	**-20.50**	**46.96**	**31.56**	**-16.75**	**-4.47**	**35.35**
α = 0.2	L.T.=1	**-2.62**	**17.65**	**15.16**	**-8.49**	**5.90**	**9.40**
	L.T.=3	**-15.34**	**35.73**	**22.54**	**-15.13**	-0.23	**22.75**
	L.T.=5	**-21.16**	**46.46**	**30.17**	**-17.04**	**-4.88**	**34.39**

Table 11.4. PBt stock results, P_2 = 0.95

The results demonstrate the superior performance of the Approximation method in terms of number of units kept in stock. The Approximation method performs significantly better than Croston's method, EWMA and MA(13) at the 1% significance level for all the combinations of the control parameter values[2].

EWMA is the least desirable estimator and this is what intuitively one should expect from the results presented in table 11.2. The EWMA estimator results in the highest CSL but at the expense of the highest number of units kept in stock. We have more to say on this particular issue in the following section. The MA(13) performs significantly better than EWMA and Croston's method and finally Croston's method outperforms EWMA in the majority of the simulated conditions. Overall Croston's method seems to perform better for higher smoothing constant values in combination with long lead times. The APR results are presented in tables 11.5 and 11.6. for all the estimators considered in the simulation exercise.

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	19.42	7.69	6.30	6.44
	L.T.=3	18.01	7.44	5.05	6.59
	L.T.=5	12.40	7.07	4.31	6.10
α = 0.1	L.T.=1	35.12	9.24	7.76	6.91
	L.T.=3	33.92	9.20	6.14	7.01
	L.T.=5	25.07	9.98	4.87	7.79
α = 0.15	L.T.=1	53.11	10.75	9.76	8.49
	L.T.=3	51.97	11.65	7.86	9.26
	L.T.=5	37.14	12.49	6.38	9.65
α = 0.2	L.T.=1	72.54	11.94	12.13	9.93
	L.T.=3	71.18	13.99	10.76	11.85
	L.T.=5	50.22	15.38	8.90	12.03

Table 11.5. APR stock results, P_2 = 0.90

[2] In the Approximation – MA(13) pair-wise comparison there is only one case (L.T.=1, α = 0.05) where the superiority of the Approximation method is not statistically significant.

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	16.42	7.76	6.82	5.90
	L.T.=3	15.81	7.21	5.42	5.67
	L.T.=5	11.52	6.74	4.37	5.70
α = 0.1	L.T.=1	29.26	9.53	8.82	7.08
	L.T.=3	29.59	9.12	6.61	6.58
	L.T.=5	22.23	9.09	5.19	6.93
α = 0.15	L.T.=1	43.88	10.98	10.98	8.66
	L.T.=3	44.97	11.40	8.72	8.95
	L.T.=5	33.38	11.93	7.07	9.17
α = 0.2	L.T.=1	60.65	12.05	13.57	10.08
	L.T.=3	61.88	13.85	11.81	11.74
	L.T.=5	44.86	14.80	9.90	11.70

Table 11.6. APR stock results, P_2 = 0.95

When relative results are generated the MA(13) is shown to perform particularly well for one step ahead forecasts. In fact this estimator is now shown to perform better than the Approximation method for all the relevant simulated conditions. For higher lead times the Approximation method gives the lowest percentage regret. The very good forecasting performance of the Approximation method, for lead times greater than one, has already been discussed in section 10.4 of the previous chapter. The relative inventory control results presented in this section indicate that the forecasting accuracy improvements achieved when the Approximation method is used for lead times greater than one, are reflected in a very good stock control performance with respect to the average number of units kept in stock. The APR results confirm the poor performance of EWMA and Croston's method. EWMA gives in all cases the highest percentage regret followed (in all cases) by Croston's method.

11.3.1 Best α value analysis

Before we close this section we view as important to generate some comparative inventory control results considering the best α value performance of the smoothing estimators discussed in this chapter. By using the same smoothing constant value for comparison purposes we may have put some of the smoothing estimators (EWMA, Croston's method, Approximation method) at a relative advantage/disadvantage (see also sub-section 10.4.1). Therefore, we wish to conduct a best α value analysis of the results with respect to the inventory control performance of the smoothing estimators. In this chapter the smoothing constant value is not only used for obtaining an estimate of the level of demand but also for estimating the MSE. Therefore, the best α value performance can also be identified for the MA(13) estimator.

As discussed before in this section, the CSL offered by any of the estimators considered in this chapter increases with the smoothing constant value. Subsequently, we also expect the average number of units kept in stock to increase with the α value. Consideration of both CSL and stock holding results for optimisation purposes, though, is not feasible, at least not without our intervention and investigation of each of the simulated cases, due to lack of relevant information (i.e. trade-offs between CSL and stock). On average, all estimators meet, approximately, or even exceed, the specified CSL. The average CSL achieved by all estimators, for each of the smoothing constant values considered and for both P_2 values, across the three simulated lead time lengths, is presented in the following table.

		EWMA	**Croston**	**Approx.**	**MA(13)**
P_2 = 0.90	α = 0.05	0.936	0.922	0.919	0.920
	α = 0.1	0.950	0.930	0.923	0.927
	α = 0.15	0.957	0.935	0.925	0.933
	α = 0.2	0.963	0.939	0.927	0.937
P_2 = 0.95	α = 0.05	0.955	0.945	0.943	0.943
	α = 0.1	0.964	0.951	0.946	0.949
	α = 0.15	0.970	0.955	0.948	0.952
	α = 0.2	0.973	0.958	0.949	0.956

Table 11.7. Average CSL achieved

Under the P_2 managerial policy the objective is to meet a specified CSL. In that respect all smoothing constant values are equally satisfactory (with respect to the CSL), when P_2 = 0.90, and the best (overall) smoothing constant value can be identified considering the stock holding performance of the estimators. For P_2 = 0.95

all estimators, apart from EWMA, do not meet the specified CSL for low smoothing constant values (say, ≤ 0.1) and identification of the best (overall) α value is not feasible for the reasons stated above in this section. (The issue of optimisation of the smoothing constant value is further discussed in the following section). Therefore, the best α value analysis is conducted only for $P_2 = 0.90$.

In particular, we record, for each of the simulated scenarios considered in our experiment and each of the estimators discussed in this chapter, the smoothing constant value that results in the lowest average (across series) number of units in stock. The best smoothing constant value is found to be 0.05. This is true for the performance of all estimators in all simulated cases (lead time lengths) and this is what intuitively one should expect from the CSL results presented in table 11.7. As the smoothing constant value increases, the achieved CSL (for every estimator) increases as well, but so does the average number of units kept in stock. The worst α value, with respect to the average number of units kept in stock, is, in all cases, 0.2.

The "optimal" smoothing constant values identified in the previous chapter, with respect to the ME performance of the smoothing estimators, are as indicated in the following table.

		Best smoothing constant value		
		EWMA	Croston	Approximation
All points in time	L.T. = 1	0.05	0.05	0.05
	L.T. = 3	0.05	0.05	0.05
	L.T. = 5	0.05	0.05	0.1

Table 11.8. Best α value performance – lowest ME across series

The results presented in tables 11.7 and 11.8 indicate the strong linkage between the bias associated with an estimator and the estimator's performance with respect to the CSL achieved.

The results of the best smoothing constant value analysis, for $P_2 = 0.90$, are as follows:

		Croston-EWMA	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
P_2 = 0.90	L.T.=1	**7.77**	**8.46**	**15.90**	**-6.97**	**14.48**	**1.51**
	L.T.=3	**-12.03**	**28.49**	**18.09**	**-21.12**	**9.93**	**8.41**
	L.T.=5	**-19.99**	**37.77**	**21.66**	**-24.10**	**5.06**	**16.88**

Table 11.9. PBt stock results, best α value

		EWMA	**Croston**	**Approx.**	**MA(13)**
P_2 = 0.90	L.T.=1	19.42	7.69	6.30	6.44
	L.T.=3	18.01	7.44	5.05	6.59
	L.T.=5	12.40	7.07	4.31	6.10

Table 11.10. APR stock results, best α value

The results indicate the superiority of the Approximation method. The Approximation estimator performs significantly better than all the other methods when the PBt measure is considered and it also gives the lowest APR. MA(13) is the second best estimator, performing better than Croston's method and EWMA when both non-parametric and relative results are generated. Finally, Croston's method performs better than EWMA. This is not true for one step ahead forecasts, when results are generated on the PBt measure. In that case EWMA outperforms Croston's estimator.

11.3.2 Conclusions

In an inventory control context, where a target service level is specified, all the estimators discussed in this chapter perform particularly well, in terms of meeting the specified CSL. From a practitioner's perspective, exceeding the target service level is desirable only if it is not happen at the expense of an increased number of units in stock. Since all methods meet the specified target[3], comparative results have been generated in this section with respect to number of units kept in stock to support the CSL achieved by the alternative estimators. Relative and non-parametric results demonstrate the superiority of the Approximation method for lead times greater that

[3] This is not true for the Approximation method and P_2 = 0.95 in which case the achieved CSL is slightly below the theoretically specified one.

one. For one step ahead forecasts, the Approximation method performs best in terms of number of times that gives the lowest average stock, but this is not reflected in the percentage regret associated with this method's application in practice. When the latter measure is used the MA(13) performs best. The APR is a relative measure and is scale independent. Therefore, we may suppose that the effect of outliers (contained in some of the series) works in favour of the MA(13) estimator and this agrees with findings presented in the previous chapter. EWMA and Croston's method are clearly the least desirable estimators, the former estimator performing always worse that the latter.

We conclude that the approximately unbiased nature of the Approximation method (that has been empirically demonstrated in the previous chapter) is reflected in the superior inventory control performance of this estimator in a re-order interval system that operates under a service constraint. The CSL achieved by this estimator deviates slightly from the theoretically specified one and the method results in the lowest number of units in stock. The MA(13) is the second best estimator. This method also performs very well, particularly for one step ahead forecasts. Croston's method and EWMA are found to be, overall, the least desirable estimators.

11.4 Inventory control results – cost criteria

In this section the stock control performance of intermittent demand estimators is analysed in the context of a specified cost criterion. In particular, we present the "relative" and non-parametric comparison results for the B_2 policy and target value equal to 0.93. This particular target value is viewed as very realistic, from a practitioner's perspective, since it relates directly to our empirical data sample and reflects real inventory control situations reported elsewhere in the academic literature (Kwan, 1991; Sani, 1995; see also section 9.7). The rest of the target values selected for simulation purposes are analysed in sub-section 11.4.1.

In table 11.11 we present the non-parametric results with respect to CSL given by alternative estimators. Positive values of the Z-test statistic indicate differences in favour of the second estimator. Statistically significant results at the 1% level are emboldened while significance at the 5% level is presented in italics.

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**26.14**	**-27.39**	**-3.78**	**24.93**	**2.51**	**-5.91**
	L.T.=3	**20.32**	**-22.22**	**-3.77**	**20.88**	-0.98	**-2.84**
	L.T.=5	**16.11**	**-17.52**	**-2.49**	**15.25**	1.27	**-3.70**
α = 0.1	L.T.=1	**29.39**	**-30.27**	**-2.66**	**30.91**	**-5.40**	**3.47**
	L.T.=3	**23.41**	**-24.72**	**-2.60**	**26.39**	**-7.17**	**5.30**
	L.T.=5	**19.66**	**-19.98**	-0.57	**22.48**	**-6.50**	**6.10**
α = 0.15	L.T.=1	**29.09**	**-29.77**	*-1.95*	**30.62**	**-5.40**	**4.01**
	L.T.=3	**23.25**	**-24.39**	*-2.20*	**26.30**	**-7.24**	**5.67**
	L.T.=5	**20.03**	**-20.23**	-0.34	**22.82**	**-6.50**	**6.27**
α = 0.2	L.T.=1	**27.80**	**-28.39**	-1.58	**29.43**	**-5.49**	**4.37**
	L.T.=3	**21.98**	**-22.77**	-1.34	**25.70**	**-7.98**	**7.03**
	L.T.=5	**19.59**	**-19.33**	0.42	**22.58**	**-6.73**	**7.03**

Table 11.11. PBt CSL results, B_2 policy (target value = 0.93)

The average (across all series) CSL achieved by all estimators, for both B_2 and B_3 policies (and their corresponding values) and for all the control parameter combinations can be found in *Appendix 11.B* of the thesis. The CSL given by all estimators, for all the control parameter combinations, when the B_2 policy is considered and for a target value equal to 0.93, always exceeds 0.92.

The results presented in table 11.11 indicate that the CSL performance of EWMA is significantly higher than that of any other estimator. For all the control parameter combinations considered in our experiment the EWMA always results in the highest customer service level. The Z-test statistic value, for all three pair-wise comparisons involving EWMA, is highly significant. The significantly greater number of times that the EWMA gives the highest CSL results is also translated to a substantial average percentage increase of the CSL achieved when EWMA is compared with all the other estimators (see *table 11.B.1* in *Appendix 11.B*). This issue is further discussed later in this section.

The inventory control results have been generated in a re-order interval context and as such they can be directly related to the results presented in sections 10.3, 10.4 and

10.5 (of the previous chapter), for all points in time. EWMA has already been shown to be a biased estimator but this is also true for Croston's method and MA(13). In that respect, a better performance of the latter two estimators was expected when results are generated on the achieved CSL. In addition, the results presented in section 10.4 demonstrate the relatively good performance of the EWMA estimator when ME (parametric and non-parametric) results are generated.

There is some evidence to attribute the exceptionally good CSL performance of EWMA not only to its biased nature but also to the substantial variability associated with the errors produced by this method[4]. In particular, the results presented in the previous chapter demonstrate that when the empirical MSE is taken into account, and for all points in time, EWMA performs worst, i.e. the MSE_{EWMA} is greater than that of any other method[5]. That agrees also with the theory generated in chapter 6, although the EWMA was then regarded as an unbiased estimator. In that particular chapter we commented on the very high variability of the EWMA estimates, in a re-order interval context. In summary, the CSL performance of EWMA can be attributed to:

- The theoretically unexpected biased nature of the estimator

- The theoretically expected high variability of the forecast errors produced by this method.

Croston's method performs significantly better than MA(13) for $\alpha \geq 0.1$. The Approximation method performs worse than Croston's method although the superiority of the latter estimator is not statistically significant for higher lead times and/or smoothing constant values. Finally, the Approximation method performs significantly better than the MA(13) although this superiority is not reflected on an average (across series) percentage reduction of the "regret" associated with the former estimator (see table 11.12). In fact, the APR given by the Approximation method is always higher than that of the MA(13) which leads us to believe that the effect of outliers works in favour of MA(13) (see also previous section).

---

[4] As the bias and the variability of the forecast errors produced by an estimator increase, the replenishment levels also increase and so does the achieved CSL.

[5] This is not true for the EWMA – MA(13) comparison in which case no theoretical results have been developed and as such no empirical tests have been performed.

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	0.18	2.22	2.63	2.26
	L.T.=3	0.30	2.22	2.82	2.38
	L.T.=5	0.35	1.58	2.15	1.66
α = 0.1	L.T.=1	0.07	2.70	3.44	2.88
	L.T.=3	0.17	2.90	3.90	3.26
	L.T.=5	0.20	2.27	3.29	2.53
α = 0.15	L.T.=1	0.06	2.86	3.83	3.03
	L.T.=3	0.16	3.17	4.62	3.56
	L.T.=5	0.18	2.59	4.08	2.94
α = 0.2	L.T.=1	0.06	2.87	4.16	3.05
	L.T.=3	0.18	3.28	5.15	3.65
	L.T.=5	0.18	2.71	4.68	3.10

Table 11.12. APR CSL results, B_2 policy (target value = 0.93)

The Approximation method is theoretically and empirically approximately unbiased while the variability of the estimates produced by this method is, theoretically, and as demonstrated in the previous chapter, empirically very low. That may explain the relatively poor performance of the Approximation method when CSL results are generated.

When the APR results are considered, Croston’s method performs slightly better than the MA(13). In fact Croston’s method gives always a lower percentage regret, but the differences are small. This is also demonstrated in table *11.B.1* of *Appendix 11.B* where the average CSL results, across all series, are presented for all the estimators considered in this chapter. The results indicate that the CSL given by Croston’s method always exceeds that of MA(13) by no more than 0.4%. The Approximation method always gives the lowest average CSL, the difference with the CSL achieved by Croston’s method being between 0.4% - 1.8%. The EWMA always performs best. The CSL given by all methods exceeds the 92% in all simulated scenarios. We have calculated the difference between the CSL given by EWMA and the average CSL, across all three remaining estimators, achieved for all the control parameter combinations (B_2 policy, target value = 0.93). The difference is always between 1% and 3%.

In the following tables the stock holding PBt and APR results are presented for the B_2 policy and target value equal to 0.93.

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**-16.29**	**30.77**	**18.94**	**-34.36**	**23.30**	**-4.75**
	L.T.=3	**-21.60**	**34.70**	**17.16**	**-33.15**	**15.20**	*2.04*
	L.T.=5	**-21.93**	**39.48**	**22.41**	**-28.10**	**8.10**	**14.70**
α = 0.1	L.T.=1	**-19.01**	**40.00**	**26.90**	**-28.31**	**12.44**	**15.28**
	L.T.=3	**-25.82**	**43.03**	**22.28**	**-25.92**	0.13	**22.16**
	L.T.=5	**-25.53**	**48.14**	**28.88**	**-21.68**	**-5.22**	**33.43**
α = 0.15	L.T.=1	**-19.07**	**47.05**	**35.07**	**-24.04**	**6.75**	**29.36**
	L.T.=3	**-26.34**	**47.99**	**27.70**	**-22.53**	**-5.16**	**32.26**
	L.T.=5	**-26.32**	**51.96**	**32.53**	**-18.60**	**-10.41**	**41.17**
α = 0.2	L.T.=1	**-18.23**	**51.23**	**40.72**	**-22.32**	**5.58**	**36.31**
	L.T.=3	**-26.48**	**49.32**	**29.15**	**-22.06**	**-5.99**	**34.38**
	L.T.=5	**-25.60**	**52.93**	**34.54**	**-19.16**	**-8.71**	**41.68**

Table 11.13. PBt stock results, B_2 policy (target value = 0.93)

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	27.43	6.92	4.02	7.80
	L.T.=3	22.37	7.26	3.94	7.84
	L.T.=5	14.17	6.89	3.26	6.73
α = 0.1	L.T.=1	49.97	7.57	2.52	6.97
	L.T.=3	42.70	8.66	2.81	7.44
	L.T.=5	28.67	9.35	2.67	7.69
α = 0.15	L.T.=1	75.13	9.04	1.85	7.95
	L.T.=3	64.95	11.08	2.54	9.20
	L.T.=5	43.82	12.32	2.59	9.62
α = 0.2	L.T.=1	102.98	10.93	1.67	9.59
	L.T.=3	88.46	13.57	2.84	11.27
	L.T.=5	60.11	15.76	3.08	12.23

Table 11.14. APR stock results, B_2 policy (target value = 0.93)

The results show what intuitively one should expect from the analysis conducted thus far in this section. In particular, EWMA is now shown to perform significantly worse than any other estimator. EWMA offers the highest CSL at the expense of a considerable amount of units kept in stock. In fact the percentage regret given by the EWMA estimator can be as high as 100%. We cannot comment further on the trade-offs between CSL and amount of units kept in stock because of lack of relevant information.

Overall, the MA(13) performs better than Croston's method even though the PBt results are in favour of the latter estimator when the lead time is 3 or 5 periods and for $\alpha \geq 0.1$. The Approximation method performs significantly better than all other estimators.

The results presented thus far in this sub-section indicate clearly which is the best estimator when either CSL or number of units in stock are considered for comparison purposes. To facilitate the process of selecting the best estimator across both criteria the simulation output was analysed with respect to the following:

- Average percentage increase or decrease of the CSL achieved, across all series, at a pair wise comparison level;

- Average percentage increase or decrease of the number of units kept in stock, across all series, at a pair wise comparison level. In this case, the average number of units in stock obtained by estimator, say x, in a particular series, is expressed as a percentage of the average stock holding associated with another estimator, say y. The percentages are averaged across series to indicate the average percentage increase or decrease in stock obtained by employing estimator x instead of y.

In table 11.15 we present the average percentage differences for the B_2 policy (target value = 0.93) and all the corresponding control parameter combinations. For each simulated scenario we first indicate the average, across series, percentage increase or decrease in stock and then the average CSL percentage differences, for all possible pair-wise comparisons. Positive results indicate differences in favour of the second method (percentage decrease in stock or percentage increase in CSL). Negative results indicate differences in favour of the first estimator.

			Croston-EWMA	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
α = 0.05	L.T.=1	% stock	-19.57	17.40	2.49	-18.92	-1.65	2.36
		% CSL	1.78	-2.14	-0.37	1.90	-0.13	-0.24
	L.T.=3	% stock	-14.72	14.15	2.87	-14.00	-1.44	2.46
		% CSL	1.60	-2.12	-0.52	1.88	-0.28	-0.24
	L.T.=5	% stock	-7.59	8.94	3.14	-7.25	-0.74	2.44
		% CSL	1.03	-1.49	-0.46	1.17	-0.13	-0.33
α = 0.1	L.T.=1	% stock	-26.92	18.49	1.91	-28.39	0.94	0.14
		% CSL	1.73	-2.19	-0.46	1.87	-0.15	-0.32
	L.T.=3	% stock	-31.69	26.18	4.90	-32.76	0.40	3.66
		% CSL	2.38	-3.28	-0.91	2.75	-0.37	-0.53
	L.T.=5	% stock	-18.26	18.79	5.56	-19.34	0.72	4.06
		% CSL	1.74	-2.62	-0.88	1.98	-0.24	-0.64
α = 0.15	L.T.=1	% stock	-60.49	39.51	6.08	-62.31	0.57	4.97
		% CSL	2.59	-3.50	-0.91	2.76	-0.16	-0.74
	L.T.=3	% stock	-48.49	35.22	6.85	-50.87	1.08	5.34
		% CSL	2.68	-4.01	-1.33	3.04	-0.36	-0.98
	L.T.=5	% stock	-28.39	26.48	7.70	-31.05	1.64	5.67
		% CSL	2.06	-3.39	-1.33	2.36	-0.30	-1.03
α = 0.2	L.T.=1	% stock	-82.66	47.17	7.62	-85.27	0.82	6.40
		% CSL	2.62	-3.85	-1.23	2.79	-0.18	-1.06
	L.T.=3	% stock	-65.66	42.12	8.16	-69.23	1.44	6.47
		% CSL	2.79	-4.54	-1.75	3.12	-0.34	-1.41
	L.T.=5	% stock	-38.37	32.57	9.43	-42.56	2.27	7.01
		% CSL	2.21	-3.99	-1.78	2.53	-0.32	-1.46

Table 11.15. Stock and CSL % differences (B_2 policy, target value = 0.93)

Considering the pair-wise relative comparison results, Croston's method is shown to outperform MA(13), for $\alpha = 0.05$. For higher smoothing constant values we cannot determine which estimator is better.

All estimators perform, overall, better than EWMA. The CSL achieved by Croston's method, Approximation and MA(13) is in all cases above 92%. EWMA offers an increase in customer service that can be as high as 3%. Nevertheless this CSL percentage increase is supported by a tremendous increase in the amount of stock that is kept in the system (the stock percentage increase can be as high as 85%). We cannot comment further on the trade-offs between CSL and amount of units kept in stock because of lack of relevant information. Nevertheless, it is important to note that there should only be few real situations where a few points increase in the CSL (2%-3% reduction in the number of units backordered) would actually reduce a cost function when there is such a tremendous increase in stock.

The CSL achieved by the Approximation method is, for $\alpha \leq 0.1$, less than 1% lower than that achieved by Croston's method or MA(13). At the same time, the Approximation method offers a percentage decrease in stock which is between 2% and 5%. For $\alpha > 0.1$ the CSL difference is between 1% - 1.5% and the stock percentage decrease between 6% and 9%. The ratio between the annual inventory holding charge (I) and the shortage fraction per unit value short (B_2) when the target value is 0.93 is:

$$1 - \frac{IR}{B_2} = 0.93 \Leftrightarrow \frac{I}{B_2} = 0.84 \text{ (for } R = 1/12 \text{, see also section 9.7)}$$

Under the B_2 managerial policy, the expected number of units short in a replenishment cycle is given by (11.5)

$$Expected\ no.\ units\ short = \sum_{x=S+1}^{\infty} (x - S)p(x) \qquad (11.5)$$

(where $p(x)$ is the probability density function of demand x over $L+T$)

and the average number of units in stock by (11.6)

$$Average\ stock\ level = S - \mu_{L+T} + \frac{DR}{2} \qquad (11.6)$$

where μ_{L+T} is the expected demand over lead time plus one review period and D is the average annual demand in units.

The total annual inventory cost (TCI) is calculated as follows:

$$TCI = \frac{A}{R} + Ic\left(S - \mu_{L+T} + \frac{DR}{2} \right) + B_2 c \frac{1}{R} \sum_{x=S+1}^{\infty} (x - S) p(x) \tag{11.7}$$

where A is the ordering cost (£).

(For the above results please refer to section 9.6.)

The percentage decrease in stock achieved when one estimator is used instead of another relates directly to (11.6) or the second part of equation (11.7). The CSL has been recorded as percentage of demand (units) satisfied directly from stock. The percentage decrease in CSL achieved when one estimator is used instead of another can be interpreted as the percentage increase in the number of units backordered. Subsequently, it relates to equation (11.5) or the third part of equation (11.7).

Considering the above information, the Approximation method can be shown to perform best, i.e. give the lowest Total Cost of Inventory (see also sub-section 11.4.1). Croston's method outperforms MA(13) for $\alpha = 0.05$. A very similar performance is the case for the estimators under concern for higher smoothing constant values. EWMA is clearly the least desirable estimator and this is due to the high stock resulting from its implementation in practice. Further investigation is clearly required to determine performance differences in cost terms. For this to be done though, the unit cost information is required.

Moreover, we would also wish to determine actual performance differences when the best smoothing constant value (with respect to both CSL and stock) is utilised for all estimators. Unfortunately we are not able to do so due to the lack of unit cost information. That is, if the unit cost information were available, the smoothing constant value that minimises a cost function could be selected for each estimator and then comparisons could be conducted with respect to the best α value performances.

The best α value with respect to CSL achieved is, for all estimators, 0.2. This is perhaps not true for the Approximation method and lead times greater than one, in which case the CSL achieved is not very much sensitive to the smoothing constant value (see tables *11.B.1* and *11.B.2* in *Appendix 11.B* of the thesis). The estimators discussed in this chapter are expected to perform best, with respect to the number of units kept in stock, across all files, for very low smoothing constant values (say 0.05) and this has been empirically demonstrated for the P_2 criterion in the previous section. When both CSL and stock are considered a decision on the best smoothing constant value for each estimator cannot be made, at least not in an automatic way (i.e. without our intervention and investigation for each one of the simulated scenarios).

The results presented thus far in this section indicate the superiority of the Approximation method in a re-order interval context where the objective is to minimise a cost function. In particular, comparative inventory control results have been generated with respect to the B_2 criterion and a target value equal to 0.93. Considering both stock and CSL results the Approximation method is shown to outperform the other estimators considered in this chapter. Croston's method and MA(13) are found, overall, to perform very similarly. EWMA performs worst and this is due to the very high number of units kept in stock to support the CSL achieved by this estimator.

11.4.1 Other simulated conditions – summary results

The comments that have been made thus far in this section are valid for all the simulated scenarios that correspond to the B_2 policy (target value = 0.96) as well as the B_3 policy (target value = 0.95, 0.98). As such the corresponding simulation results are presented in detail in *Appendix 11.B* of the thesis and only a few summary results will be given in this sub-section. Some justification for the selection of the particular target values has been given in section 9.7. The target values considered for simulation purposes are viewed as realistic, from a practitioner's perspective, and they correspond to inventory control systems described elsewhere in the academic literature (Kwan, 1991; Sani, 1995).

EWMA is always shown to be the least desirable estimator for reasons discussed in the previous sub-section. Overall, Croston's method and MA(13) perform very similarly and this finding does not agree with the results presented in Sani and Kingsman (1997). This issue is further discussed later in this section. The Approximation method performs very well, when both stock and CSL results are taken into account. In the following table we summarise the pair-wise comparison results (similar to those presented in table 11.15) for all the pair-wise comparisons involving the Approximation method, for the B_2 (target value = 0.93, 0.96) and B_3 policy (target value = 0.95, 0.98). In particular we present the average stock and CSL percentage difference across all parameter combinations for each of the four specified managerial constraints. It is important to note that these average values reflect only the simulated conditions considered in our experiment. If other scenarios had been considered (i.e. other lead times or smoothing constant values) the results would not necessarily be the same. The ratio between the inventory holding charge and the shortage fraction (B_2 or B_3) is also given, to enable an approximate quantification of the benefits gained when the Approximation method is utilised instead of the other estimators. Positive results indicate percentage differences in favour of the Approximation method whereas negative values indicate percentage differences in favour of the other estimator.

	Approximation method compared with							
	EWMA		Croston		MA(13)		$\frac{I}{B_2}$	$\frac{I}{B_3}$
	Stock	CSL	Stock	CSL	Stock	CSL		
B_2 policy target value = 0.93	27.25	-3.09	5.56	-0.99	4.25	-0.75	0.84	
B_2 policy target value = 0.96	23.14	-2.08	4.43	-0.73	3.09	-0.53	0.48	
B_3 policy target value = 0.95	25.08	-1.80	4.53	-0.60	3.41	-0.44		0.63
B_3 policy target value = 0.98	20.03	-1.08	2.95	-0.38	1.45	-0.27		0.24

Table 11.16. Average percentage differences (B_2, B_3 policies)

For the B_3 policy, the shortage function is the same as for the B_2 criterion (equation (11.5)) while the average number of units in stock is given by (11.8)

$$Average\ stock\ level = \sum_{x=0}^{S} (S - x)\, p(x) + \frac{DR}{2} \tag{11.8}$$

The total annual inventory cost (TCI) is calculated as follows:

$$TCI = \frac{A}{R} + Ic\left(\sum_{x=0}^{S} (S - x)\, p(x) + \frac{DR}{2} \right) + B_3 c \frac{1}{R} \sum_{x=S+1}^{\infty} (x - S) p(x) \tag{11.9}$$

(For the above results please refer to section 9.6)

The percentage decrease in stock achieved when the Approximation estimator is used instead of another method relates directly to the second part of equation (11.7) and (11.9) for the B_2 and B_3 policy respectively. The percentage decrease in CSL achieved when the Approximation method is used instead of another method can be interpreted as the percentage increase in the number of units backordered/satisfied from an emergency delivery. Consequently, it relates to the third part of equation (11.7) and (11.9) for the B_2 and B_3 policy respectively.

For the particular I/B_2 and I/B_3 ratios simulated with our experiment, the Approximation method performs better than EWMA. That is, the former estimator gives a lower Total Cost of Inventory (TCI, see equations (11.7) and (11.9)) for an "average"[6] SKU. The results indicate that the Approximation method should be, generally, preferred to EWMA in a cost driven system. We cannot obviously exclude the possibility that EWMA shows a superior performance; but, intuitively, for that to happen the shortage penalty fraction should be much greater than the inventory holding charge. It is possible that there are real world systems where there is a great difference between the shortage fraction and the stock holding charge. Nevertheless, the Approximation method should be expected to perform better than EWMA (lower TCI) in the majority of real world applications.

[6] The percentage increase/decrease values presented in table 11.16 are averages obtained across all 3,000 SKUs.

The general superiority of the Approximation method is also evident in the Approximation – Croston and Approximation – MA(13) comparisons. Unless the shortage penalty fraction is much greater than the inventory holding charge, the Approximation method is, intuitively, expected to perform better (i.e. give a lower TCI) than both other estimators.

The Approximation method compares more favourably with all the other methods as the target value (for both policies considered in our experiment) decreases, i.e. as the ratio between the inventory holding charge and the shortage fraction under concern increases. No empirical results have been generated for ratio values below 0.48 (for the B_2 policy) and 0.24 (for the B_3 policy). Therefore, there is only enough empirical evidence to claim that the Approximation method should be preferred to EWMA, Croston's method and MA(13) when:

$\frac{I}{B_2} \geq 0.48$, for a cost driven system that operates under a specified shortage fraction per unit value short

$\frac{I}{B_3} \geq 0.24$, for a cost driven system that operates under a specified emergency delivery premium charge per unit value short.

The Approximation method may perform best for even lower values of the ratio between the inventory holding charge and the shortage fraction, although this is just a speculation since no simulations have been conducted in this thesis for such values. Both $I / B_2 = 0.48$ and $I / B_3 = 0.24$ have been chosen rather arbitrarily and the true cut-off values may be lower. Moreover, as stated before in this section, further investigation is clearly required to determine performance differences in cost terms. For this to be done though, the unit cost information is required.

11.5 The categorisation of "non-normal" demand patterns

The theoretical categorisation rules developed in this research (pair-wise rules and overall, regarding all methods' application, rules) have been validated, as far as forecasting is concerned, in chapter 10 of the thesis. The categorisation rule regarding all methods' performance in a re-order interval context (periodic inventory control system) is the following:

For p (average inter-demand interval) > 1.32 and/or
CV^2 (squared coefficient of variation of the demand sizes) > 0.49
use Approximation method.
In all other cases use Croston's method.

In the previous chapter we found that the Approximation method performs best for p > 1.32 and/or CV^2 > 0.49. That has been validated by using both parametric and non-parametric tests. For the remaining demand patterns, non-parametric tests showed that there is some merit in adopting Croston's method but the parametric results did not support this statement.

We now wish to assess the validity of this rule in an inventory control context. Subsequently, CSL and stock holding results are generated regarding the application of the rule in our empirical data sample and the performance of the rule is compared against the inventory control performance of both Approximation and Croston's method.

When the managerial policy is a specified percentage of demand to be satisfied directly from stock (P_2 = 0.90, 0.95) the CSL achieved when the rule is employed for stock control purposes is given, for all the control parameter combinations, in the following table.

	P_2 = 0.90				P_2 = 0.95			
	L.T.=1	L.T.=3	L.T.=5	**average**	L.T.=1	L.T.=3	L.T.=5	**average**
α =0.05	0.914	0.918	0.931	**0.921**	0.943	0.940	0.948	**0.944**
α =0.1	0.925	0.923	0.933	**0.927**	0.951	0.944	0.950	**0.949**
α =0.15	0.933	0.926	0.934	**0.931**	0.956	0.948	0.951	**0.952**
α =0.2	0.939	0.929	0.935	**0.934**	0.960	0.949	0.952	**0.954**
	overall average = 0.928				**overall average = 0.949**			

Table 11.17. Categorisation rule, CSL (P_2 policy)

The results presented in the above table should be compared with the results given in table 11.2 of this chapter. The average CSLs achieved by the Approximation method and Croston's method, for both P_2 policies are as follows:

average CSL	P_2 = 0.90	P_2 = 0.95
Approximation method	0.924	0.946
Croston's method	0.932	0.952

The CSL achieved when the categorisation rule is used (CSL_{Rule}), is always between the corresponding CSL for the Approximation method and Croston's method. In particular, the CSL_{Rule} (for any of the control parameter combinations) is always slightly under the CSL resulting from Croston's method and slightly above the CSL resulting from the Approximation method. The percentage differences are assessed in detail in the following table where we present the average percentage increase or decrease of the CSL achieved and number of units kept in stock, across all series, at a pair-wise comparison level (Rule – Approximation, Rule – Croston).

For each simulated scenario we first indicate the average, across series, percentage increase or decrease in stock and then the average CSL percentage differences, for the two pair-wise comparisons involving the categorisation rule. Positive results indicate differences in favour of the rule (percentage decrease in stock or percentage increase in CSL). Negative results indicate differences in favour of either Croston's or the Approximation method.

			Approx. - Rule	Croston - Rule	Approx. - Rule	Croston - Rule
			P_2 = 0.90		P_2 = 0.95	
α = 0.05	L.T. = 1	% stock	-0.89	0.37	-0.63	0.24
		% CSL	0.17	-0.11	0.10	-0.10
	L.T. = 3	% stock	-1.34	0.76	-1.07	0.50
		% CSL	0.26	-0.19	0.16	-0.13
	L.T. = 5	% stock	-1.43	1.00	-1.11	0.95
		% CSL	0.24	-0.15	0.20	-0.13
α = 0.1	L.T. = 1	% stock	-1.35	0.10	-0.89	-0.12
		% CSL	0.29	-0.25	0.19	-0.17
	L.T. = 3	% stock	-2.13	0.59	-1.75	0.48
		% CSL	0.44	-0.32	0.32	-0.25
	L.T. = 5	% stock	-2.73	1.71	-2.15	1.20
		% CSL	0.51	-0.32	0.38	-0.23
α = 0.15	L.T. = 1	% stock	-1.52	-0.39	-0.97	-0.70
		% CSL	0.40	-0.31	0.25	-0.26
	L.T. = 3	% stock	-2.83	0.51	-2.34	0.05
		% CSL	0.62	-0.51	0.49	-0.35
	L.T. = 5	% stock	-3.57	1.66	-3.06	1.06
		% CSL	0.68	-0.43	0.55	-0.36
α = 0.2	L.T. = 1	% stock	-1.35	-1.21	-0.76	-1.74
		% CSL	0.47	-0.39	0.27	-0.28
	L.T. = 3	% stock	-3.21	-0.45	-2.60	-0.82
		% CSL	0.77	-0.59	0.59	-0.47
	L.T. = 5	% stock	-4.28	1.06	-3.53	0.47
		% CSL	0.87	-0.56	0.65	-0.44

Table 11.18. Stock and CSL % differences (P_2 policy)

The results indicate the very small percentage increase in the CSL achieved when the categorisation rule is used instead of the Approximation method. The difference never exceeds 0.9%. The increase in CSL occurs at the expense of an increased number of units kept in stock when the rule is utilised. The stock holding percentage increase can be as high as 4%.

When the rule is compared with Croston's method the results indicate, overall, the better performance of the rule, as far as the number of units kept in stock is concerned. The decrease in the number of units kept in stock results in a decreased CSL when the rule is utilised.

For both P_2 values, the CSL achieved by the rule meets approximately, or even exceeds, the theoretically desired CSL. Therefore, we focus on the stock holding results in order to generate conclusions about the comparative performance of the categorisation rule. Overall, the rule performs better than Croston's method and this is in accordance with the theory. Nevertheless, the rule does not outperform the Approximation method. In fact the opposite is the case but this is what we were expecting based on the analysis conducted in the previous chapter. The parametric results generated in sub-section 10.5.3 indicated that the Approximation method performs better than Croston's method not only in the "non-smooth" but also in the "smooth" demand category. Therefore, it is not a surprise that the Approximation method outperforms the categorisation rule in an inventory control context.

As discussed in the previous chapter, further research is required to amend the categorisation rules proposed in chapter 6 of the thesis. In particular, in the next stage of this research it is intended to produce a theoretically coherent delineation of the "smooth" demand quadrant.

Stock and CSL percentage difference results have also been generated for both cost policies considered in our simulation experiment and they are presented in *Appendix 11.C* of the thesis. The results demonstrate that the Approximation method performs better than the rule (i.e. the Approximation method results in a lower Total Cost of Inventory) for:

$$\frac{I}{B_2} \geq 0.48 \text{ and } \frac{I}{B_3} \geq 0.24$$

for the B_2 and B_3 policy respectively. The Approximation method may perform better for even lower values of the ratio between the inventory holding charge and the cost criterion under concern although this is just a speculation. More scenarios (target values) need to be simulated in order to assess the validity of such a claim.

11.6 “Real world” applications

All the empirical results presented in this thesis have been produced in a dynamic way, i.e. we have assessed how the methods would have performed if they had been applied in “real world” situations. An exception has been the testing of the categorisation rules, in which case the methods’ performance was assessed on data points that had also been considered in order to define demand (i.e. identify theoretically which method is expected to perform better/best). This ex-post evaluation of the rules was due to our objective which was solely to check the empirical validity and utility of the rules rather than simulate their “ongoing” application in practice (at the end of every review period re-categorisation occurs and the appropriate estimation procedure is chosen).

Nevertheless, and despite the dynamic nature of our empirical analysis, we do recognise that certain problems may arise in practical applications that have not been, theoretically and/or empirically, accounted for in this thesis. In particular, we have assumed that there are no problems in correctly recording the demand data, an assumption that is often violated in practice. Moreover, we have assumed that the empirical series are stationary in the mean and from a practitioner’s perspective this assumption may seem restrictive. Both issues are now discussed in detail.

The employment of the forecasting methods discussed in this thesis necessitates the update of the mean demand level estimate at regular review (forecast revision) intervals. However, as Johnston (1993) noted, “……*in practice, it may not be possible to stick to this constant pattern, due to holidays, computer malfunction etc, whereupon it may be necessary to amalgamate the data from several periods or to handle that collected across only a fraction of a normal review interval* (p. 711)”. In the 1993 paper, Johnston showed how to modify the smoothing constant value used by the EWMA estimator in this problematic context of application. Practitioners often apply a correction factor to the observed issues to adjust them to those which would have been expected in a normal period (month in our case). EWMA is then applied to the adjusted data using the same α value. This adjustment though can be shown to lead to inappropriate re-order (or replenishment) levels.

Following Johnston's analysis, if k periods of data have to be combined together, where k can be any real number greater than zero, then the new smoothing factor, A, to be used for this combined data, Y_t, in a recursive equation

$$Y'_{t+1} = \mathrm{A}\,Y_t / k + (1 - \mathrm{A})\,Y'_t \tag{11.10}$$

(where Y'_t is the last EWMA estimate and Y'_{t+1} the current estimate of the demand level)

is computed from the relationship

$$\mathrm{A} = \frac{k(6\alpha + 3\alpha^2 (k-1))}{6 + 6\alpha(k-1) + \alpha^2\, 2\,k^2 - 3k + 1}\,. \tag{11.11}$$

No similar adjustments have been proposed in the literature regarding the application of the moving average estimator.

The work presented in this thesis has been postulated on the stationary mean model and the validity of this assumption is discussed in detail in chapter 2. Nevertheless intermittent demand data may often not be stationary. In that case Croston (1972) suggested that higher smoothing constant values (in the range 0.2 - 0.3) should be used in order to deal with the non-stationary nature of the data. Some problems though can be identified in testing the stationarity of non-normal demand patterns. Suppose for example that the data exhibit a trend movement, however this is tested (for example by considering few (two in our case) average annual demands). How can we be sure that this is a genuine, say upwards, movement of the data and not the effect of one outlier? In the former case the α value should be increased whereas in the latter case it should probably remain the same. Moreover, adaptive smoothing, that could also be considered, is known to introduce considerable instabilities (especially in the presence of outliers) in the system under concern and, generally speaking, automatic adaptive methods are not necessarily better than non-adaptive smoothing (Silver et al, 1998). Perhaps simulation of the estimator's inventory control performance (considering the managerial constraint in use and across a wide range of typical α values, say 0.05 - 0.3 step 0.05) on a regular basis (say every six periods) on the most recent data is a reasonable approach to identifying the best smoothing factor in the context of this research. Nevertheless, as Silver et al (1998) pointed out, a careful testing should be undertaken with actual time series from the organisation under study before any automatic procedure (of that nature) is adopted.

The ongoing application of the categorisation schemes developed in chapter 6 has not been simulated in the thesis. For such an application, Williams (1984) proposed using “buffer zones” so that “borderline” SKUs do not switch categories as the parameters vary from one side of the border to the other or categorisation is not that easily affected by outliers. The potential real world application of the categorisation rules developed in this thesis would also necessitate the determination of similar buffer zones. These zones (± 0.05 in Williams’ categorisation scheme) can be determined only after a thorough examination of the available demand data and the alternative methods’ simulated performance.

11.7 Experimental design limitations

Before we close this chapter and thus the empirical part of the thesis, we view as necessary to make some a posteriori comments regarding the data and the methods employed to produce our empirical results. That is, with hindsight, we would like, at this point, to discuss our reflections on the experimental design and the empirical data limitations.

The empirical data series consist of 24 demand observations. We have decided to produce out-of-sample results only on the latest 11 observations thus enabling the consideration of the MA(13). The particular moving average method has been used in practice in order to deal with the intermittent nature of the demand data series available for simulation and consequently the performance of this method is treated as a benchmark. We do not regret the consideration of this estimator but we do recognise that the exclusion of the moving average method from the simulation experiment could potentially result in a larger out-of-sample data set. By not considering the MA(13), initialisation of the smoothing estimators could occur on fewer demand data periods. A larger out-of-sample demand data set would increase our confidence in the forecasting and inventory control empirical results.

It is true to say that, in the way the simulation experiment was designed, the initialisation effect is carried forward by all estimators on all out-of-sample point estimates. Therefore, some of the empirical findings could be attributed not only to the true forecasting and inventory control performance of the methods considered but also to the small number of out-of-sample data points. As discussed above, one possible way to reduce the initialisation effect would be the exclusion of the MA(13) from the simulation experiment. Alternatively, longer demand data series would enable us to generate out-of-sample results on more demand data points. In this latter case the effect of initialisation could also be approximated by considering different out-of-sample sizes. Unfortunately, our efforts to obtain such longer demand data series have not been fruitful. Longer histories of data are not necessarily available in real world applications which means that decisions often need to be made considering samples similar to the one used for this research.

As discussed in section 10.2 our sample does not include highly intermittent demand items and therefore our theoretical results could not be assessed for such data. Series with average inter-demand interval values less than two could be viewed, from a practitioner's perspective, as non-intermittent. Johnston and Boylan (1996) re-conceptualised the term "intermittence". Their approach is taken forward in this thesis by producing theoretically sound non-normal demand definitions and according to our theory, series with $p \leq 2$ are certainly intermittent. Higher inter-demand interval values would obviously enable a more thorough investigation of the problem in hand and they would increase the validity of our results. Unfortunately, such series were not available for simulation. Intuitively, the performance differences identified in the empirical part of the thesis should be even more marked on series with higher p values. This is what one should expect from the theory presented in chapter 6. Nevertheless, this is merely speculation since no simulations have been conducted in this thesis for such values. In the next stage of research it is intended to test the performance of alternative intermittent demand estimators on data sets that reflect better the variability found in practice.

Given the small average inter-demand intervals considered in this research one might have expected the data to represent multiple orders in a period, i.e. a Poisson stream of orders, rather than a Bernoulli model of demand occurrence. This is obviously just a speculation and neither assumption has been tested on the real demand data. In our empirical experiment we do assume a Bernoulli process of demand occurrence. Such an assumption is dictated from a practical perspective because of the nature of the data available for analysis and from a methodological perspective for consistency purposes (see the theoretical part of the thesis). What we have been given is the monthly (accumulated) demand for every SKU rather than the individual transactions history. As such there is no means of checking the Poisson assumption unless some rather arbitrary adjustments are made to the discrete inter-demand interval data available for analysis so that the goodness-of-fit of the negative exponential distribution can be tested (see for example Johnston, 1975; Boylan, 1997). Moreover, we could, for simulation purposes, have assumed random arrivals (i.e. a Poisson stream) based on the calculated p value. This would be meaningful if we were confident that the Poisson assumption is valid. However, this is not the case.

Nevertheless, we do recognise that, if the order sizes and transaction times were available, our data may be fitted better by a Poisson distribution. Even in that case the empirical results generated in chapters 10 and 11 are still valid for systems that consider the aggregate order size over a unit time period and treat time as a discrete rather than a continuous variable. If the demand arrivals though can be modelled as a Poisson stream and forecasting is treated as a continuous time problem our empirical results are no longer valid. At this point it is important to note that the theoretical work conducted in chapters 4, 5 and 6 can be easily extended to cover negative exponentially rather than geometrically distributed inter-demand intervals thus reflecting better the empirical demand data that represent multiple orders in a period.

Before we close this section we should also mention that, with hindsight, it is regretted that the following equation was used in order to update the mean squared forecast error, in the empirical part of the thesis:

$$MSE_{t,T+L} = \alpha \left\{ \sum_{i=t-T-L+1}^{t} \left(Y_i - Y'_i \right) \right\}^2 + \left(1-\alpha\right) MSE_{t-1,T+L} \tag{11.12}$$

(see also sub-section 9.7.4 and section 11.3) where

Y_i and Y'_i are the actual demand and the forecast (produced at the end of period $i-1$) corresponding to period i respectively, and
t is the current time period.

Assuming a stationary mean model:

$$Y'_i = Y'_{i+1} = Y'_{i+2} = \dots\dots\dots = Y'_{i+T+L-1}. \qquad (11.13)$$

Therefore a more realistic calculation of the MSE would be the following:

$$MSE_{t,T+L} = \alpha \left\{ \sum_{i=t-T-L+1}^{t} Y_i - (L+T)*Y'_{t-T-L+1} \right\}^2 + (1-\alpha) MSE_{t-1,T+L} \qquad (11.14)$$

where (*) means multiplication.

Using (11.12) to update the $MSE_{t,T+L}$ is not a truly ex-ante procedure to calculating the variability of the estimates over $T+L$. In practice the differences between (11.12) and (11.14) should not be great. The former equation though is, intuitively at least, expected to lead to consistently lower MSEs.

11.8 Conclusions

In this chapter EWMA, Croston's method, Approximation method and MA(13) have been compared with respect to their empirical inventory control performance in a re-order interval context. Results have been generated for three possible managerial constraints: a service criterion (specified customer service level) and two cost policies (a shortage fraction per unit value short and an emergency delivery fraction per unit value short). The accuracy measures chosen for comparison purposes were the Percentage Best (PBt) and the Average Percentage Regret (APR). The latter is a relative measure that indicates the "regret" associated with using a particular estimator, considering the best possible attainable performance. Pair-wise percentage difference results have also been generated to enable us to select the best estimator when both CSL and stock differences are considered.

The results demonstrate the superiority of the Approximation method when both average number of units in stock and CSL results are taken into account. In the context of a service driven system the MA(13) also performs very well. In fact, MA(13) outperforms the Approximation method when the smoothing constant value is set to 0.05. Similar results were reported in the previous chapter (section 10.4) where the estimators were compared with respect to the forecasting accuracy resulting from their implementation in practice. EWMA and Croston's method are the least desirable estimators, the latter performing, overall, better than the former.

In a cost driven system the Approximation method can be shown to perform best (i.e. give the lowest Total Cost of Inventory) when the ration between the inventory holding charge and the shortage fraction is greater than 0.48 and 0.24 for the B_2 and B_3 policy respectively. MA(13) and Croston's method perform, overall, approximately the same and EWMA is found to be the worst estimator.

EWMA gives in all cases the highest CSL. This result can be attributed to the biased nature of this estimator and the high variability of the EWMA forecast errors. The very high CSL given by EWMA occurs at the expense of a considerable amount of stock. Even though no unit cost information is available to enable a detailed assessment of the trade-offs between CSL achieved and inventory cost associated with the implementation of EWMA in practice, the results clearly indicate that EWMA is the least desirable estimator. Similar findings have been reported in Sani and Kingsman (1997).

Croston's method is outperformed by MA(13) in the context of a service driven approach but this is not the case when a cost constraint is imposed on the system. This result does not agree with results presented in Sani and Kingsman (1997) where the superiority of the moving average method is very well marked. It is important to note that the moving average length considered in that case was different than the one employed for the purposes of our research. In particular a 26 period MA was considered where each time period was two weeks. Nevertheless, a better performance of the MA(13) estimator was expected than that found in practice.

In this chapter the empirical utility of the "overall" (re-order interval) categorisation rule developed in chapter 6 has also been tested. The rule performs, overall, better than Croston's method but is outperformed by the Approximation method. Even though this finding does not agree with the theory it is what we may have expected based on the empirical analysis conducted in the previous chapter. In the next stage of this research it is intended to amend the categorisation rules proposed in chapter 6 of the thesis.

The conclusions of this chapter can be summarised as follows:

- The unbiased nature of the Approximation method is reflected in the superior inventory control performance (in a re-order interval context) of this estimator when compared with other intermittent demand methods.

- EWMA has been found in the previous chapter to perform better than Croston's method, as far as forecasting accuracy is concerned, in a re-order interval context. This superiority is not reflected in the periodic inventory control results.

- The MA(13) performs very well for low (0.05) smoothing constant values[7].

- EWMA is, from a periodic inventory control perspective, the least desirable estimator. This agrees with findings of other researchers (Sani and Kingsman, 1997).

- The inventory control performance of the MA(13) is, overall, very similar to that of Croston's method and this does not agree with results presented elsewhere in the academic literature (Sani and Kingsman, 1997).

- The demand – forecast fluctuations in an intermittent demand context can cause significant discrepancies between the achieved and theoretically specified CSL. It

[7] The smoothing constant is used to estimate the variability of the MA(13) forecast errors.

may be more cost-effective, in an intermittent demand context, to employ a cost policy as opposed to a service driven one (see also Watson, 1987).

Because of lack of the unit cost information we have not been able to fully demonstrate the empirical utility of our theoretical findings (i.e. generate inventory cost results). Performance differences have been identified but clearly further examination is required to quantify those differences in cost terms.

CHAPTER 12

Contributions and Extensions of the Thesis

12.1 Introduction

In this chapter the main issues addressed in our research are briefly discussed and our contributions are summarised. Moreover, the limitations of our theoretical and empirical work are identified and avenues for further research are suggested.

This research aspires to take forward the current state of knowledge on forecasting intermittent demand.

The objectives of this research as stated in chapter 1 of the thesis are as follows:

1. To identify some of the causes of the unexpected poor performance of Croston's method
2. To develop new intermittent demand estimation procedures
3. To derive results for the mean squared forecast error of a range of intermittent demand estimates
4. To propose theoretically coherent categorisation rules that distinguish between intermittent and non intermittent demand
5. To identify appropriate accuracy measures for application in an intermittent demand context
6. To test the empirical validity and utility of the theoretical results on a large set of real world data.

All the objectives have been achieved and the contributions of the thesis are summarised in the following section.

12.2 Contributions

Our contributions can be summarised as follows:

- Croston's method is shown to be biased mathematically. This is confirmed by simulation on theoretically generated data and on empirical data, thus explaining one factor underlying high Mean Squared Errors (MSEs). The bias of Croston's method is also reflected in the achievement of Customer Service Levels (CSLs) above that required by the target CSL. (*Objectives 1 and 6*)

- A new method, the Approximation method, is proposed based on Croston's concept of building demand estimates from constituent elements. The Approximation method is approximately unbiased and this is proven mathematically and by means of experimentation on both theoretically generated and empirical data. (*Objectives 2 and 6*)

- The Approximation method is shown on real data to perform significantly better (more accurately) than Croston's method, EWMA and a 13 period Moving Average. The inventory control implications of this increased forecasting accuracy are also assessed in this thesis by considering the case of a periodic order-up-to-level inventory control system. The Approximation method outperforms the other estimators in either a service or a cost driven system. (*Objectives 2 and 6*)

- The approximate variance of Croston's method estimates is corrected. Moreover, the approximate sampling error of the mean is derived for all the estimators discussed in the theoretical part of the thesis, thus enabling quantification of one factor underlying MSEs. The approximate variance expressions are validated by means of simulation on theoretically generated data. (*Objective 3*)

- Assuming a stationary mean model, the lead time MSE is approximated for a range of intermittent demand estimates moving beyond the classical assumption of independence of the forecast errors over a fixed lead time. The accuracy of the approximate MSEs is validated by means of experimentation on simulated data. (*Objective 3*)

- A theoretical framework is proposed in this thesis that facilitates the conceptual categorisation of demand patterns that cannot be represented by the normal distribution (“*non-normal”* demand patterns). Subsequently, theoretically coherent and universally applicable categorisation schemes are proposed to assist the process of selecting estimators to deal with non-normal demand patterns. The rules are derived based on a mathematical analysis of the MSE associated with alternative estimators. The rules have been validated on theoretically generated data but also on real data using both parametric and non-parametric tests. (*Objectives 4 and 6*)

- A range of accuracy measures is specified to enable an objective accuracy comparison of alternative estimators in an intermittent demand context. Our empirical work demonstrates the importance of the selection process of accuracy measures in an intermittent demand context. (*Objectives 5 and 6*)

12.3 Summary of the theoretical part of the thesis

In this section the main theoretical issues explored in this Ph.D. research are drawn together and our conclusions are discussed in more detail.

12.3.1 The bias of intermittent demand estimates

Intermittent demand appears at random with some time periods showing no demand at all. Moreover, demand, when it occurs, may not be for a single unit or a constant size. Consequently, intermittent demand creates significant problems in the manufacturing and supply environment as far as forecasting an inventory control are concerned.

Exponentially Weighted Moving Averages (EWMA) and simple Moving Averages (MA) are very often used in practice to deal with intermittent demand. EWMA and MA consider the aggregate demand (demand per unit time period) and estimate how that moves through time. Both methods have been shown to perform well on empirical intermittent demand data. Nevertheless, the “standard” forecasting method for intermittent demand items is considered to be Croston’s method (Croston, 1972; see for example Silver et al, 1998). Croston built demand estimates from constituent elements, namely the demand size, when demand occurs, and the inter-demand interval. The method is, intuitively at least, superior to EWMA and MA. Croston's method is currently used by a best-selling statistical forecasting software package (Forecast Pro) and it has motivated a substantial amount of research work over the years.

Croston's method has been claimed to be of great value to organisations forecasting intermittent demand. Nevertheless, empirical evidence (Willemain et al, 1994) suggests modest gains in performance when compared with less sophisticated techniques; some evidence even suggests losses in performance (Sani and Kingsman, 1997). The model used by Croston in developing his method is based on the following assumptions:

1. Stationary Mean Model (SMM) for the demand sizes
2. SMM for the inter-demand intervals
3. No cross-correlation between demand sizes and inter-demand intervals
4. Geometrically distributed inter-demand intervals
5. Normally distributed demand sizes

The last assumption is the only one that can be "relaxed", in the sense that it does not affect the results given by the forecasting method. The above discussed model has also been assumed for the purposes of our research.

In an effort to identify the causes of Croston's method's forecast inaccuracy, as a first step towards improving this estimator, a mistake was found in Croston's mathematical derivation of the expected estimate of demand. That mistake contributes towards the unexpectedly modest benefits of the method when applied in practice.

According to Croston's method, separate exponential smoothing estimates of the average size of the demand and the average interval between demand incidences are made after demand occurs. If no demand occurs, the estimates remain the same. If we let:

p'_t = the exponentially smoothed inter-demand interval, updated only if demand occurs in period t so that $\mathrm{E}(p'_t) = \mathrm{E}(p_t) = p$, and

z'_t = the exponentially smoothed size of demand, updated only if demand occurs in period t so that $\mathrm{E}(z'_t) = \mathrm{E}(z_t) = z$

then following Croston's estimation procedure, the forecast, Y'_t for the next time period is given by:

$$Y'_t = \frac{z'_t}{p'_t} \tag{12.1}$$

and, according to Croston, the expected estimate of demand per period in that case would be:

$$\mathrm{E}(Y'_t) = \mathrm{E}\left(\frac{z'_t}{p'_t}\right) = \frac{\mathrm{E}(z'_t)}{\mathrm{E}(p'_t)} = \frac{\mu}{p} \tag{12.2}$$

(i.e. the method is unbiased.)

In this thesis we prove mathematically that Croston's method is biased. Using Taylor's theorem the bias is approximated (to the second order term) by:

$$Bias_{Croston} \approx \frac{\alpha}{2-\alpha}\mu\frac{(p-1)}{p^2} \tag{12.3}$$

where α is the smoothing constant value used.

We also show by means of simulation on theoretically generated data, that for $\alpha \leq 0.2$, the difference between (12.3) and the simulated bias lies within the specified 99% confidence limits. The bias approximation is particularly accurate when the estimator is applied in a re-order interval context but also in a re-order level context for low average inter-demand interval values. Finally, Croston's method is also shown to be biased on real intermittent demand data series, by employing both parametric and non-parametric tests.

Since Croston's method is biased we consider applying a factor to the estimates produced by his method so that the second order bias term is directly eliminated. The factor λ is:

$$\lambda = \frac{1-\frac{\alpha}{2}}{1-\frac{\alpha}{2p}} \tag{12.4}$$

and the resulting, approximately unbiased, estimation procedure is called the λ Approximation method. As a special case of this method, a heuristic is also proposed, the Approximation method. The updating procedure for the Approximation method is as follows:

$$Y'_t = \left(1-\frac{\alpha}{2}\right)\frac{z_t^{/}}{p_t^{/}} \tag{12.5}$$

The heuristic provides a reasonable approximation of the actual demand per period especially for the cases of very low α values and large p inter-demand intervals. When those conditions are not satisfied, a small bias is expected. That bias is approximated by

$$Bias_{APPROXIMATION} \approx -\frac{\alpha}{2}\frac{\mu}{p^2} \tag{12.6}$$

The negative sign in the above approximation indicates the fact that bias is in the opposite direction to that associated with Croston's method. Croston's method overestimates the mean demand level, whereas the Approximation method slightly underestimates it. For $\alpha \le 0.2$, the difference between the bias (lack of bias) approximations and the simulated bias is found, for both Approximation and λ Approximation method, to lie within the specified 99% confidence limits. When issue points only are considered, the accuracy of our approximation to Croston's bias deteriorates for average inter-demand intervals greater than two review periods. The simulation results have shown that this decline in accuracy affects the bias of the

λ Approximation and Approximation method in the corresponding simulated scenarios.

The λ Approximation method is approximately unbiased but the variability of the estimates produced by this method is relatively large. This is proven theoretically and by means of experimentation on simulated data. One of the objectives of this thesis is to propose new estimators that, at least theoretically, outperform Croston's method. The λ Approximation method does not outperform any of the other estimators that are taken into account in the theoretical part of the thesis and, consequently, it has not been considered in our empirical analysis.

12.3.2 The variance of intermittent demand estimates

The issue of the variability of intermittent demand estimates has also been explicitly addressed in this thesis.

We first correct the expression for the variance of Croston's estimates that appeared in Croston's paper (1972). Croston assumed that the inter-demand intervals follow the geometric distribution including the first success (i.e. demand occurring period). We show, that by not correctly estimating the variance of inter-demand intervals and by assuming no bias in the estimates produced by his method, Croston fails to produce an accurate expression for the variance of those estimates. Subsequently a corrected approximation to the variance is derived by applying Taylor's theorem.

The approximated variances of λ Approximation method and Approximation method are also derived. Both approximations are accurate to the second order term in a Taylor series. By means of simulation on theoretically generated data we show that the difference between the simulated variance and the corresponding theoretically expected variance lies, for all methods, within a 99% confidence interval of $\pm 17\%$ of the simulated variance. The accuracy of the variance expressions is not affected by the bias-associated problems discussed in the previous sub-section.

12.3.3 The categorisation of "non-normal" demand patterns

Unless demand for an item occurs at every inventory review period and is of a fairly constant size, it is, in the majority of cases, expected to cause significant problems as far as forecasting and inventory control are concerned. Infrequent demand occurrences and/or irregular demand sizes, when demand occurs, do not allow demand per unit time period or lead time demand to be represented by the normal distribution and demand in these cases is referred to as *non-normal* for the purpose of our research.

A certain confusion has been noticed in the academic literature as far as the definitions of the alternative non-normal demand patterns are concerned. Different authors use different criteria in order to define a specific demand pattern. Those criteria are hardly ever assessed against theoretical and practical considerations that should not be ignored if meaningful decision rules are to be constructed. Moreover, arbitrary cut-off values are, in the majority of cases, assigned to those criteria making their application to a more general context problematic.

Some work has appeared in the area of categorisation for non-normal demand patterns but this work has either been simulation based, lacking empirical validation (Johnston and Boylan, 1996) or it has been conducted on real data, but lacking universal applicability (Williams, 1984). We approach the categorisation problem as follows:

We first construct a theoretical framework that facilitates the conceptual categorisation of non-normal demand patterns. Non-normal demand patterns can now be, formally, defined. The definitions developed are given below:

- An intermittent demand item is an item whose demand is zero in some time periods.
- An erratic demand item is an item whose demand size is (highly) variable.
- A lumpy demand item is an item whose demand is zero in some time periods. Moreover demand, when it occurs, is (highly) variable.
- A slow moving item is an item whose average demand per period is low. This may be due to infrequent demand arrivals, low average demand sizes or both.

Williams (1984) suggested, implicitly, some theoretical and practical requirements to be considered when developing rules for the purpose of distinguishing between alternative demand patterns. Williams' criteria are first discussed in detail, then modified and finally are drawn together to a set of requirements which is as follows:

1. The categorisation scheme should suggest in what different ways to treat the resulting categories. The objective in categorising demand patterns is the identification of the most appropriate forecasting and inventory control methods to be applied to the different demand categories. As such, categorisation schemes should explicitly suggest which methods should be used under which circumstances.
2. The criteria considered in developing the rules should be dimensionless so that categorisation decisions regarding a SKU are independent of the product's unit of measurement or of demand over any time period other than the lead time or the review period.
3. Sensitivity to outliers should be taken into account. The categorisation scheme should not allow products to move from one category to another when few extreme observations are recorded.
4. The amount of data required to reliably classify demand patterns should also be considered. That is, the decision rules should take into account the limited number of demand occurrences that characterise any intermittent demand pattern.
5. Logical inconsistencies should not allow demand for a SKU to be classified in an unintended category.
6. Determination of the cut-off values should be non-arbitrary thereby enabling the general applicability of the categorisation scheme.

Williams' modified criteria are taken into account when we construct our categorisation rules.

The ultimate objective of categorising demand patterns is to select the "best" estimator and/or inventory control model[1] for each one of the resulting categories. In that respect it seems more logical indeed to:

1. Compare alternative estimation procedures
2. Identify the regions of superior performance for each one of them
3. Define the demand patterns based on the method's comparative performance

rather than arbitrarily defining demand patterns and then testing which estimation procedure performs best on each particular demand category.

The approach discussed above, appeared in Johnston and Boylan (1996) and is the one adopted in this thesis. Because of its mathematically tractable nature, the lead time Mean Square Error (MSE) is chosen for performing direct comparisons between existing and newly developed estimation procedures. (The MSE is similar to the statistical measure of the variance of the forecast errors but not identical since bias is also taken into account.) The results, presented in the form of cut-off values assigned to the mean inter-demand interval and the squared coefficient of variation, enable us to specify regions of superior performance for each one of the methods considered. Non-normal demand patterns can then be defined based on the results.

In many short term forecasting systems the cumulative lead time MSE is taken as the sum of the MSEs of the individual forecast intervals but that implies independence of the forecast errors. In this thesis we argue that the auto-correlation terms cannot be neglected. Subsequently we derive the lead time MSE expression assuming auto-correlated errors:

$$MSE_{L.T.} = L\left\{L\mathrm{Var}\left(Estimates\right) + L\,Bias^2 + Var\left(Actual\ \mathrm{Demand}\right)\right\} .$$

[1] The issue of inventory control has not been explicitly addressed in this thesis, from a theoretical perspective. As such, no theoretical results have been generated regarding the performance of alternative stock control models.

We first construct pair-wise categorisation rules and then we synthesise the results to produce categorisation schemes that are valid across all estimators. The categorisation rules are assessed against Williams' modified criteria. The categorisation of demand patterns, in case that issue point only estimates are considered, takes the form that is indicated below:

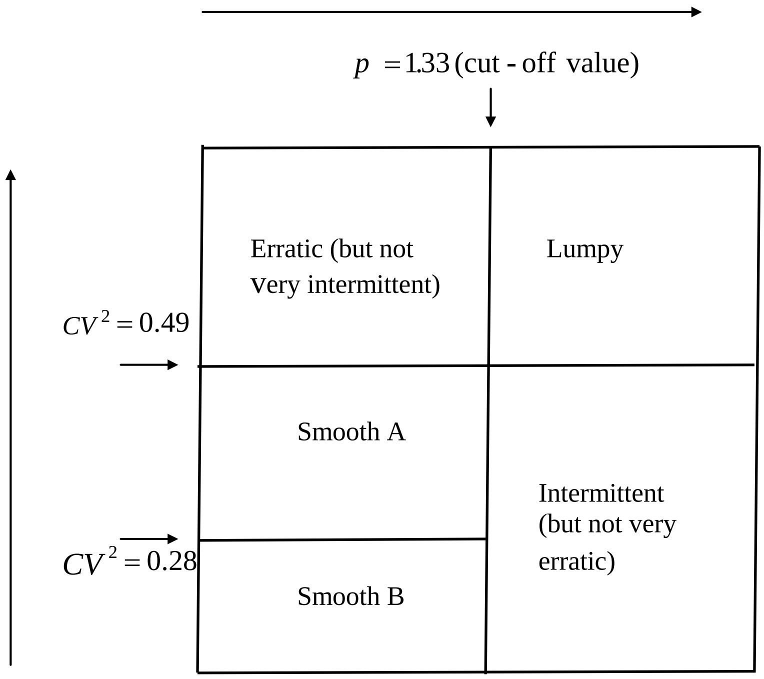

Figure 12.1. Categorisation of demand patterns (re-order level systems)

where p is the average inter-demand interval and CV^2 is the squared coefficient of variation of the sizes of demand.

The recommended estimation procedures are as follows:

Erratic:	Approximation method
Lumpy:	Approximation method
Intermittent:	Approximation method
Smooth A:	Croston's method
Smooth B:	EWMA

When all points in time are considered (i.e. in the context of a re-order interval inventory control system) the Approximation method performs better than all the other methods for $p > 1.32$ unit time periods and/or $CV^2 > 0.49$. For $p \le 1.32$ and $CV^2 \le 0.49$ Croston's method is theoretically expected to perform better than all the other methods.

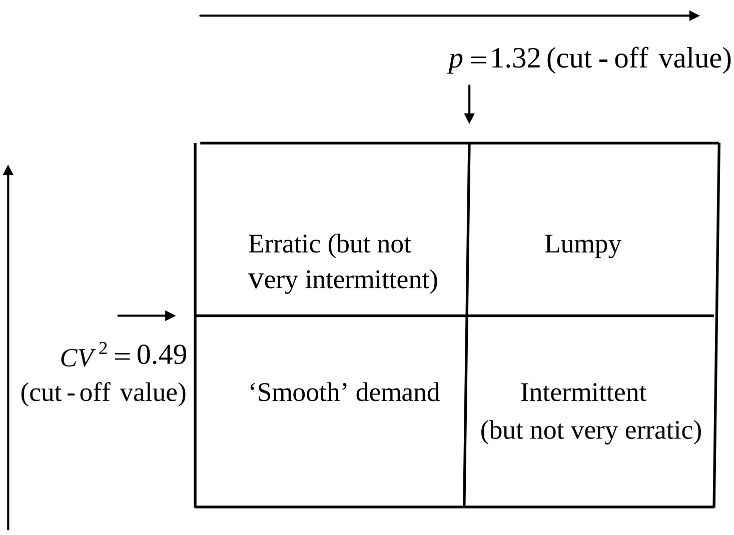

Figure 12.2. Categorisation of demand patterns (re-order interval systems)

The recommended estimation procedures are as follows:

Erratic:	Approximation method
Lumpy:	Approximation method
Intermittent:	Approximation method
Smooth:	Croston's method

For the non-smooth demand categories the Approximation method can be shown theoretically always to perform best. For the smooth demand patterns, the recommended estimator is an approximate solution since no method can be shown always to perform best. Both schemes are validated by means of simulation on theoretically generated data. When empirical demand data series are considered the schemes are validated when non-parametric tests are used. When parametric results are generated the Approximation method performs best in both smooth and non-smooth demand categories. The same is the case when the utility of the re-order interval rule (figure 12.2) is assessed in a periodic inventory control context.

From a practitioner's perspective, the "smooth" demand category does not raise any significant difficulties as far as forecasting and inventory control are concerned. Moreover, this research is not concerned with improving the management of "smooth" demand items. What will be required, most probably, in a real world system is a rule according to which the "non-smooth" demand patterns can be identified and an estimator, other than the one already in place, can be recommended to deal with the "non-normal" nature of the corresponding demand data series. In that respect, the overall categorisation rules have been empirically validated by means of parametric tests also. Clearly, further research is required to "refine" the cut-off points, but at this stage we can claim that the Approximation method performs best in a re-order interval context when:

p > 1.32 and/or CV^2 > 0.49

and in re-order level context when:

p > 1.33 and/or CV^2 > 0.49.

12.4 Summary of the empirical part of the thesis

In this section the main issues considered in the empirical part of the thesis are discussed and our detailed conclusions are summarised. The empirical analysis has been conducted on 3,000 intermittent demand data series coming from the automotive industry.

12.4.1 Accuracy measures

In order to test the empirical validity of our theoretical findings, some accuracy measures need to be selected for application on real demand data. It is the very nature of intermittent (demand) data, and in particular the existence of some zero demand time periods, that creates significant difficulties in selecting an appropriate accuracy measure. Nevertheless, those special properties of intermittent demand series seem to have been underestimated or in fact completely ignored in the past by both practitioners and academicians.

The objectives of the empirical analysis and the experimental structure of the simulation are clearly defined in order to enable the identification of specific accuracy measures and methods that collectively capture the information required.

The objectives of the simulation study are as follows:

1. To check in practice the validity of the theoretical results on bias
2. To check in practice the validity of the theoretical results on Mean Square Error
3. To generate results on the conditions under which one method is more accurate than others
4. To determine which is the most accurate forecasting method.

In order to further facilitate the selection process the most common accuracy measures (descriptive and non parametric) are categorised based on the following scheme:

- Absolute
- Relative to a base (most commonly the forecast obtained by the naïve 1 method)
- Relative to another method
- Relative to the series (in that case the error could be expressed as a percentage of the actual demand, the forecast or an arithmetic, equally weighted, average of both).

The accuracy measures within the categories that are relevant to an intermittent demand context are further discussed and evaluated and specific measures are finally selected.

The accuracy measures used for generating results are the following:

- Mean signed Error (ME)
- Wilcoxon Rank Sum Statistic (RSS)
- Mean Square Forecast Error (MSE)
- Relative Geometric Root Mean Square Error (RGRMSE)
- Percentage of times Better (PB)

- Percentage of times Best (PBt).

The empirical analysis demonstrates the importance of the selection process of accuracy measures in an intermittent demand context. The following findings are also considered to be of practical importance:

- The Mean signed Error (ME), generated across series, does not particularly suffer from the scale dependence problem. The ME though, in a single series evaluation, it is particularly sensitive to the presence of extreme observations (outliers)

- Different accuracy measures can lead to different conclusions, in an intermittent demand context

- The Percentage Best measure should be preferred to the Percentage Better measure for large scale comparison exercises

- The Relative Geometric Root Mean Square Error (RGRMSE) is a very well behaved accuracy measure in an intermittent demand context.

12.4.2 Empirical analysis – validity of the theory

By simulating the forecasting performance of Croston's method, Approximation method and EWMA on real data, we are able to test the empirical validity of the theory developed in this research. The performance of a 13 period Moving Average (MA(13)) is also simulated on the real data. No theoretical results have been developed in this thesis regarding the application of MA(13) in an intermittent demand context. Nevertheless, the MA(13) is the estimator that has been used in practice to forecast demand for the SKUs covered in our empirical sample. As such the MA(13) performance can be viewed as a benchmark against which the performance of the other estimators can be compared.

The conclusions drawn from the empirical part of the thesis reflect the synthesis of our empirical findings with respect to all the accuracy measures used for generating results, and they can be summarised as follows:

- Croston's method is biased.

- The Approximation method is approximately unbiased over the range of α values 0.05 to 0.2, showing slight bias (in opposite directions) at the extremes of this range. The Approximation method is the least biased of the four estimators examined in this chapter.

- EWMA and MA(13) are biased. The biased nature of both estimators in a re-order interval context is not what was theoretically expected and this issue requires further examination.

- Where bias had been theoretically anticipated, the sign of the bias is for all methods the theoretically expected one.

- All the pair-wise categorisation rules developed in chapter 6 have been validated, with the exception of the EWMA – Croston rule in a re-order level context and when the lead time is one or three periods.

- The Approximation method performs best in the "non-smooth" demand category (p > 1.32 and/or CV^2 > 0.49 in a re-order interval context; p > 1.33 and/or CV^2 > 0.49 in a re-order level context).

- The unbiased nature of the Approximation method is reflected on the superior forecasting accuracy of this estimator when compared with the other methods considered in this experiment.

- EWMA performs better than Croston's method in a re-order interval context. When issue points only are considered the comparison results are inconclusive.

- The MA(13) compares favourably with the smoothing methods for low smoothing constant values. This estimator is also very robust to the presence of outliers.

12.4.3 Empirical analysis – utility of the theory

To assess the empirical utility of our theoretical findings, an inventory control system needs to be specified for simulation purposes. The system is of the periodic order-up-to-level nature. The periodic nature of our model can be justified theoretically, but it is also dictated by the nature of the real demand data files available for this research. To specify the model in more detail we consider:

(a) the nature of our real demand data files
(b) additional information available for each one of those files and
(c) the objectives of the simulation experiment.

The system used for simulation purposes has the following characteristics: all demand not satisfied directly from stock is backordered and met from the next scheduled replenishment quantity or from an emergency delivery on the following day; the variability of demand over lead time plus review period is estimated by using the smoothed MSE approach; the demand over lead time plus one review period is approximated by the NBD; the managerial constraints imposed on the system are:

- a specified shortage fraction per unit value short
- a specified emergency delivery fraction per unit value short
- a specified Customer Service Level (CSL).

Comparative results are generated with respect to volume differences, regarding the average number of units in stock for each of the estimators considered, and the CSL achieved. No inventory cost results are generated due to the limited information available for our empirical data sample. Two accuracy measures are employed for comparison purposes: the Percentage Best (PBt) and the Average Percentage Regret (APR).

The empirical results can be summarised as follows:

- The Approximation method should be preferred to EWMA, Croston's method and MA(13) in a service driven system but also in a cost driven system when:

  $\frac{I}{B_2} \geq 0.48$, for a cost driven system that operates under a specified shortage fraction per unit value short (B_2)

  $\frac{I}{B_3} \geq 0.24$, for a cost driven system that operates under a specified emergency delivery premium charge per unit value short (B_3)

  where I is the annual inventory holding charge.

- EWMA is found to perform better than Croston's method, as far as forecasting accuracy is concerned, in a re-order interval context. This superiority is not reflected in the periodic inventory control results.

- The MA(13) performs very well for low (0.05) smoothing constant values[2].

- EWMA is, from a periodic inventory control perspective, the least desirable estimator. This agrees with findings of other researchers (Sani and Kingsman, 1997).

- The inventory control performance of the MA(13) is, overall, very similar to that of Croston's method and this does not agree with results presented elsewhere in the academic literature (Sani and Kingsman, 1997).

[2] The smoothing constant is used to estimate the variability of the MA(13) forecast errors.

12.4.4 Practical implications

In this thesis we have demonstrated the theoretical coherence of using the average inter-demand interval and the squared coefficient of variation of the demand sizes for categorisation purposes. The recommended cut-off values have been shown to be reasonably accurate, although it is important to note that these values have been calculated considering only the MSE performance of four particular estimators (EWMA, Croston's method, λ Approximation and Approximation method).

Moreover, all our conclusions, from the empirical part of the thesis, arise from the following assumptions:

- Choosing from a specific set of forecasting methods (EWMA, Croston's method, MA(13) and Approximation method)
- Use of particular forecasting error measures (ME, MSE, RGRMSE, PB, PBt)
- Use of a periodic order-up-to-level stock control system
- Use of particular managerial policies (P_2, B_2, B_3) with specific values assigned to them.

The Approximation method has been shown, empirically, to perform best. Nevertheless, we cannot claim that this superior performance will be reflected in other data samples when one or more of the above prerequisites do not hold. Moreover, even if all our assumptions are valid for the practitioner we cannot be sure that the organisation's data will reflect these assumptions as well as the data sample used for this research did in the empirical part of the thesis.

Considering the above, we recommend:

1. Re-designing categorisation systems to take account of the average inter-demand interval and squared coefficient of variation of the demand sizes (when demand occurs)

2. Setting the cut-off points in accordance with the recommendations made in this thesis (although we do accept that they may not be exact if other estimators are introduced and/or other accuracy measures are used)

3. Simulating the performance of the Approximation method in the “intermittent”, “erratic” and “lumpy” demand quadrants and then using the Approximation method in practice if it outperforms the other candidate estimator(s) in terms of the accuracy measure(s) used and the accepted service/cost criteria.

12.5 Further research

In this section we summarise the limitations of the thesis and, where appropriate, we suggest avenues for further research.

12.5.1 Theoretical work

Our thesis is built around the model proposed by Croston (1972). Croston’s method has been claimed to be of great value to manufacturers forecasting intermittent demand. Nevertheless when the method is tested on real demand data it shows very modest benefits. This thesis focuses on correcting and improving Croston’s approach. For this to be done, Croston’s method and not the model, upon which the method was developed, is examined in detail. No attempts have been made in this thesis to experiment with different assumptions, other than those considered by Croston (1972).

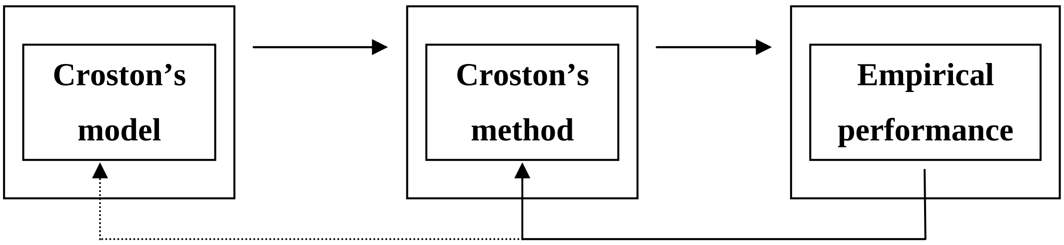

Figure 12.3. The research focus

In that respect, an interesting avenue for further research may be the performance of Croston’s method (or any other intermittent demand estimator) assuming a Poisson or condensed Poisson rather than Bernoulli demand arrival process. Some work has been conducted in this area by Johnston and Boylan (1996). Further research, though, could prove useful in explaining some of the results in chapter 6 of the thesis that were not in accordance with results presented in the 1996 paper.

Croston assumed a Stationary Mean Model (SMM) but he proposed updating the size and interval estimates using EWMA. The estimator under concern is unbiased for the SMM but it is not the optimal predictor. This inconsistency between Croston's model and Croston's method leads to two lines of investigation:

1. Examination of intermittent demand estimators that employ exponential smoothing (or even EWMA itself) under the Steady State Model (SSM) rather than SMM assumption
2. Examination of the theoretical properties of averaging methods (estimators that do not discount older data) that would be theoretically optimal for Croston's model.

The issue of modelling for intermittent demand is further discussed in the following sub-section.

In this thesis a theoretical distribution of demand per unit time period/lead time had to be selected in order to generate empirical inventory control results. The distribution chosen for that purpose was the Negative Binomial Distribution (NBD) (see chapter 9). The NBD though, implies a Poisson arrival stream while all our theoretical derivations have been based on the assumption of a Bernoulli demand arrival process.

With demand occurring as a Bernoulli process (i.e. the demand incidences follow the binomial distribution) and an arbitrary distribution of the demand sizes, the resulting distribution of total demand over a fixed lead time is compound binomial. When the order sizes are assumed to follow the Logarithmic-Poisson distribution (which is not the same with the Poisson-Logarithmic process that raises a NBD) then the resulting distribution of total demand per period is the log-zero-Poisson (lzP). No other compound binomial distributions have appeared in the academic literature for inventory control purposes. Considering that the lzP is unlikely to be used by practitioners, the next step is the mathematical (or even intuitive) justification of another compound Bernoulli distribution.

In the empirical part of our work forecasting and inventory control results were generated for a 13 period Moving Average (MA(13)) estimator. This particular estimation procedure has been used in practice to forecast demand requirements for the SKUs covered in our empirical data sample. Nevertheless, no theoretical results have been developed in this thesis, nor in the academic literature, regarding the performance of the moving average estimator in an intermittent demand context. Considering the widespread usage of moving average estimators to deal with intermittence (see for example Sani, 1995) this is obviously a very interesting area for further research.

In chapter 6 the validity of our theoretical derivations was checked by means of simulation on theoretically generated data. The demand data were developed based on the same assumptions considered in the theoretical part of the thesis. The approximate bias expressions developed in this thesis were found to be accurate in a re-order interval context. This was not the case when results were generated on issue points only. The theoretical bias of intermittent demand estimators in a re-order level context is an issue that requires further examination.

The pair-wise categorisation rules were developed in chapter 6 in a way that one estimation procedure always (theoretically) performs better in the “non-smooth” data set[3]. The estimator selected for the “smooth” data set[4] was an approximate solution, since in the case that both criteria (p and CV^2) take a value below their corresponding cut-off point, no estimator can be shown, theoretically, to perform better in all cases.

The overall rules were constructed by synthesising the pair-wise rules. Theoretically there are no doubts as to which estimator performs best in the “non-smooth” demand category but there is still uncertainty governing the area formed when both p and

[3] This is the area that corresponds to quadrants 1, 2 and 4 in any of the categorisation schemes developed in chapter 6.
[4] This is the area that corresponds to quadrant 3 in any of the categorisation schemes developed in chapter 6.

CV^2 take a value below their specified cut-off points. The theoretically coherent delineation of the "smooth" demand quadrant in the categorisation schemes discussed in chapter 6 may be of great practical importance (see also chapter 10 and 11) and is a very promising area for further research.

Moreover, it is important to note that the cut-off values assigned to both p and CV^2 criteria, when the overall rules are considered, have been calculated taking into account only the MSE performance of four particular estimators (EWMA, Croston's method, λ Approximation and Approximation method; see also sub-section 12.4.4). If other (more) estimators and/or accuracy measures had been considered, for theoretical comparison purposes, the cut-off values would not be necessarily the same.

Finally, we should also mention that in this thesis we approached the categorisation problem from a forecasting rather than inventory control point of view. Categorisation schemes can potentially also be constructed by considering alternative inventory control models rather than forecasting methods.

12.5.2 Intermittent demand models

This thesis has been postulated on Croston's model, with the exception of the normality assumption for representing the demand sizes. The assumptions upon which the model is based have not been tested on real data. The purpose of this thesis is to research the estimator used to forecast intermittent demand requirements rather than the hypothesised model. The empirical evidence, such as it is, is inconclusive regarding Croston's assumptions. Consequently, we cannot claim that the findings of this thesis apply to all intermittent demand situations.

The normality assumption is the only one that does not affect the method's performance and consequently is the only one that has been modified in this thesis for simulation purposes (on theoretically generated data). The lognormal distribution was decided, in chapter 7, to the most appropriate for representing the size of demand (in the demand occurring periods), in terms of its flexibility and the empirical evidence that exists in its support. Nevertheless, the lognormal is a continuous distribution and therefore can be only an approximation to the true demand size. Empirical experimentation with discrete demand size distributions is an interesting avenue for further research and it would contribute significantly in the process of specifying more realistic intermittent demand models.

Croston's model is a model representing the demand per unit time period. The model refers to the size of demand in the demand occurring periods rather than the size of the order in each individual transaction. In practice the demand size may be the cumulative order size in the corresponding period or the number of units required in one single transaction. Croston's model does not enable us to see the "real time" interval between two consecutive transactions since time is treated as a discrete variable. In those respects, a model of the demand generation process (see below) is a more realistic representation of the process by which orders arrive on a stockist.

The Size-Interval model proposed by Johnston and Boylan (1996) is a model for representing the actual demand generation process. The model considers the order sizes and the real time (continuous) intervals between consecutive transactions. The demand arrival stream is modelled as Poisson and it follows that the inter-order intervals are negative exponentially distributed. The demand per unit time period is now modelled as the sum of a stochastic number of order sizes. The Size-Interval model is more useful than that proposed by Croston in terms of understanding how demand is actually generated.

The Bernoulli process can be seen as the discrete time variant of the Poisson process. For small time units the Bernoulli process is an approximation to the Poisson (see also sub-section 9.5.5). Nevertheless, in the case of our research, demand has been recorded monthly and as such we cannot claim that this approximation is necessarily valid. Theory and empirical evidence suggest that the two processes are indistinguishable for very low probabilities of demand occurrence (Poisson arrival rates). Therefore highly intermittent demands can be modelled equally well by the Bernoulli or the Poisson stream. The average inter-demand interval for the series contained in our empirical data sample never exceeds two periods. Therefore, the difference in goodness-of-fit may be more marked.

For less sporadic demand SKUs (like those considered for the purposes of our research) the empirical evidence suggests that both assumptions are realistic (see sub-section 7.3.1). Therefore, we would expect some of the series used in our empirical experiment to be best fitted by the Poisson stream and some others by the Bernoulli model of demand occurrence. All our theoretical claims have been validated by means of experimentation on the empirical data and in that respect our theoretical results appear to be fairly robust to the underlying demand process, if time is treated, for forecasting purposes, as a discrete variable.

That is, if the transactions, rather than the demand, history was available for the SKUs considered in the thesis, the Poisson assumption could be tested and, intuitively at least, it would be found to fit very well some of the series. The theoretical results developed in chapters 4, 5 and 6 were based on the Bernoulli assumption and they have been validated on the empirical data. Therefore, the theoretical results appear to be reasonable approximations even for Poisson processes, which though are modelled in practice based on Croston's suggestions.

"*Models can be thought of as linking some underlying principles and the observations from the real world*. *By organising experience and data, a model provides a clear way of viewing the world and creates a basis for decision making*. Johnston et al (1986), p. 229" It is both the realism in representing the real world and the objective of the modelling process that need to be taken into account when alternative models are evaluated. As far as the first issue is concerned, the Size-Interval model is certainly more realistic. But the purpose of modelling in the context of our research is to support a forecasting mechanism. In that respect, if organisations tend to record data in specific time "buckets" (monthly in the case of this research) and/or if time intervals (inter-demand intervals in our case) are recorded as discrete numbers then a hypothesised Bernoulli model would make more sense, from a practical point of view.

If the underlying process is Poisson and forecasting takes place considering the order sizes and the actual (real time) intervals the theoretical results developed in this thesis cannot be expected to be accurate. The variability of the estimates is clearly different under the Poisson assumption and considering the order sizes rather than the demand sizes. The categorisation rules should also be different. This provides a potential explanation for the inconsistency between some of the results developed in chapter 6 of the thesis and simulation findings presented in Johnston and Boylan (1996).

This entire thesis could have been postulated on the Poisson rather than Bernoulli assumption. The bias of EWMA given the Size-Interval model can be shown to be the same as that implicitly incorporated in the method's estimates when Croston's model is used, replacing the probability of demand occurrence by the Poisson demand arrival rate. The bias of Croston's method under the Size-Interval model can be found following the approach developed in chapter 4 of the thesis accounting for a negative exponential rather than geometric distribution of the inter-order intervals. Theoretically informed corrections can then be introduced (similar to the ones used by the λ Approximation and Approximation method) in order to produce unbiased estimates of the mean demand level.

The variability of the EWMA estimates under the Poisson assumption can be found taking into account the variance of the inter-order intervals. The variability of the Size-Interval method's estimates remains as suggested in Johnston and Boylan (op. cit.) since they calculated the variance of a stochastic sum of random variables (using Clark's equations, 1957) rather than that of the ratio of two independent random variables as in the case of Croston's paper.

Subsequently, the analysis conducted in chapter 6 would be performed for the theoretically expected MSEs of the alternative estimators and rules, similar to those proposed in chapter 6, would be developed to suggest which estimator should be used, under which circumstances.

Our analysis could have also been conducted for the moving average rather than EWMA estimator either applied as a forecasting method on its own or in conjunction with Croston's approach. The adjustment required in that case is the replacement of the $(2-\alpha)/\alpha$ latest observations that are effectively considered by EWMA in updating the level (under the stationary mean model assumption, see Brown, 1963) by N, where N is the length of the moving average.

12.5.3 Empirical work

Examination (or further examination) of the following issues may be of great benefit to practitioners dealing with intermittent demand series.

- The performance of the Approximation method (or any method that builds demand estimates from constituent elements) in conjunction with a relevant inventory control method, i.e. one that considers explicitly the demand sizes and inter-demand intervals (see for example Dunsmuir and Snyder, 1989).

- The inconsistency between the theoretically unbiased nature of EWMA and Moving Average method and their biased performance in practice in an intermittent demand re-order interval context.

- The relationship between the desired and achieved CSL in an intermittent demand context, for any of the estimators discussed in the empirical part of the thesis.

No attempts have been made in this thesis to assess the empirical validity of the assumptions upon which our theory was developed. Some work has been done in this area (see for example Willemain et al, 1994) but clearly more empirical evidence is required regarding the validity of Croston's model.

Our empirical data sample consists of the demand histories for 3,000 SKUs. The range of the squared coefficient of variation values is very wide. The demand files are well-suited to the testing of the categorisation results developed in this thesis since each of the categories are well represented in the sample. However, the sample does not include highly intermittent demand items and therefore our results could not be assessed for such data.

Moreover, because of lack of the unit cost information we have not been able to fully assess the empirical utility of our theoretical findings (i.e. generate inventory cost results). Performance differences have been identified between the estimators considered in the empirical part of the thesis but more information is required to express those differences in cost terms.

In the next stage of research it is intended to

- amend the categorisation rules developed in chapter 6 and test them on real data

- assess the performance of the Approximation method in more empirical demand data series (considering assumptions other than those specified in the empirical part of the thesis)

- identify performance differences between the estimators considered in the empirical part of the thesis in cost terms.

Bibliography[1]

Adam, N.R. and Dogramaci, A. (1979) Statistical and analytical models in simulation (chapter 13). In: *Current issues in computer simulation*, Adam, N.R. and Dogramaci, A. (eds). Academic Press, New York.

Adelson, R.M. (1966) Compound Poisson distributions. *Opl Res Q*, **17**, 73-75.

Albright, S.C., Winston, W.L. and Zappe, C. (1999) *Data analysis and decision making with Microsoft Excel*. Duxbury Press, Pacific Grove, California.

Armstrong, J.S. (1978) Forecasting with econometric methods: folklore versus fact. *Journal of Business*, **51**, 549-562.

Armstrong, J.S. and Collopy, F. (1992) Error measures for generalising about forecasting methods: empirical comparisons, *Int J Forecasting*, **8**, 69-80.

Atkinson, A.C. and Pearce, M.C. (1976) The computer generation of Beta, Gamma and Normal random variables. *J R Statist Soc* A, **139** (part 4), 431-460.

Bagchi, U., Hayya, J.C. and Ord, J.K. (1983) The Hermite distribution as a model of demand during lead time for slow-moving items. *Decision Sciences*, **14**, 447-466.

Bagchi, U. (1987) Modelling lead-time demand for lumpy demand and variable lead time. *Naval Res Logist Q*, **34**, 687-704.

Bakir, M.A. (1994) *A hybrid analytic/simulation modelling approach to production planning*. Unpublished Ph.D. thesis, University of Nottingham, UK.

Bartezzaghi, E. and Verganti, R. (1995) Managing demand uncertainty through order overplanning. *Int J Production Economics*, **40**, 107-120.

Bartezzaghi, E., Verganti, R. and Zotteri, G. (1996) *A framework for managing uncertain lumpy demand*. Presented at the 9th International Symposium on Inventories, August (Budapest, Hungary).

Boonin E. (1996) *Using Excel Visual Basic for applications*. Que Corporation, Indianapolis.

Box, G.E.P. and Muller, M.E. (1958) A note on the generation of random normal variates. *Ann Math Stat*, **29**, 610-611.

Boylan, J.E. (1997) *The centralisation of inventory and the modelling of demand*. Unpublished Ph.D. thesis, University of Warwick, UK.

Bretschneider, S. (1986) Estimating forecast variance with exponential smoothing: some new results. *Int J Forecasting*, **2**, 349-355.

[1] The method of referencing used in this thesis follows the Harvard system.

Brown, R.G. (1959) *Statistical forecasting for inventory control*. McGraw-Hill, New York.

Brown, R.G. (1963) *Smoothing forecasting and prediction of discrete time series*. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Brown, R.G. (1967) *Decision rules for inventory management*. Dryden Press, Hinsdale, Illinois.

Brown, R.G. (1977) *Materials management systems*. John Wiley & Sons, Inc., New York.

Brown, R.G. (1982) *Advanced service parts inventory control*. Second edition, Materials Management Systems Inc., Norwich, Vermont.

Burgin, T.A. (1975) The gamma distribution and inventory control. *Opl Res Q*, **26**, 507-525.

Burgin T.A. and Wild, A.R. (1967) Stock control experience and usable theory. *Opl Res Q*, **18**, 35-52.

Business Forecast Systems, Inc. (1997) *Forecast Pro for Windows – Users manual*. Software version: Extended (XE) no. 3., Belmont, MA.

Chatfield, C. (1988) Editorial – Apples, oranges and mean square error. *Int J Forecasting*, **4**, 515-518.

Chatfield, C. (1992) A commentary on error measures. *Int J Forecasting*, **8**, 99-111.

Chatfield, C. and Goodhardt, G.J. (1973) A consumer purchasing model with Erlang inter-purchase times. *J Amer Statist Ass*, **68**, 828-835.

Clark, C.E. (1957) Mathematical analysis of an inventory case. *Opns Res*, **5**, 627-643.

Cox, D.R. (1962) *Renewal theory*. Methuen, London.

Croston, J.D. (1972) Forecasting and stock control for intermittent demands. *Opl Res Q*, **23**, 289-304.

Croston, J.D. (1974) Stock levels for slow-moving items. *Opl Res Q*, **25**, 123-130.

Dunsmuir, W.T.M and Snyder, R.D. (1989) Control of inventories with intermittent demand. *Eur J Opl Res*, **40**, 16-21.

Efron, B. (1979) Bootstrap methods: Another look at the Jackknife. *Annals of Statistics*, **7**, 1-26.

Efron, B. and Tibshirani, R. (1986) Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, **1**, 54-77.

Ehrhardt, R. (1979) The power approximation for computing (s, S) inventory policies. *Mgmt Sci*, **25**, 777-786.

Ehrhardt, R. and Mosier, C. (1984) A revision of the power approximation for computing (s, S) inventory policies. *Mgmt Sci*, **30**, 618-622.

Evans, M., Hastings, N. and Peacock, B. (1993) *Statistical Distributions*. Second edition, John Wiley & Sons, Inc., New York.

Fildes, R. (1983) A posteriori opinions of a Bayesian forecaster. *Journal of Forecasting*, **2**, 289-291.

Fildes, R. and Beard C. (1991) Production and inventory control. *IJOPM*, **12**, 4-27.

Fildes, R. (1992) The evaluation of extrapolative forecasting methods. *Int J Forecasting*, **8**, 81-98.

Fisher, M.L., Hammond, J.H., Obermeyer, W.R. and Raman, A. (1994) Making supply meet demand in an uncertain world. *HBR*, May-June, 83-93.

Fisher, M.L. and Raman, A. (1996) Reducing the cost of demand uncertainty through accurate response to early sales. *Opns Res*, **44**, 87-99.

Friend, J.K. (1960) Stock control with random opportunities for replenishment. *Opl Res Q*, **11**, 130-136.

Gallagher, D.J. (1969) Two periodic review inventory models with backorders and stuttering Poisson demands. *AIIE Transactions*, **1**, 164-171.

Gardner, E.S. (1983) The trade-offs in choosing a time series method. *Journal of Forecasting*, **2**, 263-267.

Gelders, L.F. and Van Looy, P.M. (1978) An inventory policy for slow and fast movers in a petrochemical plant: a case study. *J Opl Res Soc*, **29**, 867-874.

Gilchrist, W. (1976) *Statistical forecasting*. John Wiley & Sons, Inc., London.

Goodwin, P. and Lawton, R. (1999) On the asymmetry of the symmetric MAPE. *Int J Forecasting*, **15**, 405-408.

Grimmett, G.R. and Stirzaker, D.R (1992) *Probability and random process*es. Second edition, Oxford University Press, Oxford.

Hadley, G. and Whitin, T.M. (1963) *Analysis of inventory systems*. Prentice-Hall, Inc, Englewood Cliffs, N. J.

Harnett, D.L. and Soni, A.K. (1991) *Statistical methods for business and economics*. Fourth edition, Addison Wesley, New York.

Harrison, P.J. (1967) Exponential smoothing and short term sales forecasting. *Mgmt Sci*, **13**, 821-842.

Heyvaert, A.C. and Hurt, A. (1956) Inventory management of slow moving parts. *Opns Res*, **4**, 572-580.

Hollier, R.H. (1980) The distribution of spare parts. *Int J Prod Res*, **18**, 665-675.

Hoover, S.V. and Perry, R.F. (1989) *Simulation: a problem solving approach*. Addison-Wesley, New York.

Hughes, M.C. (2001) Forecasting practice: organisational issues. *J Opl Res Soc*, **52**, 143-149.

Ignall, E. and Kolesar, P. (1979) On using simulation to extend OR/MS theory: The symbiosis of simulation and analysis (chapter 15). In: *Current issues in computer simulation*, Adam, N.R. and Dogramaci, A. (eds). Academic Press, New York.

Iyer, A.V. and Schrage, L.E. (1992) Analysis of the deterministic (s, S) inventory problem. *Mgmt Sci*, **38**, 1299-1313.

Janssen, F.B.S.L.P. (1998) *Inventory management systems; control and information issues*. Published Ph.D. thesis, Centre for Economic Research, Tilburg University, The Netherlands.

Janssen, F.B.S.L.P., Heuts, R.M.J. and de Kok, T. (1998) On the (R, s, Q) inventory model when demand is modelled as a compound Bernoulli process. *Eur J Opl Res*, **104**, 423-436.

Johnston, F.R. (1973a) *Two approximations for inverting an incomplete gamma function and the incomplete standard normal*. Working paper no. 11, School of Industrial and Business Studies, University of Warwick.

Johnston, F.R. (1973b) *A note on the expected lost sales function*. Working paper no. 12, School of Industrial and Business Studies, University of Warwick.

Johnston, F.R. (1975) *Forecasting in a real time stock control system*. Working paper no. 24, School of Industrial and Business Studies, University of Warwick.

Johnston, F.R. (1980) An interactive stock control system with a strategic management role. *J Opl Res Soc*, **31**, 1069-1084.

Johnston, F.R. and Harrison, P.J. (1986) The variance of lead-time demand. *J Opl Res Soc*, **37**, 303-308.

Johnston, F.R., Harrison P.J., Marshall, A.S. and France, K.M. (1986) Modelling and the estimation of changing relationships. *The Statistician*, **35**, 229-235.

Johnston, F.R. (1993) Exponentially Weighted Moving Average (EWMA) with irregular updating periods. *J Opl Res Soc*, **44**, 711-716.

Johnston, F.R. and Boylan, J.E. (1994) How far ahead can an EWMA model be extrapolated? *J Opl Res Soc*, **45**, 710-713.

Johnston, F.R. and Boylan, J.E. (1996) Forecasting for items with intermittent demand. *J Opl Res Soc*, **47**, 113-121.

Johnston, F.R., Boylan, J.E., Shale, E. and Meadows M. (1999a) A robust forecasting system based on the combination of two simple moving averages. *J Opl Res Soc*, **50**, 1199-1204.

Johnston, F.R., Boylan, J.E., Meadows M. and Shale, E. (1999b) Some properties of a simple moving average when applied to forecasting a time series. *J Opl Res Soc*, **50**, 1267-1271.

Johnston, F.R. (2000) Viewpoint. Lead time demand adjustment or when is a model not a model. *J Opl Res Soc*, **51**, 1107-1110.

Kahn, K.B. (1998-1999) Benchmarking sales forecasting performance measures. *Journal of Business Forecasting*, 17, 19-23.

Kemp, C.D. and Kemp, A.W. (1965) Some properties of the Hermite distribution. *Biometrica*, **52**, 381-394.

Kerke, S., Morton, T.E. and Smunt, T.L. (1990) Forecasting using partially known demands. *Int J Forecasting*, **6**, 115-125.

Kleijnen, J. and Van Groenendaal, W. (1992) *Simulation. A statistical perspective*. John Wiley & Sons, Inc., Chichester.

Kogetsidis, H. and Mathews B. (1998) *Forecasting practices in British industry: An empirical assessment*. Presented at the 18th International Symposium on Forecasting, June 10-13 (Edinburgh, Scotland).

Kurawarwala, A.A. and Matsuo, H. (1996) Forecasting and inventory management of short life-cycle products. *Opns Res*, **44**, 131-150.

Kwan, H.W. (1991) *On the demand distributions of slow moving items*. Unpublished Ph.D. thesis, Lancaster University, UK.

Law, A.M. and Kelton, W.D. (2000) *Simulation modelling and analysis*. Third edition, McGraw-Hill, New York.

Makridakis, S. and Hibon, M. (1979) Accuracy of forecasting: An empirical investigation (with discussion). *J R Statist Soc* (A), **142** part 2, 97-145.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982) The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, **1**, 111-153.

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

Makridakis, S. (1993) Accuracy measures: theoretical and practical concerns. *Int J Forecasting*, **9**, 527-529.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K. and Simmons, L.F (1993) The M2-Competiton: a real time judgementally based forecasting study. *Int J Forecasting*, **9**, 5-22.

Makridakis, S. and Hibon, M. (1995) *Evaluating accuracy (or error) measures*. Working paper 95/18/TM, INSEAD.

Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (1998) *Forecasting. Methods and Applications*. Third edition, John Wiley & Sons, Inc., New York.

Makridakis, S. and Hibon, M. (2000) The M3-Competition: results, conclusions and implications. *Int J Forecasting*, **16**, 451-476.

Marsaglia, G. and Bray, T.A. (1964) A convenient method for generating normal variables. *SIAM Review*, **6**, 260-264.

Marucheck, A.S. and McClelland M.K. (1986) Strategic issues in Make-To-Order manufacturing. *APICS*, **27**, 82-95.

Meier, R.C., Newell, W.T. and Pazer, H.L. (1969) *Simulation in business and economics*. Prentice-Hall, Inc., Englewood Cliffs, N. J.

Mitchell, G.H. (1962) Problems of controlling slow-moving engineering spares. *Opl Res Q*, **13**, 23-39.

Naddor, E. (1975) Optimal and heuristic decisions in single and multi-item inventory systems. *Mgmt Sci*, **21**, 1234-1249.

Nahmias, S. and Demmy, W.S. (1982) The logarithmic Poisson gamma distribution: A model for lead time demand. *Naval Res Logist Q*, **29**, 667-677.

Newbold, P. (1983) The competition to end all competitions. *Journal of Forecasting*, **2**, 276-279.

Pearson, R.L. and Wallace, W.A. (1999) Appropriate measures of forecast error for business decision making. Presented at 19th International Symposium on Forecasting, June 27-30 (Washington, DC, USA).

Peterson, R. and Silver, E.A. (1979) *Decision systems for inventory management and production planning*. First edition, John Wiley & Sons, Inc., New York.

Popper, K.R. (1972) *The logic of scientific discovery*. Hutchinson, London.

Porteus, E.L. (1985) Numerical comparisons of inventory policies for periodic review systems. *Opns Res*, **33**, 134-152.

Prichard, J.W. and Eagle, R.H. (1965) *Modern inventory management*. John Wiley & Sons, Inc., New York.

Presutti, V.J. and Trepp, R.C. (1970) More ado about EOQ. *Naval Res Logist Q*, **17**, 243-251.

Quenouille, M.H. (1949) A relation between the logarithmic, Poisson and negative binomial series. *Biometrics,* **5**, 162-164.

Rao, A.V. (1973) A comment on: Forecasting and stock control for intermittent demands. *Opl Res Q*, **24**, 639-640.

Ritchie, E. and Kingsman, B.G. (1985) Setting stock levels for wholesaling: Performance measures and conflict of objectives between supplier and stockist. *Eur J Opl Res,* **20**, 17-24.

Sanders, N.R. and Manrodt, K.B. (1994) Forecasting practices in US corporations: Survey results. *INTERFACES*, **24**, 92-100.

Sani, B. (1995) *Periodic inventory control systems and demand forecasting methods for low demand items*. Unpublished Ph.D. thesis, Lancaster University, UK.

Sani, B. and Kingsman, B.G. (1997) Selecting the best periodic inventory control and demand forecasting methods for low demand items. *J Opl Res Soc*, **48**, 700-713.

Saunders, M., Lewis, P. and Thornhill, A. (2000) *Research methods for business students*. Second edition, Pitman Publishing, London.

Schultz, C.R. (1987) Forecasting and inventory control for sporadic demand under periodic review. *J Opl Res Soc*, **38**, 453-458.

Segerstedt, A. (1994) Inventory control with variation in lead times, especially when demand is intermittent. *Int J Production Economics*, **35**, 365-372.

Shanthikumar, J.G. and Sargent, R.G. (1983) A unifying view of hybrid simulation/analytic models and modelling. *Opns Res*, **31**, 1030-1052.

Silver, E.A. (1970) Some ideas related to the inventory control of items having erratic demand patterns. *CORS Journal*, **8**, 87-100.

Silver, E.A. and Peterson, R. (1985) *Decision systems for inventory management and production planning*. Second edition, John Wiley & Sons, Inc., New York.

Silver, E.A., Pyke, D.F. and Peterson, R. (1998) *Inventory management and production planning and scheduling*. Third edition, John Wiley & Sons, Inc., New York.

Slack, N., Chambers, S., Harland, C., Harrison, A. and Johnston, R. (1998) *Operations Management*. Second edition, Pitman Publishing, London.

Slovic, P. (1972) Psychological study of human judgement: implications for investment decision making. *Journal of Finance*, **27**, 779-799.

Smith, B.E. and Vemuganti, R.R. (1969) A learning model for inventory of slow-moving items. *AIIE Transactions*, **1**, 274-277.

Snyder, R.D. (1984) Inventory control with the gamma probability distribution. *Eur J Opl Res,* **17**, 373-381.

Snyder, R.D. (1999) *Forecasting sales of slow and fast moving inventories*. Working paper 7/99, Department of Econometrics and Business Statistics, Monash University.

Snyder, R.D., Koehler, A.B. and Ord, J.K. (1999a) *Forecasting for inventory control with exponential smoothing*. Working paper 10/99, Department of Econometrics and Business Statistics, Monash University.

Snyder, R.D., Koehler, A.B. and Ord, J.K. (1999b) Lead time demand for simple exponential smoothing: an adjustment factor for the standard deviation. *J Opl Res Soc,* **50**, 1079-1082.

Snyder, R.D., Koehler, A.B. and Ord, J.K. (2000) *Forecasting for inventory control with exponential smoothing*. Presented at the 20th International Symposium on Forecasting, June 21-24 (Lisbon, Portugal).

Stekler, H.O. (1991) Macroeconomic forecast evaluation techniques. *Int J Forecasting*, **7**, 375-384.

Strijbosch, L.W.G. and Heuts, R.M.J. (1992) Modelling (s, Q) inventory systems: Parametric versus non-parametric approximations for the lead time demand distribution. *Eur J Opl Res*, **63**, 86-101.

Strijbosch, L.W.G., Heuts, R.M.J. and van der Schoot, E.H.M (2000) A combined forecast-inventory control procedure for spare parts. *J Opl Res Soc*, **51**, 1184-1192.

Stuart, A. and Ord J. K. (1994) *Kendall's Advanced Theory of Statistics* (Volume 1, Distribution Theory). Sixth edition, Edward Arnold, London.

Syntetos, A.A., Boylan, J.E. and Croston, J.D. (2000) *Developments in forecasting intermittent demand*. Presented at 20th International Symposium on Forecasting, June 21-24 (Lisbon, Portugal).

Syntetos, A.A. and Boylan, J.E. (2001) On the bias of intermittent demand estimates. *Int J Production Economics*, **71**, 457-466.

Tavares, L.V. and Almeida, L.T. (1983) A binary decision model for the stock control of very slow moving items. *J Opl Res Soc*, **34**, 249-252.

Taylor, C.J. (1961) The application of the negative binomial distribution to stock control problems. *Opl Res Q*, **12**, 81-88.

Veinott, A.F. and Wagner, H.M (1965) Computing optimal (s, S) inventory policies. *Mgmt Sci*, **11**, 525-552.

Verganti, R. (1996) *Order Overplanning with uncertain lumpy demand: A simplified theory*. Quaderni, Dipartimento di Economia e Produzione (DEP), no.1, Politecnico di Milano.

Vereecke, A. and Verstraeten, P. (1994) An inventory management model for an inventory consisting of lumpy items, slow movers and fast movers. *Int J Production Economics*, **35**, 379-389.

Wagner, H.M. (1975) *Principles of management science (with applications to executive decisions)*. Prentice-Hall, Inc., Englewood Cliffs, N. J.

Ward, J.B. (1978) Determining re-order points when demand is lumpy. *Mgmt Sci*, **24**, 623-632.

Watson, R.B. (1987) The effects of demand-forecast fluctuations on customer service and inventory cost when demand is lumpy. *J Opl Res Soc*, **38**, 75-82.

Watson, H.J. and Blackstone, J.H. (1989) *Computer simulation*. Second edition, John Wiley & Sons, Inc. Chichester.

Wilcoxon, F. (1947) Probability tables for individual comparisons by ranking methods. *Biometrics*, **3**, 119-122.

Wilcoxon, F. (1949) *Some rapid approximate statistical procedures*. American Cyanamid Company, New York.

Willemain, T.R., Smart, C.N., Shockor, J.H. and DeSautels, P.A. (1994) Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *Int J Forecasting*, **10**, 529-538.

Willemain, T.R., Smart, C.N and Schwarz, H.F. (2000) *Forecasts of intermittent demand using the bootstrap*. Presented at the 42nd conference of the Operational Research Society, September 12-14 (University of Wales, Swansea, UK).

Williams, T.M. (1982) Reorder levels for lumpy demand. *J Opl Res Soc*, **33**, 185-189.

Williams, T.M. (1983) Table of stock-outs with lumpy demand. *J Opl Res Soc*, **34**, 431-435.

Williams, T.M. (1984) Stock control with sporadic and slow-moving demand. *J Opl Res Soc*, **35**, 939-948.

APPENDICES

APPENDIX 4.A

The expectation of the mean demand estimate for the Approximation method

$$\mathrm{E}(Y'_t) = \mathrm{E}\left(\left(1-\frac{\alpha}{2}\right)\frac{z'_t}{p'_t}\right)$$

$$\approx \left(1-\frac{\alpha}{2}\right)\frac{\mu}{p} + \frac{2-\alpha}{2}\frac{\alpha}{2-\alpha}\mu\left(\frac{p-1}{p^2}\right)$$

$$= \left(1-\frac{\alpha}{2}\right)\frac{\mu}{p} + \frac{2-\alpha}{2}\frac{\alpha}{2-\alpha}\mu\left(\frac{1}{p}-\frac{1}{p^2}\right)$$

$$= \left(1-\frac{\alpha}{2}\right)\frac{\mu}{p} + \frac{\alpha}{2}\frac{\mu}{p} - \frac{\alpha}{2}\frac{\mu}{p^2}$$

$$= \frac{\mu}{p} - \frac{\alpha}{2}\frac{\mu}{p} + \frac{\alpha}{2}\frac{\mu}{p} - \frac{\alpha}{2}\frac{\mu}{p^2}$$

$$= \frac{\mu}{p} - \frac{\alpha}{2}\frac{\mu}{p^2}$$

This proves the result given by equation (4.38).

APPENDIX 5.A

A correct approximation to the variance of Croston's estimates

We apply Taylor's theorem to a function of two variables, $g(x)$

where:

x is the vector: $x = (x_1, x_2)$ and

$g(x) = g(x_1, x_2) = \frac{x_1}{x_2}$

with $\mathrm{E}(x_1) = \theta_1$ and $\mathrm{E}(x_2) = \theta_2$.

The vector θ is defined as: $\theta = (\theta_1, \theta_2)$, with $g(\theta) = g(\theta_1, \theta_2) = \frac{\theta_1}{\theta_2}$

$$g(x) = g(\theta) + \left[\frac{\partial g}{\partial \theta_1}(x_1 - \theta_1) + \frac{\partial g}{\partial \theta_2}(x_2 - \theta_2) \right]$$

$$+ \frac{1}{2} \left[\frac{\partial^2 g}{\partial \theta_1^{\,2}}(x_1 - \theta_1)^2 + 2 \frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1 - \theta_1)(x_2 - \theta_2) + \frac{\partial^2 g}{\partial \theta_2^{\,2}}(x_2 - \theta_2)^2 \right] + \ldots\ldots\ldots \quad (5.A.1)$$

where $g(\theta) = \frac{\theta_1}{\theta_2}$ is just the first term in the Taylor series and not the population expected value.

$$\frac{\partial g}{\partial \theta_1} = \frac{1}{\theta_2} \quad (5.A.2)$$

$$\frac{\partial g}{\partial \theta_2} = -\frac{\theta_1}{\theta_2^{\,2}} \quad (5.A.3)$$

$$\frac{\partial^2 g}{\partial \theta_1^{\,2}} = 0 \quad (5.A.4)$$

$$\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2} = -\frac{1}{\theta_2^{\ 2}} \qquad (5.A.5)$$

$$\frac{\partial^2 g}{\partial \theta_2^{\ 2}} = -\theta_1 \left(-\frac{2}{\theta_2^{\ 3}} \right) = \frac{2\theta_1}{\theta_2^{\ 3}} \qquad (5.A.6)$$

therefore, considering (5.A.4), (5.A.1) becomes:

$$g(x) = g(\theta) + \left[\frac{\partial g}{\partial \theta_1}(x_1 - \theta_1) + \frac{\partial g}{\partial \theta_2}(x_2 - \theta_2) \right]$$

$$+ \left[\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1 - \theta_1)(x_2 - \theta_2) + \frac{1}{2}\frac{\partial^2 g}{\partial \theta_2^{\ 2}}(x_2 - \theta_2)^2 \right] + \ldots\ldots\ldots\ldots \qquad (5.A.7)$$

We set:

$x_1 = z'_t$, the estimate of demand size, with $\mathrm{E}(z'_t) = \mu$

and $x_2 = p'_t$, the estimate of the inter-demand interval, with $\mathrm{E}(p'_t) = p$

so that $g(x) = Y'_t$,

It has been proven, in chapter 4, that:

$$\mathrm{E}(Y'_t) = \mathrm{E}[g(x)] \approx g(\theta) + \frac{1}{2}\frac{\partial^2 g}{\partial \theta_2^{\ 2}} \mathrm{E}(x_2 - \theta_2)^2$$

considering the first three terms in the Taylor series.

Therefore:

$$Var(Y'_t) = Var[g(x)] = \mathrm{E}\left[g(x) - \mathrm{E}[g(x)]\right]^2$$

$$\approx \mathrm{E}\left\{\begin{array}{l}\frac{\partial g}{\partial \theta_1}(x_1-\theta_1)+\frac{\partial g}{\partial \theta_2}(x_2-\theta_2)+\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1-\theta_1)(x_2-\theta_2)+\frac{1}{2}\frac{\partial^2 g}{\partial {\theta_2}^2}(x_2-\theta_2)^2 \\ -\frac{1}{2}\frac{\partial^2 g}{\partial {\theta_2}^2}\mathrm{E}(x_2-\theta_2)^2\end{array}\right\}^2$$

$$= \mathrm{E}\left\{\left(\frac{\partial g}{\partial \theta_1}\right)^2 (x_1-\theta_1)^2 + \left(\frac{\partial g}{\partial \theta_2}\right)^2 (x_2-\theta_2)^2 + \left(\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}\right)^2 (x_1-\theta_1)^2 (x_2-\theta_2)^2\right.$$

$$+ \frac{1}{4}\left(\frac{\partial^2 g}{\partial \theta_2}\right)^2 (x_2-\theta_2)^4 + \frac{1}{4}\left(\frac{\partial^2 g}{\partial \theta_2}\right)^2 \left[\mathrm{E}(x_2-\theta_2)^2\right]^2 + 2\frac{\partial g}{\partial \theta_1}(x_1-\theta_1)\frac{\partial g}{\partial \theta_2}(x_2-\theta_2)$$

$$+ 2\frac{\partial g}{\partial \theta_1}(x_1-\theta_1)\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1-\theta_1)(x_2-\theta_2) + \frac{\partial g}{\partial \theta_1}(x_1-\theta_1)\frac{\partial^2 g}{\partial {\theta_2}^2}(x_2-\theta_2)^2$$

$$- \frac{\partial g}{\partial \theta_1}(x_1-\theta_1)\frac{\partial^2 g}{\partial {\theta_2}^2}\mathrm{E}(x_2-\theta_2)^2 + 2\frac{\partial g}{\partial \theta_2}(x_2-\theta_2)\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1-\theta_1)(x_2-\theta_2)$$

$$+ \frac{\partial g}{\partial \theta_2}\frac{\partial^2 g}{\partial {\theta_2}^2}(x_2-\theta_2)^3 - \frac{\partial g}{\partial \theta_2}\frac{\partial^2 g}{\partial {\theta_2}^2}(x_2-\theta_2)\mathrm{E}(x_2-\theta_2)^2$$

$$+ \frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1-\theta_1)(x_2-\theta_2)\frac{\partial^2 g}{\partial {\theta_2}^2}(x_2-\theta_2)^2 - \frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1-\theta_1)(x_2-\theta_2)\frac{\partial^2 g}{\partial {\theta_2}^2}\mathrm{E}(x_2-\theta_2)^2$$

$$\left.- \frac{1}{2}\left(\frac{\partial^2 g}{\partial {\theta_2}^2}\right)^2 (x_2-\theta_2)^2 \mathrm{E}(x_2-\theta_2)^2\right\} \quad (5.A.8)$$

Assuming that x_1, x_2 are independent:

$$Var[g(x)] \approx \left(\frac{\partial g}{\partial \theta_1}\right)^2 \mathrm{E}\left(x_1-\theta_1\right)^2 + \left(\frac{\partial g}{\partial \theta_2}\right)^2 \mathrm{E}\left(x_2-\theta_2\right)^2 + \left(\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}\right)^2 \mathrm{E}\left(\left(x_1-\theta_1\right)^2\left(x_2-\theta_2\right)^2\right)$$

$$+\frac{\partial g}{\partial \theta_2}\frac{\partial^2 g}{\partial \theta_2^{\,2}}\mathrm{E}\left(x_2-\theta_2\right)^3+\frac{1}{4}\left(\frac{\partial^2 g}{\partial \theta_2^{\,2}}\right)^2 \mathrm{E}\left(x_2-\theta_2\right)^4-\frac{1}{4}\left(\frac{\partial^2 g}{\partial \theta_2^{\,2}}\right)^2\left[\mathrm{E}\left(x_2-\theta_2\right)^2\right]^2 \tag{5.A.9}$$

considering equations (5.A.2), (5.A.3), (5.A.5) and (5.A.6)

$$Var[g(x)] \approx \frac{Var(x_1)}{\theta_2^{\,2}}+\frac{\theta_1^{\,2}Var(x_2)}{\theta_2^{\,4}}+\frac{1}{\theta_2^{\,4}}Var(x_1)Var(x_2)-\frac{2\theta_1^{\,2}\mathrm{E}\left(x_2-\theta_2\right)^3}{\theta_2^{\,5}}+\frac{\theta_1^{\,2}\mathrm{E}\left(x_2-\theta_2\right)^4}{\theta_2^{\,6}}$$

$$-\frac{\theta_1^{\,2}\left[\mathrm{E}\left(x_2-\theta_2\right)^2\right]^2}{\theta_2^{\,6}}$$

$$=\frac{Var(x_1)}{\left(\mathrm{E}(x_2)\right)^2}+\frac{\left(\mathrm{E}(x_1)\right)^2 Var(x_2)}{\left(\mathrm{E}(x_2)\right)^4}+\frac{Var(x_1)Var(x_2)}{\left(\mathrm{E}(x_2)\right)^4}-\frac{2\left(\mathrm{E}(x_1)\right)^2\mathrm{E}\left(x_2-\theta_2\right)^3}{\left(\mathrm{E}(x_2)\right)^5}$$

$$+\frac{\left(\mathrm{E}(x_1)\right)^2\mathrm{E}\left(x_2-\theta_2\right)^4}{\left(\mathrm{E}(x_2)\right)^6}-\frac{\left(\mathrm{E}(x_1)\right)^2\left[Var\, x_2\right]^2}{\left(\mathrm{E}(x_2)\right)^6} \tag{5.A.10}$$

In *Appendix 5.B* it is proven that for $n = 2, 3$:

$$\mathrm{E}\left[x'_t-\mathrm{E}(x)\right]^n = \frac{\alpha^n}{1-\left(1-\alpha\right)^n}\mathrm{E}\left[x_t-\mathrm{E}(x)\right]^n \tag{5.A.11}$$

and also that:

$$\mathrm{E}\left[x'_t-\mathrm{E}(x)\right]^4 = \frac{\alpha^4}{1-\left(1-\alpha\right)^4}\mathrm{E}\left[x_t-\mathrm{E}(x)\right]^4+\frac{\alpha^4}{\left(1-\left(1-\alpha\right)^2\right)^2}\left[Var\left(x_t\right)\right]^2 \tag{5.A.12}$$

where:

x_t represents the demand size (z_t) or inter-demand interval (p_t),

x'_t is their exponentially smoothed estimate (z'_t , p'_t) and

$\mathrm{E}(x)$ is the population expected value for either series.

Consideration of (5.A.11) and (5.A.12) necessitates the adoption of the following assumptions:

- no auto-correlation for the demand size and inter-demand interval series
- homogeneous moments about the mean for both series
- same smoothing constant value is used for both series

Taking also into account that:

$Var(z_t) = \sigma^2$ and $Var(p_t) = p(p-1)$

(5.A.10) becomes:

$$Var\left(\frac{z'_t}{p'_t}\right) \approx \frac{\alpha}{2-\alpha}\frac{\sigma^2}{p^2} + \frac{\alpha}{2-\alpha}\mu^2\frac{p(p-1)}{p^4} + \left(\frac{\alpha}{2-\alpha}\right)^2\frac{\sigma^2 p(p-1)}{p^4}$$

$$-\frac{\alpha^3}{1-(1-\alpha)^3}\frac{2\mu^2}{p^5}\mathrm{E}(p_t - p)^3 + \frac{\alpha^4}{1-(1-\alpha)^4}\frac{\mu^2}{p^6}\mathrm{E}(p_t - p)^4$$

$$+\frac{\alpha^4}{\left(1-(1-\alpha)^2\right)^2}\frac{\mu^2}{p^6}p^2(p-1)^2 - \left(\frac{\alpha}{2-\alpha}\right)^2\frac{\mu^2}{p^6}p^2(p-1)^2 \qquad (5.A.13)$$

The third moment about the mean in the geometric distribution, where: $\frac{1}{p}$ is the probability of success in each trial, is calculated as:

$$\mathrm{E}\left(p_t-p\right)^3=\left(1-\frac{1}{p}\right)\left(1+1-\frac{1}{p}\right)\frac{1}{p^3}=\frac{p-1}{p}\frac{1}{p^3}\left(2-\frac{1}{p}\right)=\frac{p-1}{p}\frac{1}{p^3}\frac{2p-1}{p}$$

$$=\frac{(p-1)(2p-1)}{p^5} \tag{5.A.14}$$

and the fourth moment:

$$\mathrm{E}\left(p_t-p\right)^4=\frac{9\left(1-\frac{1}{p}\right)^2}{\frac{1}{p^4}}+\frac{1-\frac{1}{p}}{\frac{1}{p^2}}=\left(1-\frac{1}{p}\right)\left[\frac{9\left(1-\frac{1}{p}\right)}{\frac{1}{p^4}}+\frac{1}{\frac{1}{p^2}}\right]$$

$$=\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^4+p^2\right]=p^2\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2+1\right] \tag{5.A.15}$$

If we consider (5.A.14) and (5.A.15) , and also the fact that:

$\frac{\alpha^4}{\left(1-(1-\alpha)^2\right)^2}=\left(\frac{\alpha}{2-\alpha}\right)^2$, (5.A.13) becomes:

$$Var\left(\frac{z'_t}{p'_t}\right)\approx\frac{\alpha}{2-\alpha}\frac{\sigma^2}{p^2}+\frac{\alpha}{2-\alpha}\mu^2\frac{p(p-1)}{p^4}+\left(\frac{\alpha}{2-\alpha}\right)^2\frac{\sigma^2\, p(p-1)}{p^4}$$

$$-\frac{\alpha^3}{1-(1-\alpha)^3}\frac{2\mu^2(p-1)(2p-1)}{p^{10}}+\frac{\alpha^4}{1-(1-\alpha)^4}\frac{\mu^2}{p^4}\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2+1\right] \tag{5.A.16}$$

Since the fourth part of approximation (5.A.16) becomes almost zero even for quite low average inter-demand intervals, finally the variance is approximated by (5.A.17):

$$Var\left(\frac{z'_t}{p'_t}\right)\approx\frac{a}{2-a}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]+\frac{\alpha^4}{1-(1-\alpha)^4}\frac{\mu^2}{p^4}\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2+1\right] \tag{5.A.17}$$

This proves the result given by equation (5.19).

APPENDIX 5.B

The 2nd, 3rd and 4th moment about the mean for exponentially smoothed estimates

If we define:

$$x' = \sum_{j=0}^{\infty} \alpha (1-\alpha)^j x_{t-j} \text{ (i.e. the EWMA estimate)} \tag{5.B.1}$$

$$\mathrm{E}(x') = \sum_{j=0}^{\infty} \alpha (1-\alpha)^j \mathrm{E}(x_{t-j}) =$$

assuming $\mathrm{E}(x_{t-j}) = \mathrm{E}(x)$ for all $j \geq 0$

$$\alpha \mathrm{E}(x) \sum_{j=0}^{\infty} (1-\alpha)^j = \frac{\alpha}{1-(1-\alpha)} \mathrm{E}(x) = \mathrm{E}(x) \tag{5.B.2}$$

Therefore we can write:

$$x' - \mathrm{E}(x) = \sum_{j=0}^{\infty} \alpha (1-\alpha)^j \left[x_{t-j} - \mathrm{E}(x)\right]$$

$$\left[x' - \mathrm{E}(x)\right]^n = \left\{ \sum_{j=0}^{\infty} \alpha (1-\alpha)^j \left[x_{t-j} - \mathrm{E}(x)\right] \right\}^n$$

and

$$\mathrm{E}\left[x' - \mathrm{E}(x)\right]^n = \mathrm{E}\left\{ \sum_{j=0}^{\infty} \alpha (1-\alpha)^j \left[x_{t-j} - \mathrm{E}(x)\right] \right\}^n \tag{5.B.3}$$

Assuming series is not auto-correlated, for $n = 2, 3$ we then have:

$$\mathrm{E}\left[x'-\mathrm{E}(x)\right]^n = \sum_{j=0}^{\infty}\alpha^n(1-\alpha)^{nj}\,\mathrm{E}\left[x_{t-j}-\mathrm{E}(x)\right]^n \tag{5.B.4}$$

and assuming $\mathrm{E}\left[x_{t-j}-\mathrm{E}(x)\right]^n = \mathrm{E}\left[x-\mathrm{E}(x)\right]^n$ for all $j \geq 0$, i.e. homogeneous moments of order n

$$\mathrm{E}\left[x'-\mathrm{E}(x)\right]^n = \frac{\alpha^n}{1-(1-\alpha)^n}\,\mathrm{E}\left[x-\mathrm{E}(x)\right]^n \tag{5.B.5}$$

For $n = 2$

$$Var(x') = \frac{\alpha^2}{1-(1-\alpha)^2}Var(x)$$

and for $n = 3$

$$\mathrm{E}\left[x'-\mathrm{E}(x)\right]^3 = \frac{\alpha^3}{1-(1-\alpha)^3}\,\mathrm{E}\left[x-\mathrm{E}(x)\right]^3$$

For $n = 4$, equation (5.B.3) becomes:

$$\mathrm{E}\left[x'-\mathrm{E}(x)\right]^4 = \mathrm{E}\left\{\sum_{j=0}^{\infty}\alpha(1-\alpha)^j\left[x_{t-j}-\mathrm{E}(x)\right]\right\}^4$$

assuming no auto-correlation

$$= \sum_{j=0}^{\infty}\alpha^4(1-\alpha)^{4j}\,\mathrm{E}\left[x_{t-j}-\mathrm{E}(x)\right]^4$$

$$+ \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\alpha^2(1-\alpha)^{2j}\,\mathrm{E}\left[x_{t-j}-\mathrm{E}(x)\right]^2\alpha^2(1-\alpha)^{2i}\,\mathrm{E}\left[x_{t-i}-\mathrm{E}(x)\right]^2$$

assuming homogeneous moments about the mean

$$= \alpha^4 \mathrm{E}\left[x_{t-j} - \mathrm{E}(x)\right]^4 \sum_{j=0}^{\infty}(1-\alpha)^{4j} + \alpha^4\left[Var(x)\right]^2 \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}(1-\alpha)^{2j}(1-\alpha)^{2i} \qquad (5.B.6)$$

$$\sum_{j=0}^{\infty}(1-\alpha)^{4j} = \frac{1}{1-(1-\alpha)^4} \qquad (5.B.7)$$

$$\begin{aligned}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}(1-\alpha)^{2j}(1-\alpha)^{2i} = 1 &+ (1-\alpha)^2 + (1-\alpha)^4 + (1-\alpha)^6 + \ldots\ldots\ldots\ldots \\ &+ (1-\alpha)^2 + (1-\alpha)^4 + (1-\alpha)^6 + \ldots\ldots\ldots\ldots \\ &+ (1-\alpha)^4 + (1-\alpha)^6 + \ldots\ldots\ldots\ldots \\ &\cdot \\ &\cdot \\ &\cdot \\ &\cdot \end{aligned} \qquad (5.B.8)$$

$$\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}(1-\alpha)^{2j}(1-\alpha)^{2i} = 1 + 2(1-\alpha)^2 + 3(1-\alpha)^4 + 4(1-\alpha)^6 + \ldots\ldots\ldots\ldots \qquad (5.B.9)$$

If we multiply the first and the second part of equation (5.B.9) with $(1-\alpha)^2$, we then have:

$$(1-\alpha)^2 \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}(1-\alpha)^{2j}(1-\alpha)^{2i} = (1-\alpha)^2 + 2(1-\alpha)^4 + 3(1-\alpha)^6 + \ldots\ldots\ldots\ldots \qquad (5.B.10)$$

Subtracting (5.B.9) - (5.B.10)

$$\left[1-(1-\alpha)^2\right]\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}(1-\alpha)^{2j}(1-\alpha)^{2i} = 1 + (1-\alpha)^2 + (1-\alpha)^4 + (1-\alpha)^6 + \ldots\ldots\ldots\ldots = \frac{1}{1-(1-\alpha)^2}$$

Therefore:

$$\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}(1-\alpha)^{2j}(1-\alpha)^{2i} = \frac{1}{\left(1-(1-\alpha)^2\right)^2} \qquad (5.B.11)$$

Considering equations (5.B.7) and (5.B.11), equation (5.B.6) becomes:

$$\mathrm{E}\left[x' - \mathrm{E}(x)\right]^4 = \frac{\alpha^4}{1-(1-\alpha)^4}\mathrm{E}\left[x - \mathrm{E}(x)\right]^4 + \frac{\alpha^4}{\left(1-(1-\alpha)^2\right)^2}\left[Var(x)\right]^2 \tag{5.B.12}$$

This proves the results given by equations (5.A.11) and (5.A.12).

APPENDIX 5.C

The variance of the λ Approximation methods estimates

We set, for the problem under concern:

$$x_1 = \left(1 - \frac{\alpha}{2}\right) z'_t$$

with expected value:

$$\mathrm{E}(x_1) = \theta_1 = \mathrm{E}\left[\left(1 - \frac{\alpha}{2}\right) z'_t\right] = \left(1 - \frac{\alpha}{2}\right)\mathrm{E}(z'_t) = \left(1 - \frac{\alpha}{2}\right)\mu \tag{5.C.1}$$

and variance:

$$Var(x_1) = Var\left[\left(1 - \frac{\alpha}{2}\right) z'_t\right] = \left(1 - \frac{\alpha}{2}\right)^2 Var(z'_t) = \left(1 - \frac{\alpha}{2}\right)^2 \frac{\alpha}{2-\alpha} Var(z_t)$$

$$= \left(1 - \frac{\alpha}{2}\right)^2 \frac{\alpha}{2-\alpha} \sigma^2 \tag{5.C.2}$$

and

$$x_2 = p'_t - \frac{\alpha}{2}$$

with expected value:

$$\mathrm{E}(x_2) = \theta_2 = \mathrm{E}\left(p'_t - \frac{\alpha}{2}\right) = \mathrm{E}(p'_t) - \frac{\alpha}{2} = p - \frac{\alpha}{2} \tag{5.C.3}$$

and variance:

$$Var(x_2) = Var\left(p'_t - \frac{\alpha}{2} \right) = Var(p'_t) = \frac{\alpha}{2-\alpha} Var(p_t) = \frac{\alpha}{2-\alpha}\, p(p-1) \tag{5.C.4}$$

(for the variance derivations consider also *Appendix 5.B*)

The third and the fourth moments about the mean for the x_2 variable (consider also *Appendices 5.A* and *5.B*) are calculated as follows:

$$\mathrm{E}(x_2 - \theta_2)^3 = \mathrm{E}\left(p'_t - \frac{\alpha}{2} - p + \frac{\alpha}{2} \right)^3 = \mathrm{E}(p'_t - p)^3 = \frac{\alpha^3}{1-(1-\alpha)^3} \mathrm{E}(p_t - p)^3$$

$$= \frac{\alpha^3}{1-(1-\alpha)^3} \frac{(p-1)(2p-1)}{p^5} \tag{5.C.5}$$

$$\mathrm{E}(x_2 - \theta_2)^4 = \mathrm{E}\left(p'_t - \frac{\alpha}{2} - p + \frac{\alpha}{2} \right)^4 = \mathrm{E}(p'_t - p)^4$$

$$= \frac{\alpha^4}{1-(1-\alpha)^4} \mathrm{E}(p_t - p)^4 + \frac{\alpha^4}{\left(1-(1-\alpha)^2\right)^2} [Var(p_t)]^2$$

$$= \frac{\alpha^4}{1-(1-\alpha)^4}\, p^2 \left(1 - \frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2 + 1\right] + \left(\frac{\alpha}{2-\alpha}\right)^2 p^2 (p-1)^2 \tag{5.C.6}$$

(assuming that the same smoothing constant value is used for both x_1 and x_2 series and that both series are not auto-correlated and have homogeneous moments about the mean).

Consequently we apply Taylor's theorem to a function of two variables, $g(x) = \frac{x_1}{x_2}$ with:

$$\frac{\partial g}{\partial \theta_1} = \frac{1}{\theta_2} \tag{5.C.7}$$

$$\frac{\partial g}{\partial \theta_2} = -\frac{\theta_1}{\theta_2{}^2} \tag{5.C.8}$$

$$\frac{\partial^2 g}{\partial \theta_1{}^2} = 0 \tag{5.C.9}$$

$$\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2} = -\frac{1}{\theta_2{}^2} \tag{5.C.10}$$

$$\frac{\partial^2 g}{\partial \theta_2{}^2} = -\theta_1\left(-\frac{2}{\theta_2{}^3}\right) = \frac{2\theta_1}{\theta_2{}^3} \tag{5.C.11}$$

If we consider only the first three terms, we have:

$$g(x) = g(\theta) + \left[\frac{\partial g}{\partial \theta_1}(x_1 - \theta_1) + \frac{\partial g}{\partial \theta_2}(x_2 - \theta_2)\right]$$

$$+\left[\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1 - \theta_1)(x_2 - \theta_2) + \frac{1}{2}\frac{\partial^2 g}{\partial \theta_2{}^2}(x_2 - \theta_2)^2\right] + ... \tag{5.C.12}$$

with:

$$\mathrm{E}[g(x)] = \mathrm{E}\left(\frac{x_1}{x_2}\right) = \mathrm{E}\left(\frac{\left(1-\frac{\alpha}{2}\right)z'_t}{p'_t - \frac{\alpha}{2}}\right) \approx \frac{\mu}{p} \tag{5.C.13}$$

$$g(\theta) = \frac{\theta_1}{\theta_2} = \frac{\left(1-\frac{\alpha}{2}\right)\mu}{p - \frac{\alpha}{2}} \neq \frac{\mu}{p}$$

and (consider also *Appendix 5.A*)

$$Var[g(x)] = \mathrm{E}\left[g(x) - \mathrm{E}[g(x)]\right]^2$$

$$\approx \mathrm{E}\left\{\begin{matrix}\frac{\theta_1}{\theta_2} + \frac{\partial g}{\partial \theta_1}(x_1-\theta_1) + \frac{\partial g}{\partial \theta_2}(x_2-\theta_2) + \frac{\partial^2 g}{\partial \theta_1 \partial \theta_2}(x_1-\theta_1)(x_2-\theta_2) \\ + \frac{1}{2}\frac{\partial^2 g}{\partial {\theta_2}^2}(x_2-\theta_2)^2 - \mathrm{E}\left(\frac{x_1}{x_2}\right)\end{matrix}\right\}^2 \qquad (5.C.14)$$

In order to simplify somewhat the derivation of the variance of the λ Approximation method estimates and consequently facilitate the MSE comparisons (see chapter 6) we approximate $\mathrm{E}[g(x)]$ by:

$$\mathrm{E}[g(x)] \approx g(\theta) + \frac{1}{2}\frac{\partial^2 g}{\partial {\theta_2}^2}\mathrm{E}(x_2-\theta_2)^2$$

(based on (5.C.12) and considering chapter 4, sub-section 4.5.2)

Considering (5.A.10) and assuming that x_1, x_2 are independent, (5.C.14) becomes:

$$Var[g(x)] \approx \frac{Var(x_1)}{(\mathrm{E}(x_2))^2} + \frac{(\mathrm{E}(x_1))^2 Var(x_2)}{(\mathrm{E}(x_2))^4} + \frac{Var(x_1)Var(x_2)}{(\mathrm{E}(x_2))^4} - \frac{2(\mathrm{E}(x_1))^2\mathrm{E}(x_2-\theta_2)^3}{(\mathrm{E}(x_2))^5}$$

$$+ \frac{(\mathrm{E}(x_1))^2\mathrm{E}(x_2-\theta_2)^4}{(\mathrm{E}(x_2))^6} - \frac{(\mathrm{E}(x_1))^2[Var\, x_2]^2}{(\mathrm{E}(x_2))^6} \qquad (5.C.15)$$

Finally the variance of the estimates of the λ Approximation method is calculated as follows:

$$Var\left[\frac{\left(1-\frac{\alpha}{2}\right)z'_t}{p'_t-\frac{\alpha}{2}}\right] \approx \left(\frac{\alpha}{2-\alpha}\right)\frac{\left(1-\frac{\alpha}{2}\right)^2\sigma^2}{\left(p-\frac{\alpha}{2}\right)^2}+\left(\frac{\alpha}{2-\alpha}\right)\frac{\left(1-\frac{\alpha}{2}\right)^2\mu^2 p(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}$$

$$+\frac{\left(1-\frac{\alpha}{2}\right)^2\left(\frac{\alpha}{2-\alpha}\right)^2\sigma^2 p(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-\frac{\left(1-\frac{\alpha}{2}\right)^2\mu^2}{\left(p-\frac{\alpha}{2}\right)^5}\frac{\alpha^3}{1-(1-\alpha)^3}\frac{(p-1)(2p-1)}{p^5}$$

$$+\frac{\alpha^4}{1-(1-\alpha)^4}\frac{\left(1-\frac{\alpha}{2}\right)^2\mu^2}{\left(p-\frac{\alpha}{2}\right)^6}p^2\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2+1\right]$$

$$+\left(\frac{\alpha}{2-\alpha}\right)^2\frac{\left(1-\frac{\alpha}{2}\right)^2\mu^2}{\left(p-\frac{\alpha}{2}\right)^6}p^2(p-1)^2-\left(\frac{\alpha}{2-\alpha}\right)^2\frac{\left(1-\frac{\alpha}{2}\right)^2\mu^2}{\left(p-\frac{\alpha}{2}\right)^6}p^2(p-1)^2 \quad (5.C.16)$$

Since the fourth part of approximation (5.C.16) becomes almost zero even for quite low average inter-demand intervals, and the last two terms cancel each other, finally the variance is approximated by (5.C.17):

$$Var\left[\frac{\left(1-\frac{\alpha}{2}\right)z'_t}{p'_t-\frac{\alpha}{2}}\right] \approx \left(\frac{\alpha}{2-\alpha}\right)\left(1-\frac{\alpha}{2}\right)^2\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2+p(p-1)\mu^2+\frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4}$$

$$+\frac{\alpha^4}{1-(1-\alpha)^4}\frac{\left(1-\frac{\alpha}{2}\right)^2\mu^2}{\left(p-\frac{\alpha}{2}\right)^6}p^2\left(1-\frac{1}{p}\right)\left[9\left(1-\frac{1}{p}\right)p^2+1\right] \quad (5.C.17)$$

This proves the result given by equation (5.23).

APPENDIX 6.A

MSE Croston's Method – MSE Approximation

$MSE_{CROSTON} > MSE_{APPROXIMATION} \Leftrightarrow$

$$\frac{a}{2-a}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]+\left[\frac{\alpha}{2-\alpha}\mu\frac{(p-1)}{p^2}\right]^2 >$$

$$\frac{\alpha(2-\alpha)}{4}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]+\left[\frac{\alpha}{2}\frac{\mu}{p^2}\right]^2 \Leftrightarrow$$

$$\left(\frac{\alpha}{2-\alpha}\right)^2\mu^2\frac{(p-1)^2}{p^4}-\left[\frac{\alpha}{2}\frac{\mu}{p^2}\right]^2 > \frac{a}{2-a}\left[\left(1-\frac{\alpha}{2}\right)^2-1\right]\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]\Leftrightarrow$$

$$\left(\frac{\alpha}{2-\alpha}\right)^2\mu^2\frac{(p-1)^2}{p^4}-\left[\frac{\alpha}{2}\frac{\mu}{p^2}\right]^2 > \frac{a}{2-a}\left(1+\frac{\alpha^2}{4}-\alpha-1\right)\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]\Leftrightarrow$$

$$\frac{1}{p^4}\left[\left(\frac{\alpha}{2-\alpha}\right)^2\mu^2(p-1)^2-\left(\frac{\alpha}{2}\right)^2\mu^2\right]>$$

$$\frac{1}{p^4}\left\{\left(\frac{\alpha}{2-\alpha}\right)\left(\frac{\alpha^2-4\alpha}{4}\right)\left[p(p-1)\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\sigma^2p^2\right]\right\}\Leftrightarrow$$

$$\left(\frac{\alpha}{2}\frac{2\alpha}{\alpha(2-\alpha)}\right)^2\mu^2(p-1)^2-\left(\frac{\alpha}{2}\right)^2\mu^2 > \left(\frac{\alpha^2}{4}\right)\left(\frac{\alpha-4}{2-\alpha}\right)\left[p(p-1)\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\sigma^2p^2\right]\Leftrightarrow$$

$$\left(\frac{\alpha}{2}\right)^2\left[\left(\frac{2\alpha}{\alpha(2-\alpha)}\right)^2\mu^2(p-1)^2-\mu^2\right] > \left(\frac{\alpha}{2}\right)^2\left(\frac{\alpha-4}{2-\alpha}\right)\left[p(p-1)\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\sigma^2p^2\right]\Leftrightarrow$$

$$\left(\frac{2\alpha}{\alpha(2-\alpha)}\right)^2 \mu^2 (p-1)^2 - \frac{2-\alpha}{2-\alpha}\ \mu^2 - \frac{\alpha-4}{2-\alpha} p(p-1)\mu^2 > \frac{1}{2-\alpha}(\alpha-4)\left[p(p-1)\frac{\alpha}{2-\alpha}\sigma^2 + \sigma^2 p^2\right] \Leftrightarrow$$

$$\frac{1}{2-\alpha}\left[\frac{4\alpha^2}{\alpha^2(2-\alpha)}\mu^2 (p-1)^2 - (2-\alpha)\mu^2 - (\alpha-4)p(p-1)\mu^2\right] >$$

$$\frac{1}{2-\alpha}(\alpha-4)\left[p(p-1)\frac{\alpha}{2-\alpha}\sigma^2 + \sigma^2 p^2\right] \Leftrightarrow$$

$$\mu^2\left[\frac{4}{(2-\alpha)}(p-1)^2 - (2-\alpha) - (\alpha-4)p(p-1)\right] > \sigma^2(\alpha-4)\left[p(p-1)\frac{\alpha}{2-\alpha} + p^2\right] \Leftrightarrow$$

$$\mu^2\left\{(p-1)\left[\frac{4(p-1)}{(2-\alpha)} - \frac{2-\alpha}{p-1} - (\alpha-4)p\right]\right\} > \sigma^2(\alpha-4)\left(p^2\frac{\alpha}{2-\alpha} - p\frac{\alpha}{2-\alpha} + p^2\right) \Leftrightarrow$$

$$\mu^2\left\{(p-1)\left[\frac{4(p-1)}{(2-\alpha)} - \frac{2-\alpha}{p-1} - (\alpha-4)p\right]\right\} > \sigma^2(\alpha-4)\left(p^2\frac{2}{2-\alpha} - p\frac{\alpha}{2-\alpha}\right) \Leftrightarrow$$

$$\mu^2\left\{(p-1)\left[\frac{4(p-1)}{(2-\alpha)} - \frac{2-\alpha}{p-1} - (\alpha-4)p\right]\right\} > \sigma^2(\alpha-4)\frac{p}{2-\alpha}(2p-\alpha)$$

At this stage we need to consider that

$$(\alpha-4)\frac{p}{2-\alpha}(2p-\alpha) < 0 \text{ for } p>1,\ 0\le\alpha\le1$$

Therefore, $MSE_{CROSTON} > MSE_{APPROXIMATION}$ if and only if:

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[\frac{4(p-1)}{2-\alpha} - \frac{2-\alpha}{p-1} - p(\alpha-4)\right]}{\frac{p(\alpha-4)(2p-\alpha)}{2-\alpha}},\ \text{for } p>1,\ 0\le\alpha\le1.$$

This proves the inequality (6.32).

APPENDIX 6.B

MSE EWMA – MSE Approximation (issue points)

$$MSE_{EWMA} > MSE_{APPROXIMATION} \Leftrightarrow$$

$$\alpha^2\sigma^2 + \frac{\alpha\beta^2}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right] + \left[\mu\left(\alpha + \frac{\beta}{p}\right) - \frac{\mu}{p}\right]^2 >$$

$$\frac{\alpha(2-\alpha)}{4}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] + \left[\frac{\alpha}{2}\frac{\mu}{p^2}\right]^2 \Leftrightarrow$$

$$\alpha^2\sigma^2 + \frac{\alpha\beta^2}{2-\alpha}\frac{\sigma^2}{p} + \frac{\alpha\beta^2}{2-\alpha}\frac{p-1}{p^2}\mu^2 + \mu^2\left(\alpha + \frac{1-\alpha-1}{p}\right)^2 >$$

$$\mu^2\left[\left(1-\frac{\alpha}{2}\right)^2\frac{\alpha}{2-\alpha}\frac{(p-1)}{p^3} + \frac{\alpha^2}{4p^4}\right] + \sigma^2\left[\left(1-\frac{\alpha}{2}\right)^2\left(\frac{\alpha}{2-\alpha}\right)^2\frac{(p-1)}{p^3} + \left(1-\frac{\alpha}{2}\right)^2\frac{\alpha}{2-\alpha}\frac{1}{p^2}\right] \Leftrightarrow$$

$$\mu^2\left[\frac{\alpha\beta^2}{2-\alpha}\frac{p-1}{p^2} + \alpha^2\left(1-\frac{1}{p}\right)^2 - \left(1-\frac{\alpha}{2}\right)^2\frac{\alpha}{2-\alpha}\frac{(p-1)}{p^3} - \frac{\alpha^2}{4p^4}\right] >$$

$$\sigma^2\left[\left(1-\frac{\alpha}{2}\right)^2\left(\frac{\alpha}{2-\alpha}\right)^2\frac{(p-1)}{p^3} + \left(1-\frac{\alpha}{2}\right)^2\frac{\alpha}{2-\alpha}\frac{1}{p^2} - \alpha^2 - \frac{\alpha(1-\alpha)^2}{2-\alpha}\frac{1}{p}\right] \Leftrightarrow$$

$$\mu^2\left[\frac{\alpha(1-\alpha)^2}{2-\alpha}\frac{p-1}{p^2} + \alpha^2\frac{(p-1)^2}{p^2} - \frac{(2-\alpha)^2}{4}\frac{\alpha}{2-\alpha}\frac{(p-1)}{p^3} - \frac{\alpha^2}{4p^4}\right] >$$

$$\sigma^2\left[\frac{(2-\alpha)^2}{4}\frac{\alpha^2}{(2-\alpha)^2}\frac{(p-1)}{p^3} + \frac{(2-\alpha)^2}{4}\frac{\alpha}{2-\alpha}\frac{1}{p^2} - \alpha^2 - \frac{\alpha(1-\alpha)^2}{2-\alpha}\frac{1}{p}\right] \Leftrightarrow$$

$$\frac{1}{p^2}\mu^2\left[\frac{\alpha(1-\alpha)^2}{2-\alpha}(p-1)+\alpha^2(p-1)^2-\frac{\alpha(2-\alpha)(p-1)}{4p}-\frac{\alpha^2}{4p^2}\right]>$$

$$\frac{1}{p^2}\sigma^2\left[\frac{\alpha^2(p-1)}{4p}+\frac{\alpha(2-\alpha)}{4}-\alpha^2 p^2-\frac{\alpha(1-\alpha)^2 p}{2-\alpha}\right]\Leftrightarrow$$

$$\alpha\mu^2\left\{(p-1)\left[\frac{(1-\alpha)^2}{2-\alpha}+\alpha(p-1)-\frac{(2-\alpha)}{4p}-\frac{\alpha}{4p^2(p-1)}\right]\right\}>$$

$$\alpha\sigma^2\left[\frac{\alpha(p-1)}{4p}+\frac{2-\alpha}{4}-\alpha p^2-\frac{(1-\alpha)^2 p}{2-\alpha}\right]$$

At this stage we need to consider that

$$\frac{\alpha(p-1)}{4p}+\frac{2-\alpha}{4}-\alpha p^2-\frac{(1-\alpha)^2 p}{2-\alpha}<0 \text{ for } p>1,\ 0\le\alpha\le 1.$$

This can be proven as follows:

$$f(p,\alpha)=\frac{\alpha(p-1)}{4p}+\frac{2-\alpha}{4}-\alpha p^2-\frac{(1-\alpha)^2 p}{2-\alpha}$$

$$=\frac{\alpha(2-\alpha)(p-1)+p(2-\alpha)^2-\alpha p^2 4p(2-\alpha)-4p^2(1-\alpha)^2}{4p(2-\alpha)}$$

$$Denominator=4p(2-\alpha)>0\ \ p>1,\ 0\le\alpha\le 1$$

$$Numerator=\alpha(2-\alpha)(p-1)+p(2-\alpha)^2-\alpha p^2 4p(2-\alpha)-4p^2(1-\alpha)^2$$

$$=\alpha(2p-2-\alpha p+\alpha)+p(4-4\alpha+\alpha^2)-4\alpha p^3(2-\alpha)-4p^2(1-2\alpha+\alpha^2)$$

$$=2\alpha p-2\alpha-\alpha^2 p+\alpha^2+4p-4\alpha p+\alpha^2 p-8\alpha p^3+4\alpha^2 p^3-4p^2+8\alpha p^2-4\alpha^2 p^2$$

$$= -2\alpha p - 2\alpha + \alpha^2 - 8\alpha\, p^2 (p-1) - 4p(p-1) - 4\alpha^2 p^2 (1-p)$$

$$= \alpha(-2p - 2 + \alpha) + (p-1)\left(-8\alpha\, p^2 - 4p + 4\alpha^2 p^2\right)$$

$$= \alpha(-2p - 2 + \alpha) + 4p(p-1)\{\alpha p(\alpha - 2) - 1\} < 0 \text{ for } p > 1,\ 0 \le \alpha \le 1$$

Therefore, $MSE_{EWMA} > MSE_{APPROXIMATION}$ if and only if:

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[\frac{(1-\alpha)^2}{2-\alpha} + \alpha(p-1) - \frac{\alpha}{4(p-1)\,p^2} - \frac{(2-\alpha)}{4p}\right]}{\frac{\alpha(p-1)}{4p} + \frac{2-\alpha}{4} - \alpha\, p^2 - \frac{(1-\alpha)^2 p}{2-\alpha}}, \text{ for } p > 1,\ 0 \le \alpha \le 1.$$

This proves the inequality (6.33).

APPENDIX 6.C

MSE EWMA – MSE Croston's Method (issue points)

$$MSE_{EWMA} > MSE_{CROSTON} \Leftrightarrow$$

$$\alpha^2\sigma^2 + \frac{\alpha\beta^2}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right] + \left[\mu\left(\alpha + \frac{\beta}{p}\right) - \frac{\mu}{p}\right]^2 >$$

$$\frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] + \left[\frac{\alpha}{2-\alpha}\mu\frac{(p-1)}{p^2}\right]^2 \Leftrightarrow$$

$$\alpha^2\sigma^2 + \frac{\alpha(1-\alpha)^2}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right] + \mu^2\alpha^2\left(1-\frac{1}{p}\right)^2 >$$

$$\frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2 + \frac{\alpha}{2-\alpha}\sigma^2\right) + \frac{\sigma^2}{p^2}\right] + \left(\frac{\alpha}{2-\alpha}\right)^2\frac{\mu^2}{p^2}\left(1-\frac{1}{p}\right)^2 \Leftrightarrow$$

$$\alpha^2\sigma^2 + \frac{\alpha(1-\alpha)^2}{2-\alpha}\frac{\sigma^2}{p} + \mu^2\left[\alpha^2\left(1-\frac{1}{p}\right)^2 + \frac{\alpha(1-\alpha)^2}{2-\alpha}\frac{p-1}{p^2}\right] >$$

$$\left(\frac{\alpha}{2-\alpha}\right)^2\frac{(p-1)}{p^3}\sigma^2 + \frac{\alpha}{2-\alpha}\frac{\sigma^2}{p^2} + \mu^2\left[\left(\frac{\alpha}{2-\alpha}\right)^2\frac{1}{p^2}\left(1-\frac{1}{p}\right)^2 + \frac{\alpha}{2-\alpha}\frac{(p-1)}{p^3}\right] \Leftrightarrow$$

$$\mu^2\left[\alpha^2\frac{(p-1)^2}{p^2} + \frac{\alpha(1-\alpha)^2}{2-\alpha}\frac{p-1}{p^2} - \left(\frac{\alpha}{2-\alpha}\right)^2\frac{1}{p^2}\frac{(p-1)^2}{p^2} - \frac{\alpha}{2-\alpha}\frac{(p-1)}{p^3}\right] >$$

$$\sigma^2\left[\left(\frac{\alpha}{2-\alpha}\right)^2\frac{(p-1)}{p^3} + \frac{\alpha}{2-\alpha}\frac{1}{p^2} - \alpha^2 - \frac{\alpha(1-\alpha)^2}{2-\alpha}\frac{1}{p}\right] \Leftrightarrow$$

$$\frac{\alpha}{2-\alpha}\mu^2\left[\alpha(2-\alpha)\frac{(p-1)^2}{p^2}+(1-\alpha)^2\frac{(p-1)}{p^2}-\frac{\alpha}{2-\alpha}\frac{1}{p^2}\frac{(p-1)^2}{p^2}-\frac{(p-1)}{p^3}\right]>$$

$$\frac{\alpha}{2-\alpha}\sigma^2\left[\frac{\alpha}{2-\alpha}\frac{(p-1)}{p^3}+\frac{1}{p^2}-\alpha(2-\alpha)-(1-\alpha)^2\frac{1}{p}\right]\Leftrightarrow$$

$$\mu^2\left[\alpha(2-\alpha)\frac{(p-1)^2}{p^2}+(1-\alpha)^2\frac{(p-1)}{p^2}-\frac{\alpha}{2-\alpha}\frac{1}{p^2}\frac{(p-1)^2}{p^2}-\frac{(p-1)}{p^3}\right]>$$

$$\frac{1}{p}\sigma^2\left[\frac{\alpha}{2-\alpha}\frac{(p-1)}{p^2}+\frac{1}{p}-\alpha(2-\alpha)p-1+\alpha(2-\alpha)\right]\Leftrightarrow$$

$$\mu^2\left[\alpha(2-\alpha)\frac{(p-1)^2}{p^2}+(1-\alpha)^2\frac{(p-1)}{p^2}-\frac{\alpha}{2-\alpha}\frac{1}{p^2}\frac{(p-1)^2}{p^2}-\frac{(p-1)}{p^3}\right]>$$

$$\frac{1}{p}\sigma^2\left[\frac{\alpha}{2-\alpha}\frac{(p-1)}{p^2}-\frac{p-1}{p}-\alpha(2-\alpha)(p-1)\right]\Leftrightarrow$$

$$\frac{p-1}{p}\mu^2\left[\alpha(2-\alpha)\frac{p-1}{p}+\frac{(1-\alpha)^2}{p}-\frac{\alpha}{2-\alpha}\frac{p-1}{p^3}-\frac{1}{p^2}\right]>$$

$$\frac{p-1}{p}\sigma^2\left[\frac{\alpha}{2-\alpha}\frac{1}{p^2}-\frac{1}{p}-\alpha(2-\alpha)\right]\Leftrightarrow$$

$$\mu^2\left[\alpha(2-\alpha)-\frac{\alpha(2-\alpha)}{p}+\frac{(1-\alpha)^2}{p}-\frac{1}{p^2}-\frac{\alpha}{(2-\alpha)p^2}+\frac{\alpha}{(2-\alpha)p^3}\right]>\sigma^2\left[\frac{\alpha}{2-\alpha}\frac{1}{p^2}-\frac{1}{p}-\alpha(2-\alpha)\right]\Leftrightarrow$$

$$\mu^2\left[\alpha(2-\alpha)-\frac{\alpha(2-\alpha)}{p}+\frac{(1-\alpha)^2}{p}-\frac{2}{2-\alpha}\frac{1}{p^2}+\frac{\alpha}{2-\alpha}\frac{1}{p^3}\right]>\sigma^2\left[\frac{\alpha}{2-\alpha}\frac{1}{p^2}-\frac{1}{p}-\alpha(2-\alpha)\right]\Leftrightarrow$$

$$\frac{1}{p^3}\mu^2\left[\alpha(2-\alpha)p^3+\left(1-4\alpha+2\alpha^2\right)p^2-\frac{2}{2-\alpha}p+\frac{\alpha}{2-\alpha}\right]>\frac{1}{p^3}\sigma^2\left[\frac{\alpha}{2-\alpha}p-p^2-\alpha(2-\alpha)p^3\right]$$

At this stage we need to consider that

$$\frac{\alpha}{2-\alpha}p-p^{2}-\alpha(2-\alpha)p^{3}<0 \text{ for } p>1,\ 0\leq\alpha\leq1$$

This can be proven as follows:

$$f(p,\alpha)=\frac{\alpha}{2-\alpha}p-p^{2}-\alpha(2-\alpha)p^{3}=p\left(\frac{\alpha}{2-\alpha}-p\right)-\alpha(2-\alpha)p^{3}<0 \text{ for } p>1,\ 0\leq\alpha\leq1$$

Therefore, $MSE_{EWMA}>MSE_{CROSTON}$ if and only if:

$$\frac{\sigma^{2}}{\mu^{2}}>\frac{\alpha(2-\alpha)p^{3}+\left(1-4\alpha+2\alpha^{2}\right)p^{2}-\frac{2}{2-\alpha}p+\frac{\alpha}{2-\alpha}}{\frac{\alpha}{2-\alpha}p-p^{2}-\alpha(2-\alpha)p^{3}}\text{, for } p>1,\ 0\leq\alpha\leq1.$$

This proves the inequality (6.34).

APPENDIX 6.D

MSE EWMA – MSE Approximation (all points in time)

$$MSE_{EWMA.} > MSE_{APPROXIMATION} \Leftrightarrow$$

$$\frac{\alpha}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2+\frac{\sigma^2}{p}\right] > \frac{\alpha(2-\alpha)}{4}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{a}{2-a}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]+\left[\frac{\alpha}{2}\frac{\mu}{p^2}\right]^2 \Leftrightarrow$$

$$\frac{\alpha}{2-\alpha}\frac{1}{p^4}\left[p^2(p-1)\mu^2+\sigma^2 p^3\right]>$$

$$\frac{1}{p^4}\left\{\frac{\alpha(2-\alpha)}{4}\left[p(p-1)\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\sigma^2 p^2\right]+\frac{\alpha^2}{4}\mu^2\right\}\Leftrightarrow$$

$$\frac{\alpha}{4}\frac{1}{2-\alpha}\left[4p^2(p-1)\mu^2+4\sigma^2 p^3\right]>\frac{\alpha}{4}\left\{(2-\alpha)\left[p(p-1)\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\sigma^2 p^2\right]+\alpha\mu^2\right\}\Leftrightarrow$$

$$\frac{4p^2(p-1)\mu^2}{2-\alpha}-(2-\alpha)p(p-1)\mu^2-\alpha\mu^2 > (2-\alpha)p(p-1)\frac{\alpha}{2-\alpha}\sigma^2+(2-\alpha)\sigma^2 p^2-\frac{4\sigma^2 p^3}{2-\alpha}\Leftrightarrow$$

$$\mu^2\left\{p(p-1)\left[\frac{4p}{2-\alpha}-(2-\alpha)\right]-\alpha\right\} > \sigma^2\left[p(p-1)\alpha+(2-\alpha)p^2-\frac{4p^3}{2-\alpha}\right]\Leftrightarrow$$

$$\mu^2\frac{1}{2-\alpha}\left\{p(p-1)\left[4p-(2-\alpha)^2\right]-\alpha(2-\alpha)\right\}> \sigma^2\frac{1}{2-\alpha}\left\{(2-\alpha)\left[p(p-1)\alpha+(2-\alpha)p^2\right]-4p^3\right\}\Leftrightarrow$$

$$\mu^2\left\{p(p-1)\left[4p-(2-\alpha)^2\right]-\alpha(2-\alpha)\right\}> \sigma^2\left[(2-\alpha)\left(2p^2-\alpha p\right)-4p^3\right]$$

At this stage we need to consider that $(2-\alpha)\left(2p^2-\alpha p\right)-4p^3<0$ for $p>1$, $0\le\alpha\le1$.

This can be proven as follows:

$$f(p,\alpha) = (2-\alpha)\left(2\,p^2 - \alpha p\right) - 4\,p^3 = 4\,p^2 - 2\alpha p - 2\alpha\,p^2 + \alpha^2\,p - 4\,p^3$$

$$= 4\,p^2(1-p) + \alpha p(\alpha-2) - 2\alpha\,p^2 < 0 \text{ for } \quad p > 1,\ 0 \le \alpha \le 1 .$$

Therefore, $MSE_{EWMA} > MSE_{APPROXIMATION}$ if and only if:

$$\frac{\sigma^2}{\mu^2} > \frac{p(p-1)\left[4p-(2-\alpha)^2\right] - \alpha(2-\alpha)}{(2-\alpha)\left(2\,p^2-\alpha p\right) - 4\,p^3},\ \text{for} \quad p > 1,\ 0 \le \alpha \le 1 .$$

This proves the inequality (6.35).

APPEMDIX 6.E

MSE EWMA – MSE Croston's method (all points in time)

$MSE_{EWMA} > MSE_{CROSTON} \Leftrightarrow$

$$\frac{\alpha}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2+\frac{\sigma^2}{p}\right] > \frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]+\left[\frac{\alpha}{2-\alpha}\mu\frac{(p-1)}{p^2}\right]^2 \Leftrightarrow$$

$$\frac{\alpha}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2+\frac{\sigma^2}{p}\right] > \frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}+\frac{\alpha}{2-\alpha}\mu^2\frac{(p-1)^2}{p^4}\right] \Leftrightarrow$$

$$\mu^2\left[\frac{p-1}{p^2}-\frac{p(p-1)}{p^4}-\frac{\alpha}{2-\alpha}\frac{(p-1)^2}{p^4}\right] > \sigma^2\left[\frac{\alpha}{2-\alpha}\frac{p(p-1)}{p^4}+\frac{1}{p^2}-\frac{1}{p}\right] \Leftrightarrow$$

$$\mu^2\frac{1}{p^4}\left[p^2(p-1)-p(p-1)-\frac{\alpha}{2-\alpha}(p-1)^2\right] > \sigma^2\frac{1}{p^4}\left[\frac{\alpha}{2-\alpha}p(p-1)+p^2-p^3\right] \Leftrightarrow$$

$$\mu^2(p-1)\left[p^2-p-\frac{\alpha}{2-\alpha}(p-1)\right] > \sigma^2\left[p(p-1)\frac{\alpha}{2-\alpha}-p^2(p-1)\right] \Leftrightarrow$$

$$\mu^2(p-1)(p-1)\left(p-\frac{\alpha}{2-\alpha}\right) > \sigma^2 p(p-1)\left(\frac{\alpha}{2-\alpha}-p\right)$$

At this stage we need to consider that $\frac{\alpha}{2-\alpha}-p<0$ for $p>1$, $0\le\alpha\le1$

so we have: $\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left(p-\frac{\alpha}{2-\alpha}\right)}{p\left(\frac{\alpha}{2-\alpha}-p\right)} \Leftrightarrow \frac{\sigma^2}{\mu^2} > \frac{1-p}{p}$.

Therefore, $MSE_{EWMA} > MSE_{CROSTON}$ if and only if: $\frac{\sigma^2}{\mu^2} > \frac{1-p}{p}$, for $p>1$, $0\le\alpha\le1$.

This proves the inequality (6.36).

APPENDIX 6.F

MSE EWMA – MSE λ Approximation (issue points)

$$MSE_{EWMA} > MSE_{\lambda APPROXIMATION} \Leftrightarrow$$

$$\alpha^2\sigma^2 + \frac{\alpha\beta^2}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right] + \left[\mu\left(\alpha+\frac{\beta}{p}\right) - \frac{\mu}{p}\right]^2 >$$

$$\frac{\alpha(2-\alpha)}{4}\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2 + p(p-1)\mu^2 + \frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4} \Leftrightarrow$$

$$\alpha^2\sigma^2 + \frac{\alpha\beta^2}{2-\alpha}\frac{p-1}{p^2}\mu^2 + \frac{\alpha\beta^2}{2-\alpha}\frac{\sigma^2}{p} + \frac{\alpha^2(p-1)^2\mu^2}{p^2} >$$

$$\left(\frac{\alpha}{2-\alpha}\right)\left(1-\frac{\alpha}{2}\right)^2\frac{\sigma^2}{\left(p-\frac{\alpha}{2}\right)^2} + \left(\frac{\alpha}{2-\alpha}\right)\left(1-\frac{\alpha}{2}\right)^2\frac{p(p-1)\mu^2}{\left(p-\frac{\alpha}{2}\right)^4} + \left(\frac{\alpha}{2-\alpha}\right)^2\left(1-\frac{\alpha}{2}\right)^2\frac{p(p-1)\sigma^2}{\left(p-\frac{\alpha}{2}\right)^4} \Leftrightarrow$$

$$\alpha^2\sigma^2 + \frac{p-1}{p^2}\mu^2\left[\frac{\alpha\beta^2}{2-\alpha} + \alpha^2(p-1)\right] + \frac{\alpha\beta^2}{2-\alpha}\frac{\sigma^2}{p} >$$

$$\left(\frac{\alpha}{2-\alpha}\right)\left(1-\frac{\alpha}{2}\right)^2\frac{\sigma^2}{\left(p-\frac{\alpha}{2}\right)^2} + \left(\frac{\alpha}{2-\alpha}\right)\left(1-\frac{\alpha}{2}\right)^2\frac{p(p-1)\mu^2}{\left(p-\frac{\alpha}{2}\right)^4} + \left(\frac{\alpha}{2-\alpha}\right)^2\left(1-\frac{\alpha}{2}\right)^2\frac{p(p-1)\sigma^2}{\left(p-\frac{\alpha}{2}\right)^4} \Leftrightarrow$$

$$\mu^2\frac{p-1}{p^2}\left[\frac{\alpha\beta^2}{2-\alpha} + \alpha^2(p-1) - \left(\frac{\alpha}{2-\alpha}\right)\left(1-\frac{\alpha}{2}\right)^2\frac{p^3}{\left(p-\frac{\alpha}{2}\right)^4}\right] >$$

$$\sigma^2\left[\left(\frac{\alpha}{2-\alpha}\right)\left(1-\frac{\alpha}{2}\right)^2\frac{1}{\left(p-\frac{\alpha}{2}\right)^2}+\left(\frac{\alpha}{2-\alpha}\right)^2\left(1-\frac{\alpha}{2}\right)^2\frac{p(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-\alpha^2-\frac{1}{p}\frac{\alpha\beta^2}{2-\alpha}\right]\Leftrightarrow$$

$$\mu^2\frac{p-1}{p^2}\left[\frac{\alpha+\alpha^3-2\alpha^2}{2-\alpha}+\alpha^2(p-1)-\frac{\alpha(2-\alpha)p^3}{4\left(p-\frac{\alpha}{2}\right)^4}\right]>$$

$$\sigma^2\left[\frac{\alpha(2-\alpha)}{4\left(p-\frac{\alpha}{2}\right)^2}+\frac{\alpha^2}{4}\frac{p(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-\alpha^2-\frac{1}{p}\frac{\alpha\beta^2}{2-\alpha}\right]\Leftrightarrow$$

$$\mu^2\alpha\frac{p-1}{p^2}\left[\frac{1}{2-\alpha}+\alpha\left(p-1+\frac{\alpha-2}{2-\alpha}\right)-\frac{(2-\alpha)p^3}{4\left(p-\frac{\alpha}{2}\right)^4}\right]>$$

$$\sigma^2\alpha\left[\frac{(2-\alpha)}{4\left(p-\frac{\alpha}{2}\right)^2}+\frac{\alpha}{4}\frac{p(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-\alpha-\frac{1}{p}\frac{\beta^2}{2-\alpha}\right]\Leftrightarrow$$

$$\mu^2\frac{p-1}{p^2}\left[\frac{1}{2-\alpha}+\alpha(p-2)-\frac{(2-\alpha)p^3}{4\left(p-\frac{\alpha}{2}\right)^4}\right]>\sigma^2\frac{1}{p}\left[\frac{(2-\alpha)p}{4\left(p-\frac{\alpha}{2}\right)^2}+\frac{\alpha}{4}\frac{p^2(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-\alpha p-\frac{\beta^2}{2-\alpha}\right]$$

At this stage we need to consider that

$$f(p,\alpha)=p\left[\frac{(2-\alpha)p}{4\left(p-\frac{\alpha}{2}\right)^2}+\frac{\alpha}{4}\frac{p^2(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-\alpha p-\frac{\beta^2}{2-\alpha}\right]<0 \text{ for } 1<p\leq 10,\ 0\leq\alpha\leq 1.$$

Unfortunately we are not able to prove theoretically the above statement. Nevertheless, it can be illustrated, graphically, that for fixed α values, $f(p,\alpha)$ is monotonic decreasing in p over the range [1,10], $f(1,\alpha)\leq 0$. It can also be illustrated, graphically, that for fixed p values, $f(p,\alpha)$ is monotonic decreasing in α over the range [0,1], $f(p,0)\leq 0$. In figure 6.F.1 the $f(p,\alpha)$ values are presented for p = 1 – 10 step 1 and α being 0, 0.5 and 1.

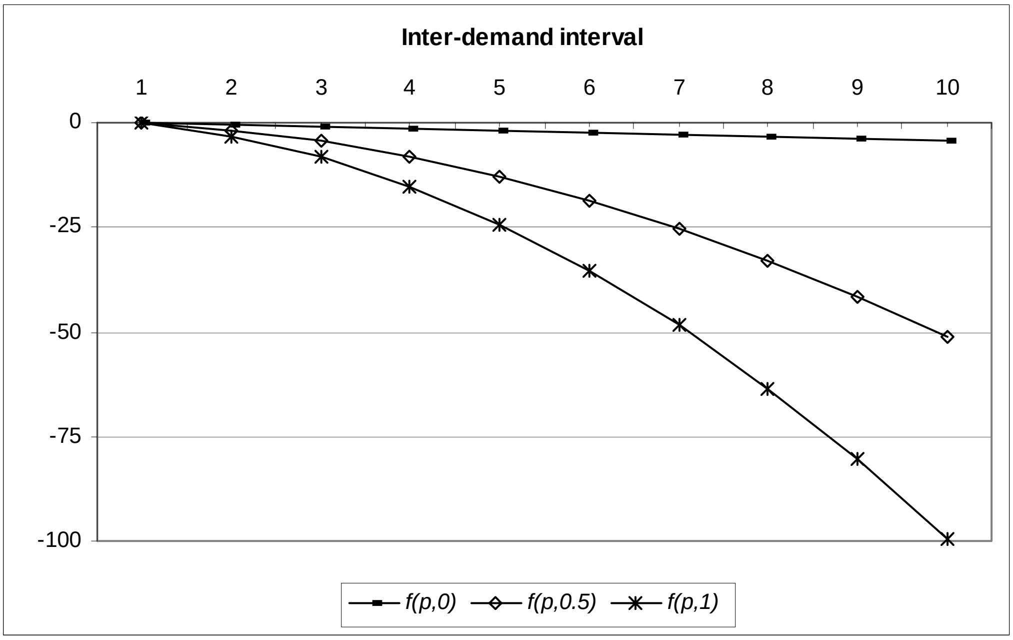

Figure 6.F.1. MSE EWMA – MSE λ Approximation (issue points)

Therefore, we conjecture that $MSE_{EWMA} > MSE_{\lambda APPROXIMATION}$ if and only if[1]:

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[\frac{1}{2-\alpha}+\alpha(p-2)-\frac{(2-\alpha)p^3}{4\left(p-\frac{\alpha}{2}\right)^4}\right]}{p\left[\frac{(2-\alpha)p}{4\left(p-\frac{\alpha}{2}\right)^2}+\frac{\alpha}{4}\frac{p^2(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-\alpha p-\frac{\beta^2}{2-\alpha}\right]}$$, for $1 < p \leq 10$, $0 \leq \alpha \leq 1$.

[1] For $p = 1$, the right hand side of the inequality is defined only for some α values.

APPENDIX 6.G

MSE EWMA – MSE λ Approximation (all points in time)

$MSE_{EWMA.} > MSE_{\lambda APPROXIMATION} \Leftrightarrow$

$$\frac{\alpha}{2-\alpha}\left[\frac{p-1}{p^2}\mu^2+\frac{\sigma^2}{p}\right] > \frac{\alpha(2-\alpha)}{4}\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2+p(p-1)\mu^2+\frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4} \Leftrightarrow$$

$$\frac{p-1}{p^2}\mu^2+\frac{\sigma^2}{p} > \frac{\left(1-\frac{\alpha}{2}\right)^2\sigma^2}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{\left(1-\frac{\alpha}{2}\right)^2 p(p-1)\mu^2}{\left(p-\frac{\alpha}{2}\right)^4}+\frac{\left(1-\frac{\alpha}{2}\right)^2\frac{\alpha}{2-\alpha}p(p-1)\sigma^2}{\left(p-\frac{\alpha}{2}\right)^4} \Leftrightarrow$$

$$\mu^2\left[\frac{p-1}{p^2}-\frac{\left(1-\frac{\alpha}{2}\right)^2 p(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}\right] > \sigma^2\left[\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^4}\frac{\alpha}{2-\alpha}p(p-1)-\frac{1}{p}\right] \Leftrightarrow$$

$$\mu^2\left\{(p-1)\left[\frac{1}{p^2}-\frac{\left(1-\frac{\alpha}{2}\right)^2 p}{\left(p-\frac{\alpha}{2}\right)^4}\right]\right\} > \sigma^2\left\{\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^2}\left[1+\frac{1}{\left(p-\frac{\alpha}{2}\right)^2}\frac{\alpha}{2-\alpha}p(p-1)\right]-\frac{1}{p}\right\}$$

At this stage we need to consider that

$$\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^2}\left[1+\frac{1}{\left(p-\frac{\alpha}{2}\right)^2}\frac{\alpha}{2-\alpha}p(p-1)\right]-\frac{1}{p}<0 \text{ for } 1<p\le 10,\ 0\le\alpha\le 1$$

This can be proven as follows (for $0 \le \alpha < 0.66$):

$$f(p,\alpha) = \frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^2}\left[1 + \frac{1}{\left(p-\frac{\alpha}{2}\right)^2}\frac{\alpha}{2-\alpha}\, p(p-1)\right] - \frac{1}{p}$$

$$= \frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^2} + \frac{\left(1-\frac{\alpha}{2}\right)^2 \alpha p(p-1)}{\left(p-\frac{\alpha}{2}\right)^4 (2-\alpha)} - \frac{1}{p} = \frac{(2-\alpha)^2}{(2p-\alpha)^2} + \frac{4(2-\alpha)^2 \alpha p(p-1)}{(2p-\alpha)^4(2-\alpha)} - \frac{1}{p}$$

$$= \frac{(2-\alpha)^2(2p-\alpha)^2 p + 4\alpha p^2(p-1)(2-\alpha) - (2p-\alpha)^4}{p(2p-\alpha)^4}$$

$Denominator = p(2p-\alpha)^4 > 0$ for $p > 1$, $0 \le \alpha \le 1$

$Numerator = (2-\alpha)^2(2p-\alpha)^2 p + 4\alpha p^2(p-1)(2-\alpha) - (2p-\alpha)^4$

$$= (2p-\alpha)^2\left[(2-\alpha)^2 p - (2p-\alpha)^2\right] + 4\alpha p^2(p-1)(2-\alpha)$$

$$= (2p-\alpha)^2\left(4p + \alpha^2 p - 4\alpha p - 4p^2 - \alpha^2 + 4\alpha p\right) + 4\alpha p^2(p-1)(2-\alpha)$$

$$= (2p-\alpha)^2\left[-4p(p-1) + \alpha^2(p-1)\right] + 4\alpha p^2(p-1)(2-\alpha)$$

$$= (2p-\alpha)^2(p-1)\left(\alpha^2 - 4p\right) + 4\alpha p^2(p-1)(2-\alpha)$$

$$= (p-1)\left[(2p-\alpha)^2\left(\alpha^2 - 4p\right) + 4\alpha p^2(2-\alpha)\right]$$

$$= (p-1)\left(4\alpha^2 p^2 - 16p^3 + \alpha^4 - 4\alpha^2 p - 4\alpha^3 p + 16\alpha p^2 + 8\alpha p^2 - 4\alpha^2 p^2\right)$$

$$= (p-1)\left[-16p^3 + 24\alpha p^2 + \alpha^4 - 4\alpha^2 p(1+\alpha)\right]$$

$$= (p-1)\left\{8p^2(3\alpha - 2p) - \alpha^2\left[4p(1+\alpha) - \alpha^2\right]\right\} < 0 \text{ for } p > 1,\ 0 \le \alpha \le 0.66.$$

Unfortunately we are not able to prove theoretically that: $f(p,\alpha)<0$, for $p>1$ and α values greater than 0.66. Nevertheless, we can illustrate graphically that $f(p,\alpha)\le 0$ for $1\le p\le 10$ and $\alpha \ge 0.66$. We do so in figure 6.G.1 where the $f(p,\alpha)$ values are presented for p = 1 – 10 step 1 and α being 0.66 and 1.

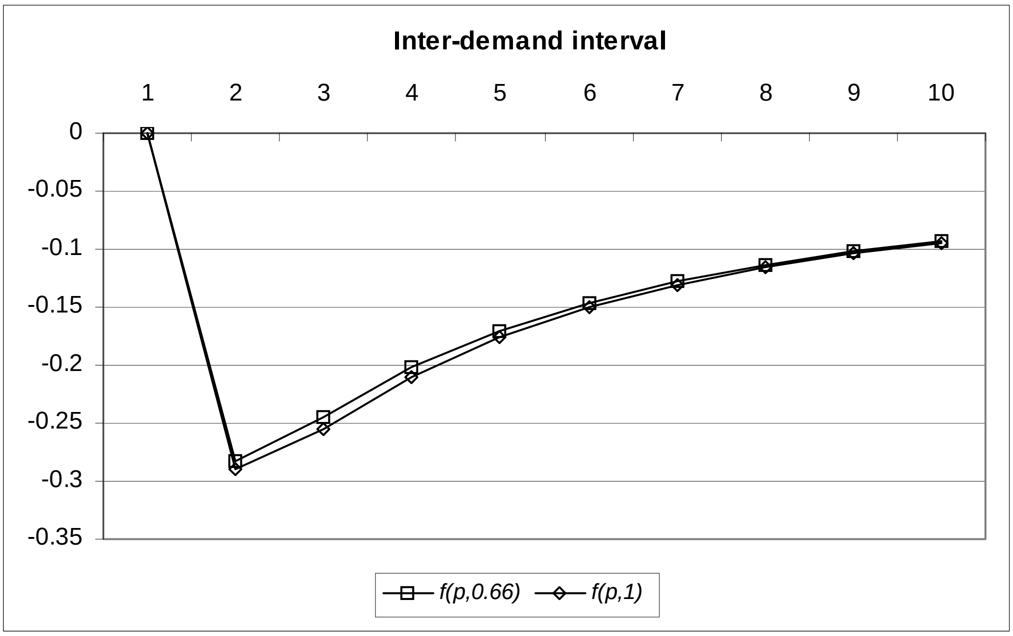

Figure 6.G.1. MSE EWMA – MSE λ Approximation (all points in time)

Therefore, we conjecture that $MSE_{EWMA} > MSE_{\lambda APPROXIMATION}$ if and only if[1]:

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[\frac{1}{p^2} - \frac{\left(1-\frac{\alpha}{2}\right)^2 p}{\left(p-\frac{\alpha}{2}\right)^4}\right]}{\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^2}\left[1+\frac{1}{\left(p-\frac{\alpha}{2}\right)^2}\frac{\alpha}{2-\alpha}p(p-1)\right]-\frac{1}{p}}, \text{ for } 1<p\le 10,\ 0\le\alpha\le 1.$$

[1] For $p=1$, the right hand side of the inequality cannot be defined, for any $0\le\alpha\le 1$.

APPENDIX 6.H

MSE Croston's Method – MSE λ Approximation

$$MSE_{CROSTON} > MSE_{\lambda APPROXIMATION} \Leftrightarrow$$

$$\frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]+\left[\frac{\alpha}{2-\alpha}\mu\frac{(p-1)}{p^2}\right]^2 >$$

$$\frac{\alpha(2-\alpha)}{4}\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2+p(p-1)\mu^2+\frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4} \Leftrightarrow$$

$$\frac{\alpha}{2-\alpha}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}+\frac{\alpha}{2-\alpha}\mu^2\frac{(p-1)^2}{p^4}\right] >$$

$$\left(\frac{\alpha}{2-\alpha}\right)\left(1-\frac{\alpha}{2}\right)^2\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2+p(p-1)\mu^2+\frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4} \Leftrightarrow$$

$$\frac{(p-1)}{p^3}\mu^2+\frac{\alpha}{2-\alpha}\mu^2\frac{(p-1)^2}{p^4}-\frac{\left(1-\frac{\alpha}{2}\right)^2 p(p-1)\mu^2}{\left(p-\frac{\alpha}{2}\right)^4} >$$

$$\frac{\left(1-\frac{\alpha}{2}\right)^2\sigma^2}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^4}p(p-1)\frac{\alpha}{2-\alpha}\sigma^2-\frac{(p-1)}{p^3}\frac{\alpha}{2-\alpha}\sigma^2-\frac{\sigma^2}{p^2} \Leftrightarrow$$

$$\mu^2 \frac{1}{p^4}\left[p(p-1)+\frac{\alpha}{2-\alpha}(p-1)^2-\frac{\left(1-\frac{\alpha}{2}\right)^2 p^5(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}\right]>$$

$$\sigma^2 \frac{1}{p^4}\left[\frac{\left(1-\frac{\alpha}{2}\right)^2 p^4}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^4} p^3(p-1)\frac{\alpha}{2-\alpha}-p(p-1)\frac{\alpha}{2-\alpha}-p^2\right]\Leftrightarrow$$

$$\mu^2 (p-1)\left[p+\frac{\alpha}{2-\alpha}(p-1)-\frac{\left(1-\frac{\alpha}{2}\right)^2 p^5}{\left(p-\frac{\alpha}{2}\right)^4}\right]>$$

$$\sigma^2 \left\{\frac{\left(1-\frac{\alpha}{2}\right)^2 p^4}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{\alpha}{2-\alpha}p(p-1)\left[\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^4}p^2-1\right]-p^2\right\}$$

At this stage we need to consider that

$$f(p,\alpha)=\frac{\left(1-\frac{\alpha}{2}\right)^2 p^4}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{\alpha}{2-\alpha}p(p-1)\left[\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^4}p^2-1\right]-p^2<0 \text{ for } 1<p\le 10,\ 0<\alpha\le 1.$$

Unfortunately we are not able to prove theoretically the above statement. Nevertheless, it can be illustrated, graphically, that for fixed α values greater than zero (when $\alpha=0$, $f(p,\alpha)=0$ for any $p\ge 1$), $f(p,\alpha)$ is monotonic decreasing in p over the range [1,10], $f(1,\alpha)\le 0$. It can also be illustrated, graphically, that for fixed p values, $f(p,\alpha)$ is monotonic decreasing in α over the range [0.05,1], $f(p,0.05)\le 0$. In figure 6.H.1 the $f(p,\alpha)$ values are presented for p = 1 – 10 step 1 and α being 0.05, 0.5 and 1.

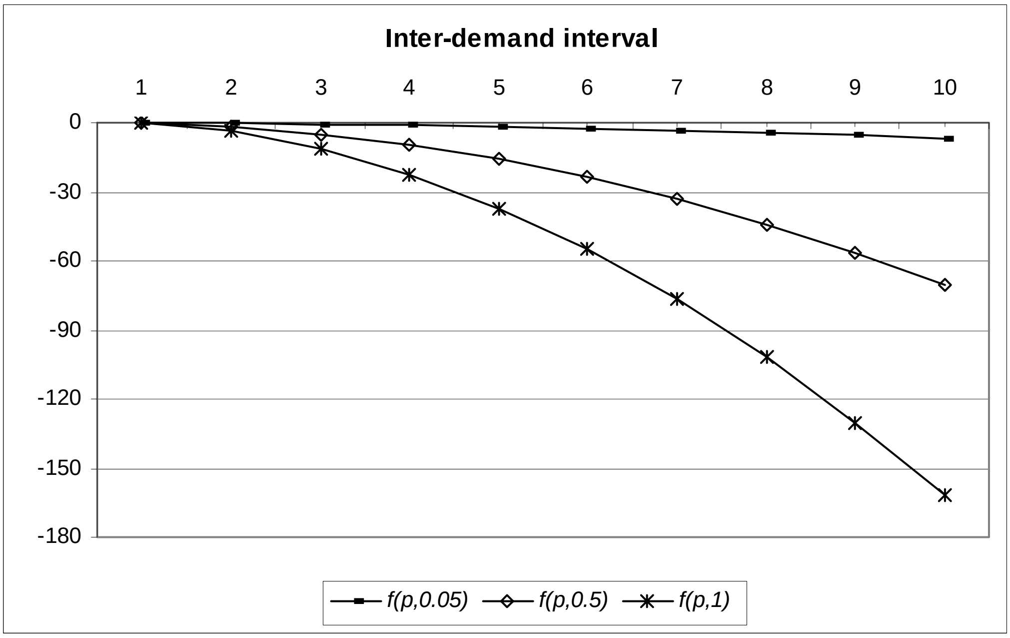

Figure 6.H.1. MSE Croston's Method – MSE λ Approximation

Therefore, we conjecture that $MSE_{CROSTON} > MSE_{\lambda APPROXIMATION}$ if and only if[1]:

$$\frac{\sigma^2}{\mu^2} > \frac{(p-1)\left[p + \frac{\alpha}{2-\alpha}(p-1) - \frac{\left(1-\frac{\alpha}{2}\right)^2 p^5}{\left(p-\frac{\alpha}{2}\right)^4}\right]}{\frac{\left(1-\frac{\alpha}{2}\right)^2 p^4}{\left(p-\frac{\alpha}{2}\right)^2} + \frac{\alpha}{2-\alpha}\, p(p-1)\left[\frac{\left(1-\frac{\alpha}{2}\right)^2}{\left(p-\frac{\alpha}{2}\right)^4} p^2 - 1\right] - p^2}, \text{ for } 1 < p \le 10,\ 0 < \alpha \le 1.$$

[1] For $p = 1$, the right hand side of the inequality is defined only for some α values.

APPENDIX 6.I

MSE Approximation – MSE λ Approximation

$$MSE_{APPROXIMATION} > MSE_{\lambda APPROXIMATION} \Leftrightarrow$$

$$\frac{\alpha(2-\alpha)}{4}\left[\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)+\frac{\sigma^2}{p^2}\right]+\left[\frac{\alpha}{2}\frac{\mu}{p^2}\right]^2 >$$

$$\frac{\alpha(2-\alpha)}{4}\frac{\left[\left(p-\frac{\alpha}{2}\right)^2\sigma^2+p(p-1)\mu^2+\frac{\alpha}{2-\alpha}p(p-1)\sigma^2\right]}{\left(p-\frac{\alpha}{2}\right)^4}\Leftrightarrow$$

$$\left[\frac{\alpha}{2}\frac{\mu}{p^2}\right]^2 >$$

$$\frac{\alpha(2-\alpha)}{4}\left[\frac{\sigma^2}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{p(p-1)\mu^2}{\left(p-\frac{\alpha}{2}\right)^4}+\frac{\frac{\alpha}{2-\alpha}p(p-1)\sigma^2}{\left(p-\frac{\alpha}{2}\right)^4}-\frac{(p-1)}{p^3}\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)-\frac{\sigma^2}{p^2}\right]\Leftrightarrow$$

$$\frac{\alpha^2}{4}\mu^2\frac{1}{p^4} >$$

$$\frac{\alpha(2-\alpha)}{4}\frac{1}{p^4}\left[\frac{\sigma^2 p^4}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{p^5(p-1)\mu^2}{\left(p-\frac{\alpha}{2}\right)^4}+\frac{\frac{\alpha}{2-\alpha}p^5(p-1)\sigma^2}{\left(p-\frac{\alpha}{2}\right)^4}-p(p-1)\left(\mu^2+\frac{\alpha}{2-\alpha}\sigma^2\right)-\sigma^2 p^2\right]\Leftrightarrow$$

$$\frac{\alpha^2}{4}\mu^2 >$$

$$\frac{\alpha(2-\alpha)}{4}\left\{\sigma^2\left[\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{\frac{\alpha}{2-\alpha}p^5(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-p(p-1)\frac{\alpha}{2-\alpha}-p^2\right]+\mu^2\left[p(p-1)\left(\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^4}-1\right)\right]\right\}\Leftrightarrow$$

$$\mu^2\left\{\frac{\alpha^2}{4}-\frac{\alpha(2-\alpha)}{4}\left[p(p-1)\left(\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^4}-1\right)\right]\right\}>$$

$$\sigma^2\frac{\alpha(2-\alpha)}{4}\left[\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^2}+\frac{\frac{\alpha}{2-\alpha}p^5(p-1)}{\left(p-\frac{\alpha}{2}\right)^4}-p(p-1)\frac{\alpha}{2-\alpha}-p^2\right]\Leftrightarrow$$

$$\mu^2\left\{\frac{\alpha^2}{4}-\frac{\alpha(2-\alpha)}{4}\left[p(p-1)\left(\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^4}-1\right)\right]\right\}>$$

$$\sigma^2\frac{\alpha(2-\alpha)}{4}\left\{\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^2}+p(p-1)\frac{\alpha}{2-\alpha}\left[\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^4}-1\right]-p^2\right\}$$

At this stage we need to consider that

$$\frac{\alpha(2-\alpha)}{4}\left\{\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^2}+p(p-1)\frac{\alpha}{2-\alpha}\left[\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^4}-1\right]-p^2\right\}>0 \text{ for } p>1,\ 0\le\alpha\le1.$$

This can be proven as follows:

$$f(p,\alpha)=\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^2}+p(p-1)\frac{\alpha}{2-\alpha}\left[\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^4}-1\right]-p^2$$

$$=\frac{4p^4}{(2p-\alpha)^2}+p(p-1)\frac{\alpha}{2-\alpha}\frac{16p^4}{(2p-\alpha)^4}-p(p-1)\frac{\alpha}{2-\alpha}-p^2$$

$$=\frac{4p^4(2-\alpha)(2p-\alpha)^2+p(p-1)\alpha 16p^4-p(p-1)\alpha(2p-\alpha)^4-p^2(2p-\alpha)^4(2-\alpha)}{(2p-\alpha)^4(2-\alpha)}$$

$Denominator=(2p-\alpha)^4(2-\alpha)>0$ for $p>1,\ 0\le\alpha\le 1$

$$Numerator=4p^4(2-\alpha)(2p-\alpha)^2+p(p-1)\alpha 16p^4-p(p-1)\alpha(2p-\alpha)^4-p^2(2p-\alpha)^4(2-\alpha)$$

$$=(2p-\alpha)^2\left[4p^4(2-\alpha)-p^2(2p-\alpha)^2(2-\alpha)\right]+\alpha p(p-1)\left[16p^4-(2p-\alpha)^4\right]$$

$$=(2p-\alpha)^2\left[8p^4-4\alpha p^4-\left(2p^2-\alpha p^2\right)\left(4p^2+\alpha^2-4\alpha p\right)\right]+\alpha p(p-1)\left[(2p)^4-(2p-\alpha)^4\right]$$

$$=(2p-\alpha)^2\left(8p^4-4\alpha p^4-8p^4-2\alpha^2p^2+8\alpha p^3+4\alpha p^4+p^2\alpha^3-4\alpha^2p^3\right)$$

$$+\alpha p(p-1)\left[(2p)^4-(2p-\alpha)^4\right]$$

$$=(2p-\alpha)^2\left[4\alpha p^3(2-\alpha)-p^2\alpha^2(2-\alpha)\right]+\alpha p(p-1)\left[(2p)^4-(2p-\alpha)^4\right]$$

$$=(2p-\alpha)^2(2-\alpha)\left(4\alpha p^3-p^2\alpha^2\right)+\alpha p(p-1)\left[(2p)^4-(2p-\alpha)^4\right]$$

$$=\alpha p^2(2p-\alpha)^2(2-\alpha)(4p-\alpha)+\alpha p(p-1)\left[(2p)^4-(2p-\alpha)^4\right]>0$$ for $p>1,\ 0\le\alpha\le 1$.

Therefore, $MSE_{APPROXIMATION} > MSE_{\lambda APPROXIMATION}$ if and only if:

$$\frac{\sigma^2}{\mu^2} < \frac{\frac{\alpha^2}{4} - \frac{\alpha(2-\alpha)}{4} p(p-1)\left(\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^4} - 1\right)}{\frac{\alpha(2-\alpha)}{4}\left\{\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^2} + p(p-1)\frac{\alpha}{2-\alpha}\left[\frac{p^4}{\left(p-\frac{\alpha}{2}\right)^4} - 1\right] - p^2\right\}}, \text{ for } p > 1,\ 0 \le \alpha \le 1.$$

This proves the inequality (6.40).

APPENDIX 7.A

Demand size distributions

Three continuous demand size distributions

The Erlang variate is a Gamma variate $\gamma : b, c$ (scale parameter $b > 0$) with shape parameter $c > 0$ an integer. The probability density function is given by (7.A.1):

$$f(x) = \frac{(x/b)^{c-1}[\exp(-x/b)]}{b\Gamma(c)} \tag{7.A.1}$$

where $\Gamma(c)$ is the Gamma function with integer argument c: $\Gamma(c) = (c-1)!$

The mean is calculated as: bc, whereas the variance is: $b^2 c$

The squared coefficient of variation (CV^2) is given by (7.A.2).

$$CV^2 = \frac{b^2 c}{b^2 c^2} = \frac{1}{c} \tag{7.A.2}$$

Considering integer values of c only, (7.A.2) is always less than or equal to 1.

If the Gamma rather than the Erlang distribution was considered then (7.A.2) could also take values greater than 1.

The negative exponential variate $\mathrm{E} : b$ is the Gamma variate $\gamma : b, c$ corresponding to shape parameter $c = 1$. In that case $CV^2 = 1$.

The rectangular (uniform) continuous variate $R : a, b$ with $a \le \mathrm{x} \le b$ has the following probability density function:

$$f(x)=\frac{1}{b-a} \tag{7.A.3}$$

The mean is: $\frac{a+b}{2}$ and the variance: $\frac{(b-a)^2}{12}$ and as such the CV^2 is calculated as follows:

$$CV^2=\frac{(b-a)^2}{3(b+a)^2} \tag{7.A.4}$$

From (7.A.4) it is obvious that CV^2 is always less than 1.

Two discrete demand size distributions

A truncated probability density function $f_T(x)$ is calculated by dividing the original density function, $f(x)$, by one minus the probability that the variable x will take a value (in the variate's X range) that we do not want to consider. Both Poisson and Pascal are discrete distributions with range: $0 \le x < \infty$. Since we refer to demand sizes, rather than demand per unit time period, a size $x=0$ would not make sense. Therefore we want to consider the original distributions but excluding the zero "class" probability.

The Poisson density function is as follows:

$$f(x)=\frac{\lambda^x \exp(-\lambda)}{x!} \tag{7.A.5}$$

where $\lambda > 0$ is the mean and variance in this distribution.

$$f(0)=\exp(-\lambda) \tag{7.A.6}$$

and

$$f_T(x) = \frac{f(x)}{1 - f(0)} = \frac{\lambda^x \exp(-\lambda)}{x!\left[1 - \exp(-\lambda)\right]} \tag{7.A.7}$$

The mean and variance of the distribution are now calculated as: $\frac{\lambda}{1 - \exp(-\lambda)}$.

Consequently the CV^2 is given by (7.A.8).

$$CV^2 = \frac{1 - \exp(-\lambda)}{\lambda} \tag{7.A.8}$$

For $\lambda = 1$, (7.A.8) becomes $CV^2 = 1 - \frac{1}{e} = 0.632$.

CV^2 is a decreasing function of λ with

$$\underset{\lambda \to \infty}{Lim} \frac{1 - \exp(-\lambda)}{\lambda} = 0 \tag{7.A.9}$$

Therefore $CV^2 < 1$, for all $\lambda \geq 1$.

The Pascal variate is the Negative Binomial variate $NB : x, p$ $(0 < p < 1)$ with x ($0 \leq x < \infty$) being an integer. (It was the truncated Pascal rather than the truncated Negative Binomial distribution that was considered in Kwan's thesis.)

The Pascal density function is given by (7.A.10):

$$f(x) = \frac{(x + y - 1)!}{(x - 1)!\,y!} p^x q^y \tag{7.A.10}$$

where y is the quantile, an integer with range: $0 \leq y < \infty$ and $q = 1 - p$

The mean of the distribution is: xq / p whereas the variance is: xq / p^2

$$f(0) = p^x \tag{7.A.11}$$

Considering (7.A.10) and (7.A.11) we then have:

$$f_T(x) = \frac{(x+y-1)!}{(x-1)!\,y!}\frac{p^x q^y}{1-p^x} \tag{7.A.12}$$

with

$$CV^2 = \frac{1-p^x}{xq} \tag{7.A.13}$$

For $x = 1$, $CV^2 = 1$

For $x > 1$, $1 - p^x$ increases exponentially with a limit equal to 1 so that $CV^2 < 1$.

For x not restricted to being an integer, i.e. truncated Negative Binomial distribution, the squared coefficient of variation can also take values greater than 1, if $x < 1$.

APPENDIX 7.B

Theoretically generated demand data

Binomially distributed random variables

For a specified average inter-demand interval p, the Bernoulli probability of demand occurrence is $\frac{1}{p}$.

A uniformly distributed in $(0,1)$ random variate (R) can then be generated so that:

If $R < \frac{1}{p}$ *then*

"Return" 1

Else

"Return" 0

End If

Cells with the value 1 indicate demand occurrence, whereas cells with the value 0 non-occurrence of demand. The number of consecutive zeroes can then be calculated so that:

Inter-demand interval = Number of consecutive zeroes + 1

The value 1 is added because we refer to the geometric distribution including the first success. As such inter-demand intervals equal to 1 indicate two consecutive demand occurring periods.

Lognormally distributed random variables

The relationship of the lognormal variate $L: m, \sigma_N$ to the unit normal variate $N: 0,1$ gives:

$$L : m, \sigma_N \sim \exp\{\mu_N + \sigma_N (N : 0,1)\} \tag{7.B.1}$$

By specifying the mean (μ) and the variance (σ^2) of the lognormally distributed demand sizes, the μ_N and σ_N values can then be derived as follows (Kleijnen and Van Groenendaal, 1992):

$$\sigma_N = \sqrt{\log(1+\delta)} \tag{7.B.2}$$

where $\delta = \dfrac{\sigma^2}{\mu^2}$

and

$$\mu_N = \log m \tag{7.B.3}$$

where m is calculated using (7.2) and (7.B.2).

Two methods are recommended in the academic literature for generating standard normal random variables. The first is the Box -Muller method (1958).

If R_1, R_2 are two independent uniform variables distributed in $(0,1]$, then two independent standard normal variables are generated by:

$$N_1(0,1) = \sqrt{-2\log R_1}\,\sin(2\pi\,\mathrm{R}_2) \tag{7.B.4}$$

$$N_2(0,1) = \sqrt{-2\log R_1}\,\cos(2\pi\,\mathrm{R}_2) \tag{7.B.5}$$

This particular method is very easy to program but it is rather slow in execution as it requires the calculation of a logarithm, a square root and a trigonometric function. Marsaglia and Bray (1964) improved the operating speed of the Box-Muller approach by the following device (see also Kleijnen and Van Groenendaal, 1992):

If Z_1, Z_2 are uniformly distributed in $(0, 1)$ subject to the condition $\mathrm{w} = w_1^{\,2} + w_2^{\,2} \leq 1$, (where $\mathrm{w}_1 = 2Z_1 - 1$ and $\mathrm{w}_2 = 2Z_2 - 1$) then two independent standard normal variables are generated by:

$$N_1(0,1) = w_1\left(\frac{-2\log w}{w}\right)^{\frac{1}{2}} \tag{7.B.6}$$

$$N_2(0,1) = w_2\left(\frac{-2\log w}{w}\right)^{\frac{1}{2}} \tag{7.B.7}$$

This method is faster than the Box-Muller method since no calculation of a trigonometric function is required (see also Atkinson and Pearce, 1976). Consequently we use it for generating pairs of independent standard normal variables.

For each particular demand time period, the binary decision variable that indicates whether or not demand occurs in that period and the lognormal variable that corresponds to that period are multiplied. The resulting series is that of demand per unit time period.

APPENDIX 7.C

The simulation control parameters

The p (average inter-demand interval) cut-off values for all the categorisation rules developed in chapter 6 were in the range 1.17 – 1.65. It is this area therefore that needs to be explored in detail in order to reach definitive conclusions about the validity of the categorisation rules. The p values selected for that purpose are 1.1 to 1.9 step 0.2. The step value chosen reflects the compromise that we try to achieve between the detailed investigation of the problem concerned and the size of the simulation experiment (see table 7.C.2). For $p \geq 2$ the MSE superiority of one method over another or over all other methods is very well marked but nevertheless we still wish to assess the sensitivity of all the results, specifically related to each one of the methods, to higher inter-demand intervals. Consideration of inter-demand intervals equal or less than 10 can be justified based on the academic literature (Watson, 1987; Willemain et al 1994; Johnston and Boylan, 1996). In addition to the p values discussed above a wide range of other p values is also considered in order to account for most of the "real world" scenarios $p = 2$ to 10 step 2.

As far as the second control parameter is concerned, and in order to account for a wide range of "real world" situations, the following conditions need to be simulated: constant demand sizes (or more generally a variance that is set close to zero); variance less than the mean; variance equal or approximately equal to the mean; variance greater than the mean.

The squared coefficient of variation (CV^2) cut-off values for all the categorisation rules developed in the chapter 6 were in the range 0.22 – 1.17. The final categorisation schemes developed (regarding all methods) were as follows:

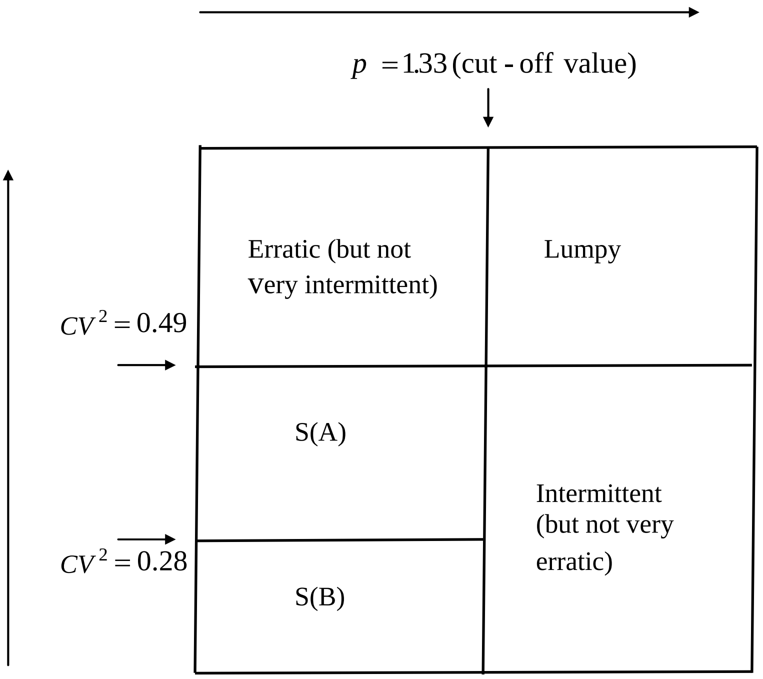

Figure 7.C.1. Categorisation of demand patterns (re-order level systems)

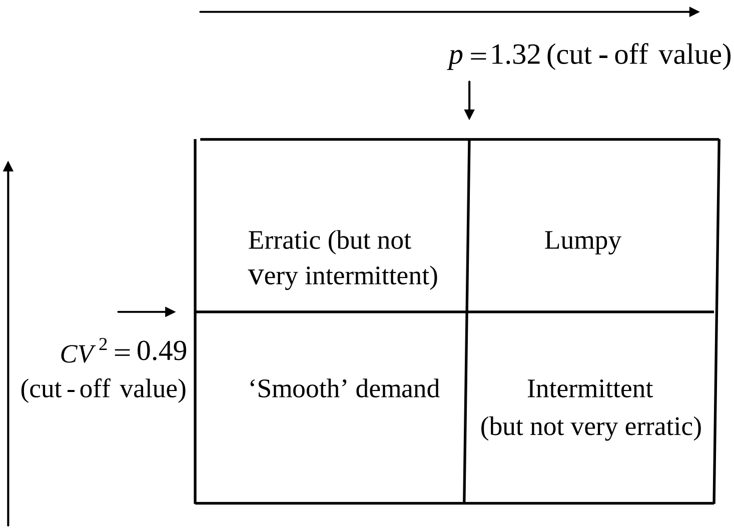

Figure 7.C.2. Categorisation of demand patterns (re-order interval systems)

In order to check the categorisation rules we require, as a minimum, to simulate demand sizes with a squared coefficient of variation in the ranges: 0 – 0.28; 0.28 – 0.49; > 0.49. For $CV^2 > 0.49$ (and in order to account for many "real world" scenarios) we wish to simulate, as mentioned above, situations where the squared coefficient of variation is close to 1 and greater than 1.

Willemain et al (1994) generated demand data for the purpose of comparing EWMA and Croston's method. The demand sizes were assumed to be lognormally distributed. The four main scenarios that they considered in generating data were the following:

1. Independent inter-demand intervals, independent sizes, independence between sizes and intervals
2. Independent intervals, auto-correlated demand sizes, independence between sizes and intervals
3. Auto-correlated intervals, independent sizes, independence between sizes and intervals
4. Independent intervals, auto-correlated sizes, cross-correlation between sizes and intervals

The assumptions associated with the first scenario are the same as those employed in our simulation experiment. The conditions considered in that case, in Willemain et al (1994), were the following:

Demand intervals	Demand sizes
Geometric ($p = 3$)	Lognormal, mean (μ) standard deviation (σ)
	1. $\mu = 2$, $\sigma = 3$, $CV^2 = 2.25$
	2. $\mu = 10$, $\sigma = 3$, $CV^2 = 0.09$
	3. $\mu = 2$, $\sigma = 0.25$, $CV^2 = 0.015625$
	4. $\mu = 10$, $\sigma = 0.25$, $CV^2 = 0.000625$

All values were selected as representative of "real world" scenarios. As such it has been decided to replicate all the above combinations. The first combination refers to the area: $CV^2 > 0.49$. The last three combinations refer to the area: $0 < CV^2 \leq 0.28$. Combination no. 3 looks the most realistic for slow moving items. In addition combination no.2 can also be used in order to ensure that "faster" moving items are represented. Finally combination no. 4 will test for all SKUs whose demand sizes are distributed with a very low variability.

We still nevertheless need to consider a combination for the decision area $0.28 < CV^2 \leq 0.49$. We have no reason to consider demand sizes other than those recommended by Willemain et al (1994) and a mean size of 10 is chosen (since we mainly test for faster moving items, “S(A)” category, see section 6.9) with a standard deviation equal to 6. Moreover, to generate CV^2 close or equal to 1 a mean demand size of 10 will be used (since the size 2 has already been considered for large values of CV^2) and the standard deviation will be also set to 10. Finally it is viewed as appropriate to generate results for $0.50 \leq CV^2 < 1$. One of the objectives of this simulation experiment is to assess the sensitivity of the categorisation rules to different CV^2 values. The CV^2 values equal to 1 and 2.25 that have already been selected are considerably higher than 0.49 to account for all cases that $CV^2 \geq 0.5$. As such, the combination L (2, 1.5) is introduced at this stage with $CV^2 = 0.5625$. The demand size was set to 2 to compensate for the higher usage of the size 10 in the experiment.

The sets of values finally selected are the following:

Mean (μ)	St. dev. (σ)	CV	CV^2	σ^2 / μ
2	3	1.50	2.25	4.50
10	10	1	1	10
2	1.50	0.75	0.5625	1.125
10	6	0.60	0.36	3.60
10	3	0.30	0.09	0.90
2	0.25	0.125	0.015625	0.03125
10	0.25	0.025	0.000625	0.00625

Table 7.C.1. Lognormally distributed demand sizes

Depending on what type of stock control system is utilised, not necessarily all estimates produced by the forecasting methods under concern are of interest to us. That is, if a continuous (re-order level) stock replenishment system is in place we are interested only in the estimates produced just after a demand occurrence (issue point) since only those estimates will be considered for replenishment purposes. On the other hand if a periodic (re-order interval or product group review) system is employed, all demand estimates are viewed as important. Therefore intermittent demand forecasting methods should be evaluated with respect to the accuracy of their estimates of the mean demand level for all points in time and for issue points only.

A wide range of lead time values has also been selected to cover for a variety of “real world” scenarios. If we assume that each review period is one month, the lead times that we have selected range from 1 period (consistent with the recent co-managed inventory schemes introduced in the automotive industry) to a lead time equal to 3 and 5 periods and finally 12 periods (an assumption still valid for many military inventory control applications).

One of the objectives of our simulation experiment is to assess the sensitivity of our theoretical results to different smoothing constant values. In chapter 2 we argued that, in an intermittent demand context, a realistic range of α smoothing constant values is 0.05 – 0.20 (see also Burgin and Wild, 1967, Croston, 1972; Willemain et al, 1994; Sani, 1995; Johnston and Boylan, 1996). As such all the categorisation rules developed in chapter 6 refer to that particular range of smoothing constant values. Moreover this is the range of values to be considered for the purpose of generating simulation results in this chapter. Four representative α smoothing constant values have been chosen from that range: 0.05 to 0.2 step 0.05. Those values cover cases that are generally encountered in practice.

The values assigned to all the control parameters are presented in the summary table 7.C.2.

Factors	System parameters: Levels	Number of levels

Inter-demand interval distribution	Geometric (mean: p) 1.1, 1.3, 1.5, 1.7, 1.9, 2, 4, 6, 8, 10	10
Demand size distribution	Lognormal L (μ, σ) L (2, 3), L (10, 10), L (2, 1.5), L (10, 6), L (10, 3) L (2, 0.25), L (10, 0.25)	7
Points in time considered	All – Issue points in time	2
Lead time length	1, 3, 5, 12	4
α smoothing constant value	0.05, 0.1, 0.15, 0.2	4

Table 7.C.2. The simulated conditions

APPENDIX 8.A

Definition of accuracy measures

Mean Absolute Error (MAE)

$$MAE = \frac{\Sigma\left|Y_{t+L} - Y'_{t,L}\right|}{n} = \frac{\Sigma\left|e_{t+L}\right|}{n}, \; t = 1, 2, \ldots\ldots., n \qquad (8.A.1)$$

Median Absolute Percentage Error (MdAPE)

The Median Absolute Percentage Error is similar to MAPE but instead of summing the Absolute Percentage Errors (APEs) and then computing their average we find their median. That is, all the APEs are sorted from the smallest to the largest and the APE in the middle (in case that there is an even number of APEs then the average of the middle two is computed) is used to denote the median.

Theil's U Statistic (U Statistic)

$$U \text{ STATISTIC} = \sqrt{\sum \frac{\left(\frac{Y_{t+L} - Y'_{t,L}}{Y_{t+L}}\right)^2}{\left(\frac{Y_{t+L} - F_{t,L}}{Y_{t+L}}\right)^2}}, \; t = 1, 2, \ldots\ldots., n \qquad (8.A.2)$$

where $F_{t,L}$ is some benchmark forecast, usually the latest available value (i.e. the forecast given by the naïve 1 forecasting method). The accuracy measure is interpreted as follows:

A value of 1 means that the accuracy of the method being used is the same as that of the benchmark method. A value smaller than 1 means that the method is better than the benchmark while a value greater than 1 means the opposite.

McLaughlin's Batting Average (Batting Average)

$$BATTING\ \text{AVERAGE} = \left[4 - \sqrt{\sum \frac{\left| \frac{Y_{t+L} - Y'_{t,L}}{Y_{t+L}} \right|}{\left| \frac{Y_{t+L} - F_{t,L}}{Y_{t+L}} \right|}} \right] (100), \quad t = 1, 2, \ldots\ldots, \text{n} \tag{8.A.3}$$

where F_t is as in the previous case the benchmark (naïve 1) forecast. In this case 300 will mean similar performance to the benchmark, 300 to 400 better performance than the benchmark and less than 300 the opposite.

Geometric Mean Relative Absolute Error (GMRAE)

The Geometric Mean Relative Absolute Error is found by calculating the geometric mean of the Relative Absolute Errors (RAE) per period.

$$RAE = \frac{\left| \frac{Y_{t+L} - Y'_{t,L}}{Y_{t+L}} \right|}{\left| \frac{Y_{t+L} - F_{t,L}}{Y_{t+L}} \right|}, \quad t = 1, 2, \ldots\ldots, \text{n} \tag{8.A.4}$$

$$GMRAE = \left(\prod_{t=1}^{n} \left[\frac{\left| \frac{Y_{t+L} - Y'_{t,L}}{Y_{t+L}} \right|}{\left| \frac{Y_{t+L} - F_{t,L}}{Y_{t+L}} \right|} \right] \right)^{\frac{1}{n}} \tag{8.A.5}$$

Median Relative Absolute Error (MdRAE)

The Median Relative Absolute Error is found by ordering the RAEs computed in (8.A.4) from the smallest to the largest and using the middle value (or the average of the middle two values if n is an even number) as the median.

APPENDIX 10.A

The bias of intermittent demand estimates

α = 0.05	*All Points in Time*					
	L.T.=1		L.T.=3		L.T.=5	
	Avg.	Var.	Avg.	Var.	Avg.	Var.
EWMA	0.08	7.46	0.33	83.00	0.67	266.50
Croston	0.12	8.02	0.45	87.32	0.88	275.35
Approx.	0.009	7.91	0.11	86.27	0.31	272.49
MA (13)	0.10	5.79	0.37	71.25	0.62	239.95
	Issue Points Only					
	L.T.=1		L.T.=3		L.T.=5	
	Avg.	Var.	Avg.	Var.	Avg.	Var.
EWMA	0.13	8.01	0.67	82.70	1.35	241.83
Croston	0.12	8.46	0.66	86.16	1.34	249.12
Approx.	0.02	8.35	0.32	84.83	0.78	244.93
MA (13)	0.17	6.36	0.74	71.07	1.35	214.76

Table 10.A.1. ME results, α = 0.05

α = 0.1	*All Points in Time*					
	L.T.=1		L.T.=3		L.T.=5	
	Avg.	Var.	Avg.	Var.	Avg.	Var.
EWMA	0.09	5.39	0.40	68.20	0.83	241.78
Croston	0.15	6.13	0.54	74.35	1.05	254.46
Approx.	-0.08	6.05	-0.15	73.22	-0.09	250.62
	Issue Points Only					
	L.T.=1		L.T.=3		L.T.=5	
	Avg.	Var.	Avg.	Var.	Avg.	Var.
EWMA	0.19	6.09	0.84	68.90	1.67	218.93
Croston	0.14	6.60	0.71	73.21	1.46	228.13
Approx.	-0.08	6.54	0.02	71.61	0.32	221.88

Table 10.A.2. ME results, α = 0.1

α = 0.15	*All Points in Time*					
	L.T.=1		L.T.=3		L.T.=5	
	Avg.	Var.	Avg.	Var.	Avg.	Var.
EWMA	0.10	4.12	0.44	58.72	0.95	227.69
Croston	0.17	4.92	0.62	65.94	1.23	242.95
Approx.	-0.17	4.93	-0.41	65.24	-0.50	238.96
	Issue Points Only					
	L.T.=1		L.T.=3		L.T.=5	
	Avg.	Var.	Avg.	Var.	Avg.	Var.
EWMA	0.23	4.98	1.00	60.59	1.98	206.97
Croston	0.15	5.41	0.76	64.84	1.58	216.46
Approx.	-0.18	5.46	-0.26	63.52	-0.14	209.24

Table 10.A.3. ME results, α = 0.15

α = 0.2	*All Points in Time*					
	L.T.=1		L.T.=3		L.T.=5	
	Avg.	Var.	Avg.	Var.	Avg.	Var.
EWMA	0.09	3.29	0.47	52.27	1.04	219.60
Croston	0.19	4.11	0.69	60.27	1.39	237.17
Approx.	-0.27	4.27	-0.69	60.35	-0.93	233.57
	Issue Points Only					
	L.T.=1		L.T.=3		L.T.=5	
	Avg.	Var.	Avg.	Var.	Avg.	Var.
EWMA	0.28	4.31	1.15	55.50	2.27	201.47
Croston	0.17	4.62	0.82	59.18	1.71	210.40
Approx.	-0.28	4.83	-0.56	58.56	-0.60	202.94

Table 10.A.4. ME results, α = 0.2

APPENDIX 10.B

The accuracy of intermittent demand estimates. Percentage Better results

Percentage Better results (ME)

α = 0.05		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**22.02** (2103 out of *3000**)	**-6.17** (1331)	**3.94** (1608)	**-16.83** (1039)	**20.33** (2057)	**-19.02** (979)
	L.T.=3	**18.43** (1976 out of *2951**)	**-3.10** (1415)	**4.34** (1619)	**-13.51** (1130)	**16.39** (1949)	**-14.71** (1097)
	L.T.=5	**13.59** (1816 out of *2900**)	0.29 (1508)	**5.14** (1641)	**-10.18** (1221)	**12.30** (1837)	**-10.62** (1209)
Issue Points Only	L.T.=1	**18.21** (1983 out of *2973**)	*1.86* (1541)	**4.63** (1627)	**-15.19** (1084)	**16.76** (1959)	**-15.04** (1088)
	L.T.=3	**15.46** (1884 out of *2931**)	**5.18** (1642)	**7.15** (1696)	**-11.46** (1186)	**13.58** (1872)	**-10.55** (1211)
	L.T.=5	**9.48** (1666 out of *2828**)	**7.37** (1702)	**8.29** (1727)	**-8.36** (1271)	**9.78** (1768)	**-6.64** (1318)

Table 10.B.1. Percentage Better results (ME), α = 0.05

*The number of ties does not change with the smoothing constant value. As such, in the following tables (PB results for $\alpha > 0.05$), and for the Croston – EWMA comparison, we present, for each of the simulated conditions, only the number of files that $ME_{EWMA} < ME_{Croston}$. The total number of files considered (3,000 minus (-) no. of ties) is the same as in the corresponding simulated scenario in the above table.

α = 0.1		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**21.65** (2093)	**-10.01** (1226)	**2.88** (1579)	*1.75* (1548)	**8.65** (1737)	**-8.22** (1275)
	L.T.=3	**17.32** (1946)	**-5.51** (1349)	**3.72** (1602)	*2.12* (1558)	**7.01** (1692)	**-4.64** (1373)
	L.T.=5	**12.33** (1782)	**-0.55** (1485)	**4.49** (1623)	0.07 (1502)	**5.04** (1638)	**-3.69** (1399)
Issue Points Only	L.T.=1	**14.65** (1886)	-0.04 (1499)	**3.61** (1599)	1.35 (1537)	**4.38** (1620)	**-2.56** (1430)
	L.T.=3	**11.95** (1789)	**3.29** (1590)	**5.81** (1659)	*2.26* (1562)	*2.23* (1561)	0.62 (1517)
	L.T.=5	**6.28** (1581)	**6.21** (1670)	**7.85** (1715)	-0.18 (1495)	**2.37** (1565)	**2.70** (1574)

Table 10.B.2. Percentage Better results (ME), α = 0.1

α = 0.15		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**20.78** (2069)	**-11.98** (1172)	1.17 (1532)	**15.77** (1932)	**-3.76** (1397)	0.51 (1514)
	L.T.=3	**15.78** (1904)	**-6.94** (1310)	**3.18** (1587)	**12.27** (1836)	**-3.61** (1401)	**2.85** (1578)
	L.T.=5	**11.51** (1760)	-1.06 (1471)	**4.13** (1613)	**6.57** (1680)	-0.66 (1482)	**2.78** (1576)
Issue Points Only	L.T.=1	**9.85** (1755)	*1.97* (1554)	*1.65* (1545)	**11.58** (1817)	**-7.63** (1291)	**5.00** (1637)
	L.T.=3	**7.70** (1674)	**5.40** (1648)	**5.37** (1647)	**9.68** (1765)	**-7.70** (1289)	**7.67** (1710)
	L.T.=5	**2.56** (1482)	**7.63** (1709)	**7.30** (1700)	**4.67** (1628)	**-3.43** (1406)	**7.12** (1695)

Table 10.B.3. Percentage Better results (ME), α = 0.15

α = 0.2		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**19.72** (2040)	**-12.82** (1149)	-0.95 (1474)	**22.42** (2114)	**-12.01** (1171)	**4.60** (1626)
	L.T.=3	**14.45** (1868)	**-7.49** (1295)	1.42 (1539)	**16.14** (1942)	**-9.46** (1241)	**5.99** (1664)
	L.T.=5	**9.92** (1717)	-1.10 (1470)	**3.76** (1603)	**8.87** (1743)	**-4.31** (1382)	**4.02** (1610)
Issue Points Only	L.T.=1	**5.01** (1623)	1.42 (1539)	-0.29 (1492)	**14.97** (1910)	**-15.92** (1064)	**6.46** (1677)
	L.T.=3	**2.57** (1535)	**6.17** (1669)	**4.45** (1622)	**12.01** (1829)	**-13.18** (1139)	**10.01** (1774)
	L.T.=5	-0.68 (1396)	**8.62** (1736)	**6.46** (1677)	**5.26** (1644)	**-7.16** (1304)	**8.25** (1726)

Table 10.B.4. Percentage Better results (ME), α = 0.2

Percentage Better results (GRMSE)

α = 0.05		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**5.00** (1637 out of *3000**)	**4.49** (1623)	**11.39** (1812)	**10.04** (1775)	**-7.74** (1288)	**9.93** (1772)
	L.T.=3	**2.45** (1542 out of *2951**)	**6.24** (1671)	**10.63** (1791)	**6.13** (1668)	**-4.82** (1368)	**6.97** (1691)
	L.T.=5	*1.67* (1495 out of *2900**)	**6.06** (1666)	**8.51** (1733)	**4.13** (1613)	**-2.67** (1427)	**6.68** (1683)
Issue Points Only	L.T.=1	**-3.10** (1402 out of *2973*)	**10.44** (1786)	**10.33** (1783)	**8.18** (1724)	**-8.14** (1277)	**10.08** (1776)
	L.T.=3	**-4.23** (1351 out of *2931**)	**11.43** (1813)	**10.26** (1781)	**5.22** (1643)	**-5.44** (1351)	**7.41** (1703)
	L.T.=5	*-2.14* (1357 out of *2828**)	**11.47** (1814)	**9.20** (1752)	**2.81** (1577)	**-3.18** (1413)	**6.43** (1676)

Table 10.B.5. Percentage Better results (GRMSE), α = 0.05

*The number of ties does not change with the smoothing constant value. As such, in the following tables (PB results for $\alpha > 0.05$), and for the Croston – EWMA comparison, we present, for each of the simulated conditions, only the number of files that $GRMSE_{EWMA} < GRMSEE_{Croston}$. The total number of files considered (3,000 minus (-) no. of ties) is the same as in the corresponding simulated scenario in the above table.

α = 0.1		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**2.41** (1566)	**7.89** (1716)	**12.49** (1842)	**9.68** (1765)	**-7.78** (1287)	**13.51** (1870)
	L.T.=3	*1.90* (1527)	**7.67** (1710)	**12.42** (1840)	**2.81** (1577)	**-2.67** (1427)	**8.58** (1735)
	L.T.=5	-1.30 (1415)	**8.14** (1723)	**9.09** (1749)	1.24 (1534)	-1.39 (1462)	**5.22** (1643)
Issue Points Only	L.T.=1	**-5.30** (1342)	**12.82** (1851)	**10.92** (1799)	**7.30** (1700)	**-9.82** (1231)	**14.13** (1887)
	L.T.=3	**-5.67** (1312)	**13.33** (1865)	**12.23** (1835)	1.20 (1533)	**-3.07** (1416)	**8.76** (1740)
	L.T.=5	**-6.06** (1253)	**13.07** (1858)	**10.26** (1781)	-0.07 (1498)	-1.53 (1458)	**7.52** (1706)

Table 10.B.6. Percentage Better results (GRMSE), α = 0.1

α = 0.15		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	*2.15* (1559)	**8.54** (1734)	**13.00** (1856)	**7.45** (1704)	**-6.65** (1318)	**14.53** (1898)
	L.T.=3	-0.83 (1453)	**7.49** (1705)	**11.79** (1823)	-0.47 (1487)	1.28 (1535)	**5.55** (1692)
	L.T.=5	*-2.27* (1389)	**8.40** (1730)	**9.42** (1758)	*-2.19* (1440)	0.22 (1506)	**3.61** (1599)
Issue Points Only	L.T.=1	**-7.35** (1286)	**13.84** (1879)	**11.87** (1825)	**3.54** (1597)	**-7.38** (1298)	**13.58** (1872)
	L.T.=3	**-9.99** (1195)	**13.55** (1871)	**10.44** (1786)	**-3.25** (1411)	0.22 (1506)	**6.10** (1667)
	L.T.=5	**-8.24** (1195)	**13.25** (1863)	**10.01** (1774)	**-2.45** (1433)	0.04 (1501)	**4.09** (1612)

Table 10.B.7. Percentage Better results (GRMSE), α = 0.15

α = 0.2		**Croston-EWMA**	**EWMA-Approx.**	**Croston-Approx.**	**MA-EWMA**	**Croston-MA**	**MA-Approx.**
All Points in Time	L.T.=1	**2.92** (1580)	**9.09** (1749)	**15.08** (1913)	**5.73** (1657)	**-4.75** (1370)	**13.99** (1833)
	L.T.=3	-1.31 (1440)	**10.08** (1776)	**12.45** (1841)	**-4.56** (1375)	**3.98** (1609)	**2.78** (1576)
	L.T.=5	**-4.42** (1331)	**10.41** (1785)	**8.80** (1741)	**-7.05** (1307)	**3.47** (1595)	0.00 (1500)
Issue Points Only	L.T.=1	**-7.17** (1291)	**14.57** (1899)	**13.29** (1864)	*1.94* (1553)	**-4.45** (1378)	**13.36** (1866)
	L.T.=3	**-11.03** (1167)	**14.46** (1896)	**11.10** (1804)	**-5.59** (1347)	**2.45** (1567)	**2.67** (1573)
	L.T.=5	**-11.96** (1096)	**13.66** (1874)	**9.79** (1768)	**-6.24** (1329)	**2.99** (1582)	0.37 (1510)

Table 10.B.8. Percentage Better results (GRMSE), α = 0.2

APPENDIX 10.C

The accuracy of intermittent demand estimates.

Percentage Best results

No. of files (Percentage Best - GRMSE)

α = 0.05		EWMA	Croston	Approx.	MA (13)	Ties
All Points in Time	L.T.=1	697	479	861	963	0
	L.T.=3	626	494	869	1006	5
	L.T.=5	595	477	873	1033	22
Issue Points only	L.T.=1	543	435	1008	1007	7
	L.T.=3	511	441	987	1051	10
	L.T.=5	507	422	960	1075	36

Table 10.C.1. No. of files (superior performance), GRMSE, α = 0.05

α = 0.1		EWMA	Croston	Approx.	MA (13)	Ties
All Points in Time	L.T.=1	697	530	1038	735	0
	L.T.=3	666	480	1024	820	10
	L.T.=5	616	553	923	897	11
Issue Points only	L.T.=1	567	452	1195	776	10
	L.T.=3	524	425	1148	890	13
	L.T.=5	510	459	1099	903	29

Table 10.C.2. No. of files (superior performance), GRMSE, α = 0.1

α = 0.15		EWMA	Croston	Approx.	MA (13)	Ties
All Points in Time	L.T.=1	713	544	1080	663	0
	L.T.=3	697	491	968	838	6
	L.T.=5	638	516	944	891	11
Issue Points only	L.T.=1	547	447	1242	756	8
	L.T.=3	512	435	1128	917	8
	L.T.=5	509	414	1121	926	30

Table 10.C.3. No. of files (superior performance), GRMSE, α = 0.15

α = 0.2		EWMA	Croston	Approx.	MA (13)	Ties
All Points in Time	L.T.=1	748	509	1119	624	0
	L.T.=3	633	481	975	902	9
	L.T.=5	577	511	922	979	11
Issue Points only	L.T.=1	564	411	1285	731	9
	L.T.=3	488	419	1116	963	14
	L.T.=5	464	394	1097	1018	27

Table 10.C.4. No. of files (superior performance), GRMSE, α = 0.2

<u>No. of files (Percentage Best – ME)</u>

α = 0.05		EWMA	Croston	Approx.	MA (13)	Ties
All Points in Time	L.T.=1	391	292	550	1767	0
	L.T.=3	435	324	600	1641	0
	L.T.=5	461	378	664	1486	11
Issue Points only	L.T.=1	393	193	717	1696	1
	L.T.=3	437	206	807	1549	1
	L.T.=5	477	232	871	1405	15

Table 10.C.5. No. of files (superior performance), ME, α = 0.05

α = 0.1		EWMA	Croston	Approx.	MA (13)	Ties
All Points in Time	L.T.=1	734	406	754	1106	0
	L.T.=3	714	413	813	1052	8
	L.T.=5	700	443	826	1017	14
Issue Points only	L.T.=1	669	260	1081	984	6
	L.T.=3	674	262	1118	935	11
	L.T.=5	649	246	1151	927	27

Table 10.C.6. No. of files (superior performance), ME, α = 0.1

α = 0.15		EWMA	Croston	Approx.	MA (13)	Ties
All Points in Time	L.T.=1	963	490	880	667	0
	L.T.=3	859	488	936	707	10
	L.T.=5	782	484	931	789	14
Issue Points only	L.T.=1	838	269	1284	602	7
	L.T.=3	766	262	1332	628	12
	L.T.=5	686	256	1294	736	28

Table 10.C.7. No. of files (superior performance), ME, α = 0.15

α = 0.2		EWMA	Croston	Approx.	MA (13)	Ties
All Points in Time	L.T.=1	993	579	929	499	0
	L.T.=3	873	543	987	586	11
	L.T.=5	751	511	991	734	13
Issue Points only	L.T.=1	889	298	1319	485	9
	L.T.=3	789	290	1395	512	14
	L.T.=5	685	270	1345	672	28

Table 10.C.8. No. of files (superior performance), ME, α = 0.2

APPENDIX 10.D

The categorisation of "non-normal" demand patterns

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
		α (smoothing constant value) = 0.05					
L.T.=1	non-smooth	0.99	0.99	0.98	0.93	1.06	0.92
	smooth	0.99	0.99	0.99	0.95	1.05	0.94
L.T.=3	non-smooth	0.99	0.99	0.98	0.97	1.02	0.95
	smooth	0.99	0.99	0.98	0.98	1.01	0.97
L.T.=5	non-smooth	0.98	0.99	0.98	0.97	1.01	0.97
	smooth	1.00	0.98	0.98	0.99	1.01	0.97
		α (smoothing constant value) = 0.1					
L.T.=1	non-smooth	0.99	0.97	0.96	0.93	1.06	0.91
	smooth	1.00	0.99	0.99	0.94	1.06	0.93
L.T.=3	non-smooth	0.98	0.97	0.95	0.96	1.02	0.94
	smooth	0.99	0.98	0.97	0.98	1.01	0.96
L.T.=5	non-smooth	0.99	0.97	0.96	0.98	1.01	0.94
	smooth	1.00	0.98	0.98	0.99	1.01	0.97
		α (smoothing constant value) = 0.15					
L.T.=1	non-smooth	0.99	0.96	0.95	0.93	1.06	0.90
	smooth	1.00	0.97	0.97	0.95	1.05	0.92
L.T.=3	non-smooth	0.98	0.96	0.94	0.96	1.02	0.93
	smooth	1.00	0.97	0.97	1.00	1.00	0.97
L.T.=5	non-smooth	1.01	0.94	0.95	0.99	1.02	0.93
	smooth	1.01	0.97	0.98	1.01	1.00	0.98
		α (smoothing constant value) = 0.2					
L.T.=1	non-smooth	0.98	0.95	0.94	0.93	1.05	0.89
	smooth	0.99	0.97	0.96	0.95	1.04	0.92
L.T.=3	non-smooth	0.99	0.94	0.93	0.98	1.01	0.92
	smooth	1.00	0.96	0.96	1.02	0.98	0.98
L.T.=5	non-smooth	1.02	0.91	0.93	1.02	1.00	0.93
	smooth	1.01	0.95	0.96	1.04	0.97	1.00

Table 10.D.1. RGRMSE, all points in time

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
		α (smoothing constant value) = 0.05					
L.T.=1	non-smooth	0.06	0.78	0.84	**2.77**	**-2.71**	**3.56**
	smooth A	0.19	0.43	0.61	*2.23*	*-2.04*	**2.67**
	smooth B	-0.17	0.50	0.33	**3.14**	**-3.30**	**3.65**
L.T.=3	non-smooth	-0.24	1.11	0.86	0.97	-1.21	*2.10*
	smooth A	0.05	0.67	0.72	0.63	-0.58	1.32
	smooth B	0.13	0.47	0.60	*2.05*	*-1.92*	**2.51**
L.T.=5	non-smooth	0.12	0.77	0.89	0.85	-0.72	*1.65*
	smooth A	0.14	0.53	0.67	0.08	0.07	0.61
	smooth B	0.01	1.09	1.10	1.27	-1.26	**2.38**
		α (smoothing constant value) = 0.1					
L.T.=1	non-smooth	-0.37	*1.78*	1.41	**2.59**	**-2.98**	**4.40**
	smooth A	-0.13	1.07	0.94	**2.47**	**-2.60**	**3.56**
	smooth B	0.09	-0.39	-0.30	**4.26**	**-4.16**	**3.85**
L.T.=3	non-smooth	-0.35	*1.97*	1.62	0.90	-1.25	**2.89**
	smooth A	0.12	1.19	1.32	0.47	-0.35	*1.67*
	smooth B	0.06	1.30	1.37	1.60	-1.55	**2.92**
L.T.=5	non-smooth	-0.17	*1.96*	*1.79*	0.71	-0.89	**2.71**
	smooth A	-0.16	0.91	0.74	-0.28	0.11	0.64
	smooth B	0.04	1.44	1.48	1.30	-1.26	**2.78**

Table 10.D.2. *t*-test (GRMSE, difference between population means), issue points only (α = 0.05, 0.1)

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
		α (smoothing constant value) = 0.15					
L.T.=1	non-smooth	-0.58	**2.35**	*1.76*	**2.41**	**-3.01**	**4.83**
	smooth A	-0.28	*2.23*	*1.95*	*1.91*	*-2.20*	**4.18**
	smooth B	-0.22	0.51	0.29	**3.47**	**-3.70**	**3.97**
L.T.=3	non-smooth	-0.49	**2.59**	*2.10*	0.76	-1.26	**3.40**
	smooth A	-0.50	*1.66*	1.17	-0.30	-0.21	1.39
	smooth B	-0.45	1.55	1.10	0.56	-1.02	*2.14*
L.T.=5	non-smooth	-1.18	**3.35**	*2.16*	0.01	-1.21	**3.42**
	smooth A	-0.42	1.52	1.09	-1.01	0.58	0.51
	smooth B	-0.44	1.10	0.64	0.82	-1.26	*1.94*
		α (smoothing constant value) = 0.2					
L.T.=1	non-smooth	-0.51	**2.75**	*2.26*	*2.22*	**-2.76**	**5.10**
	smooth A	-0.40	**3.25**	**2.85**	1.52	*-1.93*	**4.84**
	smooth B	-0.03	0.84	0.80	**3.31**	**-3.33**	**4.13**
L.T.=3	non-smooth	-1.32	**4.01**	**2.68**	-0.19	-1.15	**3.87**
	smooth A	-0.53	*1.77*	1.25	-1.12	0.59	0.67
	smooth B	-0.01	*2.29*	*2.27*	-0.34	0.33	*2.00*
L.T.=5	non-smooth	*-1.82*	**4.61**	**2.76**	-1.19	-0.65	**3.47**
	smooth A	-0.69	*2.24*	1.54	**-2.41**	*1.71*	-0.19
	smooth B	-0.12	*1.81*	*1.69*	-0.23	0.11	1.61

Table 10.D.3. *t*-test (GRMSE, difference between population means), issue points only (α = 0.15, 0.2)

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
		α (smoothing constant value) = 0.05					
L.T.=1	non-smooth	1.00	0.98	0.98	0.93	1.07	0.92
	smooth A	1.00	0.99	0.99	0.96	1.04	0.95
	smooth B	1.00	0.99	0.99	0.94	1.07	0.93
L.T.=3	non-smooth	1.01	0.97	0.98	0.97	1.03	0.95
	smooth A	1.00	0.98	0.98	0.99	1.01	0.97
	smooth B	1.00	0.99	0.99	0.96	1.04	0.95
L.T.=5	non-smooth	1.00	0.98	0.97	0.98	1.02	0.95
	smooth A	1.00	0.99	0.98	1.00	1.00	0.99
	smooth B	1.00	0.97	0.97	0.97	1.03	0.94
		α (smoothing constant value) = 0.1					
L.T.=1	non-smooth	1.01	0.96	0.96	0.94	1.08	0.90
	smooth A	1.00	0.98	0.98	0.95	1.05	0.93
	smooth B	1.00	1.01	1.01	0.92	1.08	0.93
L.T.=3	non-smooth	1.01	0.95	0.96	0.98	1.03	0.93
	smooth A	1.00	0.97	0.97	0.99	1.01	0.96
	smooth B	1.00	0.97	0.97	0.97	1.03	0.94
L.T.=5	non-smooth	1.00	0.94	0.95	0.98	1.03	0.93
	smooth A	1.00	0.98	0.98	1.01	1.00	0.98
	smooth B	1.00	0.97	0.97	0.97	1.03	0.94

Table 10.D.4. RGRMSE results, issue points only (α = 0.05, 0.1)

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
		α (smoothing constant value) = 0.15					
L.T.=1	non-smooth	1.01	0.94	0.96	0.94	1.08	0.89
	smooth A	1.01	0.95	0.96	0.96	1.05	0.92
	smooth B	1.00	0.99	0.99	0.94	1.07	0.93
L.T.=3	non-smooth	1.01	0.93	0.95	0.98	1.03	0.92
	smooth A	1.01	0.96	0.97	1.01	1.00	0.97
	smooth B	1.01	0.97	0.98	0.99	1.02	0.96
L.T.=5	non-smooth	1.03	0.91	0.94	1.00	1.03	0.91
	smooth A	1.01	0.96	0.97	1.03	0.99	0.99
	smooth B	1.01	0.97	0.99	0.98	1.03	0.96
		α (smoothing constant value) = 0.2					
L.T.=1	non-smooth	1.01	0.93	0.94	0.95	1.07	0.88
	smooth A	1.01	0.93	0.94	0.97	1.04	0.91
	smooth B	1.00	0.98	0.98	0.94	1.07	0.92
L.T.=3	non-smooth	1.04	0.90	0.93	1.00	1.03	0.90
	smooth A	1.01	0.96	0.97	1.03	0.99	0.99
	smooth B	1.00	0.95	0.95	1.01	0.99	0.96
L.T.=5	non-smooth	1.05	0.88	0.92	1.03	1.02	0.91
	smooth A	1.02	0.95	0.96	1.06	0.96	1.00
	smooth B	1.00	0.96	0.96	1.01	1.00	0.96

Table 10.D.5. RGRMSE results, issue points only (α = 0.15, 0.2)

APPENDIX 11.A

Inventory control results (P_2 policy)

		EWMA	Croston	Approx.	MA(13)	Ties
		Customer Service Level				
α = 0.05	L.T. = 1	693	42	54	32	2179
	L.T. = 3	515	51	48	36	2350
	L.T. = 5	350	39	55	37	2519
α = 0.1	L.T. = 1	792	56	75	4	2073
	L.T. = 3	598	52	74	0	2276
	L.T. = 5	454	39	67	0	2440
α = 0.15	L.T. = 1	751	50	81	6	2112
	L.T. = 3	597	55	81	0	2267
	L.T. = 5	442	44	82	0	2432
α = 0.2	L.T. = 1	690	49	93	4	2164
	L.T. = 3	552	53	106	0	2289
	L.T. = 5	428	45	104	0	2423
		Number of units in stock				
α = 0.05	L.T. = 1	446	253	704	655	942
	L.T. = 3	159	438	1042	744	617
	L.T. = 5	50	495	1258	649	548
α = 0.1	L.T. = 1	489	335	928	629	619
	L.T. = 3	176	494	1307	618	405
	L.T. = 5	37	571	1641	451	300
α = 0.15	L.T. = 1	418	413	1066	647	456
	L.T. = 3	167	543	1446	539	305
	L.T. = 5	38	561	1732	395	274
α = 0.2	L.T. = 1	343	528	1091	737	301
	L.T. = 3	140	615	1431	579	235
	L.T. = 5	35	553	1694	422	296

Table 11.A.1. Best performance (number of files), P_2 = 0.90

		EWMA	Croston	Approx.	MA(13)	Ties
		Customer Service Level				
α = 0.05	L.T. = 1	504	31	50	31	2384
	L.T. = 3	390	35	57	28	2490
	L.T. = 5	270	36	47	25	2622
α = 0.1	L.T. = 1	545	39	67	7	2342
	L.T. = 3	461	42	68	0	2429
	L.T. = 5	339	37	63	0	2561
α = 0.15	L.T. = 1	515	39	78	3	2365
	L.T. = 3	423	46	76	0	2455
	L.T. = 5	347	42	70	0	2541
α = 0.2	L.T. = 1	459	35	79	2	2425
	L.T. = 3	393	46	81	0	2480
	L.T. = 5	321	40	83	0	2556
		Number of units in stock				
α = 0.05	L.T. = 1	572	245	717	667	799
	L.T. = 3	202	403	1042	788	565
	L.T. = 5	65	525	1267	690	453
α = 0.1	L.T. = 1	620	347	926	606	501
	L.T. = 3	242	481	1288	661	328
	L.T. = 5	62	524	1604	461	349
α = 0.15	L.T. = 1	550	419	1047	659	325
	L.T. = 3	232	531	1412	602	223
	L.T. = 5	65	545	1731	418	241
α = 0.2	L.T. = 1	451	526	1042	711	270
	L.T. = 3	198	602	1428	595	177
	L.T. = 5	71	581	1717	439	192

Table 11.A.2. Best performance (number of files), P_2 = 0.95

APPENDIX 11.B

Inventory control results (B_2 , B_3 policies)

		B_2 = 0.93				B_2 = 0.96			
		EWMA	Croston	Approx.	MA(13)	EWMA	Croston	Approx.	MA(13)
α =0.05	L.T.=1	0.947	0.930	0.926	0.928	0.962	0.951	0.949	0.950
	L.T.=3	0.942	0.926	0.921	0.923	0.956	0.944	0.941	0.943
	L.T.=5	0.944	0.934	0.929	0.932	0.957	0.949	0.946	0.948
α =0.1	L.T.=1	0.961	0.937	0.931	0.936	0.972	0.957	0.953	0.956
	L.T.=3	0.954	0.931	0.922	0.927	0.965	0.949	0.943	0.946
	L.T.=5	0.954	0.937	0.928	0.934	0.965	0.952	0.945	0.950
α =0.15	L.T.=1	0.969	0.943	0.934	0.941	0.978	0.962	0.955	0.960
	L.T.=3	0.961	0.934	0.921	0.931	0.971	0.952	0.943	0.950
	L.T.=5	0.959	0.939	0.926	0.936	0.969	0.954	0.944	0.951
α =0.2	L.T.=1	0.973	0.947	0.935	0.945	0.981	0.965	0.957	0.964
	L.T.=3	0.965	0.937	0.920	0.934	0.974	0.955	0.942	0.952
	L.T.=5	0.963	0.941	0.923	0.937	0.972	0.955	0.942	0.953
Average		**0.958**	**0.936**	**0.926**	**0.934**	**0.969**	**0.954**	**0.947**	**0.952**

Table 11.B.1. Customer Service Level results (B_2 policy)

		B_3 = 0.95				B_3 = 0.98			
		EWMA	Croston	Approx.	MA(13)	EWMA	Croston	Approx.	MA(13)
α =0.05	L.T.=1	0.965	0.954	0.952	0.952	0.977	0.972	0.971	0.971
	L.T.=3	0.965	0.956	0.953	0.953	0.976	0.971	0.969	0.969
	L.T.=5	0.966	0.960	0.958	0.958	0.977	0.973	0.971	0.972
α =0.1	L.T.=1	0.973	0.959	0.955	0.957	0.983	0.976	0.973	0.975
	L.T.=3	0.971	0.958	0.953	0.956	0.980	0.973	0.969	0.971
	L.T.=5	0.971	0.961	0.956	0.960	0.980	0.974	0.970	0.973
α =0.15	L.T.=1	0.978	0.963	0.957	0.961	0.986	0.979	0.975	0.978
	L.T.=3	0.974	0.960	0.953	0.959	0.983	0.974	0.969	0.973
	L.T.=5	0.973	0.962	0.954	0.960	0.982	0.974	0.969	0.973
α =0.2	L.T.=1	0.981	0.966	0.958	0.965	0.988	0.981	0.976	0.980
	L.T.=3	0.976	0.962	0.952	0.961	0.984	0.975	0.969	0.974
	L.T.=5	0.974	0.963	0.952	0.961	0.983	0.975	0.968	0.974
Average		**0.972**	**0.960**	**0.954**	**0.959**	**0.982**	**0.975**	**0.971**	**0.974**

Table 11.B.2. Customer Service Level results (B_3 policy)

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**21.17**	**-22.37**	**-3.10**	**20.27**	*1.73*	**-4.64**
	L.T.=3	**18.07**	**-19.00**	*-1.83*	**17.11**	1.60	**-3.37**
	L.T.=5	**14.09**	**-15.05**	-1.51	**14.56**	-0.72	-0.79
α = 0.1	L.T.=1	**23.52**	**-24.19**	*-1.68*	**25.19**	**-5.21**	**4.01**
	L.T.=3	**19.96**	**-20.09**	-0.24	**22.45**	**-6.10**	**5.93**
	L.T.=5	**16.87**	**-16.58**	0.45	**19.71**	**-6.18**	**6.50**
α = 0.15	L.T.=1	**23.25**	**-23.61**	-0.84	**25.05**	**-5.40**	**4.81**
	L.T.=3	**19.20**	**-19.65**	-0.76	**22.35**	**-6.88**	**6.35**
	L.T.=5	**16.30**	**-16.64**	-0.53	**19.83**	**-6.96**	**6.58**
α = 0.2	L.T.=1	**21.79**	**-21.91**	-0.26	**23.74**	**-5.49**	**5.30**
	L.T.=3	**18.10**	**-17.83**	0.41	**21.36**	**-6.88**	**7.17**
	L.T.=5	**15.74**	**-15.18**	0.81	**19.16**	**-6.73**	**7.31**

Table 11.B.3. PBt CSL results, B_2 policy (target value = 0.96)

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	0.14	1.39	1.67	1.50
	L.T.=3	0.27	1.66	2.05	1.71
	L.T.=5	0.29	1.22	1.66	1.28
α = 0.1	L.T.=1	0.06	1.70	2.17	1.86
	L.T.=3	0.16	2.05	2.74	2.30
	L.T.=5	0.20	1.68	2.46	1.92
α = 0.15	L.T.=1	0.06	1.80	2.47	1.94
	L.T.=3	0.16	2.20	3.27	2.50
	L.T.=5	0.19	1.91	3.06	2.21
α = 0.2	L.T.=1	0.06	1.75	2.58	1.91
	L.T.=3	0.16	2.22	3.58	2.51
	L.T.=5	0.20	2.05	3.52	2.37

Table 11.B.4. APR CSL results, B_2 policy (target value = 0.96)

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**-21.01**	**29.70**	**11.52**	**-33.80**	**16.73**	**-5.40**
	L.T.=3	**-24.61**	**35.27**	**13.97**	**-32.33**	**10.19**	**3.86**
	L.T.=5	**-23.80**	**39.76**	**20.41**	**-27.94**	**5.43**	**15.22**
α = 0.1	L.T.=1	**-24.63**	**38.10**	**17.64**	**-28.94**	**5.79**	**12.04**
	L.T.=3	**-28.60**	**42.89**	**18.57**	**-25.72**	**-3.87**	**22.23**
	L.T.=5	**-27.16**	**48.12**	**26.86**	**-22.22**	**-6.68**	**32.79**
α = 0.15	L.T.=1	**-25.58**	**42.71**	**22.19**	**-26.06**	0.65	**21.58**
	L.T.=3	**-29.10**	**46.49**	**22.43**	**-22.16**	**-9.31**	**30.92**
	L.T.=5	**-27.14**	**51.15**	**30.57**	**-19.13**	**-10.77**	**39.70**
α = 0.2	L.T.=1	**-25.92**	**44.76**	**24.29**	**-24.71**	*-1.64*	**25.79**
	L.T.=3	**-29.43**	**47.65**	**23.45**	**-21.98**	**-9.99**	**32.50**
	L.T.=5	**-26.72**	**51.62**	**31.64**	**-20.34**	**-8.61**	**38.94**

Table 11.B.5. PBt stock results, B_2 policy (target value = 0.96)

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	22.21	6.08	3.85	5.95
	L.T.=3	19.33	6.69	3.84	6.75
	L.T.=5	13.00	6.61	3.50	6.09
α = 0.1	L.T.=1	40.43	6.87	2.88	5.56
	L.T.=3	36.58	8.13	3.35	6.78
	L.T.=5	25.71	9.04	3.31	7.13
α = 0.15	L.T.=1	60.74	8.05	2.73	6.48
	L.T.=3	55.13	10.27	3.51	8.41
	L.T.=5	38.72	11.64	3.58	8.92
α = 0.2	L.T.=1	83.43	9.38	2.90	7.74
	L.T.=3	75.23	12.72	4.35	10.64
	L.T.=5	52.44	14.63	4.60	11.32

Table 11.B.6. APR stock results, B_2 policy (target value = 0.96)

			Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	% stock	1.24	0.49	2.05	1.18	-0.89	1.64
		% CSL	-0.17	-0.11	-0.28	-0.04	-0.12	-0.16
	L.T.=3	% stock	-12.46	12.24	2.44	-12.14	-0.87	1.96
		% CSL	1.19	-1.53	-0.33	1.32	-0.13	-0.21
	L.T.=5	% stock	-6.82	7.87	2.69	-6.73	-0.43	1.77
		% CSL	0.77	-1.14	-0.37	0.89	-0.12	-0.25
α = 0.1	L.T.=1	% stock	-31.46	24.71	3.39	-32.99	0.78	2.04
		% CSL	1.50	-1.93	-0.43	1.65	-0.15	-0.28
	L.T.=3	% stock	-26.68	22.65	3.91	-27.86	0.60	2.67
		% CSL	1.65	-2.27	-0.62	1.91	-0.26	-0.37
	L.T.=5	% stock	-15.96	16.62	4.67	-17.28	0.93	3.00
		% CSL	1.25	-1.93	-0.68	1.48	-0.22	-0.45
α = 0.15	L.T.=1	% stock	-48.56	33.21	4.39	-50.68	1.09	2.95
		% CSL	1.62	-2.25	-0.63	1.75	-0.13	-0.50
	L.T.=3	% stock	-40.79	30.66	5.24	-42.91	1.12	3.79
		% CSL	1.82	-2.80	-0.98	2.10	-0.28	-0.70
	L.T.=5	% stock	-24.71	23.34	6.17	-27.26	1.65	4.14
		% CSL	1.48	-2.52	-1.04	1.75	-0.27	-0.77
α = 0.2	L.T.=1	% stock	-67.27	40.40	5.14	-69.83	1.17	3.72
		% CSL	1.58	-2.37	-0.79	1.73	-0.16	-0.63
	L.T.=3	% stock	-55.40	36.98	5.99	-58.17	1.29	4.53
		% CSL	1.87	-3.13	-1.26	2.12	-0.25	-1.01
	L.T.=5	% stock	-33.28	28.54	7.06	-36.89	2.10	4.82
		% CSL	1.63	-2.97	-1.35	1.90	-0.27	-1.08

Table 11.B.7. Stock and CSL % differences (B_2 policy, target value = 0.96)

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**22.68**	**-23.64**	**-2.70**	**21.81**	*1.82*	**-4.34**
	L.T.=3	**18.85**	**-19.89**	*-2.11*	**19.19**	-0.64	-1.48
	L.T.=5	**14.48**	**-15.36**	-1.46	**14.96**	-0.77	-0.70
α = 0.1	L.T.=1	**25.64**	**-26.40**	*-2.20*	**27.05**	**-4.91**	**3.32**
	L.T.=3	**20.10**	**-21.17**	*-1.98*	**23.10**	**-6.81**	**5.40**
	L.T.=5	**15.79**	**-17.22**	*-2.25*	**19.51**	**-7.10**	**5.49**
α = 0.15	L.T.=1	**24.64**	**-25.35**	*-1.90*	**26.20**	**-5.11**	**3.75**
	L.T.=3	**19.21**	**-19.84**	-1.04	**22.63**	**-7.24**	**6.50**
	L.T.=5	**15.05**	**-15.68**	-0.93	**18.98**	**-7.24**	**6.58**
α = 0.2	L.T.=1	**23.43**	**-23.85**	-0.98	**25.22**	**-5.40**	**4.70**
	L.T.=3	**18.20**	**-17.88**	0.49	**21.68**	**-7.17**	**7.52**
	L.T.=5	**14.77**	**-14.14**	0.88	**18.57**	**-7.03**	**7.65**

Table 11.B.8. PBt CSL results, B_3 policy (target value = 0.95)

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	0.11	1.23	1.48	1.40
	L.T.=3	0.17	1.17	1.46	1.38
	L.T.=5	0.20	0.86	1.17	1.05
α = 0.1	L.T.=1	0.06	1.57	2.04	1.76
	L.T.=3	0.13	1.52	2.07	1.74
	L.T.=5	0.15	1.22	1.79	1.37
α = 0.15	L.T.=1	0.05	1.69	2.33	1.85
	L.T.=3	0.13	1.64	2.45	1.82
	L.T.=5	0.15	1.36	2.24	1.52
α = 0.2	L.T.=1	0.06	1.70	2.49	1.83
	L.T.=3	0.15	1.71	2.79	1.86
	L.T.=5	0.18	1.46	2.57	1.59

Table 11.B.9. APR CSL results, B_3 policy (target value = 0.95)

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**-20.88**	**29.66**	**11.68**	**-33.94**	**17.14**	**-5.66**
	L.T.=3	**-24.75**	**35.38**	**13.91**	**-31.34**	**8.71**	**5.30**
	L.T.=5	**-24.25**	**39.49**	**19.41**	**-27.38**	**4.08**	**15.50**
α = 0.1	L.T.=1	**-23.89**	**39.07**	**19.81**	**-28.36**	**6.01**	**14.05**
	L.T.=3	**-27.92**	**43.45**	**20.14**	**-25.65**	**-3.05**	**23.00**
	L.T.=5	**-27.53**	**47.79**	**25.99**	**-22.27**	**-7.10**	**32.33**
α = 0.15	L.T.=1	**-24.71**	**43.99**	**24.84**	**-25.07**	0.50	**24.39**
	L.T.=3	**-28.54**	**46.41**	**23.04**	**-23.15**	**-7.24**	**29.66**
	L.T.=5	**-27.16**	**50.59**	**29.86**	**-20.40**	**-9.12**	**37.70**
α = 0.2	L.T.=1	**-23.96**	**47.23**	**29.67**	**-23.81**	-0.20	**29.85**
	L.T.=3	**-28.67**	**46.91**	**23.49**	**-23.35**	**-7.15**	**30.01**
	L.T.=5	**-27.21**	**50.57**	**29.78**	**-21.47**	**-7.74**	**36.50**

Table 11.B.10. PBt stock results, B_3 policy (target value = 0.95)

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	23.18	6.11	3.86	6.37
	L.T.=3	19.00	6.38	3.57	6.63
	L.T.=5	12.82	6.02	3.18	6.02
α = 0.1	L.T.=1	42.66	6.87	2.68	6.00
	L.T.=3	36.40	7.92	2.94	6.71
	L.T.=5	25.21	8.31	2.82	6.87
α = 0.15	L.T.=1	64.26	8.01	2.39	6.76
	L.T.=3	55.32	9.92	3.13	8.10
	L.T.=5	38.03	10.80	3.11	8.56
α = 0.2	L.T.=1	88.68	9.57	2.39	8.20
	L.T.=3	75.68	12.24	3.79	10.21
	L.T.=5	51.24	13.53	3.98	10.62

Table 11.B.11. APR stock results, B_3 policy (target value = 0.95)

			Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	% stock	-16.42	14.79	1.95	-16.25	-0.95	1.57
		% CSL	1.02	-1.26	-0.24	1.21	-0.19	-0.05
	L.T.=3	% stock	-12.41	12.36	2.44	-11.98	-0.99	2.17
		% CSL	0.90	-1.16	-0.26	1.13	-0.23	-0.03
	L.T.=5	% stock	-7.10	8.13	2.46	-6.65	-0.78	2.10
		% CSL	0.58	-0.86	-0.28	0.77	-0.19	-0.09
α = 0.1	L.T.=1	% stock	-33.58	26.21	3.60	-34.63	0.35	2.62
		% CSL	1.42	-1.86	-0.44	1.60	-0.19	-0.25
	L.T.=3	% stock	-26.74	23.09	4.15	-27.86	0.52	3.02
		% CSL	1.28	-1.78	-0.50	1.50	-0.22	-0.28
	L.T.=5	% stock	-16.17	16.94	4.56	-17.18	0.63	3.30
		% CSL	0.96	-1.48	-0.52	1.10	-0.14	-0.38
α = 0.15	L.T.=1	% stock	-51.92	35.07	4.70	-53.68	0.78	3.54
		% CSL	1.54	-2.16	-0.61	1.70	-0.16	-0.46
	L.T.=3	% stock	-41.42	31.29	5.38	-43.57	1.12	3.93
		% CSL	1.40	-2.15	-0.75	1.57	-0.18	-0.57
	L.T.=5	% stock	-25.02	23.69	6.04	-27.17	1.37	4.34
		% CSL	1.09	-1.90	-0.81	1.24	-0.14	-0.66
α = 0.2	L.T.=1	% stock	-71.79	42.58	5.83	-74.10	0.93	4.61
		% CSL	1.55	-2.31	-0.76	1.68	-0.12	-0.64
	L.T.=3	% stock	-56.49	37.83	6.25	-59.28	1.29	4.77
		% CSL	1.45	-2.47	-1.02	1.59	-0.14	-0.88
	L.T.=5	% stock	-33.59	28.94	6.96	-36.86	1.90	4.93
		% CSL	1.17	-2.21	-1.04	1.28	-0.11	-0.93

Table 11.B.12. Stock and CSL % differences (B_3 policy, target value = 0.95)

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**17.20**	**-17.44**	-0.52	**16.81**	0.78	-1.30
	L.T.=3	**14.93**	**-15.90**	*-1.70*	**16.15**	*-2.20*	0.51
	L.T.=5	**11.91**	**-12.37**	-0.70	**12.19**	-0.41	-0.29
α = 0.1	L.T.=1	**18.55**	**-17.97**	1.18	**20.00**	**-4.37**	**5.21**
	L.T.=3	**15.09**	**-16.50**	*-2.24*	**18.66**	**-6.81**	**5.21**
	L.T.=5	**11.91**	**-12.76**	-1.19	**15.60**	**-6.43**	**5.58**
α = 0.15	L.T.=1	**16.62**	**-16.25**	0.64	**18.84**	**-5.30**	**5.76**
	L.T.=3	**13.51**	**-14.50**	-1.42	**17.49**	**-7.03**	**6.02**
	L.T.=5	**11.05**	**-11.71**	-0.85	**15.43**	**-7.03**	**6.43**
α = 0.2	L.T.=1	**15.21**	**-13.89**	*2.03*	**17.49**	**-5.21**	**6.66**
	L.T.=3	**12.40**	**-12.71**	-0.42	**16.49**	**-6.96**	**6.66**
	L.T.=5	**10.28**	**-10.04**	0.30	**14.86**	**-7.10**	**7.31**

Table 11.B.13. PBt CSL results, B_3 policy (target value = 0.98)

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	0.10	0.70	0.84	0.84
	L.T.=3	0.15	0.75	0.95	0.93
	L.T.=5	0.17	0.61	0.80	0.73
α = 0.1	L.T.=1	0.06	0.86	1.12	0.96
	L.T.=3	0.13	0.94	1.31	1.11
	L.T.=5	0.14	0.82	1.22	0.94
α = 0.15	L.T.=1	0.07	0.89	1.26	0.99
	L.T.=3	0.14	1.04	1.56	1.17
	L.T.=5	0.16	0.95	1.55	1.07
α = 0.2	L.T.=1	0.08	0.87	1.30	0.96
	L.T.=3	0.15	1.11	1.76	1.21
	L.T.=5	0.18	1.01	1.80	1.12

Table 11.B.14. APR CSL results, B_3 policy (target value = 0.98)

		Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	**-21.94**	**29.77**	**10.01**	**-32.24**	**13.11**	**-3.17**
	L.T.=3	**-26.50**	**34.93**	**10.99**	**-31.19**	**6.15**	**4.89**
	L.T.=5	**-25.27**	**41.15**	**20.27**	**-26.85**	*2.07*	**18.30**
α = 0.1	L.T.=1	**-24.73**	**35.53**	**13.94**	**-29.58**	**6.33**	**7.71**
	L.T.=3	**-29.12**	**42.79**	**17.78**	**-26.08**	**-4.08**	**21.65**
	L.T.=5	**-28.13**	**47.67**	**25.11**	**-22.69**	**-7.32**	**31.70**
α = 0.15	L.T.=1	**-25.68**	**38.87**	**17.02**	**-28.28**	**3.43**	**13.70**
	L.T.=3	**-29.68**	**44.34**	**19.02**	**-23.99**	**-7.63**	**26.16**
	L.T.=5	**-28.15**	**50.09**	**28.04**	**-20.87**	**-9.79**	**36.60**
α = 0.2	L.T.=1	**-27.04**	**40.66**	**17.64**	**-27.88**	**1.11**	**16.57**
	L.T.=3	**-30.28**	**44.94**	**18.97**	**-23.59**	**-8.94**	**27.33**
	L.T.=5	**-27.67**	**49.79**	**28.27**	**-21.63**	**-8.14**	**35.40**

Table 11.B.15. PBt stock results, B_3 policy (target value = 0.98)

		EWMA	Croston	Approx.	MA(13)
α = 0.05	L.T.=1	17.80	5.59	3.84	4.46
	L.T.=3	15.97	5.92	3.86	5.22
	L.T.=5	11.58	6.01	3.50	5.44
α = 0.1	L.T.=1	32.19	6.17	3.53	4.34
	L.T.=3	30.24	7.51	3.81	5.75
	L.T.=5	22.02	8.01	3.69	6.27
α = 0.15	L.T.=1	48.91	7.31	3.78	5.45
	L.T.=3	45.62	9.38	4.62	7.33
	L.T.=5	33.09	10.55	4.44	8.18
α = 0.2	L.T.=1	68.39	8.55	4.58	6.70
	L.T.=3	62.45	11.51	5.98	9.48
	L.T.=5	44.34	13.15	5.98	10.37

Table 11.B.16. APR stock results, B_3 policy (target value = 0.98)

			Croston-EWMA	EWMA-Approx.	Croston-Approx.	MA-EWMA	Croston-MA	MA-Approx.
α = 0.05	L.T.=1	% stock	-11.82	10.78	1.50	-12.93	0.54	0.09
		% CSL	0.56	-0.68	-0.12	0.69	-0.14	0.01
	L.T.=3	% stock	-10.06	9.80	1.73	-10.43	-0.06	0.79
		% CSL	0.55	-0.73	-0.18	0.73	-0.18	0.00
	L.T.=5	% stock	-5.97	6.81	2.14	-5.98	-0.26	1.35
		% CSL	0.38	-0.56	-0.17	0.50	-0.12	-0.05
α = 0.1	L.T.=1	% stock	-24.60	19.38	2.17	-26.56	1.34	0.39
		% CSL	0.75	-1.00	-0.25	0.85	-0.10	-0.15
	L.T.=3	% stock	-21.53	18.68	2.94	-23.11	1.05	1.38
		% CSL	0.74	-1.08	-0.34	0.91	-0.17	-0.18
	L.T.=5	% stock	-13.59	14.07	3.43	-14.82	0.88	1.95
		% CSL	0.61	-0.98	-0.37	0.72	-0.11	-0.26
α = 0.15	L.T.=1	% stock	-38.60	26.84	2.75	-40.87	1.41	1.05
		% CSL	0.77	-1.13	-0.35	0.88	-0.10	-0.25
	L.T.=3	% stock	-33.32	25.46	3.43	-35.49	1.35	1.80
		% CSL	0.83	-1.32	-0.49	0.96	-0.13	-0.36
	L.T.=5	% stock	-20.91	19.86	4.47	-23.03	1.45	2.69
		% CSL	0.72	-1.29	-0.57	0.83	-0.11	-0.46
α = 0.2	L.T.=1	% stock	-54.76	33.49	2.83	-57.29	1.41	1.19
		% CSL	0.75	-1.16	-0.42	0.83	-0.08	-0.33
	L.T.=3	% stock	-45.79	31.07	3.45	-48.18	1.30	2.01
		% CSL	0.89	-1.50	-0.61	0.98	-0.09	-0.52
	L.T.=5	% stock	-28.06	24.08	4.59	-30.84	1.75	2.70
		% CSL	0.77	-1.51	-0.74	0.86	-0.10	-0.65

Table 11.B.17. Stock and CSL % differences (B_3 policy, target value = 0.98)

APPENDIX 11.C

Inventory control performance of a categorisation rule

	B_2 = 0.93				B_2 = 0.96			
	L.T.=1	L.T.=3	L.T.=5	**average**	L.T.=1	L.T.=3	L.T.=5	**average**
α =0.05	0.928	0.923	0.932	**0.927**	0.950	0.942	0.948	**0.947**
α =0.1	0.934	0.926	0.933	**0.931**	0.955	0.946	0.949	**0.950**
α =0.15	0.938	0.927	0.933	**0.932**	0.958	0.947	0.949	**0.952**
α =0.2	0.940	0.928	0.932	**0.933**	0.960	0.948	0.949	**0.953**
	overall average = 0.931				**overall average = 0.950**			

Table 11.C.1. Categorisation rule, CSL (B_2 policy)

	B_3 = 0.95				B_3 = 0.98			
	L.T.=1	L.T.=3	L.T.=5	**average**	L.T.=1	L.T.=3	L.T.=5	**average**
α =0.05	0.953	0.954	0.959	**0.955**	0.971	0.969	0.972	**0.971**
α =0.1	0.957	0.956	0.959	**0.957**	0.974	0.971	0.972	**0.972**
α =0.15	0.959	0.956	0.958	**0.958**	0.976	0.972	0.971	**0.973**
α =0.2	0.961	0.956	0.957	**0.958**	0.978	0.972	0.971	**0.974**
	overall average = 0.957				**overall average = 0.972**			

Table 11.C.2. Categorisation rule, CSL (B_3 policy)

			Approx. Rule	Croston Rule	Approx. Rule	Croston Rule
			B_2 = 0.93		B_2 = 0.96	
α = 0.05	L.T. = 1	% stock	-1.28	1.33	-1.04	1.01
		% CSL	0.18	-0.19	0.10	-0.15
	L.T. = 3	% stock	-1.57	1.45	-1.40	1.18
		% CSL	0.25	-0.27	0.17	-0.16
	L.T. = 5	% stock	-1.66	1.64	-1.51	1.33
		% CSL	0.25	-0.21	0.21	-0.15
α = 0.1	L.T. = 1	% stock	-2.28	2.35	-1.87	1.80
		% CSL	0.30	-0.38	0.18	-0.25
	L.T. = 3	% stock	-2.89	2.44	-2.51	1.84
		% CSL	0.46	-0.45	0.31	-0.31
	L.T. = 5	% stock	-3.15	2.89	-2.76	2.40
		% CSL	0.48	-0.40	0.37	-0.31
α = 0.15	L.T. = 1	% stock	-3.26	3.35	-2.58	2.32
		% CSL	0.40	-0.50	0.26	-0.37
	L.T. = 3	% stock	-4.14	3.55	-3.47	2.61
		% CSL	0.62	-0.71	0.48	-0.50
	L.T. = 5	% stock	-4.65	4.03	-3.91	3.22
		% CSL	0.71	-0.62	0.57	-0.46
α = 0.2	L.T. = 1	% stock	-4.20	4.26	-3.10	2.82
		% CSL	0.52	-0.72	0.31	-0.48
	L.T. = 3	% stock	-5.27	4.26	-4.31	3.07
		% CSL	0.84	-0.91	0.60	-0.66
	L.T. = 5	% stock	-6.06	5.02	-4.90	3.78
		% CSL	0.94	-0.84	0.72	-0.63

Table 11.C.3. Stock and CSL % differences (B_2 policy)

			Approx. Rule	Croston Rule	Approx. Rule	Croston Rule
			B_3 = 0.95		B_3 = 0.98	
α = 0.05	L.T. = 1	% stock	-1.12	0.94	-0.87	0.73
		% CSL	0.11	-0.13	0.05	-0.07
	L.T. = 3	% stock	-1.29	1.27	-1.06	0.80
		% CSL	0.13	-0.13	0.08	-0.10
	L.T. = 5	% stock	-1.35	1.23	-1.17	1.10
		% CSL	0.13	-0.15	0.09	-0.09
α = 0.1	L.T. = 1	% stock	-1.98	1.89	-1.35	1.08
		% CSL	0.18	-0.26	0.09	-0.16
	L.T. = 3	% stock	-2.36	2.16	-1.93	1.40
		% CSL	0.23	-0.26	0.14	-0.20
	L.T. = 5	% stock	-2.56	2.39	-2.16	1.69
		% CSL	0.26	-0.26	0.17	-0.20
α = 0.15	L.T. = 1	% stock	-2.69	2.51	-1.83	1.41
		% CSL	0.25	-0.36	0.14	-0.22
	L.T. = 3	% stock	-3.33	2.79	-2.57	1.67
		% CSL	0.34	-0.41	0.22	-0.26
	L.T. = 5	% stock	-3.64	3.21	-2.97	2.41
		% CSL	0.41	-0.40	0.27	-0.30
α = 0.2	L.T. = 1	% stock	-3.38	3.21	-2.08	1.51
		% CSL	0.31	-0.46	0.16	-0.26
	L.T. = 3	% stock	-4.10	3.35	-3.05	1.74
		% CSL	0.46	-0.57	0.28	-0.34
	L.T. = 5	% stock	-4.50	3.81	-3.67	2.49
		% CSL	0.52	-0.51	0.35	-0.39

Table 11.C.4. Stock and CSL % differences (B_3 policy)