LICENTIATE THESIS

Evaluation of forecasting techniques and forecast errors

With focus on intermittent demand

Peter Wallström

LULEÅ UNIVERSITY OF TECHNOLOGY

LULEÅ UNIVERSITY OF TECHNOLOGY

LICENTIATE THESIS

EVALUATION OF FORECASTING TECHNIQUES AND FORECAST ERRORS

WITH FOCUS ON INTERMITTENT DEMAND

Peter Wallström

Luleå, May 2009

Luleå University of Technology
Department of Business Administration and Social Sciences
Division of Industrial Marketing, e-commerce and Logistics
Industrial logistics

Printed by Universitetstryckeriet, Luleå 2009

ISSN: 1402-1757
ISBN 978-91-86233-63-1

Luleå 2009

www.ltu.se

I’m just doing my rock’n’roll duty

creating a buzz buzz buzz

Mitchell/Dubois

Like the fly on the wheel, who says

"What a lot of dust we're raising"

Lee/Lifeson/Peart

Hypotesen har förflyktigats i tomma intet

Johannes Kepler

Abstract

To decide in advance the amount of resources that is required next week or next month can be both a complicated and hazardous task depending on the situation, despite the known time frame when the resources are needed. Intermittent demand, or slow-moving demand, that is when there are time periods without demand and then suddenly a time period with demand, becomes even more difficult to forecast. If the demand is underestimated it will lead to lost sales and therefore lost revenues. If the demand is overestimated, in the best case the stock is increased or in worst case, the items lie unsold until they become obsolete. The items with intermittent demand can have a value of up to 60% of the total stock value for all items.

This thesis addresses the topic of forecasting intermittent demand and how to measure the accuracy of the chosen forecast method or methods. Four forecasting methods are tested on almost 18 months of empirical demand data from a manufacturing company. The tested forecasting methods are single exponential smoothing, Croston and two modification of the Croston method, one by Syntetos and Boylan the other by Segerstedt (modified Croston). Four start values and eight smoothing constants are tested.

The methods are evaluated with different accuracy measures; variance (MSE and MAD), bias (CFE, the maximum and minimum value of CFE) and sMAPE. In addition with a new complementary measure of bias; Periods in Stock (PIS), PIS considers the time aspect, when the forecast error occurred not just the error size. Also two variants of MAD and MSE are tested. To improve the evaluation of the bias measures, the percentages of demand occasions that can not be fulfilled are used.

The relationship between the different errors for a certain method is examined with principal component analysis (PCA). The errors are also examined with logistic regression to find out if a certain forecasting method is favoured by certain accuracy measures. The logistic regression is based on descriptive statistics for time series plus the mean absolute change that considers the sequence of the time series as well as the variation. Ranking and error quotients between different methods are other applied methods.

The results of the research both confirm and contradict earlier findings. Among the confirming research results are the bias among the different methods. Croston and Modified Croston are overestimating the demand, Syntetos and Boylan's Croston variant has a tendency to underestimate the demand. Single exponential smoothing is relatively biasfree when low smoothing constants are concerned. The contradictive results are that CFE is not a suitable measure of bias at least when the number of forecasting periods is limited. The value of CFE can indicate a nonbiased forecast when both PIS and the percentage of unmet demands indicate a biased forecast. PIS is also less sensitive to transient demand events that can distort CFE. PIS is recommended as a bias measure for limited time series, especially considering intermittent demand, along with the percentage of unmet demand. Another result is that MAD is not reliable since the method in certain circumstances favours methods that underestimate the demand.

Abstract in Swedish

Att på förhand bestämma vilken mängd resurser som krävs nästa vecka eller nästa månad kan vara både en komplicerad och riskfylld uppgift beroende på situationen, trots att man känner till när resurserna behövs. Intermittent efterfrågan, eller lågrörlig efterfrågan, är när många perioder saknar efterfrågan och plötsligt sker en efterfrågan en period. Detta gör det svårare att prognostisera. Om efterfrågan underskattas kommer det att leda till förlorad försäljning och därmed förlorade intäkter. Om efterfrågan är överskattad kommer det i bästa fall att leda till ökat lager eller, i värsta fall, leda till osålda produkter och till slut inkurans. Artiklar med intermittent efterfrågan kan utgöra upp till 60 % av det totala lagervärdet för samtliga artiklar.

Denna uppsats avhandlar prognoser av intermittent efterfrågan samt hur prognosfelen ska mätas för den valda eller de valda prognosmetoderna. Fyra prognosmetoder utvärderas med nästan 18 månaders empirisk efterfrågedata från ett tillverkande företag. De utvärderade metoderna är exponentiell utjämning, Croston och två modifierade varianter av Croston; Syntetos och Boylans metod samt modifierad Croston av Segerstedt. Fyra olika startvärden och åtta utjämningskonstanter används.

Prognosmetoderna utvärderas med olika typer av prognosfel; varians (MSE och MAD), bias (CFE samt max- och minvärde av CFE) och sMAPE. Vidare sker utvärdering med ett komplimenterande mått för bias, Lagerperioder (Periods in Stock, PIS). PIS tar tidsaspekten i beaktande och inte bara storleken på prognosfelen. Dessutom undersöks två varianter av MAD och MSE. För att förbättra utvärderingen av biasmåtten undersöks procentantalet av de efterfrågetillfällen som en prognosmetod inte kan uppfylla.

Förhållandet mellan de olika prognosfelen undersöks med hjälp av principal component analysis (PCA). Prognosfelen undersöks även med binär logistisk regression för att utröna huruvida vissa prognosmetoder gynnas av vissa prognosfel. Den logistiska regressionen baseras på deskriptiv statistik för tidsserierna samt medelabsolutförändringen som tar ordningen för tidserien i beaktande såväl som variationen. Rankning och kvoter mellan olika prognosfel från olika metoder är andra tillämpade metoder.

Resultatet av forskningen både bekräftar och motsäger tidigare forskning. Bland de bekräftande resultaten är den bias olika prognosmetoder har. Croston och modifierad Croston överskattar efterfrågan, Syntetos och Boylans metod underskattar efterfrågan. Exponentiell utjämning är förhållandevis fri från bias när utjämningskonstanterna har låga värden. De avvikande resultaten är att CFE inte är lämpligt att använda när antal prognosperioder är begränsat. Värdet för CFE kan indikera att prognosen är fri från bias när både PIS och procentandelen icke mött efterfrågan. PIS är dessutom mindre känslig för transienta efterfrågehändelser som kan förvränga CFE. PIS rekommenderas som ett biasmått när tidsserien är ändlig, särskilt när det gäller intermittent efterfrågan, tillsammans med måttet procentandelen icke mött efterfrågan. Andra resultat är att MAD inte är pålitlig eftersom måttet, under vissa förhållanden, gynnar prognosmetoder som underskattar efterfrågan.

Contents

Preface

The preface, the part where profound wisdom and acknowledgement meet, so what could be better than writing the immortal words right at the end of the work with the thesis? When the wisdom, knowledge and understanding have reached heights that requires astronomical measure. The amount of wisdom alone should be enough to make the printed copy of the thesis vibrate in the bookshelf. I am not sure about the PDF-version though, but at least the lights should flicker or causing a power surge or two when opening the file.

Is this true? In theory; yes… In practice: No! No! And No! After burning the midnight oil with a flame thrower, redefining the concept of 24-7, all with sore fingers tired of the keyboard and its marathons; it is not the best of moments to come up with what even resembles immortal words. It is hard enough to come up with something that resembles ordinary words when your finger just want to go; meep, meep.

Instead of forced and futile attempts of catching an immortal thought, over to something completely different; the acknowledgements in no particular order

Anders Segerstedt, my supervisor. Diana Chronéer – for the eye with the red thread. Håkan Wallström – for the first eye, with a microscopical resolution, on the drafts, without that eye ltters and words would have been m.

Håkan Norberg, he is the one behind the template of this thesis and an invaluable resource during the construction projects, a duty he shared with Peter Simonsson, who also gets a special credit for having such a nice first name.

Terese Lantto – For all the shorter sentences, shorter paragraphs and lots of commas, despite the statement: “Matematik kan inte användas, bara missförstås”, a statement which unfortunately do not lend itself to the English tongue without getting lost in translation.

Finally, some words of wisdom to future preface writers, wisdom gained from the hard earned empirical insights; go for the fresh-brain-approach instead of the oversaturated-brain-approach. Then perhaps… but only perhaps, may the words have a chance to last as long as the pyramids.

PETER WALLSTRÖM
Luleå, May 2009

Abbreviations

-25-s	Start value that is 75% of the mean for the time series
+25-s	Start value that is 125% of the mean for the time series
APE	Absolute Percentage Error
APR	Average Percentage Regret
ASE	Absolute Scaled Error
CFE	Cumulated Forecast Error
COV	Covariance
CV	Coefficient of variance, quotient between the standard variation and the mean
D	Demand time series
Dem occ	Demand occasions
DO%	The Percentage of actual demand occasions in relation to the total number of possible demand occasions
DR	Demand Rate time series
ID	Inter-demand interval time series
MAC	Mean Average Change

MACs	Mean Average Change scaled
MAD	Mean Absolute Deviation
MAD_n	Mean Absolute Deviation summarised error from the previous demand occasion to the present demand occasion
MAPE	Mean Absolute Percentage Error
Max (max)	Maximum
MdAPE	Median Percentage Absolute Percentage Error
Mean-s	Start value that is equal to the mean of a time series
Min (min)	Minimum
ModCr	Modified Croston that uses only one forecast instead of two
MPE	Mean Percentage Error
MSE	Mean Squared Error
MSE_n	Mean Squared Error summarised error from the previous demand occasion to the present demand occasion
NOS	Number of Shortages
NOSp	Number of Shortages expressed in percent
PB	Percent Better
PBt	Percent Best
PCA	Principal Component Analysis
PIS	Periods in Stock
p-value	Probability value
RGRMSE	Relative Geometric Root Mean Square Error
RSQ	R-squared, pattern matching indicator

SES	Single Exponential Smoothing
SMA	Single Moving Average
Std	Standard deviation
SyBo	The forecast method of Syntetos and Boylan, a method based on the Croston method but with an additional bias correction

1 INTRODUCTION

The first chapter starts with a quotation from a forecasting webpage from the University of Exeter and continues with a background to forecasting and some of the problems that make forecast a difficult task. Research questions, aim and objective are introduced as well as the limitations in this thesis.

1.1 Background

"… And there shall be battles among many kingdoms. That year shall be the bloody field, and lily F.K. shall lose his crown, and therewith shall be crowned the Son of Man K.W., and the fourth year shall be preferred. And there shall be a universal peace over the whole world, and there shall be plenty of fruits; and then he shall go to the land of the Cross."

The paragraph above is the words of Mother Shipton, a 15 century prophet, whether she was a human of flesh and blood or a myth is beyond this thesis. To be able to predict the future has through time played a vital role; by knowing what was to come could help the peasants to decide when to sow and when to reap, it could also help rulers of the past to keep the crown or expand the empire. The forecasting methods differed partly depending on the available resources, some studied the movement of the birds other studied the movement of stars and the planets in the night sky. The study of the night sky turned slowly superstition into science. Tycho Brahe, part astronomer and part astrologer, studied large amount of astronomical observations. In order to minimise the influence of observation errors and other disturbances, he used simpler statistical methods in the end of the sixteenth century.

1.2 The Complexity of Forecasting

Forecasting is still important. Some might argue that in the age of pull systems, a production start of an item should not be made until the shelf in the grocery is almost empty. The idea behind the pull system is said to be how the filling of diary product was made. New milk cartons were put on the shelf first when the customers had emptied the shelf. To let the customers govern the production rate means that the risk of saturated production or transportation is diminished. The idea of letting the customer purchase trigger the production has some drawbacks. As long as there are new milk cartons an arms length away from the personnel, the pull system works. If the replenishment to the grocery from the dairy takes some time (at least ordering and transportation) some kind of forecast is necessary.

To predict the dawn or the planets position five years or ten years from now can be made with almost deterministic accuracy. To predict the price of metals or the sales of cars five or ten years from now is not done with decimal accuracy. Many organisations deal with recourses that are not possible to forecast with decimal accuracy. Sophisticated forecasting methods that work in a more stable environment have been proven to be inferior compared to simpler methods in a series of forecast competitions. (Makridakis et al, 1982; Makridakis et al, 1993; Makridakis and Hibon, 2000)

The consequences for an organisation when the forecasting precision is limited can be severe, especially if the precision continues to be limited for a longer period of time. An underestimation of a future need may result in lost sales or inability to fulfil the undertaking of the organisation for example medical supply. Overestimation on the other hand may lead to excessive capital tied to a stock that in the end is impossible to sell. In inventory control the conflicting goals between keeping the inventory low and avoiding stockouts and lost sales must be balanced.

If a company wants to be successful it is not enough to focus on the most moved items with the highest sales. According to Johnston et al (2003) approximately 75% of the item lines are moved six times or fewer in most branches. The slow moving items are responsible for over 40% of the income and require approximately 60% of the total investment in stock.

However, to forecast intermittent demand is different compared to when there is a demand in every forecasting period. Silver (1981) defines intermittent demand as when both the demand and the periods between the demands are random. Croston (1972) defines intermittent as when the demand is zero in a number of forecasting periods. When the zero demands are present some of the most used forecasting methods start to overestimate the demand. Croston (1972) suggested that the forecast should be split in two; one for the demand and the other forecast for the number of periods between the demand occasions, inter-demand. Syntetos and Boylan (2001) proved that the Croston forecasting method also had bias. The method overestimates the demand.

There have been a number of alternatives to Croston's forecasting method, some of the alternatives to Croston's method are; Segerstedt (2000) Snyder (2002), Willemain et al (2004) and Syntetos and Boylan (2005). Gardner (2006) compared five studies of forecasting methods for intermittent demand and draws the conclusion that the performance of the different methods depends on the type of demand data and the error measures. Similar conclusion for the non intermittent situation is drawn by Makridakis et al (1982), Makridakis et al (1993) and Makridakis and Hibon (2000).

When forecasting method is in use one wants to know how well the method predicts what it is supposed to predict. According to Armstrong and Yokom (1995) accuracy is the most common reason to use a certain type of forecast error. The second reason to use a certain forecast error, is that it is relatively easy to interpret the error measure. Accuracy as the only type of measure is questioned by Fildes et al (2009). Mathews and Diamantopoulos (1994) states that one type of error will not be sufficient to cover all aspects that is of interest for the user concerning forecast accuracy. Despite this fact, there is forecasting software that only measures one type of error (Wallström, 2006). Errors are usually regarded as isolated from each other, or at least the correlation between different measures is seldom discussed. None of the studied papers have considered the relationship between the errors and forecast methods when intermittent demand is concerned.

According to Fildes and Makridakis (1995) the forecasting research has in general concentrated on theoretical contributions while ignoring empirical results. Makridakis and Hibon (2001) could not understand how theories can be accepted without testing the theories on real data not just simulations. Theoretical statisticians have responded to the criticism. Clements and Henry (2001) claims that the theoretical statisticians do not ignore empirical results. The references, that Clements and Henry use to prove their point, are three references from other authors and ten references from themselves. Moon et al (2003) describes the forecasting research as a research that emphasises the development of the new forecasting methods instead of identifying the needs of the practitioners in the environment where the forecasting methods are supposed to contribute to an organisation.

Forecasting researchers are concerned with error measures that can be used on different types of series' for comparisons between different methods such as Armstrong and Collopy (1992). In general when the practitioner's needs are concerned, the emphasis is the understanding of the error, not that the error is useful in practice. Also what is easy to include in the software can be more important to the developer of the software than the needs of the practitioner, the end customers (Wallström, 2006).

1.3 Aim and Objective

The objective of this research is to find complementary error measures and an increased understanding of the relationship between different forecasting methods and errors. The aim is to test different methods of evolution that can be used by a practitioner.

1.4 Research Questions

How are bias measures affected by the short series that real demand usually have?

To what degree is the relationship between errors for a forecasting method unique compared to other forecasting methods?

Are there common error dimensions among the forecasting methods and how are the error dimensions structured for the forecasting methods?

What dimensions should a forecast error cover? Is it possible to have all the dimensions covered in one error so that the evaluation of forecasting methods can be done by using only one measure? If so, which measure?

How robust are the forecasting methods considering different errors smoothing constants and start values?

1.5 Limitations

Supply chain management, the coordination of resources and production among the organisations that are a part of the chain from raw material to a service or product for the end customer, can be affected by the accuracy of the forecast. Lee et al (1997) recognise forecasting as a major cause of the bullwhip effect, the variance of the demand increases as the distance to the end customer increases. Their solution to decrease the bullwhip effect is that forecasts should only be performed on the end demand and that more supply chains adopt the grocery industries information flow upstream from the end customer.

Sohn and Lim (2008) state that the choice forecast method has an influence on supply chain performance and that information exchange and forecasting has a major influence on short lived products. Even the parameter choice of the forecast method will influence the bullwhip effect in the supply chain according to Bayraktar et al (2008).

However, Zhao et al (2002) claim that many studies are too concerned with the bullwhip ratio (variance (order)/variance (demand)) instead of measuring the financial effects in the supply chain. Sucky (2009) is hesitant of the forecast influence of the bullwhip effect. He motivates this by using a forecast method that he consider inferior to forecasting methods used in practice and claims an inferior method should also have a larger bullwhip. The inferior method has a bullwhip effect and therefore the ‘better’ methods must have smaller bullwhip effects. Fildes et al (2009) state that forecasting is a vital part of supply planning and therefore can an accurate forecasting be a competitive advantage for a supply chain in form of saved costs.

Despite the consequences that forecasting might have in a supply chain context, this is not a supply chain management thesis. In order to make the thesis supply chain management oriented the research should consider what happened both downstream and upstream in the chain. Since no such information was available, only the demand data, no consideration has been taken regarding supply chain.

The tests for the different methods do not consider any human involvement. In practice if a method has too large errors the forecast is stopped and a new method or new parameters are chosen. In the test the forecast starts and continues for almost 18 months. Fildes et al (2003) state that a person responsible for the forecasts, tends to act according to the performance measure of that person. By not using a real forecast situation this influence of performance measure is avoided.

No other forecasting periods than one day is considered. One of the reasons is to guarantee a sufficient number of items that has a low number of demand occasions during the test period. Another reason is that there is no information concerning the lead time for the different items.

There is no adjustment of the forecast for the different methods to 'pratical' numbers. If an item has a forecast of 2.35 no rounding off is done since rounding off errors is comparable to a policy which distorts the performance of the methods. The forecast errors would then not only consist of the errors of the method but also the policy's influence on the error.

All the forecast errors are points in time. It is the last value that is used this means that for the mean errors every error value is included. The development over the test period has not been monitored.

1.6 Definitions

Intermittent demand is when there are periods where no demand occurs, regardless of the size of the demand or the variation of the demand.

2 THEORETICAL FRAMEWORK

The second chapter starts with an introduction of the further presented theories and the notations that are used in the equations. Then the four methods are presented including what previous research has concluded when the methods have been used with intermittent demand. The forecast errors are presented in a similar manner followed by theory concerning tracking signals. In addition to the 'traditional' forecast errors some complementary errors are presented. A complementary measure concerning descriptive statistics is also presented. The final part of the chapter describes some earlier evaluations regarding error dimensions, intermittent forecasting methods and choice of forecast errors.

2.1 Introduction and Notation

In demand forecasting the single exponential smoothing (SES) is one of the most adopted techniques. One of the initial advantages was the limited computational effort that was required for yesterday's computers. The method and the variants of the method have been proven to perform well compared to complicated methods. According to Makridakis (1986) less sophisticated methods do work better when the level of aggregation is low compared to the more sophisticated methods, which is one of the conclusions of the M-competitions (Makridakis et al, 1982; Makridakis et al, 1993; Makridakis and Hibon, 2000). The conclusions of the M-competitions did receive critic for the evolution, among the critics are Chatfield (1988). The more sophisticated methods are appropriate when there are complex pattern in the time series according to Kohler (2001). A method like Box-Jenkins needs at least 50 observations according to the initial research which later research downsized to 30 observations (Ord, 2001). When intermittent time series are considered, Box-Jenkins is not recommended because of the presence of zero demands (Eaves and Kingsman, 2004).

Even if SES performs better it is not necessary a better choice depending on the forecast situation, Makridakis (1986) states that SES do not try to explain economic or business phenomena which leads to a better understanding of the relationship among the variables. The automatic method chosen for demand forecasts may need to be complemented. Vokurka et al (1996) identify characteristics which need human input. The characteristics are; irrelevant early data, unusual observations, level discontinuities, trend or form in data and cycles or regular movements of the series about the basic trends. According to Fildes et al (2003) the human input may depend on the performance measure of that human. Persons tend to act according to the performance measure and not just the forecast situation.

Notations

X_t	Demand in period *t*
$\hat{X}_t$	Demand forecast in period *t*
α	Smoothing parameter, value 0-1
T_t	Inter-demand interval when the latest demand occurs in period *t* which is the difference in time periods between the latest and the previous demand
$\hat{T}_t$	Forecasted inter-demand intervals in period *t*
$\hat{d}_t$	Forecast of the demand rate in period *t*
β	Smoothing parameter for inter-demand intervals, value 0-1
t_n	The time period when the latest demand occurs
t_{n-1}	Time period for the previous demand
M	Number of demand occasions, $M \le T$

X_m	Demand in demand occasion *m*
p	The time period *t* that coincides with demand *m*
o	The time period *t* that coincides with demand *m-1*
T	Number of time periods, $T \geq M$
e_t	Forecast error in period *t*

2.2 Forecasting Methods

Some of the most common methods to forecast intermittent demand are SES and moving average. Moving average is the mean of a fixed number of the demand of the previous periods, as new demands occurs it replaces the oldest (Makridakis et al, 1998). However, the ability for SES to forecast slow moving items or an intermittent demand has been questioned. Croston (1972) presents a method that is updated only when there is a demand and therefore the forecast precision should increase. The method consists of two forecasts, one for the demand and the other for the inter-demand period. Segerstedt (2000) suggests a variant of Croston but with one forecast. Syntetos and Boylan (2005) present a version of Croston where a bias correction is added. Bias is a systematic error, when the forecast is, on the average, significantly above or below the demand during the forecasted periods

2.2.1 Single Exponential Smoothing (SES)

Single exponential smoothing is a technique applied in different fields, such as forecasting Brown (1959), and process regulation Montgomery (2005). According to Gardner (2006) the method was originally developed for antisubmarine purposes. Brown used a variant of the exponential smoothing to create a tracking model for fire-control information on the location of the submarine. Makridakis and Hibon (1991) consider SES to be a robust method that is easy to use.

In every time period the model is re-estimated with the most recent available demand data and the previous forecast. The smoothing constant, α, regulates the amount of influence the forecast error have, see equation 2.1. The forecast error is the difference between the real demand and the forecasted demand. (Montgomery et al, 1990)

$$\hat{X}_{t+1} = \hat{X}_t + \alpha (X_t - \hat{X}_t) \tag{2.1}$$

Another way of describing the function of the smoothing constant is that the different observations have weights that decrease geometrically with age. The smoothing constant regulates the influence of historical values; a low smoothing constant emphasis the past, favourable with a stable demand but then the technique is slow to react if systematic changes occur. A high smoothing constant emphasis the most recent observations, which is better suited when faster reaction is wanted, but the drawback is sensitivity to random changes. (Montgomery et al, 1990)

In a practical application different smoothing constants should be used for different classes of items, which should also be the case with SES. A smoothing constant between 0.1-0.3 is suitable for SES when forecasts are done on a monthly basis (Silver et al, 1998).

The weight given to data with number of *k* periods ago can be expressed as $\alpha(1- \alpha)^k$ which makes the average age:

$$\alpha \sum_{k=0}^{\infty} (1-\alpha)^k k = \frac{1-\alpha}{\alpha} \tag{2.2}$$

For *N*-period moving average the average age of the data is:

$$\frac{1}{N} \sum_{k=0}^{N-1} k = \frac{N-1}{2} \tag{2.3}$$

If an exponential smoothing system is defined to be equal to an N-moving average, the following equation is reached

$$\alpha = \frac{2}{N+1} \tag{2.4}$$

With a higher resolution of the forecast intervals (shorter time periods) the probability for periods with zero demand increases. If several zero demand is consecutive the forecast will decrease and eventually approach zero. This scenario is most likely to occur when the items are slow moving. The items have an intermittent demand. An alternative is to update the forecast only after a demand has occurred. This makes SES biased which is not the case when the update occurs in every period (Boylan and Syntetos, 2007).

The conclusions of SES as a forecast method of intermittent demand are varied. Croston (1972) discusses SES problem with overestimation of the demand when the forecast update takes place right after a demand. Boylan and Johnston (1996) consider SES to be suitable when the inter-demand is 1.25 periods or lower. Eaves and Kingsman (2004), on the other hand, concludes that SES can be used as a method for demand that is intermittent.

2.2.2 The Croston Method (Croston)

Croston (1972) presented a solution for slow-moving items. He suggests that there are two processes involved when the demand is intermittent, a demand process that is normally distributed and an inter-demand interval that is generated by a Bernoulli process. The two processes are assumed to be independent. Croston shows that the variance for his method has a lower variance than SES. The original paper was corrected by Rao (1973). Croston suggested that his method needs more control signals since it forecast more than just the demand but also the inter-demand periods. Croston did not consider different smoothing values for demand and inter-demand periods. The forecast consists of two parts; one for the demand and one for the inter-demand intervals. The forecast is only recalculated when there is a demand. Schultz (1987) regarded the forecasts as two different forecasts which also are valid for the smoothing constants.

If $X_t = 0$ (2.5)

$$\hat{X}_{t+1} = \hat{X}_t$$

$$\hat{T}_{t+1} = \hat{T}_t$$

$$T \leftarrow T + 1$$

If $X_t \neq 0$ (2.6)

$$\hat{X}_{t+1} = \hat{X}_t + \alpha (X_t - \hat{X}_t)$$

$$\hat{T}_{t+1} = \hat{T}_t + \beta (T_t - \hat{T}_t) \qquad \text{where } T_t = t_n - t_{n-1}$$

The two exponential smoothing forecasts are combined to estimate the mean demand per period length. When there is a demand in every period. The Croston method becomes equal to SES.

$$\hat{d}_{t+1} = \frac{\hat{X}_{t+1}}{\hat{T}_{t+1}} \tag{2.7}$$

Willemain et al (1994) questioned the distribution assumptions for demand and inter-demand and therefore also tested lognormal demand distribution and correlation between the demand and inter-demand. The Croston method does not perform equally well when the assumptions are challenged but the method is still better than SES. When real data is used instead of simulation, the margin between Croston and SES shrinks but Croston is still better. The recommendation is that inventory managers should switch to Croston instead of SES. Syntetos and Boylan (2001) proved that Croston has a bias problem, the method overestimates the demand. In a later paper Syntetos and Boylan (2005) calculated the theoretical bias of the Croston method. Gardner (2006) concludes that theoretical variance of the Croston method and its versions are approximations.

The Croston method and Croston will be used interchangeably.

2.2.3 Croston According to Syntetos Boylan (SyBo)

Syntetos and Boylan (2005) suggested a modification to the Croston method since the Croston method has a bias problem. A mistake of the expected estimate of demand in the original derivation by Croston made the method's improvement marginal. The modification can be described as a bias correcting function of the original Croston method. This modification of the Croston method will from now on be referred as SyBo. In equation 2.8 the bias corrector is added to the original Croston. According to Eaves and Kingsman (2004) the bias correction is based on a Taylor series expansion. Syntetos and Boylan (2005) recommend that the bias smoothing constant has the same value as the smoothing constant has for the inter-demand intervals. The forecast updates are the same as for the original Croston with one exception; when there is a demand in every period SyBo the method does not become equal to SES, since the bias corrector is not affected.

$$\hat{d}_{t+1} = \left(1 - \frac{\beta}{2}\right)\frac{\hat{X}_{t+1}}{\hat{T}_{t+1}} \tag{2.8}$$

Eaves and Kingsman (2004) concluded that SyBo (called the approximation method in their article) gives a significant reduction in stock-holdings while still being able to meet a specified service level. Teunter and Sani (2009) showed that SyBo is biased; the method underestimates the demand due to an overcompensating bias corrector. They also tested another not published method from Syntetos' PhD thesis. The unpublished method has a lower bias but has a higher variance than SyBo.

2.2.4 Modified Croston (ModCr)

SES uses one smoothing constant where Croston and SyBo use two smoothing constants which increase the complexity compared to SES. The Croston technique forecasts the mean demand and the mean inter-demand. Another interpretation of the quotient is that it represents the demand. Levén and Segerstedt (2004) presented a version of Croston, Modified Croston (ModCr), that forecasts the demand rate directly instead of separating the forecast into demand and inter-demand; therefore it requires one smoothing constant. The update occurs when there is a demand, but maximum is one per working day. If there are several demands in a day, the demands are summarised. The demand rate is the quotient between the demand and the inter-demand interval. In a simulation study the method was shown to perform better than SES. The method was proposed to avoid the bias problem that Croston had.

$$If\ X_t = 0 \tag{2.9}$$

$$\hat{d}_{t+1} = \hat{d}_t$$

$$If\ X_t \neq 0$$

$$\hat{d}_{t+1} = \hat{d}_t + \alpha\left(\frac{X_t}{t_n - t_{n-1}} - \hat{d}_t\right) \tag{2.10}$$

The idea behind ModCr is to avoid the decision of what method to use in a practical application, SES or Croston. A withdrawal every time period (working day) transforms equation. 2.10 to be equal to equation 2.1. When the demand takes place in every time period the ModCr is equal to SES. The smoothing constant for ModCr must have lower start values, probably 0.05-0.3, when the resolution of the time period is higher, days or weeks instead of months. (This is also valid for other forecast methods). Items with high frequency of withdrawals or demand (every day) with ModCr should have a lower constant than SES, if the forecast interval for SES is much longer (weeks, months) than for ModCr (days).

The simulation study of Levén and Segerstedt (2004) has not been confirmed in other studies. Teunter and Sani (2009) find the ModCr to have bias problem that is more severe than for Croston that also tends to overestimate the demand. This is a confirmation of the results Syntetos and Boylan (2007) presented. They also found that the bias of ModCr is not dependent of the value of the smoothing constant. The statement that ModCr is nonbiased is a relevant claim when demand rate is considered, however this is not what the method is supposed to forecast (Syntetos Boylan, 2007). Gardner (2006) stated that there was no evidence presented that motivates the statement of nonbiased.

2.3 Forecast Accuracy

In an evaluation of forecasting methods, start values or smoothing parameters, it is necessary to measure the differences between the possible alternatives. However, it is not obvious which method is the most suitable method to measure forecast errors. Different measures have their advocates. Gardner (2006) compared five studies concerning the Croston method and some of its variants. He established that the performance depended on the error measures and the type of data. Teunter and Duncan (2009) discussed that conflicting results may depend on accuracy measures that is not appropriate. To reduce the forecast errors to only one measure does not seem possible or even desirable. The different types of forecast errors have different dimensions that make it impossible to reduce them to only one dimension without sacrificing information. Behind each measurement lie different methods to calculate the errors and a possibility to obtain complementary information.

Instead of trying to reduce the number of measures it may be more informative to interpret several measures in an attempt to grasp the whole picture concerning the evaluation of the forecasting methods or an individual item. The choice of forecasting method or methods should not only be based on forecast errors but also on the consequences for the organisation the chosen method or methods have.

The choice should reflect the organisation's strategies. If the choice is between two methods (*A* and *B*) with slightly different error performance, assume method *A* has the better error performance but tends to increase the inventory compared to method *B*, then method *B* is probably preferable. According to Syntetos and Boylan (2005) a better error performance is not always equal to a better customer service level or a reduced cost of inventory. The main purpose of forecasting in most cases is to cost efficiently serve the customers with short and precise delivery times; low forecast errors are only instruments for the aim.

Another complication of forecast accuracy is when methods are evaluated with the purpose of finding one or two appropriate methods among a number of contenders. Some measures might be useful when an individual forecast is evaluated but when the errors are summarised across series with different mean demand sizes and variations some errors will prove unreliable, like scale dependent measures. Scale dependent errors will, when summarised, be dominated by the larger values from certain forecasted items (Makridakis and Fildes, 1988; Armstrong and Collopy, 1992).

2.3.1 Mean Squared Error (MSE)

A common measure for forecasting errors and its variance is Mean Squared Error (MSE). MSE is related to standard deviation of forecast errors and is therefore an appropriate error for mathematical operations. In equation 2.11 the assumption is that the mean forecast error deviates from zero and therefore is the mean error present which can be compared to the standard deviation. If instead the mean error is assumed to be equal to zero the term for mean forecast error is deleted, see equation 2.12. However due to the squared function MSE is more sensitive to outliers and errors smaller than one. (Montgomery et al, 1990)

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left(e_t - \bar{e}_t\right)^2 \tag{2.11}$$

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left(e_t\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\left(X_t - \hat{X}_t\right)^2 \tag{2.12}$$

Silver et al (1998) and Koehler (2001) recommend the use of MSE. Sani and Kingsman (1997) regard it as better than the other commonly used variance errors. Chatfield (1988) describes MSE as measure to decide how well a forecasting method fit. In a study by Carbon and Armstrong (1982) MSE was a commonly used measure of accuracy both among academicians and practitioners. Syntetos and Boylan (2005b) used MSE when the theoretical errors of forecasting methods should be calculated. Since MSE is scale dependent it is not a measure that is suitable to summarise across different time series (Makridakis and Fildes, 1988; Armstrong and Collopy, 1992).

2.3.2 Mean Absolute Deviation (MAD)

Mean Absolute Deviation (MAD) measures variance just like MSE. But unlike MSE, MAD lacks the stronger statistical relationship that MSE has. However MAD has the advantage of being easier to understand among non-specialists than MSE, partly because that the error has the same dimension as the forecast. MAD is not as sensitive to outliers as MSE due to the absolute value instead of the quadratic value for each error. In equation 2.13 the mean error is not assumed to be equal to zero. If the mean error instead is assumed to be equal to zero the calculation of MAD is performed according to equation 2.14. (Montgomery et al, 1990)

$$\text{MAD} = \frac{1}{T}\sum_{t=1}^{T}\left|e_t - \bar{e}_t\right| \tag{2.13}$$

$$\text{MAD} = \frac{1}{T}\sum_{t=1}^{T}\left|e_t\right| \; = \frac{1}{T}\sum_{t=1}^{T}\left|X_t - \hat{X}_t\right| \tag{2.14}$$

MAD is more often used than commented, at least in comparison to MSE. However when intermittent demand is considered, MAD benefits the method that underestimates demand the most, a zero demand forecast outperformed the other methods (Teunter and Duncan, 2009).

2.3.3 Cumulated Forecast Error (CFE)

The previously presented error measures do not have the ability to reveal if there is a systematic error (bias) since the sign of an individual error is removed with either the absolute operation or the quadratic operation.. One of the most common measurements of bias is *Cumulated Forecast Error* (CFE). CFE is the cumulated sum of all forecast errors and CFE_t is the cumulated forecast error from period 1 to period *t*, and CFE_T the cumulated forecast error from period 1 to period *T*, i.e. the cumulated forecast error during the whole investigated time interval, see equation 2.15. Montgomery et al (1990) describes an alternative, if just the most recent periods are of interest; just a certain number of the latest periods are included to calculate the error, which reminds of the moving average.

$$CFE_t = \sum_{i=1}^{t}(X_t - \hat{X}_t) = X_t - \hat{X}_t + CFE_{t-1}, \quad t = 1, 2, \ldots, T \tag{2.15}$$

What is close to zero is associated with the variability of the forecast errors. A larger error can deviate more from zero then a small error for CFE. Makridakis et al (1998) considers CFE to be dependent on the scale of what is forecasted. If the forecast is unbiased the CFE values should be close to zero. However Syntetos and Boylan (2001) do consider CFE scale depended to a lesser extent then MSE and MAD.

2.3.4 Symmetrical Mean Absolute Percentage Error (sMAPE)

The previous measures are all, to a various degree, dependent on the scale of the data. Therefore it is difficult to make comparisons across different time series and different time intervals, percentage errors are therefore used. When demand occurs in every period the quotient between the forecast error and the demand can be utilised. The mean percentage error (MPE) is formed by the mean of the sum of the percentage errors. The mean absolute percentage error is formed in a similar manner but with the addition of absolute values of the individual forecast errors. One of the drawbacks with these types of measures is the influence of a denominator with a low value which inflates the percentage error and causes outliers. Another problem is that a forecast larger than the actual demand results in a larger error than if the forecast is lower than the demand, see equation 2.16 and 2.17. (Makridakis, 1993)

$$\text{APE}_t = \frac{\left|X_t - \hat{X}_t\right|}{(X_t} = \frac{150 - 100}{150} = 33.33\% \tag{2.16}$$

$$\text{APE}_t = \frac{\left|X_t - \hat{X}_t\right|}{(X_t} = \frac{100 - 150}{100} = 50\% \tag{2.17}$$

To avoid the difference caused by whether the larger value is the demand or the forecast Makridakis (1993) presented symmetric Mean Absolute Value (sMAPE), see equation (2.18). It was originally named modified absolute percent error and later named sMAPE. The error can be applied when the demand is intermittent since the error can handle zero demand without approaching infinity. The measure can vary between -200 to +200 percent.

$$\text{sMAPE} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|X_t - \hat{X}_t\right|}{(X_t + \hat{X}_t)/2}\cdot 100 \qquad (2.18)$$

Goodwin and Lawton (1999) disagrees and gives an example where the demand is 100 and errors of -10 or +10 results in equal error; 10%. With sMAPE in the same situation the errors are 18.8% and 22.2% when just one observation is considered. Also the sMAPE is asymmetrical, the bound value for positive errors occurs when the forecast is zero and for negative errors minus infinity is required before the bound value is reached. The bound value is 200% in both cases. Koehler (2001) draws similar conclusions and states the sMAPE can, when only one observation is studied, have a percentage error that is twice as high as the non-symmetrical percentage error. Makridakis and Hibon (2001) find sMAPE as the better alternative to MAPE when summarising. Fildes et al (2009) could not find any difference between the two measures (sMAPE and MAPE).

2.4 Tracking Signal

The purpose of a tracking signal is to detect a systematic change in the demand or a systematic error of the forecast method. A common tracking signal is based on the quotient between CFE and MAD. The quotients are plotted in chronological order in a control chart or displayed in the forecast software. The choice of control limits is a balance between indications of an out-of-control condition when there is no obvious cause and where the forecast really is out of control but is within the control limits. (Makridakis och Wheelwright, 1989)

$$\text{Tracking signal} = \frac{\text{CFE}_\text{t}}{\text{MAD}_\text{t}}$$

One problem that can occur is an unstable quotient caused by errors that makes the denominator close to zero. To avoid this problem a smoothing function similar to SES can be applied to CFE and MAD. A disadvantage with the smoothing of MAD and CFE is the influence of the value of the smoothing constant (for MAD and CFE) which affects the possibility to detect a systematic change. (Makridakis och Wheelwright, 1989)

Another problem is the distribution of the qoutient. The assumed distribution is the normal distribution. If the real distribution is not the normal distribution, the number of signals outside of the control limits will be increased compared to a normal distribution. In such a case it might be necessary to analyse the distribution in order to find suitable control limits. (Montgomery, 2005)

2.5 Suggested Measures of Forecast Accuracy

In the initial studies a minimisation of the forecast error measures with the Excel solver was used, with smoothing constants and start values as variables. Optimization has been done previously with the Excel solver e.g. Rasmussen (2004). Since there were more periods with no demand than with demand, the solver focused on minimising the error for periods without demand. This especially occurred when the error minimising were based on MAD. When the demand is low the distance to zero is small and hence a forecast close to zero or even zero in the majority of the forecast periods will have the lowest forecast errors. This resulted in a biased forecast much lower than the actual demand for the 5-10 items that were tested, which can be compared to the results of Teunter and Duncan (2009). To optimize MSE did not result in bias forecast because the greatest errors occurred during the demand periods.

2.5.1 MAD_n and MSE_n

In the forecasting situation where there are more periods with zero demands than periods with demands. Therefore it is not always advisable to optimize MAD when slow-moving items are concerned. Furthermore; is it the resolution of the time period that shall govern the error measurements? What happens if MAD is measured when there is a demand, MAD_n? For example if period 6 is the first period with a demand, the forecast from period 1 to 6 is summarised. The absolute error is the difference between demand and the sum of the forecast 1 to 6, see equation 2.19.

$$\mathrm{MAD}_{\hat{n}} = \frac{1}{M}\sum_{i=1}^{M}\left| X_m - \sum_{t=p-o+1}^{p} \hat{X}_t \right| \tag{2.19}$$

MSE_n is calculated in a similar manner, but with a quadratic operation instead

$$\text{MSE}_{\hat{n}} = \frac{1}{M}\sum_{n=1}^{M}\left(X_m - \sum_{t=p-o+1}^{p}\hat{X}_t\right)^2 \tag{2.20}$$

2.5.2 CFE$_{min}$ and CFE$_{max}$

If the last CFE value, when the time T had occurred, happens to be zero it might be more due to random conincidences than a proof of an unbiased forecast. An earlier bias below zero might be covered with more recent errors above zero. To diminish this phenomenon two additional CFE periods are measured namely the maximum and minimum values; CFE$_{max}$ and CFE$_{min}$. CFE$_{max}$ is equal to the greatest cumulative forecast error during the forecast, see equation 2.21. CFE$_{min}$ is equal to the greatest, in absolute terms, negative cumulative forecast error, see equation 2.22. A positive error is related to shortages and a negative error is related to surplus. The reason for this is the definition of the forecast error. The forecast is subtracted from the actual demand and therefore an overly positive forecast should result in a negative error.

$$CFE_{\max} = \max_{t\in\{1,2,\ldots,T\}}\left(CFE_t\right) \tag{2.21}$$

$$CFE_{\min} = \min_{t\in\{1,2,\ldots,T\}}\left(CFE_t\right) \tag{2.22}$$

2.5.3 Bias Errors – Percentage of Number of Shortages (NOSp) and Periods In Stock (PIS)

To be able to trace whether a forecast is biased or not, a tracking signal is used. The quotient between CFE and MAD serves as a tracking signal and is based on the assumption of a quotient that has a normal distribution. If the quotient is over a previously decided number or under the negative counterpart to the decided number, then the forecast is biased. In the initial studies of the 5-10 items there was nothing that indicated a normal distribution. To substitute, in this case, or rather complement the tracking signal, two measurements are introduced, Number of Shortages and Periods in Stock.

Number of Shortages (NOS), during the investigated time interval, is the number of times CFE_t is over zero, when there is a demand. It indicates the number of shortages the method has without safety stock since a positive CFE_t is a sign of shortage. If a method has forecasted a demand that is higher than the actual demand there is a surplus. This surplus may cover a forecast that is below the actual demand and therefore no shortage arises.

NOS can be used as an indicator of bias. A situation where there are very few shortages or a lot of shortages indicates that a bias problem might exist. Few or none NOS indicates that the forecast method, or its parameter settings, is creating a stock. The reverse situation where almost every demand results in a backorder is also a sign of bias. The forecast are below the actual demand. The reason for the name Number of Shortages (NOS) instead of stock-outs is to indicate there is no control present. To make the error comparisons easier between different items a percentage version of the error is used, NOSp. The error is the quotient between NOS and the number of the demands.

$$\text{If } X_t \neq 0 \text{ and } CFE_t > 0 \text{ then } NOS \leftarrow NOS + 1, \quad t = 1, 2, \ldots, T \tag{2.23}$$

$$NOSp = \frac{NOS}{M} \cdot 100 \tag{2.24}$$

Measures concerning forecasting mostly focus on errors instead of the behaviour of the method and the consequences of the errors. To use only NOSp and/or CFE do not explicitly reveal if a method or a method's current settings is increasing the stock or systematically underestimate the real demand, which is the purpose of this next measure.

Periods In Stock (PIS) measures the total number of periods the forecasted items has spent in stock or number of stock-out periods. A period is equal to the length of the used time period. In this case a period is measured in days. To exemplify how the PIS work let us assume we forecast a total of three days (periods). Each day the forecast is one unit. In the beginning of the first period the one item is delivered to the fictitious stock (this is a simplification compared to reality). If there has been no demand during the first day, the result is plus one PIS. When a demand occurs, the demand is subtracted from the forecast. A demand of one in period 1 results in zero PIS in period 1 and

CFE of -1. If the demand is equal to zero during three periods, PIS in period 3 is equal to plus six. The item from day one has spent three days in stock, the item from the second day have spent two days in stock and the last item has spent one day in stock, see Figure 2.1.

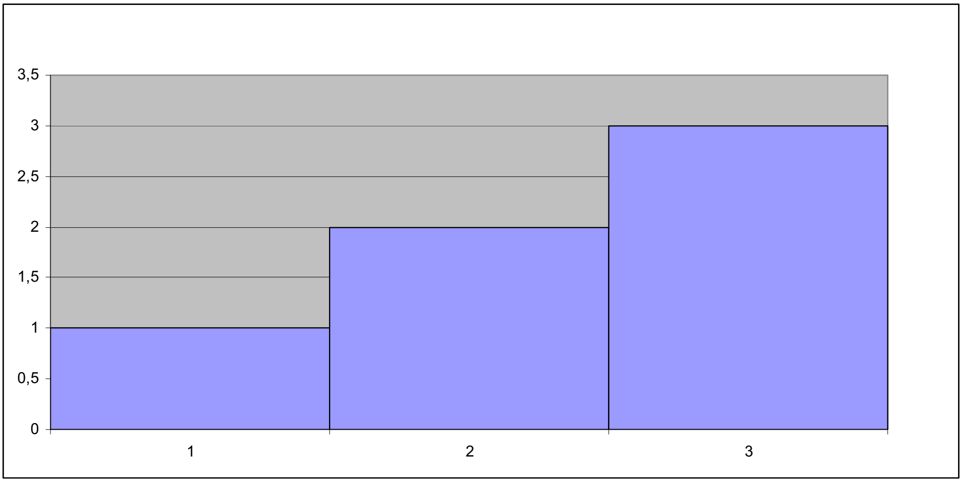


Figure 2.1 The number of forecasted items in stock for each period, the sum of the periods is equal to PIS in period 3. The x-axis represents the time periods and the y-axis is –CFE.

A positive number is a sign that the forecasting method tends to overestimate the demand. A negative number is a sign of underestimation of the demand. Therefore the error subtraction is reversed, forecast minus demand. If the forecast period is longer (e.g. a month) it is possible that the replenishment can be made on more than one occasion or in another moment besides the beginning or the end of a period. In such a case a linear replenishment version may describe the circumstances better. In the thesis immediate replenishment is assumed. PIS, immediate replenishment:

$$PIS_t = PIS_{t-1} + \sum_{i=1}^{t} (\hat{X}_t - X_t) \tag{2.25}$$

$$PIS_t = PIS_{t-1} - CFE_t \tag{2.26}$$

$$PIS_t = \sum_{i=1}^{t}\sum_{i=1}^{t}(\hat{X}_t - X_t) = -\sum_{i=1}^{t} CFE_i \qquad (2.27)$$

2.6 Additional Measure of Time Series

Coefficient of variance (CV) is a common measure to evaluate if the demand is difficult to forecast. CV is the quotient between the standard deviation and the mean and is therefore dimension free. The higher the number, the more difficult the time series is to forecast. To determine whether a time series is difficult or not to forecast, may demand more information than CV can offer. Two different time series with similar CV can be very different from a forecasting point of view. One reason for that is the CV does not consider the order of the demand. A pattern where high or low demand is mixed in a random order can have the same CV as a pattern that displays season, trend or both. To demonstrate this limitation with CV a gamma distribution with 40 observations is used (first time series; original order), see Figure 2.2.

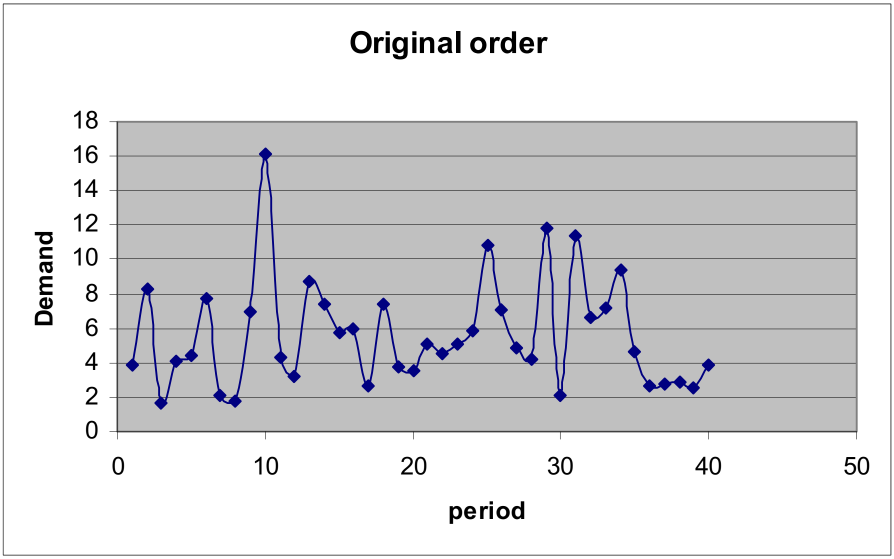


Figure 2.2 First time series with the original order for the time series, CV=0.59, MACs=0.59.

Then a new order was manually created from the original order so that the time series should have both a kind of season variability and trend (second time series), see Figure 2.3.

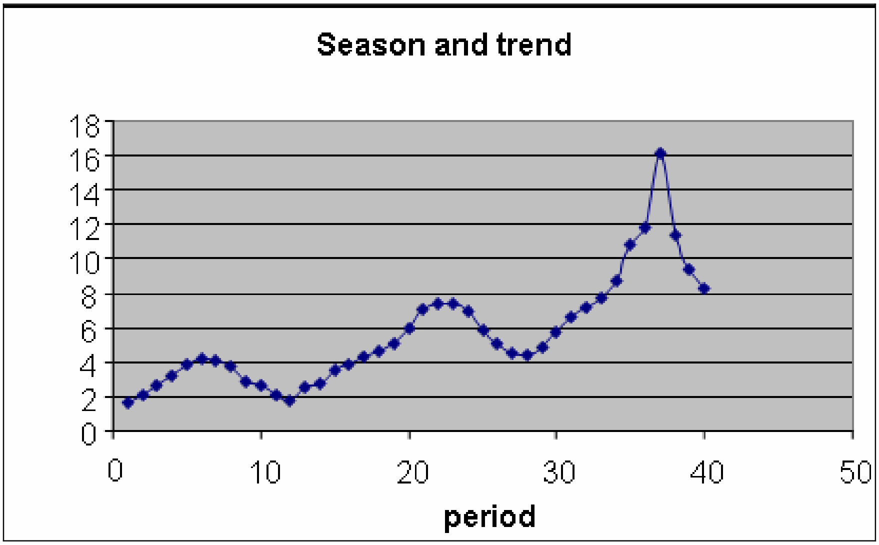


Figure 2.3 Second time series with season and trend, CV=0.56, MACs=0.15.

In Figure 2.4 the third time series has the following data; demand consists of 40 periods where one type of demand size dominates. The dominating demand size is 1. There are 2 periods where demand is 2 and 1 period where the demand is 8. Since autocorrelation measures linear covariation (Makridakis et al, 1998), there is no significant autocorrelation in the time series. The CV is 0.89 which is even larger than in the previous example which can be interpreted as that the third time series is more difficult to forecast despite the fact that 93% of the observations has a demand of one unit.

A study of the original order (first time series), 'season and trend' (second time series) and the third time series, the order reveals a crucial difference. The distances in demand are greater between consecutive observations for the original order compared to 'season and trend' or the third time series. A measure based on the distances between the observations that are in the same order as the time series should therefore be different for the three demand series.

To be certain that upward and downward random patterns do not cancel each other, the absolute values of the distances are calculated. The sum of these values is divided by the number of distances, mean absolute change (MAC), equation 2.28.

$$\mathrm{MAC} = \frac{1}{n-1}\sum_{t=1}^{n-1}\left|X_t - X_{t-1}\right| \tag{2.28}$$

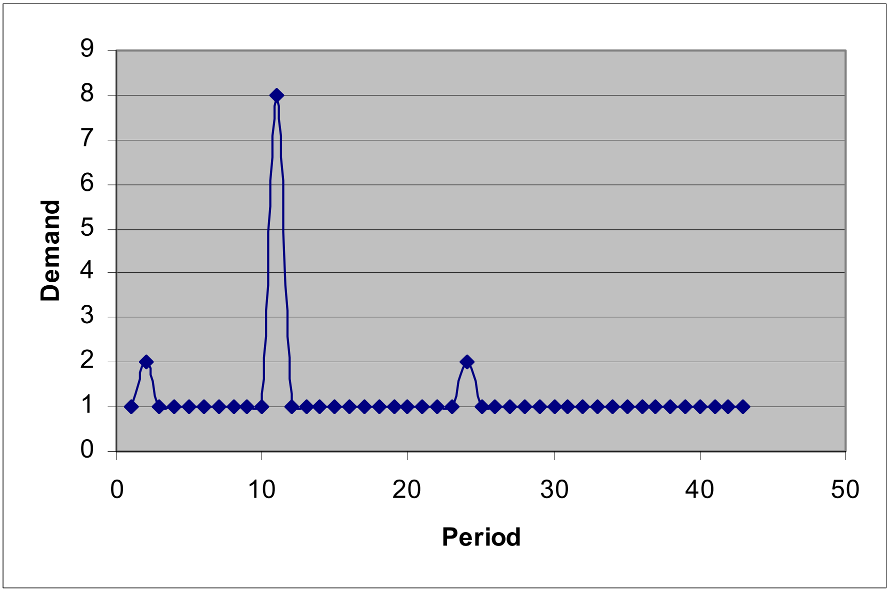


Figure 2.4 Third time series, CV=0.89, MACs=0.35.

Since MAC is scale dependent it is a good idea to make the measure scale-independent to make comparisons possible. It is done by dividing MAC with the mean of the demand; mean absolute change scaled (MACs), equation 2.29. With MACs complementing CV it may be possible to do a better classification of a demand pattern then just using CV since MACs contains additional information. Shah (1997) used discriminant analysis of descriptive statistics to select a suitable forecast method. One of the conditions for discriminant analysis is multivariate normal distribution of the data. To fulfil this condition each measure must have a univariate normal distribution (even if the measures have univariate normal distributions there is no guarantee for multivariate normal distribution). Since it is not certain that every standard descriptive measure has a normal distribution MACs may serve as an alternative.

$$\text{MACs} = \frac{\frac{1}{n-1}\sum_{t=2}^{n}\left|X_t - X_{t-1}\right|}{\frac{1}{n}\sum_{t=1}^{n}X_t} \tag{2.29}$$

In Table 2.1 are a summary of the three earlier presented time series. The first and second series share the same data but the sequence differs. The sequence affects the possibilities to an accurate forecast which MACs reflects but not CV. An autocorrelation analysis would also detect the pattern that is present. But the third series lacks the linear change that is necessary in order for the autocorrelation to work. Instead of a linear or curve-like pattern the majority of the data has the same size with a few irregular demand spikes. This is not unlike the demand for some of the items in this thesis. To be able to get information surrounding the sequence MACs can be used as a complement to CV. A none scientifically tested rule of thumb, is that a deviation between CV and MACs larger than 10-15% indicates that there are some kind of none random sequence in the time series or that a value dominates the time series.

Table 2.1 Summary of CV and MACs values for the three time series.

Series	CV	MACs
1	0.56	0.59
2	0.56	0.15
3	0.89	0.35

The idea of MACs originates from an error measure from Hyndman and Koehler (2006). Their error gave the idea to use a variant as a complement to CV. They presented an error called Absolute Scaled Error (ASE). The error is defined as

$$\text{ASE} = \frac{e_t}{\frac{1}{n-1}\sum_{t=2}^{n}\left|X_t - X_{t-1}\right|} \tag{2.30}$$

There is no claim that MACs is new but the search for this kind of descriptive statistic for time series has generated almost nothing; Makridakis et al (1979) mentions mean absolute percentage change without a definition. Answers from different persons with statistical professions have three things in common; the first is the conclusion that an error like MACs must exist, the second is the lack of idea what such a measure is called and the third is no idea whom might have the necessary knowledge.

2.7 Evaluations

Mathews and Diamantopoulos (1994)

Mathews and Diamantopoulos (1994) explore the different dimensions among 14 measures to find the underlying dimensions. The measures include forecast errors such as percentage errors, variance errors and bias errors. Also included are descriptive statistics and tracking signal. The data set based on real demand data has 691 products and spans over six quarterly periods. Products with an intermittent demand were removed. The two tested forecast methods are a naïve forecast where the last demand is the forecast for the next future demand and a variant of exponential smoothing that also has a smoothing function for the trend (Holt's exponential smoothing).

The analysis was done first with principal components analysis and later a factor analysis was applied. The results of the study are that the underlying dimensions are not dependent on the used forecast method and that the number of dimensions is four.

The first factor explains more than 33% and is the ratio-type accuracy where percentage errors and quotients of descriptive errors can be found. The second factor explains over 25% and is the volume-based accuracy measures where the variance measures (MAD, MSE and the standard deviation of error) dominates. The third factor explains a smaller part of the variability and is the bias factor where the mean error (CFE divided with number of time periods) and tracking signal have the strongest loadings. The fourth factor is the least important factor and loads only one variable; RSQ. RSQ is a coefficient of determination that is a pattern-matching indicator; this sets the RSQ measure apart from the other measures that are distance related.

Willemain et al (1994)

Willemain et al (1994) compare SES and Croston with both real demand data and simulation. The real demand data comes from four sources and the percentage of possible demand occasions varied from 0.2% to 83%. The resolution of the inter-demand varies, days, weeks and months are used. The mean demand varied from 1.2 to 7764. The relationship between demand and inter-demand proved to be correlated. The time series for both demand and inter-demand were autocorrelated. The two findings are violations of the assumptions Croston (1972) presented. The simulated study is based on a Monte Carlo comparison with geometric distribution for inter-demand interval and lognormal distributions for the demand. The simulation has four scenarios:

- Scenario 1: Demand and inter-demand are uncorrelated. The scenario Croston (1972) assumed.
- Scenario 2: Demand and inter-demand correlated. Both positive and negative crosscorrelations are tested.
- Scenario 3: Demand is autocorrelated. Both negative and positive autocorrelations are tested.
- Scenario 4: Inter-demand is autocorrelated. Both negative and positive autocorrelations are tested.

The forecasts are evaluated with mean absolute percentage error (MAPE), median absolute percentage error (MdAPE), MAD and MSE. The percentage errors were summarised before the percentage were calculated to avoid infinity of the errors.

The smoothing constants were; 0.01, 0.1 and 0.5 for the simulated data set and varied between 0.01 and 0.9 for each method according to a grid search. Two types of start values were used; the first with 0 demand for both Croston and SES and 1 as the start value for the inter-demand interval, the second start value was based on the first demand and inter-demand information.

The results of the study are that Croston is the better choice. The greatest difference between SES and Croston are when the demand and inter-demand are uncorrelated. Croston's performance is better in the simulation situation compared to the situation with real data. Croston is still better than SES but the improvement to switch from Croston to SES is smaller.

Eaves and Kingsman (2004)

Eaves and Kingsman (2004) examine SMA, SES, Croston and SyBo with the purpose of finding a suitable method for intermittent demand. The data set is real data from RAF and covers a 6-year period with the transactions aggregated into quarterly, monthly or weekly demand. 18 750 items were randomly selected. They classify the data according to Table 2.2.

Table 2.2 The classification used by Eaves and Kingsman (2004).

Variability			*Demand pattern*
Inter-demand	*Demand*	*Lead-time*	*Classification*
Low	Low		Smooth
Low	High		Irregular
High	Low		Slow moving
High	High	Low	Mildly intermittent
High	High	High	Highly intermittent

The evaluation was done with forecast errors and stock-holding consequences. The used forecast errors were MAD, RMSE (root MSE) and MAPE. The methods were optimised with MAPE and a hold-out sample. An optimisation of the stock-holdings consequences were optimised based on the costs and a hold-out sample.

The results between the best methods in each category, forecast error and stock-holdings consequences, differs. When the evaluation is based on the forecast error, no method is unambiguous the best method. In the case of one period ahead, moving average method is best when MAD is used, SyBo is best when MAPE is used and SES is best regardless of measure when forecasting demand in all periods. When the evaluation is based on the economic consequences, SyBo is best since the method, regardless of demand patterns, has the lowest stock-holding. The savings can be considerable if a more accurate forecasting method is used since the safety stock can be reduced without any major implication affecting the service levels.

Syntetos and Boylan (2005)

Syntetos and Boylan (2005) introduce SyBo and test the method against moving average (SMA) with 13 periods, SES and Croston. The smoothing methods have four different constants, 0.05, 0.10, 0.15, and 0.20. The methods are tested with real, intermittent demand data from the automotive industry. The 3000 stock keepings units have over 2-years of documented monthly data. The average inter-demand interval varies from 1.04 to 2 months (50-96% of the possible demand occasion periods). The average demand sizes, when demand occurs, vary from 1 to 194 units.

The absolute forecast errors are mean error and relative geometric root mean square error (RGRMSE). Mean error is considered less scale dependent than MAD and MSE are therefore the scale dependency is avoided. RGRMSE is used since it is not sensitive to the influence of outliers. In addition relative measures are used; the quotient between RGRMSE and percentage better (PB). Percentage better is the percentage of times a certain method performs better than one other method. In addition the percentage method is best compared to the other methods, Percentage Best (PBt). SyBo is considered to be the best method for intermittent demand and PBt is better than PB. The reason is that in practice only one method is chosen and therefore PBt resembles a real world choice.

Syntetos and Boylan (2006)

Syntetos and Boylan (2006) evaluate the stock control performance of forecasting intermittent demand with four methods; Croston, SES, SyBo and a moving average with 13 periods. Four smoothing constants are used; 0.05, 0.10, 0.15, 0.20. The data set is 3000 stock keeping units that has a monthly demand history of two years. The descriptive statistics are the same as Syntetos and Boylan (2005).

The forecasting methods are compared using the methods’ inventory control performance with three constraints; a specified customer service level and two cost policies. The accuracy measures are the Percentage best (PBt) and the average percentage regret (APR). APR is a relative measure where a method is compared to the best method and measures the deterioration of choosing an inferior method. The first 13 months are used for the initialisation and the 11 remaining months are used for the assessment of the forecasting methods.

The results of the assessment are that SyBo is the most cost effective method with the best inventory control performance. When service is concerned the moving average is also a good alternative while both SES and Croston are not as good, and SES is the worst method of them all.

Boylan and Syntetos (2007)

Boylan and Syntetos (2007) evaluate ModCr together with Croston and SES. Two smoothing constants were used; 0.1 and 0.2. The evaluation is based on simulation where five different inter-demand interval was used ranging from 1.1 to 10 (10-91%). The simulations consisted of 20 000 periods repeated 5 times. “Following Willemain et al (1994)”, as the authors states. The forecast error used is MSE.

The results of the simulations are that SyBo is the best, apart from 1.1 inter-demand intervals and that ModCr has its best accuracy performance when the inter-demand intervals are low. ModCr is worse than both SyBo and Croston concerning MSE performance.

Teunter and Sani (2008)

Teunter and Sani (2008) test the bias of Croston and some of its variants. The four tested methods are; Croston, ModCr SyBo and a previously unpublished method by Syntetos. The evaluation is done with simulation. The demand distributions used are; normal with mean 1, discrete uniform between 1 and 2, discrete uniform between 1 and 10. The probability of a demand are; 0.1, 0.3, 0.5, 0.7. The start value are the correct values and the smoothing constants are; 0.1, 0.2, 0.3. For each of the 48 experiments 10 000 periods are generated randomly.

The conclusions are that ModCr is the most overestimating method, and that SyBo can be just as biased as Croston but with an underestimation instead of an overestimation. The least biased method is the method of Syntetos, but the drawback with the method of Syntetos is a variance performance that is not as good as SyBo.

Teunter and Duncan (2009)

Teunter and Duncan (2009) study the traditional forecast errors (MAD, MSE, RGRMSE) in relation to target service levels and stock holding implications. The data set is the demand for spare parts and has 5000 items, the period is 6 years. The number of demand occasions varies from 0.5 to 3 per year (4-25%

of the possible demand occasion periods) and the mean demand varies from 1 to 1330.

The forecast methods are simple moving average (SMA), SES, Croston, SyBo and a bootstrapping method. The smoothing constant is 0.15 with a sensitivity analysis from 0.10 to 0.20 where no significant changes could be noted. Bootstrapping estimates the distribution by repeating the sampling of a certain size. The sampling can be done with or without replacements. A zero forecast is used as a benchmark technique.

They draw the conclusion that MAD favours underestimating forecasting methods since the zero forecast has the lowest MAD. The other favoured underestimating methods are SMA and SES. Therefore the stock holding implications, service levels and a bias measure must be considered. The other traditional measures had similar problems but not as severe. Contradicting results may depend on the distortion caused by traditional errors.

3 METHODOLOGY

The third chapter addresses the issue of how and why certain methods are used or not used. The chapter starts with an introduction to why a methodology is of importance and continues with the theories on which the decision for the chosen path of the research is based. Statistical methods are described and the chapter ends with a description of the experiments and the data set.

3.1 Introduction to Research Process and Methods

20^{th} September 1804 the French astronomer Pierre-François-André Méchain died of yellow fever in Spain. The reason for his stay in Spain was an attempt to correct anomalies from the observations he had done a decade earlier at Montjuïc, Barcelona. The stars he used in order to decide the positions on earth for the measuring instrument he used proved to give different results and precisions. For one of the stars, Mizar; the precision was about one tenth compared to the others. Méchain blamed himself for the anomalies. The reason for the observations was to establish the size of the earth and as an extension decide the length of one meter.

The observations in Barcelona were just a part of an expedition measuring the length from Dunkerque to Barcelona. These anomalies had haunted him since the winter of 1793-94 and to avoid that the anomalies where detected by anyone else, he spent seven years in exile in futile attempts to solve the problems with the anomalies. Méchain tried to keep the observation data to himself until he had solved the anomaly problems, but as the pressure increased from the others involved in the project he had to do something. He presented a form of mean values of the observations, a cover up that was not detected until after his death. (Adler, 2003)

For Méchain and the scientists of his time the method was important, at least the mechanical measuring methods, but the importance of the methods were nothing compared to the scientist himself (most scientists or nature philosophers were men although Méchain's wife was more than capable of performing her husbands work at the Parisian observation during his absence). The scientist was the guarantee for good science, not the methods that was used. Since Méchain had performed the observations his position and reputation as a scientist was threatened. (Adler, 2003)

In this day and age, how a problem is approached, the method is of importance. Lundahl and Skärvad (1999) consider a scientific research as research based on scientific methods with the objective to produce theoretical contribution. The method used can not be chosen arbitrary. According to Forsman (1997) the scientists in Germany during the Nazi regime could use any methods they saw fit as long as they could answer the hypothesis. However, regarding the hypothesis or research questions, the freedom was limited. A hypothesis regarding Slavic or Jewish people and their superiority would not have been allowed. The freedom in research should be in the research questions, not the methods.

Research design is an action plan to get from a starting point, with a number of initial questions, to an end point with a number of conclusions as a result of the answered questions. The design is the logical sequence based on what type of data that is available and the research questions (Yin, 1994). A part of the research design is to choose a suitable method according to the research questions and the data material that is available or possible to collect.

Description is a method that has the aim to describe the studied phenomena in a thorough manner. This makes it necessary to a continuous selection of the data which has to be categorised and sorted before the data can be of any use. The method is suitable when the purpose is to create an overview. (Ejvegård, 2003)

The case study is appropriate when the problem formulation is not yet established (Ejvegård, 2003). In an initial stage the aim of the case study is to increase the understanding of the problem rather than being explanatory. A sign of the case study is the great number of variables studied on a few objects, while in the statistical analysis the variables are few and the objects are many (Lekvall and Wahlbin, 2001).

According to Lundahl and Skärvad (1999) an experiment is when the researcher changes the values of the independent variables in order to measure the size of the effect among the dependent variables. The aim of the experiment is to measure the effect for one or more variables. The influence from non controlled sources must be eliminated. Gorad (2003) states that; "Social science research has, for too long, relied on fancy statistical manipulation of poor datasets, rather then well-designed studies".

Simulation is a variant of experiment. In order to use simulation the model of simulation must be extensive enough to resemble the phenomenon that is studied. The data in the model is simulated with different statistical distributions according what is regarded suitable. (Lekvall and Wahlbin, 2001)

In this thesis the experiment was chosen since the first task was to use a data set that had been used by a former PhD student to confirm the earlier results. The most suitable method to use the date set was to conduct experiments. Instead of just confirming the results several questions was raised regarding the evaluation methods such as; How does one measure a measure?

3.1.1 Quantitative and Qualitative Research

Quantitative research is suitable when the information can be expressed in numbers, which makes it possible to apply statistical methods and analysis in search of patterns and/or correlations (Eriksson and Wiedersheim-Paul, 2001). The relationships between different variables are one of the main issues in quantitative research, while the knowledge of the process behind the relationships is not as important (Denzin and Lincoln, 2000).

Qualitative research on the other hand is suitable when the information is best described with words rather than numbers for example interviews. In qualitative research the emphasis is on the understanding of a process (Patel and Davidsson, 2003). Another distinguishing feature is the narrative description of the problem (Remenyi et al, 1998). According to Lekvall and Wahlbin (2001) a research is rarely just qualitative or just quantitative. However, the interpretation and analysis on a higher level is always qualitative.

In this thesis the quantitative research is the foundation. There are research question that implies emphasis on the quantitative aspects such as the relationship between different errors. However of at least equal importance is to gain a deeper understanding of the process behind different relationships. The result from the different statistical methods are first analysed detached from each other and at a later stage the results of the methods are combined in order to increase the knowledge that goes beyond significant variables.

3.1.2 Induction and Deduction

The choice of methodology can be divided into two main types; induction and deduction. When induction is used the understanding is based on the empirical data. Deduction is when the understanding of a problem is founded in theory. The presentation of a problem can also be approached with induction or deduction. In the case of induction a data material is collected without a well defined problem. In the analysis the material is studied in search of interesting phenomenon that is investigated further. In the deductive approach a well formulated problem exists before the data collection starts. (Hellevik, 1984)

Induction was primarily used during the initial stage of the experiments, to avoid being influenced by answers before an understanding of the problem was reached. The initial problems were to find suitable measures and figure out what made an error measure suitable or not. To just use the forecast errors used with demand in every period led to distortion of certain forecasts in certain situations. After that a combination of induction and deduction has been used.

3.1.3 Data Collection

In a research project there are two types of data; primary and secondary data. Primary data is data collected solely for the research. Secondary data is data that is collected earlier for another purpose. (Eriksson och Wiedersheim-Paul, 2001)

The data was already collected and tests had been made by a previous PhD student. The initial experiment concerned finding the best method for intermittent demand. Since there was no documentation that could confirm the results, the experiments had to start from the beginning by choosing the items, forecast methods and forecast errors.

3.1.4 Sampling

The sampling issue is of importance to the research since it might be impossible to study every individual in a data set (in this case; every one of the initial approximately 20 000 items, author’s own comment). Therefore it is necessary to select a subset of the whole population, to choose a sample of the population. A major concern regarding sampling is that the sample is representative of the whole population in the important characteristics. If the sample is not representative it will be difficult to draw generalising conclusions. Even if the sample is non-biased the variability may differ in comparison to the population. (Remenyi et al, 1998)

The sampling technique is categorized in two classes; non-probability samples and probability samples. If non-probability samples are used then the statistical techniques may not be used in the analysis. In non-probability sampling the subjective judgements of the researchers are employed. Some of the probability samplings are; simple random sampling, systematic sampling and stratified sampling. Simple random sampling is when each member of a population has an equal chance of being selected. (Remenyi et al, 1998)

In systematic sampling the members of the population are chosen according to a certain interval. In worst case this means that a bias may be present. If the population consists of weekdays and the sixth weekday should always be chosen according to the sampling criteria, then whole sample consist of only Saturdays. (Remenyi et al, 1998)

Stratified sampling is a subdivision of the population in homogenous groups called strata before the random sampling takes place from each stratum. The stratified sampling can be made in two ways. First and most common is when each stratum has a sample size that is in proportion to the size of the stratum. The second stratified sampling is when the number of items from a stratum is determined according to the relative variability of the items within the strata. (Remenyi et al, 1998)

The sampling method in this thesis is a form of stratified sampling based on the number of demand occasions. The items where chosen according to their number of demand occasions that were generated randomly. If there was more than one item with a particular number of demand occasions, the items were numbered and simple random sampling was applied.

3.1.5 Validity

Validity in a measurement is the absence of systematic errors of measure. There are two general types of validity; external and internal validity. (Lundahl and Skärvad, 1999)

External validity concerns the possibility to generalise findings on other data then the data used in the research. If the new data corresponds to the significant characteristics of the original data, a replication should be possible. (Yin, 1994)

Internal validity is of interest when explanatory or causal studies between a set of variables are performed. The internal validity has been given a great deal of consideration in experimental and quasi-experimental research. (Yin, 1994)

According to Lekvall Wahlbin (2001) validity is to measure what is supposed to be measured. A complication with validity is the problem to decide whether a method is valid or not. In order to make a correct judgement regarding the validity a 'truthful' method must exist. If the 'truthful' method exists, there is no need for any other method. A method's validity is determined on grounds that are to various degrees subjective. Also the selection of data can affect the validity of the research if a bias is present in the data.

The continuing problem throughout the experiments has been to be as certain as possible that the measures or the methods chosen do not favour an individual forecasting method or forecast error; how should a measure be measured? To decrease these problems, different methods have been used.

3.1.6 Reliability

Reliability is the consistency of a measurement or measuring instrument according to Ejvegård (2003). Reliability is a question to what degree the results are possible to repeat (Merian, 1994). Lekvall and Wahlbin (2001) draw the conclusion that a common reason for low reliability is caused by a measurement that is not well defined. Appropriate methods to examine the reliability are:

- Test-Retest. Repeated tests on the same individuals, if the correlation between the test and the retest are high the reliability is also high

- Parallel-test. Two identical instruments measures the same individuals at two different occasions

- Split-half. The measuring is conducted in such a way that it is possible to split the population of individuals in two equal halves

The resolution will affect the reliability. The resolution of the instrument is related to the reliability. A high resolution makes it possible to detect small changes including random changes, which will decrease the reliability. If the resolution is decreased, it will increase the reliability since the small random changes will not be detected, but at the same time it decreases the usefulness of the results. (Lekvall and Wahlbin, 2001)

3.1.7 Triangulation

Triangulation is the use of multiple sources of evidence and can be used to increase the validity of the research. Yin (1994) mentions four types of triangulation:

- Data triangulation
- Investigator triangulation
- Theory triangulation
- Methodological triangulation

In an experiment the concentration is on measuring and recording the experiment rather than expanding the experiment with surveys (Yin, 1994).

Denzin and Lincoln (2000) discusses that triangulation is not a tool or strategy of validation but rather an alternative to validation. The use of triangulation, multiple methods, is an attempt to gain a deeper understanding of the studied phenomenon. An example of triangulation is to count the numbers of first places a certain forecasting method has when one forecast error is used and then examine the relative quotients between the measures to see how much better the method is or to add another error to see if the results are conflicting.

3.2 Data Analysis

In the experiments the 4 methods had 4 start values each and each start value had 8 smoothing constants for each of the 72 items, which makes the total number of forecasts equal to 9216. Ten different error measures are used which makes the total number of error data to 92160. To analyse every number of error individually takes to much time and at the same time it is nearly impossible to detect patterns. If the analysis and documentation for a single error takes 5 minutes, it takes approximately 4 years to analyse the error data. (Assuming a workday of 8 hours and 220 workday per year.)

Multivariate data can be defined as data where the researcher measures or evaluates more than one characteristics of each experimental unit. The multivariate techniques that are used for multivariate date are usually exploratory instead of confirmatory. Multivariate techniques have a tendency of motivating hypothesis, not testing hypothesis. (Johnson, 1998)

The techniques are divided into two categories; variable-directed and individual-directed. A variable-directed technique is concerned with relationships among the variables that are being measured. Correlation matrices, principal component analysis and regression analysis are examples of variable-directed methods. Individual-directed techniques are concerned with relationships between the items that are being measured. Discriminant analysis and logistic regression are examples of individual-directed methods. (Johnson, 1998)

3.2.1 Software

Three different programs have been in used during the research, first Microsoft Access in which the raw data was stored. Then the raw date was transferred to Excel for the forecasting experiments. The results of the forecast experiments were stored in Excel-files. The majority of the statistical analysis was performed with the aid of Minitab.

3.2.2 Statistical Measures

The Mean

The mean is a measure of location and is calculated by the adding of every measurement concerning one type of measure. The sum is divided by the number of measurements. A drawback with the mean is that it is influenced by outliers, very large or very small observations compared to the majority of observations. (Holme and Solvang, 1996)

$$\text{Mean} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3.1}$$

X Observation

n Number of observations

The Median

The median is another measure of location and is the observation in the middle when every observation is sequenced according to the size of the observation. If the number is even the median is the mean of the largest number of the first half of the sequenced observations and the smallest number of the second half of the sequenced observations. This makes the median insensitive of the influence of outliers and can therefore be a better choice when dealing with skewed distributions. (Remenyi et al, 1998)

Standard Deviation

Standard deviation is a measure of variance, how widely the observations are scattered in relation to the mean. (Remenyi et al, 1998)

$$\text{Standard deviation} = Std_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}} \qquad (3.2)$$

Range

The median is a simpler measure of the location compared to the mean. Under some circumstances it may be more appropriate to choose an alternative to standard deviation. Range is the difference between the largest value and the smallest value. While the median is relatively insensitive to outliers, range is not. A very large and/or small observation will affect the value. (Körner et al, 1984)

3.2.3 Statistical Methods

Regression Analysis

Regression analysis is a method to explain data in a compact form according to Hamilton (1992). The analysis can describe the relations between a response variable and one or more predictors. Despite the fact that new software and improved computer performance, that use more sophisticated statistical methods, regression analysis still proves to be a valid method for many types of problems. (Weisberg, 2005)

Before the regression analysis, graphic plots of the variables should be studied since the plots can reveal if a linear dependence exists among the variables. The simplest form, one predictor and one response variable, only need a two dimensional plot. (Hamilton, 1992)

A linear regression assumes a correlation or covariance between the response and the predictors. Covariance (COV) is a measure of the way two random variables, *X* and *Y*, vary together. If the variables are independent the covariance is equal to zero. In practice the covariance can differ from zero and the variables can still be independent. To avoid units of measurement and scale of the variation, correlation is used. The denominator of the correlation is the product between the standard deviations of *X* and *Y, Std_X and Std_Y.* The correlation coefficient has a value between -1 and 1. (Weisberg, 2005)

Makridakis et al (1998) has the following definition of the covariance and the correlation coefficient:

$$\mathrm{COV}_{\mathrm{XY}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) \tag{3.3}$$

Correlation (*r*) between *X* and *Y*:

$$r = \frac{COV_{XY}}{\frac{1}{n-1}\sqrt{\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right)\cdot\left(\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right)}} \tag{3.4}$$

Simple regression is a regression between a single Y variable and a single X variable. The linear relationship between Y and X can be described as

$$Y = a + bX + e \tag{3.5}$$

a is the intercept, *b* is the slope of the line and *e* is the deviation of the observation from the line. The slope *b* and the correlation coefficient *r* have the following relationship

$$b = \frac{COV_{XY}}{S_X^2} = r\frac{S_Y}{S_X} \tag{3.6}$$

The correlation is a widespread measure that can be helpful in an analysis. In Table 3.1 there is a rule of thumb for interpretation of the correlation coefficient. But the measure is not without drawbacks. One drawback is that it describes a linear relationship. A non linear relationship might not be traceable in the value of the correlation coefficient. Another drawback is that correlation is sensitive to influential extreme points. In Figure 3.1 the first 19 observations have a correlation coefficient of -0.148.

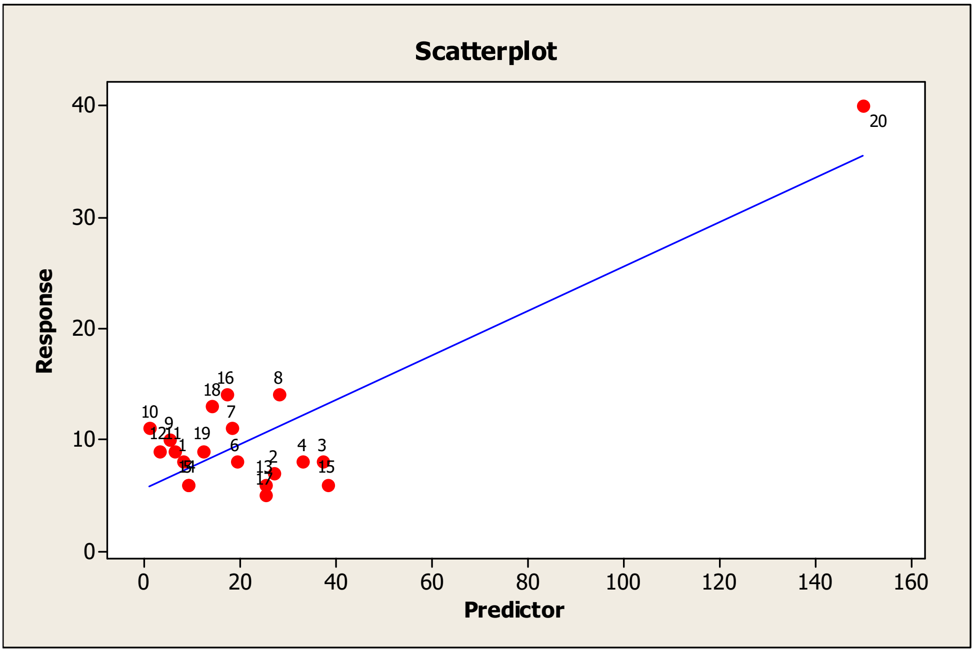


Figure 3.1 Influence of an extreme observation. Observation 20 causes an impression of a linear relationship between the predictor and the response that is described with the line in the figure.

When observation 20 is added, the correlation coefficient increases to 0.847. It is not necessary that the correlation distorting observation, like number 20, needs to be larger than the rest of the observations in order to create the distorted correlation value. It can also be smaller than the other observations.

The main issue is that the highly influential observation is sufficiently far apart from the rest of the observations in X- and Y-space.

In order for the linear regression to be valid some conditions must be fulfilled. The errors, residuals, should be randomly scattered around zero and have a normal distribution. The influential observations that affect the equation to a great deal should be omitted, like observation 20 in the previous paragraph. This in done in order to check how much the omitted observations influence the equation. Even if there is a real correlation between two variables, the amount of the correlation must be large enough. In Table 3.1 is a classification of the strength of a correlation.

Table 3.1 Guide to interpretation of the correlation coefficient. (Tersine, 1988)

Absolute Value of Correlation Coefficient	*Interpretation*
0.90-1.00	Very high correlation
0.70-0.89	High correlation
0.40-0.69	Moderate correlation
0.20-0.39	Low correlation
0.00-0.19	Very low correlation

Principal Component Analysis

Principal component analysis (PCA) is a tool for screening multivariate data, usually performed early in the analysis process to examine how different variables relate to each other. PCA uses either the covariance matrix or the correlation matrix of the variables to create new uncorrelated variables called principal components. This is done by creating new axes instead of x and y. (Johnson, 1998)

The new axes are angled so that the first axis intersects the observations in such a way that the axis explains as much as possible of the variability between the observations, see Figure 3.2. The next (second) axis is orthogonal to the first axis; the third is orthogonal to the second and so on. Every new component (axis) that is added explains less of the variance than the previous. The number of components can at maximum be equal to the number of original variables. The number of components should account for 70-90%. If the scale of the variables greatly differs it is better to use the correlation matrix since it is not scale dependent like the covariance matrix. (Johnson, 1998)

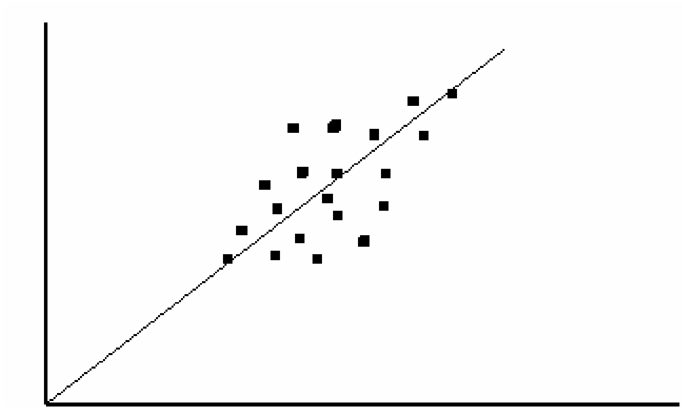


Figure 3.2 A new axis straight through the observation in order for the new axis to account for as much of the variability as possible.

The original variables are the initial axis (dimensions) before the transformation to the new axis, the principal components. The transformation is not a replacement of the original dimensions since the original dimensions (variables) are necessary to create the principal components. Some variables can be more influential than other variables in some principal components and therefore can different principal component be regarded as different dimensional subgroups. It is not always possible to interpret the principal components in practical terms as well as a subjective interpretation can find non existing patterns. (Johnson, 1998)

PCA is used to find the appropriate dimension of the original variables. When the correlation matrix is used the eigenvalue of a principal component can be of guidance, a value less than 1.0 means that there is noise present, the lower the eigenvalue the more noise content of the component. In Table 3.2 the first component accounts for 77.1% of the variability and variable 1 is dominating the first component. Variable 2 has the highest weight in the second component but since the first component accounts for 77.1% of the variation and that the second component has an eigenvalue much lower than 1.0. Only the first component will be used. The dimension of the three variables can be reduced to 1. (This is not the only possible solution to the number of dimensions). (Johnson, 1998)

Table 3.2 Example of information from a PCA.

	PC 1	PC 2	PC 3
Eigenvalue	2,312	0,616	0,071
Proportion	0,771	0,206	0,024
Cumulative	0,771	0,976	1,000

Variable	*PC 1*	*PC 2*	*PC 3*
Variable 1	0,638	-0,174	0,400
Variable 2	0,224	-0,933	0,057
Variable 2	0,124	0,40	0,87

In order for a PCA to be useful, a correlation is necessary and this means that the scatterplot between the original variables should be examined in order to avoid distorted correlation coefficients created by one influential observation. An alternative is to study the score plot of the components, see Figure 2.1.

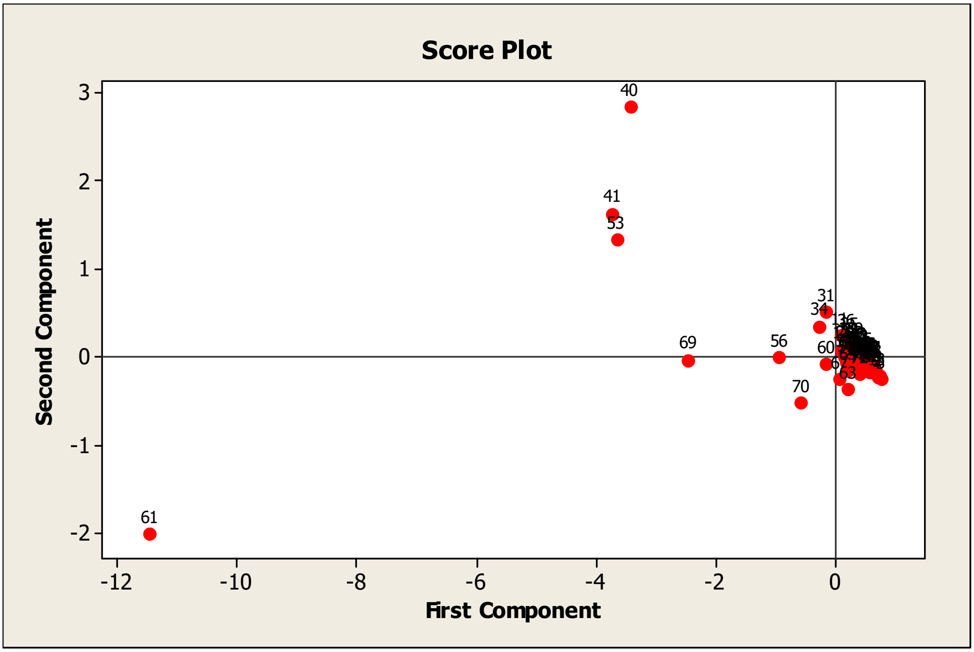


Figure 3.3 Score plot of the first two components in the PCA. The numbers in the figure are equal to the numbers of the items.

Observations that lie far away from the main cluster of observations are most likely extreme points that can influence the appearance of the PCA. In Figure 3.3 are 61, 40, 41, 53 and 69 extreme observations that can create the impression of linear relationship between variables without there actually exist a linear relationship just like correlation. A scatterplot between the variables can also reveal outliers. (Johnson, 1998)

Mathews and Diamantopoulos (1994) used factor analysis in their research where PCA also was used. In this thesis no factor analysis is used because of the criticism of the method. According to Johnston (1998) a major criticism is the nonuniqueness of the solutions in combination of the researcher subjectivity that can give an objective impression of a method based on subjective preferences. Practitioners do think factor analysis is better than its reputation.

Logistic Regression

Logistic regression is a method that classifies the observations into different groups. Logistic regression is similar to multiple regression but with the addition of a dependent variable (the response) that is binary. However the logistic regression is fitted by the use of maximum likelihood methods instead of the least square that is used for regression. Logistic regression can be used to determine whether new customers are a credit risk or not. (Johnson, 1998)

X is a data vector for an item (unit) that is randomly selected and y is the value of the binary outcome. If X comes from population 1 then y = 1 and y= 0 if X comes from population 2. If $p(y = 1 \mid X)$ equals the probability of y = 1 than the logistic regression can be described as:

$$p(1 \mid X) = \frac{e^{g(x)}}{1 + e^{g(x)}} \tag{3.7}$$

The above equation is the classification equation where p larger than 0.5 for an individual is classified as population 1 even if the individual belongs to population 0. A value of 0.5 or less means the individual is classified as population zero.

In the study of logistic regression it is usual to take the logit transformation in consideration that is performed on $p(y = 1 \mid X)$. The log of the odds y = 1 versus y = 0 is the logit transformation and can be described as:

$$g(X) = \log\{ p(y = 1 \mid X) / [1 - \{ p(y = 1 \mid X)]\} \qquad (3.8)$$

If all the variables that are included in the model have a fixed value with the exception of one variable and that variable is changed, it is possible to see the tendency of the classification accuracy. A steeper slope usually means better classification accuracy, see Figure 3.4. The changed variable is called the significant classification variable in the example.

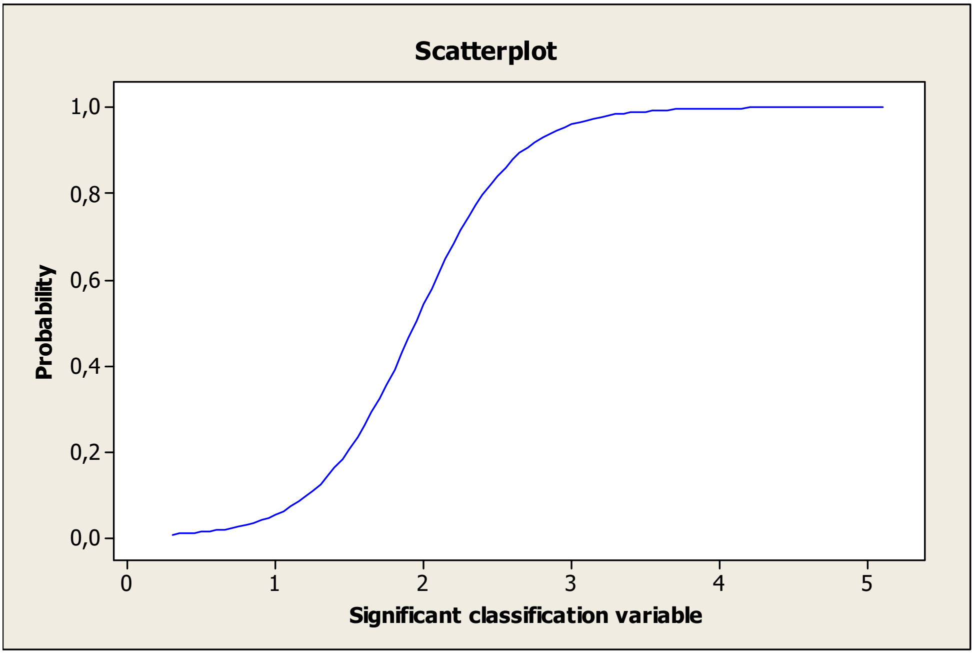


Figure 3.4 Scatterplot of model where only one variable is changed.

The logistic regression is commonly used for individual-directed analysis (Johnson, 1998). In this thesis the logistic regression is not used to study the individuals but rather the values of descriptive statistics that makes the individual items to be classified in different groups.

Two versions of logistic regression are used. One version where group A consists of items where the error increases as the smoothing constant increases, group B is the all the other items. The second version is when group A is the items that have decreasing errors as the smoothing constant increases and the rest of the items are in group B, see Table 3.3. The logistic regression is used where the mean is the start value. The idea is that the mean should be a better start value than a start value that deviates from the mean. The lowest forecast error may not be the lowest smoothing value but it should most likely be a lower smoothing constant. If the error decreases when the smoothing constant is increased from 0.025 to 0.30 it is not likely a random event since the probability is 0.0078 (0.5 to the power of seven).

Table 3.3 Summary of the subgroups used in the logistic regression.

	Smoothing constant	*Error*	*Group*
Logistic regression 1	Increasing	Increasing	A1
	All other patterns		B1
Logistic regression 2	Increasing	Decreasing	A2
	All other patterns		B2

The purpose of the logistic regression is not to find coefficients for the variables (descriptive statistics) that can be applied to a general case. It is rather to find the variables that are significant when the combination of an error and forecast method react in a certain way such as when the error decrease whiles the smoothing constant increase. Instead of coefficients the terms “high” and “low” are used. High is a positive value of the coefficient and low is a negative value of the coefficient.

To find significant variables each descriptive statistics of demand, inter-demand and demand rate is tested separately in order to avoid distortion caused by a correlation between the variables. A variable, where the logistic regression table, log-likelihood and goodness-of-fit test all corresponds to the p-value of 0.05 or less (for the logistic regression table) is considered significant. The use of multiple tests is to decrease the chance of a non-significant variables being interpreted as significant. This might result in significant variables that are rejected. The high and low values of the significant variables are then checked by performing a logistic regression of the principal components of the significant variables. To use principal components makes it possible to test all significant variables at the same time since the components are not correlated.

3.2.4 Nonparametric Statistical Methods

The nonparametric tests do not need as strong distributions assumptions as the parametric tests which are not the same as being assumption-free. When the data describes order (ranks) or counts of number of events or individuals in various categories, the nonparametric tests may be the only alternative. (Sprent and Smeeton, 2001)

The Binomial Distributed Data

The binomial distribution is relevant when the outcome only has two alternatives. The distribution has two parameters n (the total number of observations) and p (the probability that a particular outcome occurs of the two possible outcomes at any observation). The key assumptions for the binomial distribution are; the n observations (trials) are independent, only one of the two possible outcomes occurs at each observation and for every observation the probability is fixed. Therefore if the possibility p is associated with outcome A then the possibility for outcome B, q, can be written as; $q = 1 - p$. (Sprent and Smeeton, 2001)

The binomial distribution can be described with:

$$\binom{n}{x} p^x (1-p)^{n-x} \tag{3.9}$$

The binomial distribution is used to test if the numbers of first places a method have are significant.

3.3 Experimental Data

The demand data is from an anonymous manufacturing company in Europe. In the best of worlds the data would have spanned three years or more, but since it is a common practice to aggregate the original demand data after a month or so, it is difficult to attain non distorted demand data. The reason for a longer data period is that it possible is to divide the data set into two parts; one where the parameters are optimised and the second part where the actual forecast is done, Makridakis et al (1998). However, the data from the company spans over almost 18 months (April 1st 2001 to September 19th 2002). The data contains the demand quantity and date. A certain date can have more than one demand occasion. For the experiments the data was aggregated each working day, see Table 3.4.

Table 3.4 Example of the demand data summarised for every period.

Order date	Sum of Demand
2001-04-03	2
2001-04-04	7
2001-04-05	3
2001-04-06	8
2001-04-07	3
2001-04-09	12

The stratified selection was done from the 3827 items with more than 44 demand occasions, where some of the demand occasions occurred on the same day. Only items with demand over the whole 18-month period were chosen. A total of 72 items were chosen. The items are numbered according to the number of demand occasions before the 6 standard deviations limit has been applied, where item 1 has the lowest number of demand occasions. The items have periods with demand spanning from 42 to 391. Expressed in percentage the demand occasions for the different items are between 12% and 95% based on the mean of the inter-demand interval, see Figure 3.5.

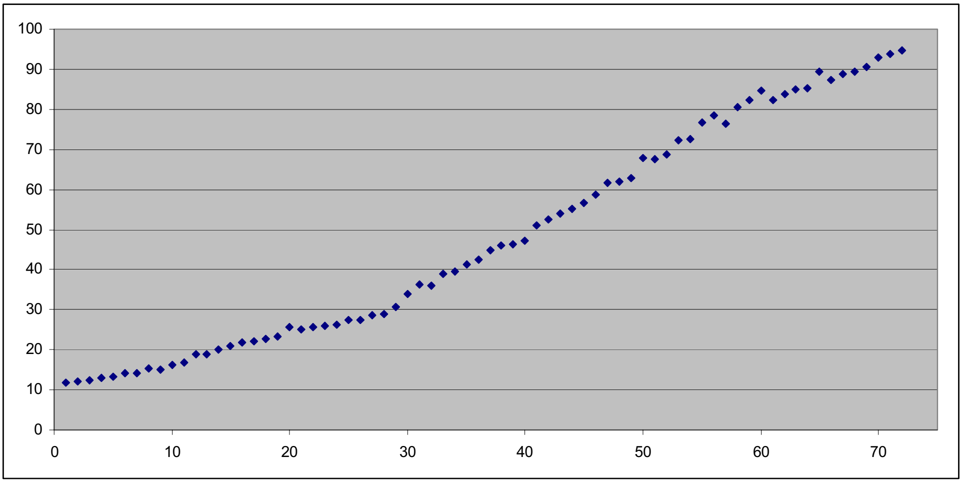


Figure 3.5 Percentage of possible demand occasions for the 72 items after the limit of 6 standard deviations has been applied

In practice the percentage can be slightly lower since the inter-demand mean do not take consider the last forecasted days unless the last day of the test period has a demand occasion. The majority of the items had a mean demand less then 10. Because of the resolution it was difficult to detect trends or seasonal variation with autocorrelation graphs in Minitab. According to Tashman (2001) the season is most prominent when the time period is a quarter or a month. Dekker et al (2004) found weekly seasonal indices. For the descriptive statistics of the experimental data see Table 3.5. A more detailed version with every item can be found in Appendix Descriptive Statistics. The relationship between mean demand with the outliers and without the outliers can be found in Figure 3.7 and Figure 3.8 (page 56).

Table 3.5 Summary of the descriptive statistic regarding the data set. A more detailed version with every item can be found in Appendix Descriptive Statistics.

	Demand/occasion				*Interdemand period*			
	Mean	*Std*	*CV*	*MACs*	*Mean*	*Std*	*CV*	*MACs*
Max	41,66	46,26	2,17	1,46	8,50	12,11	2,28	1,18
Min	1,04	0,20	0,20	0,08	1,06	0,24	0,23	0,10

The items all had distribution that where more or less skewed, in Figure 3.6 is an example of the demand observations for an item. The analysis of the distribution was carried out before any observations where removed according to the 6 standard deviation rule. The normal distribution is not suitable for any of the items or demand, inter-demand interval, demand rate. No standard distributions available in Minitab can describe the distributions of the items other than acceptable at best. Usually the distributions have kind of a log or exponential tendency. The anomalies that make the fit between the observations and standard distributions are to a high degree caused by one type of demand or inter-demand interval that dominates a series. The demand and inter-demand interval are discrete while the demand rate is continuous.

The demand and inter-demand interval are dependent to various degrees. When plotting there are a pattern for some of the items but it is not possible to use correlation in order to describe the relationship since none of the 72 items have a significant correlation.

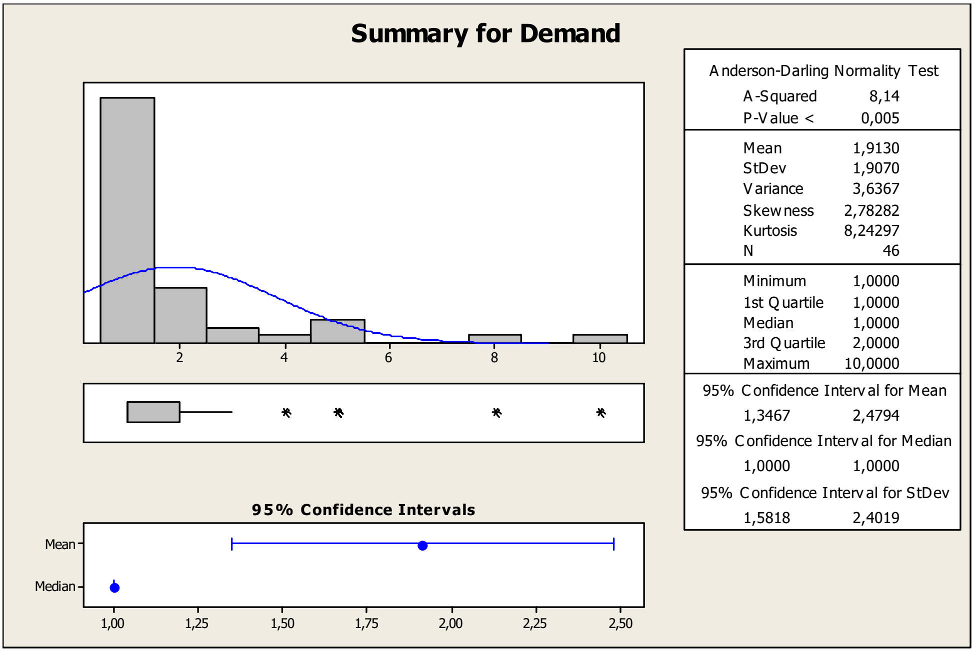


Figure 3.6 Example of a summary of observations. The summary concerns the demand of item 2.

The forecast period was every weekday, Monday to Friday, unless there was a demand on Saturday. Then the period for that week is changed to Monday to Saturday. A few demand occasions occurred at Sunday for which caused the Excel worksheet to collapse, therefore the demand on Sunday was moved to the nearest day without a demand, Saturday or Monday. If that was not possible the item was replaced with a new item from the same stratum. The reason the demand was moved only to periods without a demand, was to insure that no changes occurred in the demand pattern. The inter-demand interval pattern is not affected by the operations due to how the code in Excel was written.

For every method and every item, a limit in demand of mean demand plus 6 σ, has been used to reduce the influence of outliers. For some of the items when the demand usually consisted of 1 and the maximum demand was up to 5, no limit was used since the difference between the mean and the 6 σ was small from a practical point of view.

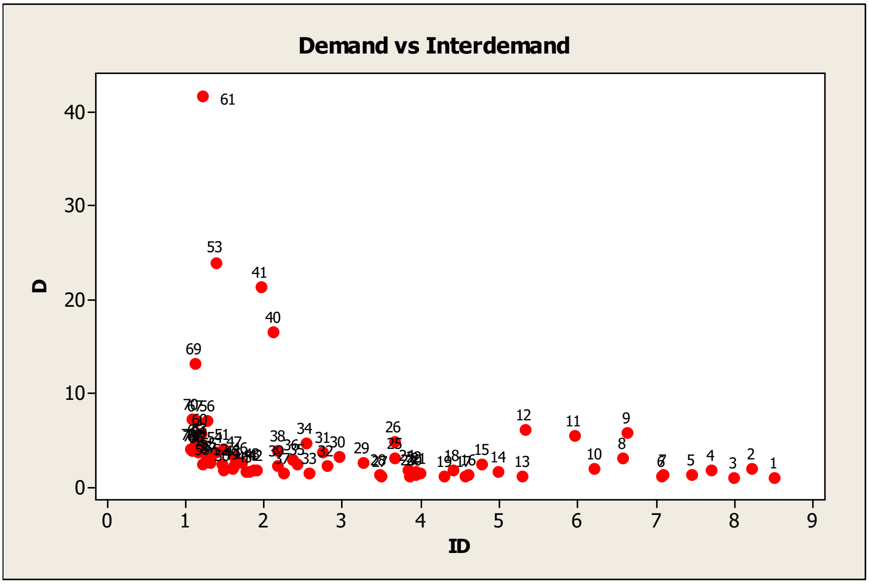


Figure 3.7 Mean demand (D) versus mean Inter-demand (ID) for the 72 items.

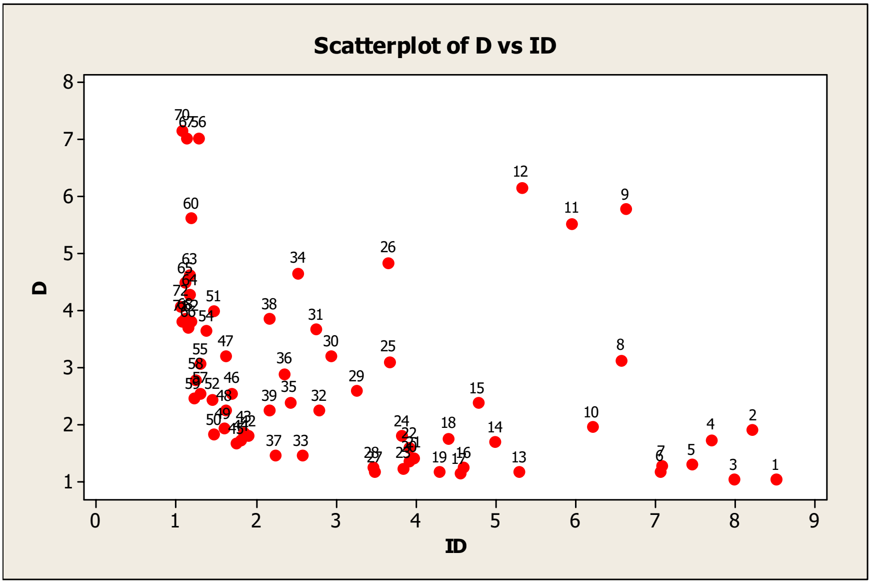


Figure 3.8 Mean demand (D) versus mean Inter-demand (ID) for the 72 items without the multivariate outliers.

Syntetos and Boylan (2005b) presented an approach to the categorization of demand patterns that are based on threshold values for the squared coefficient of variation (CV) for the demand and the average inter-demand interval. Kostenko and Hyndman (2006) suggested an adjusted version of the threshold values. The Croston method should only be used within the smooth category and SyBo in every other category. The categorisation of the 72 items is presented in Figure 3.9.

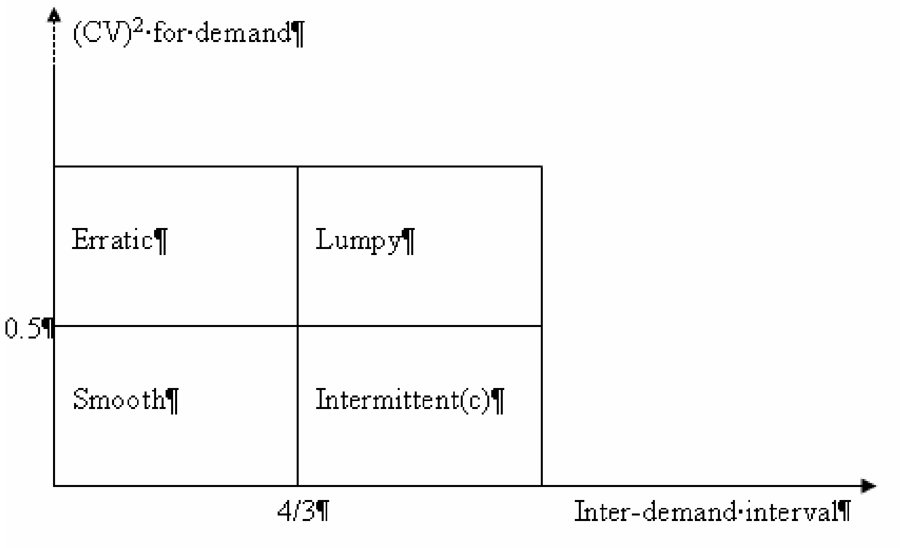


Figure 3.9 The adjusted categorization of Syntetos and Boylan (2006).

The 'Intermittent (c)' is a classification made by Syntetos and Boylan and is not equal to the definition in this thesis. According to the categorization, 18 items belong to the smooth category and 54 items belong to the intermittent (c) category. No one of the 72 items can be classified as erratic or lumpy.

3.4 The Experiments

The real demand is discrete and the forecast is not. Therefore it is not possible in the real world to replenish according to the forecast. In this study this constraint is relaxed to avoid the influence of different policies; should the discrete replenishment take place as soon as the value is larger than zero or until the value is larger than 1 or maybe somewhere in between?

The value of the smoothing constant α was varied with a number of fixed values (0.025, 0.05, 0.075, 0.10, 0.15, 0.20, 0.25 and 0.30). The choice of smoothing constants varies in the literature. Croston (1972) recommends in certain circumstances values up to between 0.2 and 0.3, Eaves and Kingsman (2004) uses 0.01 up to 0.10. To use smoothing values from 0.025 to 0.30 covers the smoothing constants used in a number of earlier tests. The second smoothing constant β was set to 0.2 for Croston and SyBo. The reason for keeping the constant fixed is the increasing number of simulations that had to be done and the fact that in reality it is not always possible to fine-tune every parameter.

For each smoothing constant and forecasting methods five different approaches to start values for the forecasts were used, the mean for the whole 18-months ±25% of the mean and naïve start value. The naïve start value is the value based on the first demand that occurs in the same manner as Willemain et al (1994). Approximately 50% of the naïve start values were larger than the mean. In a real situation when the start value should be decided one does not have the luxury of the mean for the future forecasts. But since it might be possible with an educated guess by the user, the mean increased with 25% and the mean decreased with 25% were also used. A summary of the definitions can be found in Table 3.6.

Table 3.6 Summary of the definitions of the start values.

Start value	Definition
mean-s	The mean for the whole forecast period
-25-s	75% of the mean-s
+25-s	125% of the mean-s
Naive	The first demand and inter-demand period becomes the start value for the forecast

4 methods, 4 start values, 8 smoothing constants and 72 items make the total number of forecasts equal to 9216. Ten different error measures are used which makes the total number of error data to 92160.

The initialisation values consisted of the corresponding mean value for the overall forecast period (April 2001 – September 2002). It would have been better if the mean value was calculated from the preceding 12 months period but no such data were available. A low value for the smoothing constant emphasis the history and therefore it takes longer for the forecast to adept to the actual demand. A drawback with no initialisation period is the danger of overfitting when the start value and the mean are based on the same time periods as the forecasts. An advantage is the possibility to study how dependent different forecasting methods are on a start value close to the mean or when the start value has a known deviation from the mean.

Since the data for the low demand items is too limited to divide into an 'initialisation' set for the start values, a 'test' set for the smoothing constants and a 'holdout' set, recommended by Makridakis et al (1998), no attempts have been made to optimise the methods. The data for the items with a higher demand would have been suitable to use the method suggested by Makridakis et al (1998). However this would degenerate the comparison between "items with optimisation" and "items without optimisation".

To do an optimisation with only five to ten demands is to let chance play a far too important role. It is not the value of different measures that is important but how the different methods performed compared to each other. An alternative would have been to optimise with the available data, but since the correlation between in-sample and out-of-sample data is low according to Makridakis (1986) and Chatfield (2001) emphasise that it is out-of-sample that is important and therefore no optimising or minimising attempts has been made. Also Fildes et al (1998) draw the conclusion that the method of optimising will affect the performance.

The methods were compared to each other for every item and smoothing constant for a certain error. The error methods themselves were also studied, not just the forecasting methods. Could different error methods favour different forecasting methods? According to Gardner (2006) the choices of forecast error influence which method that is considered the best method.

The method with the lowest error value received one point, the other methods received nothing. This is a variant of percentage best used by Syntetos and Boylan (2006). For every value of a smoothing constant, a method could get a total between 0 and 72 in the eight smoothing constants comparisons for a specific start value. In addition an overall best comparison was made. The lowest error for each method, regardless of the smoothing constant, was compared with the other methods lowest error. The forecast technique with the lowest error was awarded one point which sums up to a total between 0 and 72 for every method.

The number of first places can be seen as a binominal situation. The numbers of trials are mutually independent and each trial has a probability of 0.25 to the first place. The probability of the outcome was set to 0.01 in order to decrease the chance of interpreting something as significant when it actually was not significant, hence 0.01 instead of 0.05. Both a significant small number of first places and a significant large number of first places were of interest since non random behaviour of a poor performance (according to the error) could give additional information. In addition the second places were also studied in the same manner.

The use of first places as the reason to chose a forecasting method, do not consider how far apart the methods are for the individual items. A large or a very small difference will have the same outcome; the method with the lowest value will still get the point. Therefore, quotients were formed between the same smoothing constant and start value that two different methods had.

$$\text{Relative error quotient} = \frac{Error\ Method_A}{Error\ Method_B}$$

For a certain start value, 72 quotients between two methods were formed, one for every item. The basic statistical measures of the 72 items was analysed for every combination of different methods. See Table 3.7 for an example. A drawback with quotients is whether the number is the numerator or the denominator. If method A has an error value of 80 and is the numerator, method B has an error value of 100 and is the denominator the value of the quotient is 0.80. If the method B is the numerator and method A is the denominator the quotients has the value of 1.25. The quotients will not be symmetrical. This fact has been taken in consideration for the analysis.

Table 3.7 Example of a summary concerning the descriptive statistics of relative error quotient.

	0.025	*0.05*	*0.075*	*0.10*
Mean	1,043	1,066	1,079	1,089
Median	1,024	1,034	1,048	1,057
Std	0,054	0,079	0,092	0,101
Min	0,974	0,977	0,980	0,983
Max	1,276	1,344	1,367	1,447

To improve the analysis of the experiment and the understanding, different multivariate methods were applied with the aid of the statistical software. Shah (1997) used discriminant analysis of descriptive statistics for the different series to classify the appropriate forecasting method. Instead of the discriminant analysis the logistic regression was used since it does not require normal distributions for the predictor variables, (Johnson, 1998).

The classification was performed with the descriptive statistics of the demand, the inter-demand interval and the demand rate plus the percentage of actual demand occasion in relation to the total number of possible demand occasions. The classification was used to find out whether an item had an error that was increasing or decreasing as the smoothing constant was increased for a certain forecasting method. Since the descriptive statistics was correlated in various degrees, the logistic regression used the uncorrelated principal components (PCA) based on the descriptive statistics that had proven to be significant in the individually logistic regressions. The reason to use the logistic regression was to find if there was some kind of logic behind decreasing or increasing patterns that could imply a situation where certain types of error proved unreliable.

When several different dimensions of errors are measured it is not necessary, or even likely, that each measure carries unique information (Mathews and Diamantopoulus, 1994). Almost every error consists of the difference between the demand and the forecast that has been subjected to different mathematical operations. Consequently the different types of error are probably correlated in various degrees. PCA is described by Dallas (1998) as an exploratory technique that sometimes results in a better understanding of the correlation structure and a method to find the 'true' number of dimensions of the data. PCA uses either a variance-covariance matrix or a correlation matrix for the operations. To reduce the influence of the size of the variation between the error measurements the correlation matrix was used instead of a variance-covariance matrix that is scale-depended.

The previous presented methods have more focus on errors methods than the forecasting methods. Therefore some additional methods were used to analyse the forecasting methods. When the logistic was used the focus was on the change from low to high values of the smoothing constants, not with the size of the error. If a method varies greatly between different smoothing constants, then the method is sensitive to find the appropriate constant. -25-s, mean-s and +25-s was examined since the random start value of the naïve start made it harder to relate to the other start values.

The different items has various sizes of the demand, therefore a scaled measure is used. For an item's errors from the eight smoothing constants and a specific start value and method, four measures were calculated; the maximum value, the minimum value, the median and the mean. The difference between the maximum and minimum value (the range) can be seen as the difference between best case and worst case scenarios. The max-min quotient was formed and is the range divided with the median:

Max-Min quotient (Method A, Item i) =

$$= \frac{Max\ value_{0.025-0.30} - Min\ value_{0.025-0.30}}{Median_{0.025-0.30}}$$

The quotients of the mean as the denominator were also formed, but since the difference between the mean and the median proved to be in the region of one percent, only the median quotients are discussed. The mean of all the individual quotients was then used to form the value for comparison with other methods.

Apart from the smoothing constants the start value can also influence the size of the error. One thing that were examined were the tendency of increasing or decreasing number of first places when the start values were changed from -25-s to mean-s and from mean-s to +25-s. The probability, that the error increases or decreases as the start value is increased, is 0.25 (0.5^2) for an item. The probability is the same as it is with the binominal situation of forecasting methods. With the same p-value a quantity lower than 10 or higher than 26 is not random. Even if the required quantity is reached (less than 10, more than 26) it is not certain that one can expect the same behaviour for other forecasting situations outside this experiment. But it is an indication of a pattern that can occur in other situations.

A PCA was performed on every start value and smoothing constant for every method. The influential items were identified with the aid of scoreplots of the PCA. The influential items differed slightly between the methods therefore every item identified as an influential item for a method was removed from all methods to make the comparison easier. Especially the relation between CFE and CFE_{min} or CFE and CFE_{max} were studied to find indication of bias. If the loading plots indicated a bias tendency, quotients between CFE and CFE_{max} or CFE and CFE_{min} were studied. Values in the proximity of 1 indicate that it is more likely that a bias is present, especially if it is repeated over the majority of the items.

$$\text{CFE quotient (Method A, Item i)} = \frac{CFE_{\max}}{CFE} \text{ or } = \frac{CFE_{\min}}{CFE}$$

The same basic statistical measures that was used for the Max-Min quotient was also used for the CFE quotients. Since CFE is always in the denominator position of the quotient the symmetry problem of the relative error quotient do not occur in the CFE quotients. Regardless the type of bias only one quotient was used. See Table 3.8 for an example.

Table 3.8 Example of CFE quotients when the start value is -25-s.

	-25-s							
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	0,38	1,13	2,61	0,53	-0,08	2,15	0,11	0,96
Median	-0,46	1,22	1,30	1,32	1,34	1,30	1,27	1,28
Std	4,60	5,04	9,14	7,90	15,90	4,22	9,97	3,54
Min	-7,30	-22,02	-15,75	-59,30	-128,99	-4,94	-80,36	-20,11
Max	27,12	16,50	66,68	14,12	18,25	28,37	10,70	11,11

4 RESULTS AND ANALYSIS

In the fourth chapter the results of the performance concerning both forecasting methods and forecast errors are analysed. First the relationships between the traditional errors (MSE, MAD and CFE) are examined with correlation and PCA. This is followed by the results and analysis of the different errors where number of lowest errors, relative error quotients and logistic regression is used. The forecasting methods are analysed with Max-Min quotients, the relation between numbers of lowest error and start values, PCA and finally the quotients between CFE and its maximum or minimum values. The chapter ends with a forecasting example where the four methods are compared.

4.1 Relation between MSE, MAD and CFE

After reading a number of articles on the subject of forecast errors, the impression is that MSE has the most advocates. A study conducted by Armstrong and Carbone (1982) concludes that MSE is the most popular forecast error among both academicians and practitioners. The statistical properties of MSE and the relationship to variance makes this measure of accuracy one of the statistical minded's first choice. (Makridakis et al, 1998)

There are evaluations of forecasting methods based on MSE as the sole measure. To make this measure the sole measure is not without complications. To only use MSE implies its value as bias measure. If this is valid some relationship between MSE and a bias measure must exist. To examine whether such relation existed or not, the correlation between MSE, MAD and CFE was used. Also a PCA was applied. In both cases the outliers was left out.

4.1.1 Correlation between MSE, MAD and CFE

Croston has the most significant correlations for lower smoothing constants, see Table 4.1. The significant correlation for -25-s is found where the smoothing constant is 0.025, the lowest value, and the correlation value is 0.317. The value is the only significant positive value for Croston and no other smoothing constant is significant for -25-s. The reason for the positive correlation is that -25-s makes Croston more prone to underestimate the demand compared to mean-s, also the 0.025 value results in a start value that has a high influence of the future forecasts. This makes CFE generally a positive number and hence a positive correlation since MSE is always positive. Every other significant correlation is negative.

When mean-s is the start value, the correlation values have shifted to negative numbers because of the opposing signs for MSE and CFE (for Croston). The correlation is stronger for the lower smoothing constants and gets weaker as the smoothing constant gets higher. This behaviour is also present for -25-s and +25-s. However the significant correlations are generally increasing with a higher start value with the exception of 0.025.

The strongest correlation between CFE and MSE for Croston is -0.581. This occurs when the start value is +25-s and the smoothing constant is 0.025. The p-values are stronger for higher values of the smoothing constant with +25-s as a start value compared to the mean-s or -25-s. Also the smoothing constant affects the correlation. A lower value with the same start value has a higher value. The P-value for +25-s where the smoothing constant is 0.20 is 0.023. The P-value for mean-s where the smoothing constant is 0.10 is 0.026.

ModCr has a similar correlation behaviour compared to Croston, see Table 4.1. The weakest correlation can be found among the smoothing values for -25-s and the strongest among smoothing values for +25-s. For every smoothing constant the correlation is increased with a higher start value. A difference between ModCr and Croston is the positive or negative sign of the correlation, every correlation for ModCr is negative. For mean-s and +25-s the correlations get weaker with a higher value for the smoothing constant. This is not the case for -25-s; the overall weakest correlation for ModCr is for the 0.025 smoothing value.

Table 4.1 The correlation between CFE and MSE or MAD. The p-value is Pearson correlation value. The cells without additional colour have p-value < 0.01.

		-25-s		mean-s		+25-s	
		MAD	MSE	MAD	MSE	MAD	MSE
Croston	0.025	0,402	0,317	-0,313	-0,277	-0,690	-0,581
	0.05	0,092	0,043	-0,324	-0,305	-0,564	-0,498
	0.075	-0,013	-0,047	-0,305	-0,291	-0,487	-0,439
	0.10	-0,064	-0,085	-0,286	-0,272	-0,433	-0,391
	0.15	-0,116	-0,116	-0,259	-0,239	-0,361	-0,324
	0.20	-0,133	-0,127	-0,235	-0,214	-0,314	-0,280
	0.25	-0,137	-0,127	-0,213	-0,193	-0,276	-0,245
	0.30	-0,132	-0,121	-0,192	-0,172	-0,244	-0,215
ModCr	0.025	-0,260	-0,187	-0,488	-0,375	-0,640	-0,503
	0.05	-0,317	-0,247	-0,419	-0,329	-0,499	-0,395
	0.075	-0,317	-0,251	-0,385	-0,303	-0,439	-0,347
	0.10	-0,316	-0,249	-0,367	-0,288	-0,408	-0,321
	0.15	-0,315	-0,246	-0,349	-0,272	-0,376	-0,294
	0.20	-0,313	-0,243	-0,339	-0,262	-0,359	-0,278
	0.25	-0,309	-0,238	-0,330	-0,253	-0,346	-0,266
	0.30	-0,303	-0,232	-0,321	-0,244	-0,334	-0,255
SES	0.025	0,693	0,557	0,188	0,163	-0,509	-0,390
	0.05	0,662	0,493	0,277	0,179	-0,344	-0,309
	0.075	0,645	0,453	0,341	0,197	-0,179	-0,217
	0.10	0,624	0,418	0,369	0,199	-0,057	-0,146
	0.15	0,573	0,349	0,368	0,170	0,064	-0,081
	0.20	0,517	0,281	0,341	0,126	0,104	0,569
	0.25	0,464	0,218	0,309	0,082	0,114	-0,080
	0.30	0,417	0,164	0,278	0,043	0,112	-0,094
SyBo	0.025	0,898	0,703	0,843	0,653	0,725	0,561
	0.05	0,883	0,685	0,848	0,652	0,796	0,607
	0.075	0,879	0,681	0,852	0,656	0,816	0,624
	0.10	0,877	0,680	0,855	0,659	0,826	0,634
	0.15	0,873	0,679	0,856	0,662	0,834	0,643
	0.20	0,869	0,679	0,854	0,662	0,835	0,646
	0.25	0,864	0,673	0,851	0,660	0,834	0,646
	0.30	0,860	0,669	0,847	0,657	0,832	0,645
		p-value > 0.01		p-value > 0.05		p-value > 0.10	

SES has partly a different correlation pattern compared to the previous methods, see Table 4.1. The gradually increasing correlation from -25-s to +25-s is not present. For some of the smoothing constants the correlation decreases with an increasing start value. Not one of the smoothing constants for mean-s is significant. The strongest correlations are among the -25-s, the opposite of Croston and ModCr. A few correlations for the +25-s can be found among the lower smoothing values. Here the correlations get weaker with the increase of the smoothing constant. -25-s has the same tendency which is similar to the tendency for Croston and ModCr. Another similarity is that the sign goes from positive to negative. The CFE changes from a positive sign to a negative sign for majority of the items.

SyBo differs from the other methods because regardless of smoothing constant or start value there is a significant correlation. The correlation is strongest for every one of the eight smoothing constants when the lowest start value (-25-s) is used, see Table 4.1. The higher the start values are, the lower are the correlations. For the other methods the tendency is that the correlation is stronger with lower smoothing constants. This tendency is not plausible for SyBo. There is no negative correlation for SyBo.

4.1.2 Principal Component Analysis of MSE, MAD and CFE

If there is a linear dependency between MSE and CFE the principal components scores for MSE and CFE will differ only slightly. The PCA is without outliers, influential items that would increase the correlation between the errors and thereby change the principal component scores.

Croston

For Croston, when start value is -25-s and the smoothing constant 0.10, two principal components account for 96.5% of the variability, see Table 4.2. The accountability for mean-s and +25-s is in the same region. The first component is the variance errors since MSE and MAD has the highest values that are more than six times larger than the value for CFE when the absolute values are considered, see Table 4.2.

The first component accounts for 63.6% while the second component accounts for 32.9%. The second component is the bias component dominated by CFE. The pattern is valid both for mean-s and +25-s but the component score value for the bias component changes the sign. For -25-s it is positive but for mean-s and +25-s it is negative due to the increased overestimation of the demand when the start value is higher. MSE and MAD have similar component scores and in the loading plots they are close to each other, see Figure 4.1. However the variance components (for mean-s and +25-s) have a larger proportion of CFE and the bias components (for mean-s and +25-s) have a larger proportion of MSE and MAD.

Table 4.2 Eigenanalysis of the Correlation Matrix for Croston with the three start values for smoothing constant 0.10.

	-25-s			mean-s			+25-s		
	PC 1	*PC 2*	*PC 3*	*PC 1*	*PC 2*	*PC 3*	*PC 1*	*PC 2*	*PC 3*
Eigenvalue	1,908	0,988	0,104	2,046	0,851	0,103	2,185	0,715	0,101
Proportion	0,636	0,329	0,035	0,682	0,284	0,034	0,728	0,238	0,034
Cumulative	0,636	0,965	1,000	0,682	0,966	1,000	0,728	0,966	1,000

Variable	*PC 1*	*PC 2*	*PC 3*	*PC 1*	*PC 2*	*PC 3*	*PC 1*	*PC 2*	*PC 3*
MAD	-0,702	0,094	0,706	-0,663	-0,241	0,709	-0,639	-0,282	0,716
MSE	-0,703	0,070	-0,708	-0,660	-0,259	-0,705	-0,630	-0,342	-0,697
CFE	0,116	0,993	-0,017	0,353	-0,935	0,013	0,441	-0,896	0,041

For Croston the dimensions of the three errors are two rather than three. The third component (-25-s) has an eigenvalue that is closer to 0.0 than 1.0. A value less than 1.00 indicates the presence of noise. With an eigenvalue of 0.1038 the component consists mostly of noise that, if used, the third component may lead to overfitting (it will fit the actual data extremely well but not new data equally well). The eigenvalue for the second component is close to 1.0 which becomes lower as the start value increases and therefore increases the amount of noise in the second component. The first component increases the eigenvalue as the start value increases and therefore accounted variability increases from 63.6% to 72.8%. Since 63.8% is on the low side, two components are chosen with the increasing noise for higher start values in mind.

The numbers for the other smoothing constants and start values have similar values and relations between the errors. With mean-s or +25-s the Croston PCA resembles the ModCr's PCA see Figure 4.1. The most deviant behaviour can be found when the start value is -25-s and the smoothing constant is 0.025, see Table 4.3 and Figure 4.2. The first component is still the variance component but the signs for MSE and MAD are positive for the first component instead of the opposite sign that are the case for all the other of the PCA regarding Croston. Large CFE values coincide with large MSE and MAD values, which is more typical for a forecast that underestimates the demand. A scenario more likely to occur when the start value is low compared to the mean of the time series and when the smoothing constant is low so the adaptation to the mean of the time series takes longer.

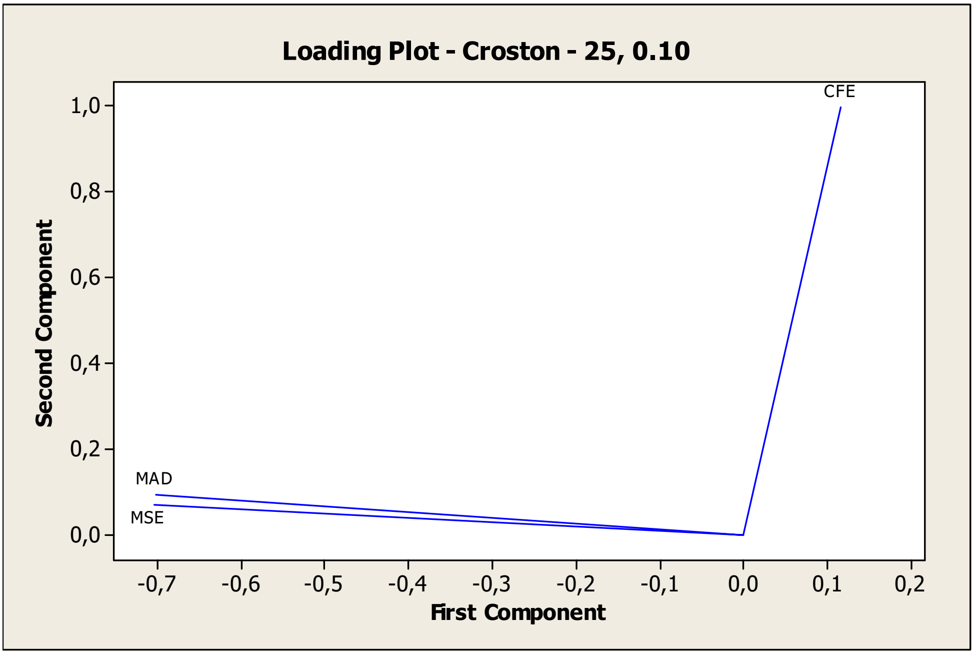


Figure 4.1 Loading plot for Croston with start value -25-s and smoothing constant 0.10.

Table 4.3 Eigenanalysis of the Correlation Matrix for Croston with start value -25-s and smoothing constant 0.025.

	PC 1	*PC 2*	*PC 3*
Eigenvalue	2,1214	0,7743	0,1044
Proportion	0,707	0,258	0,035
Cumulative	0,707	0,965	1,000
Variable	*PC 1*	*PC 2*	*PC 3*
MAD	0,653	-0,234	0,721
MSE	0,653	-0,350	-0,689
CFE	0,413	0,907	-0,08

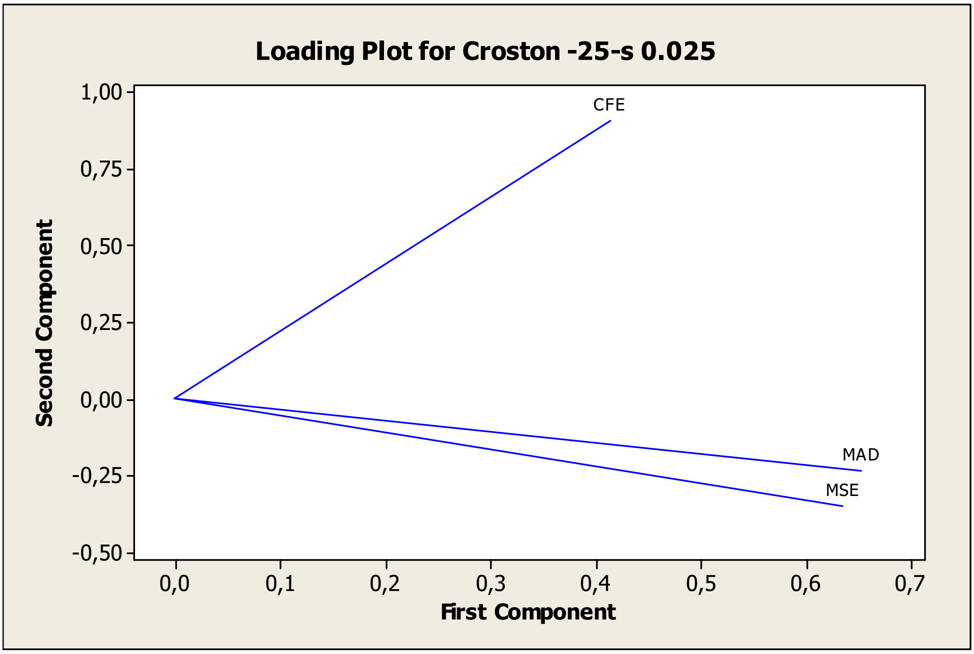


Figure 4.2 Loading plot for Croston with start value -25-s and smoothing constant 0. 025.

ModCr

The appearance for the PCA of ModCr resembles the PCA of Croston for mean-s and +25-s, see Figure 4.3. The components scores for MSE and MAD have similar values while CFE differs. In Table 4.4, with start value -25-s and smoothing constant 0.10, the first component accounts for 68.5% of the variability while the second component accounts for 28.4%, a total of 96.9%. The eigenvalue for the second component has a lower value compared to Croston but the second component can be also be chosen depending on whether the concern lies with the percentage accounted for, or to minimise introduction of additional noise. If the percentage accounted for is more important than added noise two component should be chosen.

As with Croston the first component increases its eigenvalue as the start value increases and the opposite is valid for the second component. The eigenvalue decreases as the start value increases. But the depreciation is not as large compared to Croston. The first component consists first and foremost of the variability errors, MSE and MAD, twice the absolute value of CFE. The second component is the bias component, where the score for CFE is 3 times larger compared to MSE or MAD. The interpretation of the components as a variability component for the first component and a bias component for the second component becomes degraded as the start value increases. CFE increases its value as MSE and MAD decreases in the first component and vice versa for the second component, when the start value increase. Also the relationship between MSE and MAD weakens. The number of dimensions of the errors is one or two as the third component has an eigenvalue of 0.090. Most of the variability of the errors is caused by MSE and MAD.

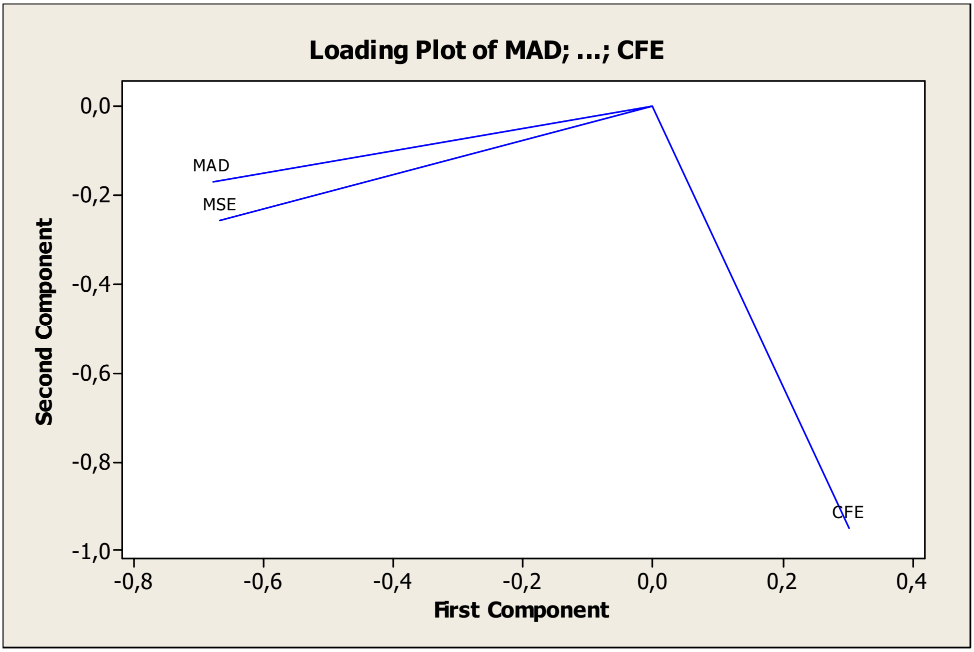


Figure 4.3 Loading plot for ModCr with start value -25-s and smoothing constant 0.10.

Table 4.4 Eigenanalysis of the Correlation Matrix for ModCr with the three start values for smoothing constant 0.10.

	-25-s			*mean-s*			*+25-s*		
	PC 1	*PC 2*	*PC 3*	*PC 1*	*PC 2*	*PC 3*	*PC 1*	*PC 2*	*PC 3*
Eigenvalue	2,056	0,851	0,093	2,100	0,809	0,092	2,139	0,771	0,090
Proportion	0,685	0,284	0,031	0,700	0,270	0,031	0,713	0,257	0,03
Cumulative	0,685	0,969	1,000	0,700	0,969	1,000	0,713	0,970	1,000
Variable	*PC 1*	*PC 2*	*PC 3*	*PC 1*	*PC 2*	*PC 3*	*PC 1*	*PC 2*	*PC 3*
MAD	-0,667	-0,209	0,715	-0,659	-0,224	0,718	-0,653	-0,234	0,721
MSE	-0,655	-0,292	-0,697	-0,644	-0,325	-0,693	-0,653	-0,350	-0,689
CFE	0,354	-0,933	0,057	0,389	-0,919	0,007	0,413	-0,907	0,08

SES

SES has a different PCA pattern than Croston and ModCr; see Table 4.5 for SES eigenvalue analysis and Figure 4.4 for the loading plot. The eigenvalue of the first component decreases with increasing start value, the opposite of Croston and ModCr. The second components value increases when the start value increases, also the opposite compared to Croston and ModCr. For -25-s it is questionable to use any more than the first component due to the fact that the second component has an eigenvalue of 0.6165 in which the noise content is high. The dimension of the errors are therefore one.

The first component is an overall error component. MSE and MAD are a bit more dominant than CFE. The component represents 77.1% of the variability. When mean-s is used the first component decreases to 68.3% and the second component's eigenvalue is close to 1.00. Therefore two components are necessary. The dimensions of the errors are two for mean-s and +25-s. As for Croston and ModCr the first component constitutes of the variability errors and the second component constitutes of bias measure. The components accounts for 97.1% of the variability. The tendency of a variability and bias component is clearer for +25-s. Regardless of which start value MSE and MAD have more reminiscent values for the component score than CFE has to either MSE or MAD. The PCA for +25-s has a resemblance closer to ModCr than for mean-s or -25-s, see Figure 4.3 and Table 4.4.

Table 4.5 Eigenanalysis of the Correlation Matrix for SES with the three start values for smoothing constant 0.10.

	-25-s			mean-s			+25-s		
	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3
Eigenvalue	2,312	0,616	0,0711	2,050	0,863	0,087	1,918	0,982	0,100
Proportion	0,771	0,206	0,024	0,683	0,288	0,029	0,64	0,327	0,033
Cumulative	0,771	0,976	1,000	0,683	0,971	1,000	0,64	0,967	1,000
Variable	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3
MAD	0,638	-0,174	0,750	0,675	-0,15	0,722	-0,695	-0,159	0,701
MSE	0,592	-0,512	-0,622	0,644	-0,357	-0,677	-0,702	-0,059	-0,709
CFE	0,492	0,841	-0,224	0,359	0,922	-0,144	0,155	-0,985	-0,071

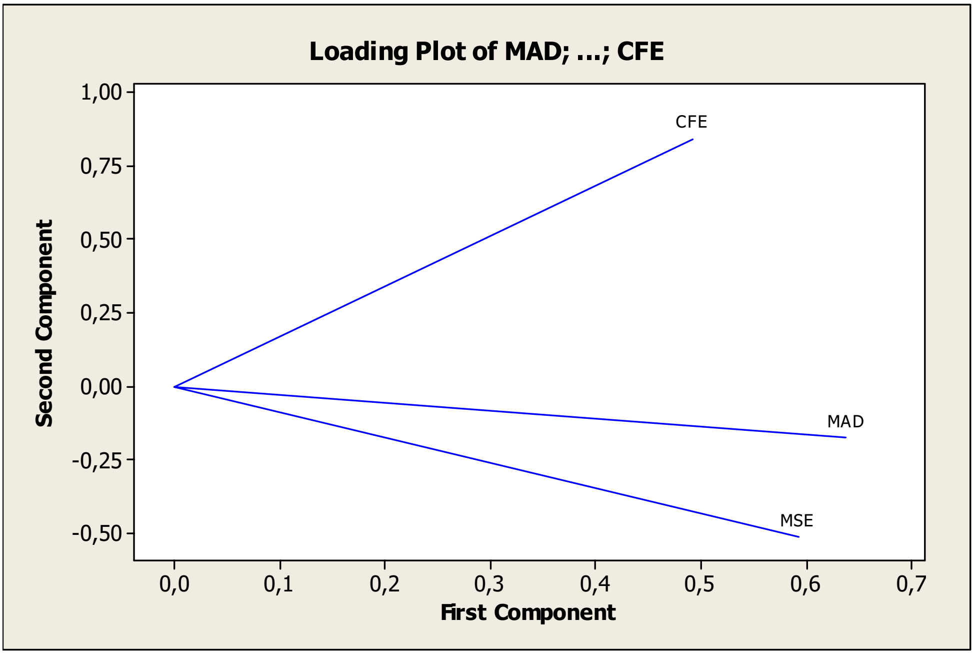


Figure 4.4 Loading plot for SES with start value -25-s and smoothing constant 0.10.

SyBo

SyBo differs from the other three forecasting methods partly because the coefficients and signs for the scores are more stable than for the other methods. Also the number of dimensions is the same regardless of the start values, see Table 4.6. The significant dimension consists of the three errors with approximately the same proportions for every start value and smoothing constant. The component represents 85.7-87.8% of the variability. The low eigenvalue (<0.3703) makes any interpretation uncertain. SyBo is the method, where CFE and MSE are closest, CFE and MSE share almost the same coefficients for the first component, see Figure 4.5.

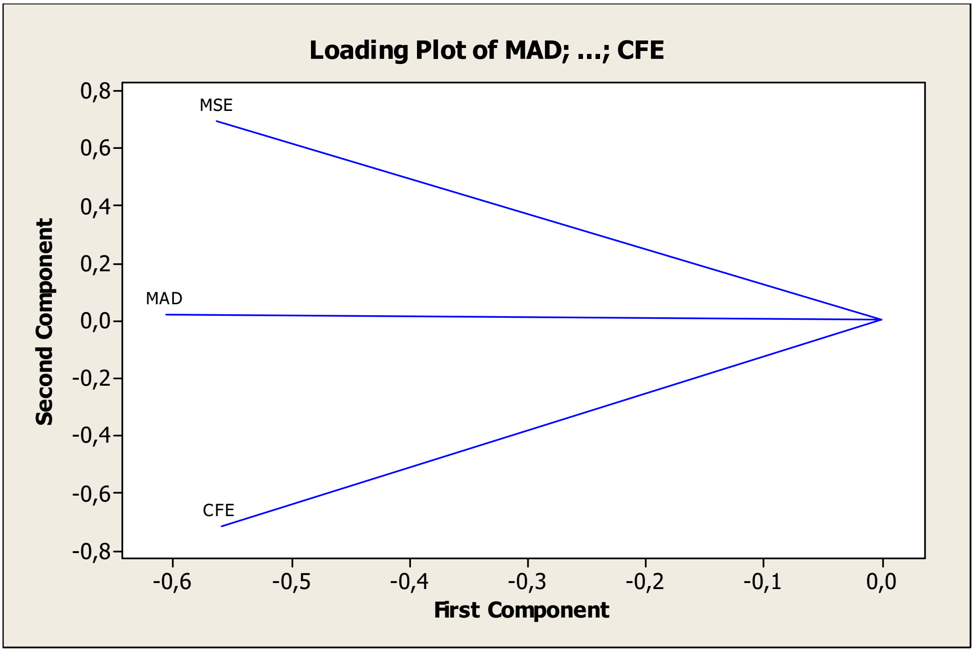


Figure 4.5 Loading plot for SyBo with start value -25-s and smoothing constant 0.10.

Table 4.6 Eigenanalysis of the Correlation Matrix for SyBo with the three start values for smoothing constant 0.10.

	-25-s			mean-s			+25-s		
	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3
Eigenvalue	2,312	0,616	0,071	2,050	0,863	0,087	1,919	0,982	0,010
Proportion	0,771	0,206	0,024	0,683	0,288	0,029	0,64	0,327	0,033
Cumulative	0,771	0,976	1,000	0,683	0,971	1,000	0,64	0,967	1,000
Variable	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3	PC 1	PC 2	PC 3
MAD	0,638	-0,174	0,750	0,675	-0,15	0,722	-0,695	-0,159	0,701
MSE	0,592	-0,512	-0,622	0,644	-0,357	-0,677	-0,702	-0,059	-0,709
CFE	0,492	0,841	-0,224	0,359	0,922	-0,144	0,155	-0,985	-0,071

4.1.3 Summary of PCA and Correlation

Both the analysis of the correlation and of the PCA shows that the linear dependency between MSE and CFE is weak in the majority of the cases. In addition, the dependency varies. For the same forecast method the linear relation is not stable. Different start values and smoothing constants affect the linear relation. Between the forecast methods there are also differences that makes MSE an uncertain error to use as the only measure or in a bias context.

The correlation and PCA are based on a linear relationship between two or more variables. The relationship between MSE and MAD should not be linear but rather some kind of non linear relationship. But the non linear relationship is linear enough to reveal that MSE and MAD have a stronger relationship than MSE and CFE.

4.2 MSE

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}\left(X_t - \hat{X}_t\right)^2 \tag{4.1}$$

Number of lowest error

The forecasting method that has the highest number of items where the MSE is lower than for the other forecasting methods is SES followed by SyBO, see Table 4.7. SES has the overall lowest error for the most items, at least 35 of the maximum 72. Overall it is the lowest error for all the smoothing constants for a certain method that is compared to the other methods lowest value. Although the performance of SES declines as the smoothing constant is increased. As the smoothing constant is increased SyBo becomes the method with most first places.

Table 4.7 Number of items with the lowest error for MSE for four different start values; -25-s, +25-s, mean-s and naïve start.

	-25-s				+25-s			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	9	7	11	45	2	2	21	47
0,025	8	12	8	44	2	1	20	49
0,05	6	6	27	33	3	2	35	32
0,075	6	2	45	19	2	2	48	20
0,10	5	2	55	10	3	1	58	10
0,15	2	3	64	3	2	3	65	2
0,20	1	3	67	1	0	1	70	1
0,25	0	1	71	0	0	0	72	0
0,30	0	0	72	0	0	0	72	0
	mean-s				*NAIVE*			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	8	4	16	44	7	6	24	35
0,025	7	7	15	43	6	9	21	36
0,05	4	3	31	34	6	10	27	29
0,075	5	1	48	18	5	2	44	21
0,10	4	1	58	9	4	2	55	11
0,15	2	3	64	3	4	2	60	6
0,20	0	3	68	1	2	2	66	2
0,25	0	0	72	0	2	2	67	1
0,30	0	0	72	0	1	0	70	1

If the second places also are accounted for, Croston and SyBo, have most second places with SES on third place. If one assumes that all the methods have equal chance to have the lowest error and that the items are independent of each other. The assumptions meet the criteria for the binomial distribution. With a probability of 0.25, 72 as number of possible successful trails and a p-value less or equal to 0.01, then values smaller than 10 or larger than 26 cannot be seen as random. Based on the binomial situation SES performance for low smoothing constants is most likely not just because of random coincidences. The same reasoning can be applied to SyBo for medium and higher smoothing constants. Also, the poor performance regarding first places of Croston and ModCr is not just bad luck.

Relative Error Quotients

To use just the number of first places alone as the criteria for choosing a forecasting method do not reveal how far apart the methods are from each other. A difference of 0.01 or 100 to one method's advantage will still have the same result; the method with the lower value will get one point. To find out the differences between the methods, the quotient between the forecast errors for method A and method B and for an item was formed. The MSE value of Croston for item 1 and smoothing constant 0.025 was divided with the MSE value of SyBo for the same item and smoothing constant.

Different statistic measures were used on the quotients for the 72 items. In Table 4.8 a summary of the mean and median values can be found. The tables for the other start values are in Appendix - Relative Error Quotients.

Table 4.8 The mean and median for the MSE quotients with mean-s as a start value. A value larger than one means that the numerator has the larger error compared to the denominator and vice versa.

	Mean							
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
ModCr/Croston	1,043	1,067	1,080	1,089	1,105	1,120	1,135	1,151
ModCr/SyBo	1,046	1,071	1,085	1,095	1,113	1,130	1,147	1,165
ModCr/SES	1,052	1,071	1,077	1,079	1,078	1,075	1,071	1,066
Croston/SyBo	1,003	1,004	1,005	1,006	1,007	1,009	1,011	1,013
Croston/SES	1,009	1,004	0,998	0,991	0,976	0,960	0,944	0,928
SyBo/SES	1,006	1,000	0,993	0,985	0,969	0,952	0,934	0,916
	Median							
	0,025	0,050	0,075	0,100	0,150	0,200	0,250	0,300
ModCr/Croston	1,024	1,034	1,048	1,057	1,067	1,077	1,092	1,098
ModCr/SyBo	1,027	1,036	1,053	1,063	1,075	1,083	1,098	1,105
ModCr/SES	1,027	1,036	1,039	1,042	1,038	1,034	1,027	1,024
Croston/SyBo	1,003	1,004	1,005	1,006	1,008	1,009	1,011	1,012
Croston/SES	1,009	1,003	0,999	0,995	0,979	0,963	0,946	0,925
SyBo/SES	1,006	0,999	0,994	0,989	0,973	0,954	0,935	0,914

The difference between Croston and SyBo mean quotient MSE performance is from 0.3% to 0.013%. The lowest difference is for the lowest smoothing constant and as the smoothing constant increases so does the quotient. The relativity small differences between the methods cannot be translated to their bias tendencies. Croston has been proven biased in previous studies (Syntetos and Boylan, 2001; Teunter and Sani, 2009). ModCr's lack of success concerning the first or second places is repeated with the quotients. Also SES deteriorating performance as the smoothing constant is increased is repeated, but now it is possible to see that the performance for SES with higher smoothing constant is worse than the performance of Croston.

Logistic Regression

To use the mean from the future period that is going to be forecasted, as a start value, is to know more than in a real forecasting situation. But with a known mean as a start value, the smoothing constant with the smallest error should likely be found among the smoothing constants with lower values. In a situation where the error decreases for every smoothing constant up to 0.30 is suspicious because of the known mean. Logistic regression has been used to help to find variables that influence among other things the error decreasing behaviour for some items when the smoothing constant increases.

Two versions of the logistic regression have been applied, one regression where one group constitutes of the errors that increase as the smoothing constant increase and the other group consist of the rest of the items. The other logistic regression constitutes of decreasing errors as the smoothing constant increases as one group and the other group consist of the rest of the items. The logistic regression has not been performed to find a model to predict other items beyond the scope of this experiment. The logistic regression has been used explanatory. To find variables from the statistic measures of the time series that makes the combination of an error and forecast method to react in a certain way, such as a decreasing error when the smoothing constant increases. Therefore there will not be any estimated coefficients from the regression of the significant variables but the terms "High" or "Low". High stands for a positive value of the estimated coefficient for a certain variable and low stands for a negative sign of the estimated coefficient for a certain variable.

The logistic regression for the forecasting methods shows some similarities between some of the methods. To consider a time series measure significant the p-values should be less than 0.05.All four methods have CV and MACs as significant variables, see Table 4.9. What differs is where MACs and CV are significant, demand inter-demand or demand rate. Croston and SES are affected by the variability quotients of the demand. An expected result, since it is the smoothing constant for the demand that is changed. Both Croston and SyBo forecast the demand separately. ModCr and SES are both affected by the quotients for the demand rate. Since the two methods have both one forecast that forecasts the demand rate, it is what is to be expected. Less expected is that SyBo is also affected by the demand rate which is probably caused by the covariation between the demand and the demand rate among a sufficiently large group of items.

Since CV and MACs have related mathematical operations it is more likely that it is not just a pure coincidence when both are significant.. The high value for CV and MACs imply that when the value for CV or MACs increases the error for MSE will also increase. It is the most likely event for MSE and its statistical properties, where the mean start value usually gives the lowest MSE error. For the same two items Croston and SyBo have errors that decrease with increasing smoothing constant. But since it is just two of the items that have the decreasing pattern, it will be difficult to use the regression where one group constitutes of less than three percent of the items. Therefore no logistic regression has been done. SES increasing deterioration of MSE is the only method affected by the percentage of demand occasion. The less demand occasions the more likely the errors will increase with higher smoothing constants. Not one of the descriptive measures of inter-demand periods is significant for any of the methods.

Table 4.9 Summery of variables for MSE. The variables that affect an increasing error with a higher demand smoothing constant. DO% stands for demand occasions in percent.

	Type	*Quantity*	*DO%*	*Demand*	*Inter-demand*	*Demand rate*
Croston	Increase	51		CV - High MACs - High		
ModCr	Increase	58				CV - High MACs - High
SES	Increase	62	Low			CV - High MACs - High
SyBo	Increase	50		CV - High MACs - High		CV - High MACs - High

4.3 MAD

$$\text{MAD} = \frac{1}{n}\sum_{t=1}^{n}\left|X_t - \hat{X}_t\right| \tag{4.2}$$

Number of Lowest Error

The same two methods that had the most first places for MSE are also the methods with the most first places regarding MAD, see Table 4.10. SyBo is the method that has the most first places. The number is affected by start value and smoothing constant. The lowest start value in combination with a low smoothing constant is the most likely combination where the highest number of first places can be found.

Table 4.10 Number of items with the lowest error for MAD for four different start values; -25-s, +25-s, mean-s and naïve start

	-25-s				+25-s			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	0	0	60	12	0	0	48	24
0,025	0	0	66	6	0	0	36	36
0,05	0	0	64	8	0	0	44	28
0,075	0	0	62	10	0	0	51	21
0,10	0	0	59	13	0	0	54	18
0,15	0	0	60	12	0	0	52	20
0,20	0	0	58	14	0	0	55	17
0,25	0	0	58	14	0	0	56	16
0,30	0	0	58	14	0	0	57	15
	mean-s				*NAIVE*			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	0	0	51	21	1	0	52	19
0,025	0	0	54	18	1	0	57	14
0,05	0	0	53	19	1	0	58	13
0,075	0	0	57	15	1	0	55	16
0,10	0	0	57	15	1	0	54	17
0,15	0	0	57	15	1	0	50	21
0,20	0	0	58	14	1	0	51	20
0,25	0	0	58	14	1	0	50	21
0,30	0	0	58	14	1	0	52	19

None of the combinations of start values and smoothing constants can most likely be seen as random for SyBo according to the binominal assumption. Contrary to SES where most of the numbers with the lowest error can not be verified statistically apart from the highest start value, +25-s, and smoothing constant 0.05 or 0.075. SES is the method with the highest number of second places that are in most cases significant. Croston is the method with the third highest number of second places. The tendency of the naïve start is similar to the other start values.

Relative Error Quotients

The size of the errors has a larger variation between the different methods for MAD compared to MSE. SyBo has lowest MAD, followed by SES, see Table 4.11. The difference between the quotients of SyBo and SES varies from 0.9% to 2.7%. Croston is closer to SES or SyBo than ModCr that has the highest MAD. The mean quotient of ModCr is at least 16% higher than SyBo and the mean quotient increases as the smoothing constant increases. The same tendency is also valid for the quotient between ModCr and Croston.

Table 4.11 The mean and median for the MAD quotients with mean-s as a start value. A value larger than one means that the numerator has the larger error compared to the denominator and vice versa.

	Mean-s							
	0,025	*0,05*	*0,075*	*0,1*	*0,15*	*0,2*	*0,25*	*0,3*
ModCr/Croston	1,117	1,153	1,168	1,176	1,184	1,189	1,193	1,197
ModCr/SyBo	1,160	1,197	1,213	1,221	1,230	1,236	1,240	1,244
ModCr/SES	1,150	1,187	1,201	1,208	1,213	1,214	1,214	1,214
Croston/SyBo	1,038	1,038	1,038	1,038	1,038	1,039	1,039	1,039
Croston/SES	1,028	1,028	1,027	1,025	1,021	1,018	1,014	1,011
SyBo/SES	0,991	0,990	0,989	0,988	0,984	0,980	0,976	0,973
	Median							
	0,025	*0,05*	*0,075*	*0,1*	*0,15*	*0,2*	*0,25*	*0,3*
ModCr/Croston	1,110	1,136	1,126	1,123	1,130	1,136	1,141	1,146
ModCr/SyBo	1,136	1,166	1,157	1,151	1,160	1,167	1,175	1,183
ModCr/SES	1,140	1,159	1,151	1,145	1,140	1,140	1,136	1,128
Croston/SyBo	1,040	1,040	1,040	1,040	1,041	1,041	1,041	1,041
Croston/SES	1,021	1,021	1,018	1,014	1,010	1,004	1,002	0,998
SyBo/SES	0,989	0,988	0,986	0,984	0,978	0,974	0,970	0,965

The MAD performance of ModCr is deteriorating when the smoothing constants are increased. The relative error quotients are only to a minor degree affected by the start value. The quotients are approximately the same for both -25-s and +25-s. The tables for 25-s and +25-s can be found in Appendix - Relative Error Quotients.

Logistic Regression

There are items that increase MAD when the smoothing constant increases as well as there are items that decrease MAD when the smoothing constant increases. Croston and SyBo have both items with increasing and decreasing MAD while ModCr and SES have items with increasing MAD, see Table 4.12.

The items with a decreasing trend are the same for Croston and SyBo. The two forecasting methods have the same significant variables in the logistic regression. An item will have a lower MAD if the smoothing constant is increased as long as the item can meet the following conditions; a low percentage of demand occasions, a low mean demand and low variation (which includes a standard deviation, CV and MACs) and finally a high mean of inter-demand periods that also has a high MACs value. The fact that only MACs is significant and not CV or the standard deviation can be either be the effect of random events or MACs' sensitivity to the sequence. An item with a low and stable demand as well as a few demand occasions with high and varied inter-demand periods will likely have a lower MAD error if the smoothing constant is higher.

When the increasing MAD is concerned the items are not the same for SyBo and Croston but the significant predictors are the same. The most surprising part is that CV is the only statistic measure that is not significant concerning the inter-demand periods. The number of decreasing items for SyBo and Croston are 9. ModCr has only one significant predictor namely MACs. This may be caused by random circumstances but more probable is that ModCr is sensitive to the order of the sequence, a high demand rate for a few periods followed by a low demand rate for a longer period of time will result in a higher forecast error than the opposite, a low demand rate for a few periods followed by a higher demand rate for a longer time period. The first case (the few periods of a high demand) will make the forecast higher than the second case and this high forecast will continue until the next demand occasion change the forecast. SES, like ModCr, only have items with an increase of MAD as the smoothing constant increases. A difference is that SES has more significant variables. CV and MACs of the demand as well as the CV of the demand rate are significant for SES.

Table 4.12 Summery of variables for MAD. The variables that affect an increasing error with a higher demand smoothing constant or the opposite.

	Type	Quantity	DO%	Demand	Inter-demand	Demand rate
ModCr	Increase	46				MACs - high
Croston	Decrease	9	low	All - low	Mean - high MACs - high	Mean - low Std - low
Croston	Increase	29	high	CV - high MACs - high	Mean - low Std - low MACs - low	
SyBo	Decrease	9	low	All - low	Mean - high MACs - high	Mean - low Std - low
SyBo	Increase	31	high	CV - high MACs - high	Mean - low Std - low MACs - low	
SES	Increase	38		CV - high MACs - high		CV - high MACs - high

There is a pattern when either SES or SyBo will have the lowest value. SES is usually better when the inter-demand interval is high, for 18 of the 45 items with the lowest percentage of demand occasions SES has the lowest errors. SyBo has a superior MAD performance if the inter-demand interval is low regardless of the start value. Quite the opposite of what one would expect since SyBo will underestimate the demand when there is a demand in nearly every period. Is MAD the most suitable, or even a suitable measure of intermittent demand? If the demand is low and the inter-demand interval is high then a method that forecasts closer to zero will have an advantage compared to a method that do not have equally low forecasts.

4.4 sMAPE

$$\text{sMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|X_t - \hat{X}_t\right|}{(X_t + \hat{X}_t)/2}\cdot 100 \qquad (4.3)$$

Number of Lowest Error

sMAPE is an accuracy measure that is different than MAD or MSE, which is also traceable in the method that has the lowest sMAPE in most cases. ModCr has the lowest sMAPE. The numbers for ModCr are the only significant number of first places when every item is concerned, see Table 4.13. The number of lowest error is also stable over the different smoothing constants and start values including naïve start. However among the 45 items with the lowest number of demand occasions, ModCr is dominant. ModCr has the lowest sMAPE regardless of smoothing constant and/or start values for at least 42 and up to every one of the 45 items.

Table 4.13 Number of items with the lowest error for sMAPE for four different start values; -25-s, +25-s, mean-s and naïve start

	-25-s				*+25-s*			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	47	2	12	11	45	2	15	10
0,025	47	2	13	10	44	2	15	11
0,05	48	2	16	6	47	2	18	5
0,075	49	2	17	4	47	3	19	3
0,10	49	3	17	3	47	4	18	3
0,15	53	2	16	1	53	1	16	2
0,20	54	4	13	1	54	4	13	1
0,25	54	5	12	1	54	5	12	1
0,30	56	2	14	0	56	2	13	1
	mean-s				*NAIVE*			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	46	3	13	10	44	3	14	11
0,025	46	3	14	9	41	5	14	12
0,05	47	2	16	7	43	5	18	6
0,075	47	3	18	4	46	4	19	3
0,10	49	3	17	3	48	4	16	4
0,15	53	2	16	1	53	1	16	2
0,20	54	4	13	1	52	5	13	2
0,25	54	5	12	1	53	5	12	2
0,30	56	2	13	1	55	3	12	2

From item 46-72, SyBo is the method with the most first places, ranging from 12 to 17 of the possible 27. SyBo is designed for intermittent demand but the method is better than the other methods when the percentage of demand occasions is high. When the demand is clearly intermittent SyBo has a higher sMAPE than ModCr. Croston is the method with the highest number of second places that are statistically significant for every start value and smoothing constant except one, -25-s and smoothing constant 0.025.

Relative Error Quotients

The variation between different error quotients of sMAPE is smaller than the variation of MAD's error quotients, see Table 4.14. The size of the error quotients are also closer to 1.00 compared to MAD. ModCr has a mean of the quotients that is 1.6-2.1% lower than Croston. The median of the quotients has almost the same difference. The difference of the mean and median is marginally larger concerning ModCr and SyBo. The largest differences between any of the methods are between SES and ModCr when the smoothing constant is 0.075 or higher. SES performance worsens as the smoothing constant increases compared to Croston or SyBo. The tendencies of mean-s is repeated for -25-s or +25-s, see Appendix - Relative Error.

Table 4.14 The mean and median for the sMAPE quotients with mean-s as a start value. A value larger than one means that the numerator has the larger error compared to the denominator and vice versa.

	Mean							
	0,025	*0,05*	*0,075*	*0,1*	*0,15*	*0,2*	*0,25*	*0,3*
ModCr/Croston	0,984	0,981	0,979	0,979	0,978	0,978	0,979	0,979
ModCr/SyBo	0,982	0,978	0,976	0,975	0,974	0,974	0,974	0,974
ModCr/SES	0,984	0,976	0,972	0,969	0,964	0,961	0,957	0,954
Croston/SyBo	0,996	0,996	0,996	0,996	0,995	0,995	0,995	0,995
Croston/SES	1,002	0,997	0,994	0,991	0,986	0,982	0,978	0,974
SyBo/SES	1,002	0,999	0,997	0,994	0,990	0,986	0,983	0,979
	Median							
	0,025	*0,05*	*0,075*	*0,1*	*0,15*	*0,2*	*0,25*	*0,3*
ModCr/Croston	0,987	0,984	0,981	0,980	0,979	0,979	0,979	0,980
ModCr/SyBo	0,978	0,977	0,976	0,976	0,975	0,975	0,975	0,975
ModCr/SES	0,984	0,977	0,972	0,968	0,961	0,957	0,954	0,949
Croston/SyBo	0,994	0,994	0,994	0,993	0,993	0,994	0,993	0,993
Croston/SES	1,001	0,998	0,994	0,993	0,988	0,984	0,981	0,976
SyBo/SES	1,002	1,000	0,998	0,996	0,992	0,986	0,982	0,979

Additional Comments

Since sMAPE behaved more like a compass needle with a 180 degrees error, no logistic regression was done. The intermittent demand situation is special, not only for the forecasting difficulties, but also in the way accuracy is measured. When the number of zero demands is increasing some measures do not have satisfactory performance. One of these measures is sMAPE which has the advantage of being able to handle zero demands without breaking down. Unfortunately, the number of zero demands appears to at best affect the resolution of sMAPE and at worst its values do not offer valuable information.

In review of the data, a low sMAPE for one method compared with the other methods for a certain item indicates a tendency for the forecasting method to overestimate the actual demand. An explanation for this behaviour is that a zero demand is at the end of the scale. The value for an individual observation has a maximum value of 200%. A study of equation 4.3 makes it possible to draw the conclusion that 200%, or close to 200%, is possible in two cases; when the forecast is large in comparison to the actual demand or when the demand is large compared to the forecast. A zero demand is therefore one of the two cases. This is a disadvantage with sMAPE, since it disregards a forecast’s distance to the zero demand in any given period, if one method forecasts 1 unit for a certain item and the other method forecasts 10 units, the outcome is 200% regardless of the method. The more periods with zero demand, the less valuable is sMAPE.

When a demand occurs it is most likely higher for the method that overestimates the demand compared to the other methods. Therefore the highest demand rate will result in the lowest error. ModCr had the lowest sMAPE for most of the items (see Table 4.13). In many cases it could be a difference of 1.0 between methods with different bias and variance performance. This is an indication of the problems involving error measures when there is zero demand and finding reliable measures. The size of the error is negatively correlated to the number of demand or the percentage of demand occasions, less demand occasions increases sMAPE. The naïve start with no outliers present had the weakest correlation factor, approximately -0.65 regardless of forecasting method. sMAPE is, as its name suggests, more suitable for symmetry than for measuring intermittent demand. At least in theory since Goodwin and Lawton (1999) and Koehler (2001) has criticised the lack of symmetry.

4.5 MAD$_n$

$$\text{MAD}_{\hat{n}} = \frac{1}{M}\sum_{i=1}^{M}\left| X_m - \sum_{t=p-o+1}^{p}\hat{X}_t \right| \tag{4.4}$$

Number of Lowest Error

SES and SyBo are the only two methods that have first places, see Table 4.15. Croston has at most 5 second places and the rest of the second places are divided between SES and SyBo. SES has most of its first places among the first half of the items (with fewer demand occasions), which is similar to sMAPE. Also similar is the second half of the items, where SyBo has the most first places.

Table 4.15 Number of items with the lowest error for MAD$_n$ for four different start values; -25-s, +25-s, mean-s and naïve start

	-25-s				+25-s			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	0	0	32	40	0	0	25	47
0,025	0	0	53	19	0	0	30	42
0,05	0	0	35	37	0	0	29	43
0,075	0	0	34	38	0	0	28	44
0,10	0	0	33	39	0	0	29	43
0,15	0	0	28	44	0	0	27	45
0,20	0	0	26	46	0	0	26	46
0,25	0	0	25	47	0	0	25	47
0,30	0	0	24	48	0	0	24	48
	mean-s				*NAIVE*			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	0	0	29	43	1	0	30	41
0,025	0	0	40	32	1	0	48	23
0,05	0	0	34	38	1	0	33	38
0,075	0	0	31	41	1	0	30	41
0,10	0	0	31	41	1	0	31	40
0,15	0	0	29	43	1	0	25	46
0,20	0	0	25	47	1	0	24	47
0,25	0	0	25	47	1	0	24	47
0,30	0	0	24	48	1	0	23	48

With an increasing smoothing constant SES increases its quantity of first places on the second half. The pattern is valid for all start values. SES performance deteriorates with a higher smoothing constant for MAD and MSE but for MADn it becomes better.

Relative Error Quotients

The quotients of MAD_n have the largest spread of all the error measures so far, see Table 4.16. The spread is smaller for the medians than for the mean, but it is still larger than for any other method. ModCr has the worst performance, at least a mean of the quotients that is 19.7% higher than for any other method. Croston has the second worst performance with quotients that is at least 6.7%, but far better than ModCr. SES and SyBo have the smallest difference of the quotients.

Table 4.16 The mean and median for the MAD_n quotients with mean-s as a start value. A value larger than one means that the numerator has the larger error compared to the denominator and vice versa.

				Mean				
	0,025	*0,05*	*0,075*	*0,1*	*0,15*	*0,2*	*0,25*	*0,3*
ModCr/Croston	1,197	1,269	1,300	1,316	1,332	1,341	1,347	1,352
ModCr/SyBo	1,265	1,342	1,375	1,392	1,410	1,419	1,425	1,431
ModCr/SES	1,283	1,388	1,444	1,484	1,550	1,609	1,663	1,716
Croston/SyBo	1,056	1,056	1,056	1,056	1,056	1,056	1,056	1,056
Croston/SES	1,067	1,083	1,096	1,109	1,137	1,166	1,192	1,219
SyBo/SES	1,010	1,025	1,037	1,050	1,076	1,103	1,128	1,153
				Median				
	0,025	*0,05*	*0,075*	*0,1*	*0,15*	*0,2*	*0,25*	*0,3*
ModCr/Croston	1,176	1,220	1,225	1,236	1,244	1,252	1,255	1,262
ModCr/SyBo	1,236	1,283	1,294	1,301	1,309	1,316	1,323	1,327
ModCr/SES	1,247	1,308	1,339	1,353	1,380	1,410	1,424	1,432
Croston/SyBo	1,057	1,058	1,058	1,058	1,058	1,057	1,057	1,057
Croston/SES	1,046	1,053	1,067	1,073	1,090	1,104	1,114	1,113
SyBo/SES	0,995	1,007	1,015	1,016	1,039	1,052	1,063	1,063

The higher the smoothing constant, the better SES performs compared to the other methods. This is contradictory to the behaviour of SES when the demand is intermittent. The forecast decreases since the forecast is updated every period regardless of demand or not. With every update the forecast decreases and eventually the forecast will approach zero, which will happen faster the larger the smoothing constant is. It is obvious that MAD_n is incapable of capture SES degrading performances with higher smoothing constants. The difference between the methods with other start values is in the same region as mean-s; see Appendix - Relative Error Quotients.

Logistic Regression

The ‘better’ performance of SES regarding the relative error quotients, in particular with higher smoothing constants, is traceable in the high quantity (35 items) with decreasing error as the smoothing constants increase, see Table 4.17. It is more likely that MAD_n decreases if the mean demand is low as well as all the other descriptive measures of the demand time series. When the mean demand is low (in the proximity of 1.0) as the smoothing constant increases and the value of the forecasts are low, it will not be a major difference between the demand and close to zero forecasts. With the summarised error over a certain amount of periods the difference decreases further. In fact SES will probably have a smaller error compared to a method that has a forecast with higher demand. It is the summarisation of time periods that causes this effect which is also affected by the inter-demand characteristics. All measures concerning the inter-demand periods are significant and furthermore, these measures should be high, in order to decrease the MAD_n of SES as the smoothing constant increases. In accordance with the high setting of inter-demand periods the percentage of demand occasions should be low.

The significant variables of SES when MAD_n increases are almost the same but the variables have opposite signs, low instead of high and vice versa. However, there are not as many significant variables, the means are not significant. This is regardless of demand, inter-demand periods or demand rate.

All the other forecasting methods have only items with increasing errors as the smoothing constant increases. The significant descriptive measures can be found among demand and demand rate where CV and MACs are in both cases significant concerning ModCr. Croston and SyBo have the same significant descriptive measures. CV and MACs are significant for demand and MACs for demand rate. No inter-demand descriptive measure is significant for Croston, ModCr or SyBo.

Table 4.17 Summery of variables for MADn. The variables that affect an increasing error with a higher demand smoothing constant or the opposite.

	Type	Quantity	DO%	Demand	Inter-demand	Demand rate
ModCr	Increase	48		CV - high MACs - high		CV - high MACs - high
Croston	Increase	31		CV - high MACs - high		MACs - high
SyBo	Increase	35	high	CV - high MACs - high		MACs - high
SES	Decrease	35	low	All - low	All - high	All - low
SES	Increase	16	high	Std - high CV - high MACs - high	Mean - low Std - low MACs - low	Std - high MACs - high

4.6 MSE$_n$

$$\mathrm{MSE}_{\hat{n}} = \frac{1}{M}\sum_{n=1}^{M}\left(X_m - \sum_{t=p-o+1}^{p}\hat{X}_t\right)^2 \tag{4.5}$$

Number of Lowest Error

MSE$_n$ is similar to MAD$_n$ concerning the number of lowest error. SES and SyBo are the only methods that have first places. The difference between the half with the lower number of the demand occasions and the half with the higher number of demand occasions is not as strong compared to MAD$_n$. SES dominates the lower half but has more first places for lower smoothing constants compared to SES number of first places for MAD$_n$. This is reflected in Table 4.18, where SES has more number one places for MSEn compared to MAD$_n$. As with MAD$_n$ the number of first places increases with the smoothing constant. The start value does not alter this trend. SyBo has the most number of second places followed by SES. Croston and ModCr have a few second places each.

Table 4.18 Number of items with the lowest error for MSE_n for four different start values; -25-s, +25-s, mean-s and naïve start

	-25-s				+25-s			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	0	0	14	58	0	0	10	62
0,025	0	0	34	38	0	0	24	48
0,05	0	0	20	52	0	0	18	54
0,075	0	0	13	59	0	0	12	60
0,10	0	0	14	58	0	0	11	61
0,15	0	0	12	60	0	0	10	62
0,20	0	0	12	60	0	0	11	61
0,25	0	0	13	59	0	0	12	60
0,30	0	0	13	59	0	0	11	61
	mean-s				*NAIVE*			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	0	0	12	60	1	0	16	55
0,025	0	0	26	46	1	2	42	27
0,05	0	0	16	56	1	0	24	47
0,075	0	0	12	60	1	0	18	53
0,10	0	0	12	60	1	0	17	54
0,15	0	0	11	61	1	0	11	60
0,20	0	0	11	61	1	0	10	61
0,25	0	0	13	59	1	0	11	60
0,30	0	0	12	60	1	0	10	61

Relative Error Quotients

The quotients of MSE_n have the largest spread of all the non bias accuracy measures, see Table 4.19. Apart from the larger size for the quotients of MSE_n, the pattern is similar to MADn. ModCr has the largest quotients followed by Croston. SES and SyBo are the two methods with the smallest difference, although the quotients increase as the smoothing constant increase. The quotient between ModCr and SES is at least 1.90 up to 4.69. When comparing the quotients of the median the quotients are not equally large which is valid for all the quotients, but the quotients still range from 1.37-2.29 between ModCr and SES. The differences between the methods are approximately the same regardless of start value, see Appendix - Relative Error Quotients.

Table 4.19 The mean and median for the MSE_n quotients with mean-s as a start value. A value larger than one means that the numerator has the larger error compared to the denominator and vice versa.

	Mean							
	0,025	*0,05*	*0,075*	*0,1*	*0,15*	*0,2*	*0,25*	*0,3*
ModCr/Croston	1,509	1,733	1,837	1,898	1,980	2,050	2,119	2,190
ModCr/SyBo	1,699	1,963	2,085	2,158	2,256	2,339	2,422	2,508
ModCr/SES	1,907	2,404	2,721	2,974	3,421	3,841	4,260	4,689
Croston/SyBo	1,104	1,105	1,106	1,107	1,109	1,111	1,113	1,115
Croston/SES	1,201	1,267	1,326	1,383	1,487	1,579	1,660	1,733
SyBo/SES	1,081	1,137	1,186	1,233	1,318	1,393	1,459	1,518
	Median							
	0,025	*0,05*	*0,075*	*0,1*	*0,15*	*0,2*	*0,25*	*0,3*
ModCr/Croston	1,257	1,355	1,451	1,483	1,495	1,544	1,593	1,629
ModCr/SyBo	1,375	1,472	1,620	1,652	1,691	1,717	1,752	1,786
ModCr/SES	1,465	1,681	1,797	1,894	2,037	2,099	2,171	2,295
Croston/SyBo	1,104	1,105	1,104	1,102	1,099	1,099	1,104	1,108
Croston/SES	1,114	1,161	1,193	1,212	1,226	1,252	1,286	1,304
SyBo/SES	1,018	1,052	1,083	1,098	1,121	1,147	1,177	1,189

Logistic Regression

As with the relative error quotients and placements, the logistic regression of MSE_n is similar to MAD_n. SES still has almost the same amount of items that decrease the error, 33 instead of 35, see Table 4.20. The items with increases errors for SES when MSE_n is concerned, has the same number of items, 16 items, as SES had for MAD_n. It is not exactly the same items that increase. 14 items are identical. The high and low values are still the same for every significant variable. No further discussion around the high and low places will therefore take place; see MAD_n logistic regression for more details.

Differences between MSE_n and MADn are the significant variables for some forecasting methods; ModCr has also the CV of the demand rate, for SES decrease the mean of demand and demand rate is no longer significant, for SES increase the percentage of demand occasions is no longer significant but the CV of demand rate is. The significant variables for SyBo and Croston are identical to MAD_n. Both Croston and SyBo have seven items that error decreases with increasing smoothing constant, but none of the used descriptive measures are significant.

Table 4.20 Summery of variables for MSE_n. The variables that affect an increasing error with a higher demand smoothing constant or the opposite.

MSEn	*Type*	*Quantity*	*DO%*	*Demand*	*Inter-demand*	*Demand rate*
ModCr	Increase	45		CV - high MACs - high		CV - high MACs - high
Croston	Increase	32		CV - high MACs - high		MACs - high
SyBo	Increase	34		CV - high MACs - high		CV - high MACs - high
SES	Decrease	33	low	Std - low CV - low MACs - low	All - high	Std - low CV - low MACs - low
SES	Increase	16	low	Std - high CV - high MACs - high		Std - high CV - high MACs - high

4.6.1 Additional Comments Regarding MADn and MSEn

It is not certain that the mean of the whole 18 months is always a suitable start value, because over the length of time what is the overall mean may not be the initial mean for the first part of the series since the mean may drift over time. Actually, the drift of the mean is one of the reasons to use forecasting. But when SES has the most items the lowest values of MADn and MSEn especially with the highest smoothing constants, then something is fundamentally wrong with the two errors. SES is the method that decreases the influence of the start value faster than any other of the three methods especially with higher smoothing constants and larger inter-demand periods.

The two errors has a bias that favours methods that underestimate the demand when it occurs and reacts when its to late, hence the good performance of SES and bad performance of ModCr. SES underestimates the demand when the smoothing constants are higher while ModCr does the opposite. The difference becomes even more extreme with the squared function. Is it sensible that ModCr has an error that is approximately between 2 and 4.5 times larger than SES?

MADn and MSEn collapses when the previous conditions for demand and inter-demand periods are met which makes them even worse than MAD as variance measure of evaluation of forecasting methods when the demand is intermittent. To use a percentage variant of MADn or MSEn will avoid the infinity problem when there is no demand but with the cost of variance measures that distorts the performance for under- or overestimating forecasts. The initial idea to use MADn and MSEn was to avoid the tendency of MAD, to have forecasts with zero demand when optimising. None of the two measures are able to do that. Instead they act like an extreme version of MAD. Also the measures have no documented relation to variance measures such as standard deviation as MSE and MAD have. The best future for MADn and MSEn is oblivion.

4.7 CFE

$$CFE_t = \sum_{i=1}^{t} (X_t - \hat{X}_t) = X_t - \hat{X}_t + CFE_{t-1}, \quad t = 1, 2, \ldots, T \tag{4.6}$$

Number of Lowest Absolute Error

When the bias measures CFE and PIS are evaluated with number of lowest error SES is the method that has most first places regardless of start value, see Table 4.21. The higher the start value the more first places. The number one items of SES have no bias regarding low or high percentage of demand occasions. This is also valid for Croston who is the second best for -25-s. The combination of a lower start value and a overestimating forecast results in a lower CFE for Croston. SyBo has most of the first places of the lower half of the items, the half with lower percentage of demand occasions. However there is one exception for +25-s, when the smoothing constant is 0.025 the number one items are spread among both low and high percentage of demand occasion. It is questionable if CFE can be regarded as a reliable measure of bias when finite time series are concerned, since SES has a tendency of increasing the number of first places as the smoothing constant increases because CFE approaches zero.

Table 4.21 Number of items with the lowest error for CFE for four different start values; -25-s, +25-s, mean-s and naïve start

	-25-s				+25-s			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	0	22	10	40	0	4	13	55
0,025	3	37	2	30	0	1	20	51
0,05	0	25	4	43	0	2	12	58
0,075	0	19	5	48	0	3	9	60
0,10	0	14	3	55	0	3	8	61
0,15	0	12	4	56	0	3	8	61
0,20	0	10	5	57	0	4	7	61
0,25	0	11	2	59	0	6	6	60
0,30	0	12	2	58	0	7	7	58
	Mean-s				*NAIVE*			
	ModCr	*Croston*	*SyBo*	*SES*	*ModCr*	*Croston*	*SyBo*	*SES*
Overall	0	7	6	59	3	20	15	34
0,025	0	9	19	44	16	14	18	24
0,05	0	7	12	53	9	18	20	25
0,075	0	4	8	60	6	16	19	31
0,10	0	4	6	62	3	17	15	37
0,15	0	7	8	57	3	11	10	48
0,20	0	6	5	61	2	11	8	51
0,25	0	8	6	58	1	10	9	52
0,30	0	9	3	60	1	7	9	55

4.8 PIS

$$PIS_t = -\sum_{i=1}^{t} CFE_i \tag{4.7}$$

Number of Lowest Error

The outcome of the number of first places for PIS is similar to CFE. SES is the method with most first places, see Table 4.22. Compared to the other methods SES is more dominating than the method was for CFE. SyBo has approximately the same number for mean-s but not for the other start values. Croston has more first places for -25-s and for that start value ModCr has some first places. The first places of ModCr are the items with the highest percentage of demand occasions. PIS shares the same problem that CFE has; the tendency of a better performance of SES when the smoothing constants are high. PIS has an even stronger tendency of this. Therefore it is doubtful if PIS alone can be used as a measure of bias when the time series consists of a finite number of periods, as in a practical case.

Table 4.22 Number of items with the lowest error for PIS for four different start values; -25-s, +25-s, mean-s and naïve start

	-25-s				+25-s			
	ModCr	Croston	SyBo	SES	ModCr	Croston	SyBo	SES
Overall	1	32	6	33	0	1	15	56
0,025	14	46	1	11	0	0	29	43
0,05	2	54	0	16	0	0	13	59
0,075	0	31	1	40	0	0	10	62
0,10	0	22	1	49	0	0	9	63
0,15	0	15	1	56	0	0	5	67
0,20	0	14	4	54	0	0	3	69
0,25	0	10	2	60	0	0	4	68
0,30	0	8	1	63	0	1	0	71
	mean-s				NAIVE			
	ModCr	Croston	SyBo	SES	ModCr	Croston	SyBo	SES
Overall	0	2	1	69	2	21	6	43
0,025	0	6	3	63	18	10	4	40
0,05	0	1	0	71	5	27	5	35
0,075	0	1	2	69	1	25	6	40
0,10	0	1	0	71	0	19	5	48
0,15	0	1	0	71	0	13	3	56
0,20	0	2	0	70	0	10	1	61
0,25	0	2	0	70	0	8	1	63
0,30	0	1	0	71	0	6	1	65

Relative Error Quotients

The relative error quotients are not as informative when the error can fluctuate around zero instead of a larger positive number which is the case for both SES and SyBo. The result of this fluctuation is quotients with a higher degree of chance that makes the quotients sometimes very large or small compared to the nearby quotients, see Table 4.23. The quotients between ModCr and SyBo for PIS and mean-s starts with -27.8 then -1.5 followed with 406.1. The interpretation and conclusion of the material becomes hazardous and therefore the relative error quotients are not used for CFE and PIS.

Table 4.23 The mean for the PIS quotients with mean-s as a start value.

	Mean							
	0,025	*0,05*	*0,075*	*0,1*	*0,15*	*0,2*	*0,25*	*0,3*
ModCr/Croston	4,0	4,3	-6,4	7,3	7,0	7,5	7,2	7,4
ModCr/SyBo	-27,9	-1,5	406,1	-2,3	2,7	-13,4	-20,5	1,9
ModCr/SES	84,8	10,9	-326,0	-320,7	-661,6	-10037	-13397	-7240
Croston/SyBo	-5,9	0,1	72,8	0,0	0,8	-1,8	-2,4	0,4
Croston/SES	17,5	7,2	-58,8	-45,1	-89,3	-1713	-1624	-1680
SyBo/SES	-29,9	31,9	21,5	31,2	117,2	-171,2	-796,4	-1025

4.9 Combined Analysis with CFE, PIS and NOSp

$$\text{If } X_t \neq 0 \text{ and } CFE_t > 0 \text{ then } NOS \leftarrow NOS + 1, \quad t = 1, 2, \ldots, T \qquad (4.8)$$

$$NOSp = \frac{NOS}{M} \cdot 100 \qquad (4.9)$$

The intention of the two maximum and minimum measures of CFE, CFE_{max} and CFE_{min}, was to indicate whether a method has a bias problem. The relation of the two measures compared to CFE should indicate if a bias existed. If CFE and one of the two errors had errors in the proximity of each other, a method could have a bias problem. If the tendency for CFE and one of the extreme point measures was repeated over a sufficient large number of items then a bias problem would exist. This assumption works satisfactory for CFE_{min} but not for CFE_{max}, since a demand cannot be less than zero. What was missing from the assumption was that the demand has a different effect on the max and min measures.

Transients (sudden peaks compared to the surrounding observations) are upwards movement not downwards movement of the CFE when the demand usually is low. A relative high demand causes a transient in the demand period 31, see Figure 4.6. This peak can be traced in the change of CFE for both ModCr and SyBo. The demand in period 31 is mainly the cause of CFE_{max} for period 1-39. When the demand is zero the CFE is decreasing slowly compared to the increase after a demand. This applies to both methods. Two high demands that are far apart, one at the end of the series and one earlier, can give the impression of a underestimating bias.

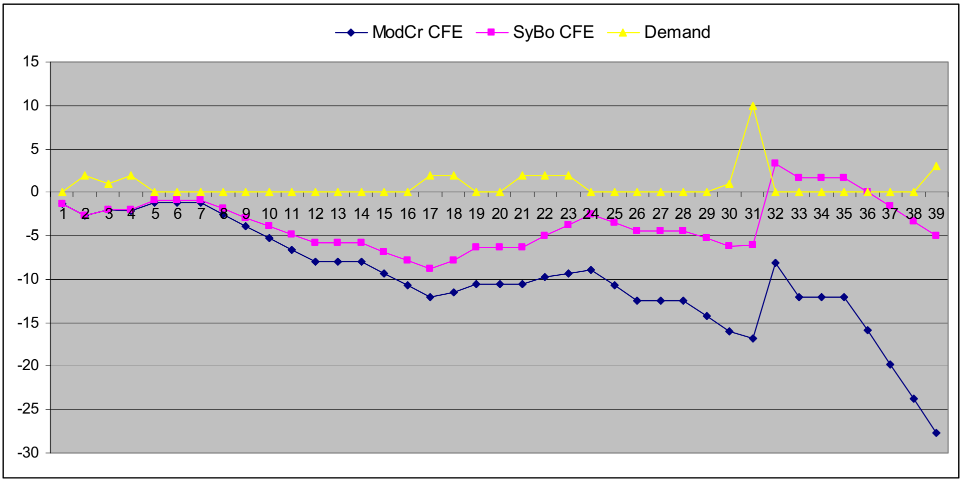


Figure 4.6 Demand and CFE for SyBo and ModCr from item 3, mean-s and smoothing constant 0.30. The three in a row values that are equal is Friday to Sunday.

The assumption of maximum and minimum can be applied if the measure is not as sensitive to transient demand as CFE. PIS is an integration of CFE considering the time aspect and is therefore slower to react to random transients, except for the start. In Figure 4.7 is the PIS data for SyBo and ModCr. The demand transient still occurs in period 31 but the transient is only to a minor extent passed on to PIS for either of the methods despite the same smoothing constant as in the previous example with CFE. A slight increase of PIS for SyBo can be detected but not as extreme to classify it as a transient. The counterpart to CFE_{max} is PIS_{min} (due to different signs) which takes place in the start of the forecast. If PIS_{min} should be in the proximity of PIS both method must begin to underestimate the demand. The PIS curves show tendencies of overestimation for both methods. However, ModCr is more biased since the slope of the PIS is steeper as well as the PIS values are higher.

The choice of smoothing constant affect the transient response, with a lower smoothing constant the transient would not have the same peak-value but the CFE would still react more than PIS.

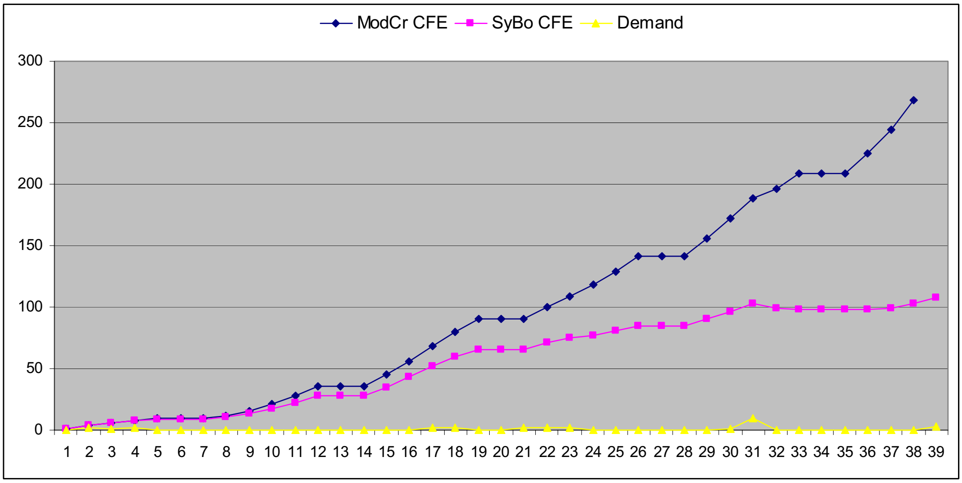


Figure 4.7 Demand and PIS for SyBo and ModCr from item 3, mean-s and smoothing constant 0.30.

CFE is not trustworthy under certain circumstances such as SES with a high smoothing constant resulting in a low CFE or peaks in the demand that causes a large change of CFE. PIS is smoother concerning sudden demand changes but like CFE it is undependable regarding SES and high smoothing constants. One method to monitoring the forecast is to use a tracking signal. In order for the tracking signal to work, a distribution assumption must be fulfilled; the assumption of a normal distribution of the tracking signal is common. If the distribution assumption is not fulfilled the tracking signal becomes undependable. An alternative to a tracking signal is to use a combination of CFE, PIS and NOSp.

An example of the possibilities of interpreting the forecasting is in Table 4.24. CFE is complemented by PIS and NOSp. ModCr is not present since all three measures (CFE, PIS and NOSp) indicate an overestimating bias. The forecast of Croston for Item 24 is overestimating the demand since the CFE is negative, PIS is positive and NOSp is 2%. That is only 2% of the demand occasions that could not be met by the accumulated forecast. Somewhere in the proximity of 50% for NOSp is excepted when the forecast is none biased. Apart from the overestimating forecast, MAD or MSE will be used to calculate a safety stock, which will further increase the bias situation in form of a larger stock.

Regarding item 31, SyBo and SES have similar values of both CFE and NOSp but SES is the ‘leaner’ since PIS is lower. For item 36 SyBo has a low CFE near zero but a very low PIS, -5454, that is a sign of underestimating the demand which is also confirmed by the value of NOSp, 97.6%. The size of the PIS values in absolute terms of Croston and SyBo is not that far apart as their respective CFE are. For the final item in this example, Croston has a positive CFE for item 38, which makes it likely that it is a minor underestimation of the demand; neither PIS nor NOSp can confirm the underestimation. Instead the two measures indicate that Croston is overestimating the demand.

Table 4.24 Example of the difference between PIS, NOSp and CFE based on the last values for each error.

	Croston			*SyBo*			*SES*		
	CFE	*PIS*	*NOSp*	*CFE*	*PIS*	*NOSp*	*CFE*	*PIS*	*NOSp*
Item 24	-15,8	4189	2,0	3,8	536	58,0	0,4	14	61,0
Item 31	-49,9	12637	15,1	5,8	1005	36,0	5,4	66	38,1
Item 36	-53,9	6008	27,1	-0,6	-5454	97,6	-13	-485	75,9
Item 38	9,0	8881	10,7	75,5	-2141	62	40,0	1263	51,4

When only NOSp is evaluated with the mean, no method appears to be non-biased, see Table 4.25. The methods are divided in two groups; overestimating and underestimating. ModCr is the most overestimating forecast method followed by Croston. The two methods are far from 50%, a value around 50% would be the most likely value in the long run if a method is non-biased. SyBo is the most underestimating method. SES does not have the same underestimating tendencies as SyBo.

Table 4.25 Values of NOSp with mean-s. A low value indicate overestimation and a high value indicates underestimation.

Smoothing constant	*ModCr*	*Croston*	*SyBo*	*SES*
0,025	12,0	22,5	76,1	59,5
0,05	7,9	19,8	76,9	60,0
0,075	6,2	18,2	77,4	61,2
0,1	5,3	17,1	77,8	62,9
0,15	4,4	15,8	78,0	65,4
0,2	3,8	15,2	78,2	68,2
0,25	3,4	15,5	78,3	70,5
0,3	3,1	15,9	78,2	72,2

With low smoothing constants SES is the method closest to 50%. All of the methods increase their respective bias as the smoothing constants are increased. For ModCr and Croston NOSp becomes smaller and for SyBo and SES NOSp becomes larger.

4.10 Croston

When the forecasting errors were studied, a part of the analysis was the eventual increase of the error as the smoothing constant increased. The focus in the analysis is with the change, rather than the size of the error. One thing that is interesting to know is the distance between the smallest and largest error as an indication of how sensitive a method is in finding a suitable smoothing constant.

For a forecasting method with a certain start value and a specific item there are eight different smoothing constants. From those eight values three were used to form the Max-Min quotient, namely the maximum value, the minimum value and the median value. The minimum error for a certain item was subtracted from the maximum error. The difference and the median of the errors from the smoothing constants (0.025-0.30) form the quotients. The mean based on the Max-Min quotients for all the items, is the measure of the relative spread.

Croston has a stable performance of the variance measures (MAD and MSE), see Table 4.26. The method is not very sensitive to what value the smoothing constant has since the mean Max-Min quotient is around 0.05 regardless of start value. If not the right smoothing constant has been found or is possible to find, Croston will still perform in the proximity of the lowest error for MAD and MSE. Obviously there will be items where this is both better and worse.

The difference for sMAPE is even smaller; to a high degree neither the smoothing constant nor the start value will affect the result. MADn and MSEn have larger Max-Min quotients but compared to the other methods Croston is one of the methods with lowest relative spread. Concerning the bias measures, it is more difficult to draw a conclusion of the quotient when zero or numbers close to zero in comparison with the difference are included. But the number of shortages (NOSp) is dependent on the smoothing constant. With a higher start value Croston get more biased towards overestimating the demand. The value of PIS +25-s indicates that the smoothing constant can, to a high degree, influence the value of PIS.

Table 4.26 Croston Max-Min quotients

	-25-s	*mean-s*	*+25-s*
MSE	0,054	0,053	0,052
MAD	0,051	0,034	0,040
sMAPE	0,019	0,019	0,022
MADn	0,074	0,046	0,055
MSEn	0,178	0,131	0,158
NOSp	2,139	0,880	1,242
CFE	-1,595	-0,437	-1,326
PIS	0,248	0,554	1,175
CFEmin	-0,659	-0,473	-0,947
CFEmax	1,234	0,413	0,843

To find whether the start value influence the size of the error three scenarios were investigated; if the error decreased when the start value was increased, if the error increased as the start value was increased or if the mean-s had a lower error than -25-s and +25-s. Each with a probability of 0.25 for every item, 72 as number of possible successful trails and a p-value less or equal to 0.01, then values smaller than 10 or larger than 26 cannot be seen as random.

Even if the required quantity is reached it is not certain that one can expect the same behaviour for other forecasting situation outside this experiment. But it is an indication of a pattern that can occur in other situations. In Table 4.27 are the values for MAD and MSE. According to the table the mean-s value is not significant but -25-s will most likely reduce the error. The same tendency is valid for MADn and MSEn.

However, for certain items of MSE the error is decreasing. The number of items with the error reduction for higher start values rises as the smoothing constant is increasing, but it is only significant for 0.30. The relation between the MSE and/or MAD performance of different start values corresponds to the overestimating bias of Croston. As the start value becomes smaller the bias errors decreases and so does MAD and MSE (to an extent).

Table 4.27 Number of items with lowest error for a certain scenario concerning MAD and MSE. 'Increasing' have the lowest error when the start value is -25-s

	MAD			MSE		
	Increasing	*Decreasing*	*mean-s*	*Increasing*	*Decreasing*	*mean-s*
0.025	71	0	0	46	1	24
0.05	69	0	2	41	6	24
0.075	67	1	3	39	9	23
0.10	68	1	2	36	13	22
0.15	67	2	2	34	22	15
0.20	64	4	3	31	22	18
0.25	63	6	2	32	25	14
0.30	60	7	4	34	27	10

The PCA of Croston is changed when 10 errors are used instead of the previous PCA where three errors were included. To represent the variability, three components are needed. The three components accounts for 85.4-87% of the variability.

The first component accounts for 54-64% depending on the start value. The component is primary a variance component with MSE, MAD, MSEn and MADn, but also CFE_{max} with opposite sign for -25-s and mean-s. With the start value +25-s CFE_{max} is replaced with CFE_{min} (also opposite sign). With a higher start value it is more likely that CFE_{max} is lower and therefore do not co-vary with the variance errors and the opposite reasoning is plausible for CFE_{min}.

The second component (12.5-21.4%) is still a bias component but now with PIS and NOSp besides CFE. The third component is a mixture between bias and sMAPE that accounts for 9-10.5% of the variability. The three loading plots of the different start values have similar appearance for the variance measures; see Figure 4.8 and Figure 4.9. The figures describe the first two components, for more information of the PCA, see Appendix – PCA.

The loadings of CFE and CFE_{min} become more and more similar as the overestimation increases when the start value increases. The proximity between CFE and CFE_{min} is an indication of bias. The value of CFE is the value for the whole period while CFE_{min} is the maximum value during the whole period. If CFE and CFE_{min} have similar values but CFE and CFE_{max} have not, the probability of a bias situation with overestimation increases.

The relationship between CFE and PIS changes with different start values. If PIS carried the same information as CFE the loadings of the two errors should have 180 degrees between them hence the opposite signs for the errors and be of equal length. This is true for mean-s. If it had been true for the other start values PIS would not be an error that was required in the PCA and could therefore be omitted. But in the loading plot of for PCA of Croston -25-s and +25-s CFE and PIS forms an angle that is not in the vicinity of 180 degrees, see Figure 4.8 and Figure 4.9. PIS and NOSp are closer to the mirror image of 180 degrees when the start value is -25-s. If there are a large percentage of shortages, the values of PIS would most likely be low. The correlation between PIS and NOSp is lost for mean-s and +25-s where PIS variability comes closer to the variance measures.

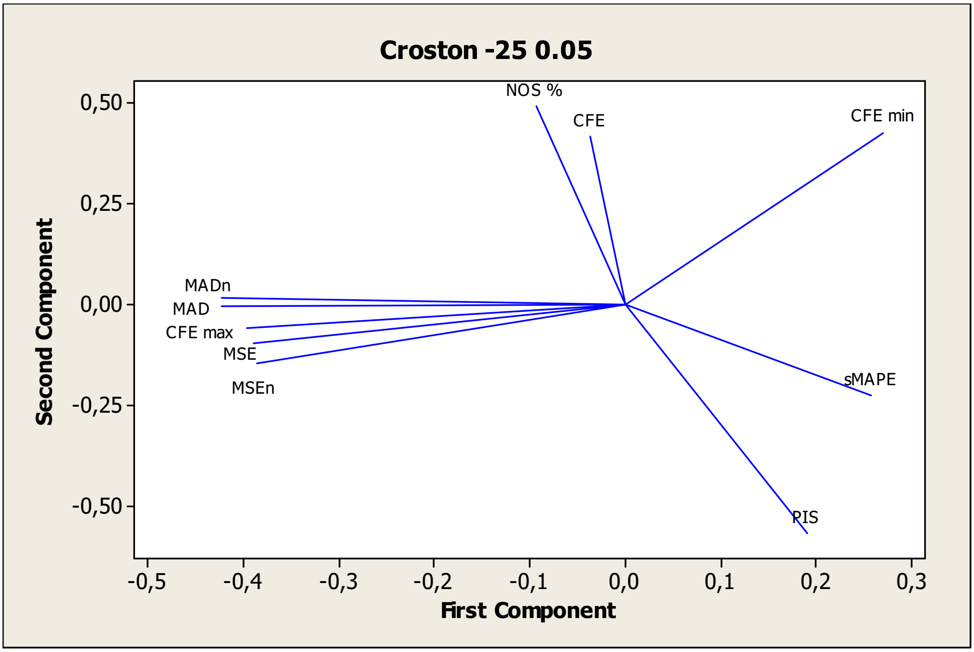


Figure 4.8 Loading plot of Croston -25-s with 10 forecast errors.

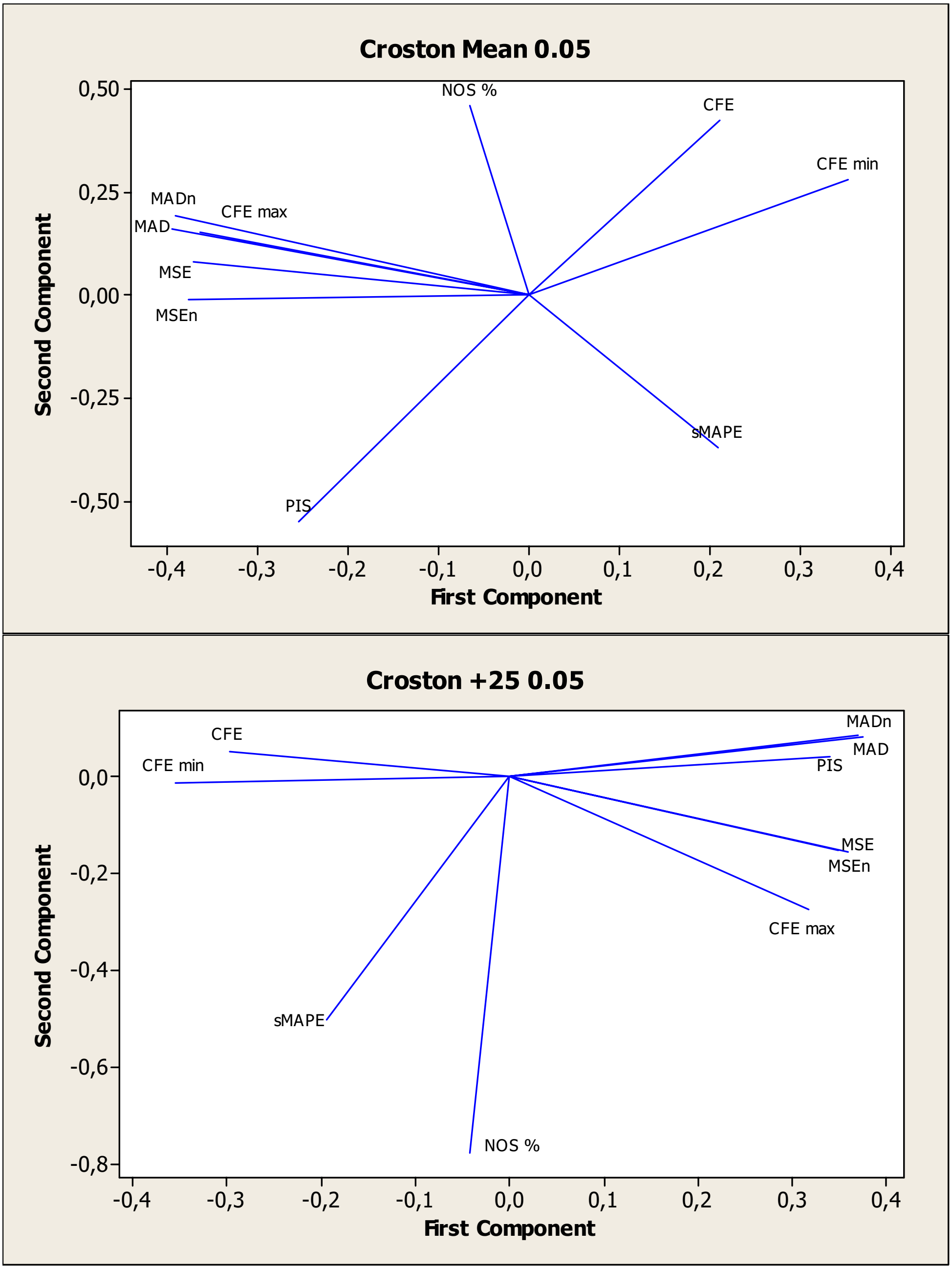


Figure 4.9 Loading plots of mean-s and +25-s of Croston

To further investigate the relationship between CFE and the momentary measures of CFE_{min} and CFE_{max} the quotient between CFE and CFE_{min} was examined as well as the quotient between CFE and CFE_{max}. In Table 4.28 the statistics are based on the quotients between CFE and CFE_{min} of every item. CFE is the denominator. A value of the quotient in the region of 1.0 implies a correlation between the two measures. The greatest change between the start values can be found in the variance which is lower for +25-s in relation to -25-s or mean-s. The median has less spread compared to the mean and similar values for mean-s and +25-s. The great spread especially with -25-s and mean-s is partly caused by the CFE vicinity to zero. The results of the PCA can be traced in the CFE_{min}/CFE quotient. No table is presented between CFE_{max}/CFE Since the quotient did not reveal any additional information to the PCA.

Table 4.28 CFE_{min}/CFE Quotients of Croston

				-25-s				
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	0,38	1,13	2,61	0,53	-0,08	2,15	0,11	0,96
Median	-0,46	1,22	1,30	1,32	1,34	1,30	1,27	1,28
Std	4,60	5,04	9,14	7,90	15,90	4,22	9,97	3,54
Min	-7,30	-22,02	-15,75	-59,30	-128,99	-4,94	-80,36	-20,11
Max	27,12	16,50	66,68	14,12	18,25	28,37	10,70	11,11
				Mean-s				
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	1,28	1,20	4,39	17,01	0,95	3,22	1,29	1,85
Median	1,27	1,29	1,31	1,29	1,29	1,28	1,28	1,28
Std	3,92	4,35	23,72	131,49	3,89	15,51	3,52	15,09
Min	-17,96	-23,09	-15,26	-12,60	-19,25	-13,05	-20,28	-90,04
Max	11,54	13,27	199,87	1116,87	10,15	130,79	12,84	71,73
				+25-s				
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	1,34	1,45	1,52	1,57	1,68	1,85	2,62	1,02
Median	1,14	1,20	1,22	1,24	1,25	1,26	1,27	1,26
Std	0,49	0,66	0,80	0,93	1,29	2,05	6,97	5,38
Min	1,00	1,00	1,00	1,00	1,00	1,00	1,00	-40,97
Max	3,54	4,13	4,82	5,58	8,35	14,19	58,54	10,44

Even though the PCA of Croston proves that some of the relations between some of the errors are more dynamic than others, the PCA of Croston shows that the errors have a relationship that is unique compared to the other methods. Croston's documented bias in other studies (Syntetos and Boylan, 2001;

Teunter and Sani, 2009) is conformed. The values of PIS for every item are positive for mean-s or +25-s. Since none of the items have negative PIS it is unlikely that this is just coincidence if the binominal situation is applied. Also the NOSp has a mean of 16-23% for mean-s and of 5-10% for +25-s.

4.11 ModCr

The Max-Min quotients of ModCr have a slightly different appearance than Croston, see Table 4.29. The variance errors have a larger range. For instance MSE has a quotient of 0.14 which is larger than the value of Croston. ModCr is generally more sensitive to an accurate smoothing constant than Croston. This makes it necessary to find an appropriate smoothing constant for ModCr so that ModCr will have a satisfactory performance.

Table 4.29 ModCr Max-Min quotients

	-25-s	mean-s	+25-s
MSE	0,150	0,146	0,137
MAD	0,110	0,093	0,076
sMAPE	0,024	0,023	0,022
MADn	0,166	0,140	0,113
MSEn	0,395	0,359	0,320
NOSp	4,836	1,756	0,794
CFE	-0,435	-0,312	-0,304
PIS	0,790	0,411	0,343
CFE min	-0,388	-0,300	-0,301
CFE max	4,557	0,547	0,613

The Bias spread is partly large and partly small. CFE_{max} and NOSp have the larger spread. CFE_{max} has a larger spread with start value -25-s than with mean-s or +25-s. The highest CFE_{max} values occur with low smoothing constants, as the smoothing constants increases the CFE_{max} values decrease. With low smoothing constants the low start value has a larger influence and makes the forecast more underestimating than compared to mean-s. A higher smoothing constant makes the transition faster from under- to overestimating. All this is the reason for the spread value of CFE_{max} and NOSp. CFE, PIS and CFE_{min} have the small spread. In the case of CFE, the spread goes from low (-25-s) to lower (+25-s). This stability in conjunction with the negative CFE spread value makes one suspicious if there is no or little random behaviour around zero. The cause of the negative number of CFE is that the minimum value is also negative and the difference between maximum and minimum is positive which makes the quotient negative when the denominator is positive. The stability, the absolute value of the quotients, is a sign of an error performance that is not changing as much as Croston. This error performance is also valid for PIS and CFE_{min}.

The start values influence on the variance errors resembles Croston, see Table 4.27 for Croston and Table 4.30 for ModCr. Both MAD and MSE increased the size of the error with an increase of the start value. The increase of MAD is comparable to Croston, but the numbers of items with an increasing error of MSE are more than in the Croston case. Both the increase of MAD and MSE is significant (p-value < 0.01). When Croston were concerned the increasing trend of variance as the start value increased, coinciding with the overestimating bias of Croston.

Table 4.30 Number of items with lowest error for a certain scenario concerning MAD and MSE.

	MAD			*MSE*		
	Increasing	*Decreasing*	*mean-s*	*Increasing*	*Decreasing*	*mean-s*
0.025	71	0	0	59	1	11
0.05	71	0	0	58	5	8
0.075	69	1	1	55	5	11
0.10	68	1	2	53	6	12
0.15	67	1	3	49	10	12
0.20	65	3	3	45	13	13
0.25	64	5	2	42	16	13
0.30	63	6	2	45	18	8

To account for the variability of the ten errors three components are required for every start value in the PCA, see Figure 4.10 and Figure 4.11 for the loading plots for various start values. The percentage of accounted variability is 88-90.5%. The first component represents the variance measures with its 52.2-57.9%. The relation CFE_{max} has to the variance measures resembles the situation for Croston and therefore the same reasoning can be applied for ModCr. The second component accounts for more variability for ModCr compared to Croston with 17.5-27.9% instead of 12.5-21.4%.

Noteworthy is the close distance and similar angle of the loadings between CFE and CFE_{min}, an indication of bias. PIS and CFE carries almost the variation information which is traceable to the loadings; almost 180 degrees angle and the same length. Under this condition PIS do not reveal any additional information than what CFE reveals.

The third component (10.5-12.6%) is a bias component dominated by NOSp and where CFE_{max} increases its weight with higher start values. The relationship between the errors in the loading plots bears a stronger resemblance for ModCr than Croston between the different start values. However in the mean-s loading plot the second component is a mirror image compared to the other two. Still, the relationship is to a high degree intact. Notice how NOSp goes from an own position to become more similar to sMAPE. When the start value is lower than the mean it is more probable that NOSp varies more than in the overestimating situations that occur with higher start values.

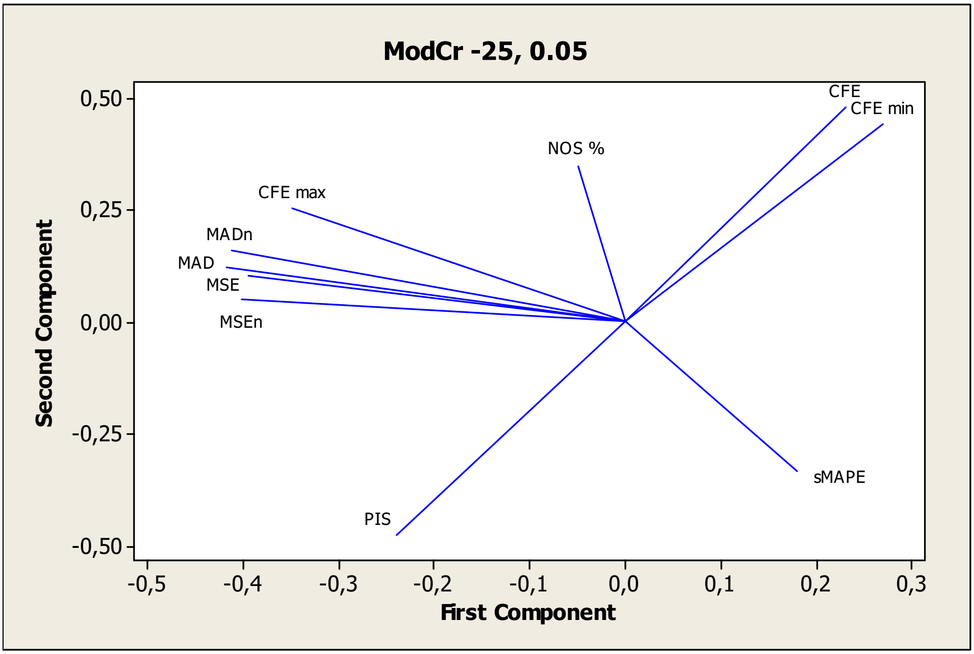
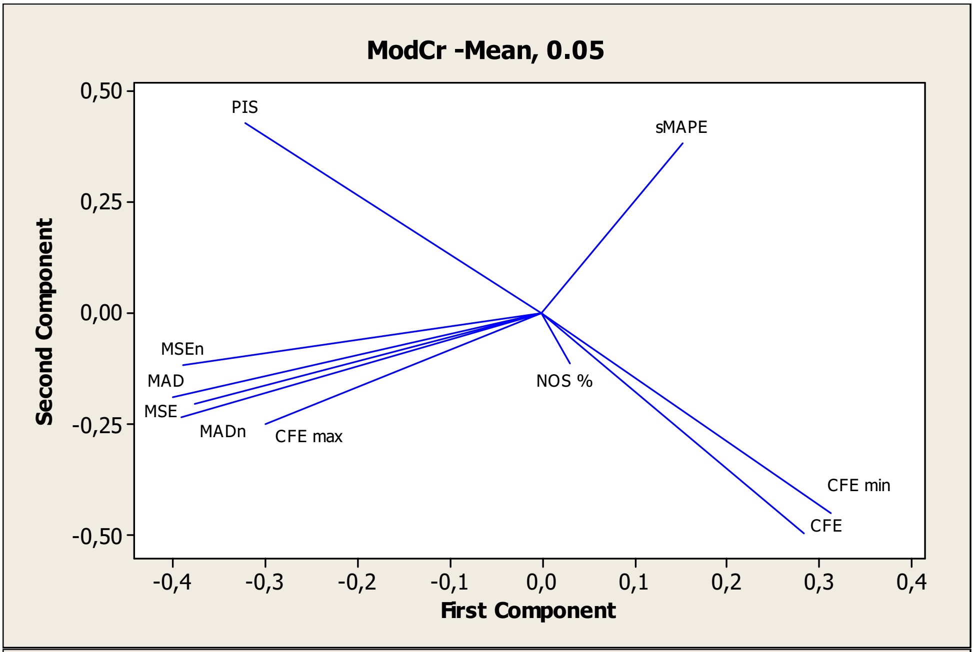


Figure 4.10 Loading plot of ModCr -25-s with 10 forecast errors.



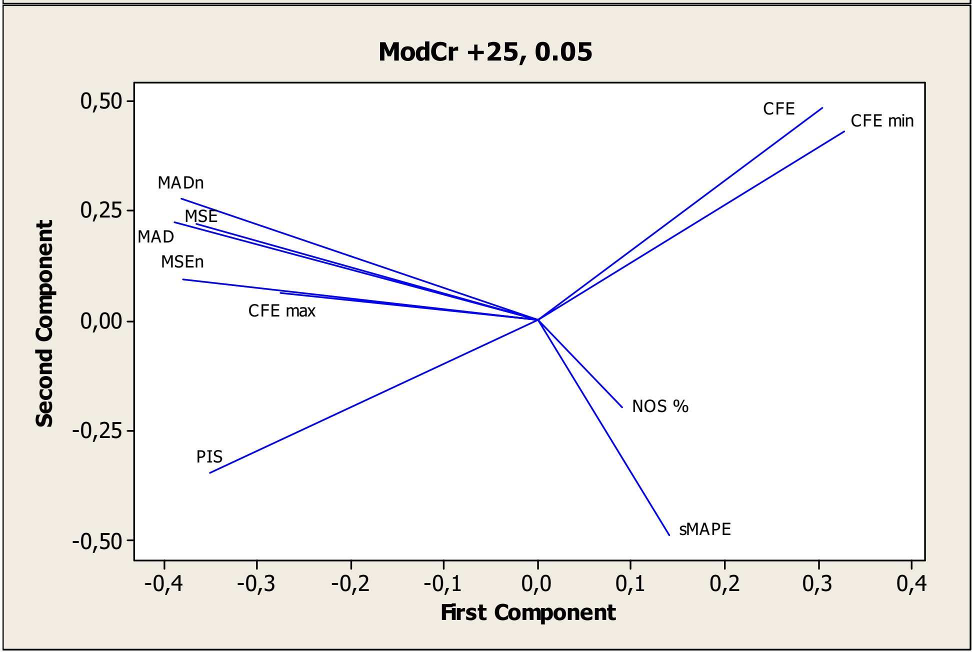


Figure 4.11 Loading plot of ModCr with mean-s and +25-s with 10 forecast errors.

In Table 4.31 the CFE_{min}/CFE quotient has both mean and median values that are close to 1.0 which suggests a resemblance. It is more likely that CFE and CFE_{min} values that are approximately equal if the two values are not to far apart in time. Besides the mean and median the variance is rather low, it is only 0.025 for start value -25-s that has a standard deviation that sets it apart from the rest. The spread is also most deviant for the same start value and smoothing constant, but not as extreme. This is another sign of bias when the variance is low and the quotient in the vicinity of 1.0. The CFE_{max}/CFE quotient did not have a mean or median close to 1.0, that and the loading plots of the PCA did not support an underestimating bias, which did not motivate a further analysis of the quotient CFE_{max}/CFE.

Table 4.31 CFEmin/CFE Quotients of ModCr

				-25-s				
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	1,08	1,08	1,06	1,06	1,05	1,04	1,04	1,03
Median	1,01	1,01	1,01	1,01	1,01	1,01	1,01	1,01
Std	1,22	0,14	0,12	0,10	0,08	0,07	0,06	0,06
Min	-7,01	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Max	7,19	1,76	1,64	1,52	1,37	1,27	1,25	1,27
				mean-s				
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	1,08	1,06	1,05	1,05	1,04	1,04	1,03	1,03
Median	1,01	1,01	1,01	1,01	1,01	1,01	1,01	1,01
Std	0,17	0,10	0,09	0,09	0,07	0,06	0,06	0,05
Min	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Max	1,99	1,46	1,45	1,40	1,31	1,24	1,23	1,25
				+25-s				
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	1,05	1,05	1,05	1,04	1,04	1,04	1,03	1,03
Median	1,01	1,01	1,01	1,01	1,01	1,01	1,01	1,01
Std	0,09	0,07	0,08	0,07	0,07	0,06	0,05	0,05
Min	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Max	1,46	1,33	1,35	1,32	1,26	1,22	1,21	1,23

The preceding methods of analysis all points in one direction; that ModCr has a bias problem which is in agreement with the findings of Boylan and Syntetos (2007) and/or Teunter and Sani (2009). The method tends to overestimate the actual demand. During the work, discrepancies were found between the demand rate and the quotient demand and inter demand period. The demand rate was higher than the quotient for every item. This is analogous to flipping a coin 72 times and the same side comes up, hardly a coincidence since the probability is less than 10^{-21}.

The deviation is from 3% up to 160%. The larger deviation occurs when the percentage of demand occasion is low. The scatterplot revealed a linear correlation between the demand rate and the quotient (demand/inter demand period). Figure 4.12 shows the plot but without the five most influential observations in X and Y space. The judgement of influential points was made based on the scatterplot of every item.

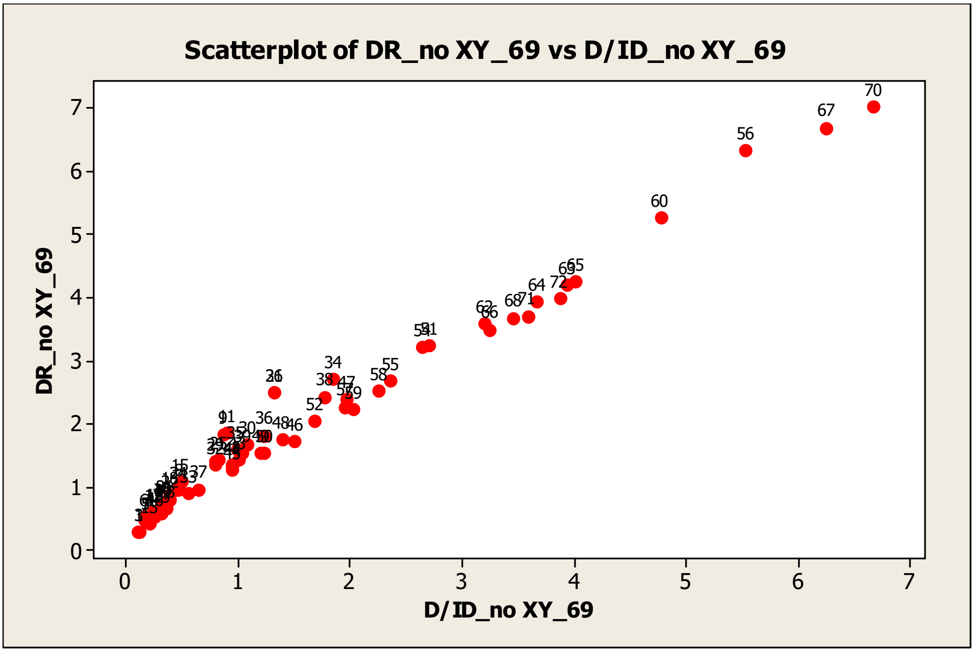


Figure 4.12 Plot of the quotient between demand, inter-demand and the demand rate without the influential observation in X- and Y-space.

The regression analysis with demand rate as the response and the quotient between demand and inter-demand period have a R-squared of 98,8%, see Table 4.32. The choice of response and predictor does not reflect any hypothesis regarding the cause and effect. By removing the most influential predictor observations, the R-squared decreases to 95.3%. If both the most influential predictor and response observations are removed the R-squared increases to 98.8%. In all three analyses the p-values for constant and predictor are less than 1% which makes them both significant. The normal probability plots of the residuals show that the right end tails are a slightly extended. Two other regression analyses were also made, one between demand rate and demand occasions in percent and the other between the deviation in percent (demand rate/(demand/inter-demand period)) and demand occasion in percent.

Table 4.32 Summary of Regression analysis of Demand rate (DR) and the quotient between demand (D) and inter- demand periods (ID)

		P-value		
	Equation	*Constant*	*D/ID*	*R-Sq*
Initial	DR = 0,319 + 1,13 D/ID	0,000	0,000	98,8%
No X outliers	DR = 0,269 + 1,16 D/ID	0,005	0,000	95,3%
No X and Y outliers	DR = 0,406 + 1,02 D/ID	0,000	0,000	98,8%

The first analysis shows that the demand occasions in percent is significant but the model has a R-squared of 17%. The other analysis has a higher R-squared, 81.2%, and the demand occasions in percent is significant but with a residual versus fit, that has a curvature. The normal probability plots for the analyses between demand rate and quotient show that a linear regression might not be the perfect model to describe the relationship between demand rate and the quotient, but with a R-squared of at least 95.3% and a demand rate always higher than the quotient it is a valid assumption that the bias of ModCr is related to the discrepancies. According to the equation without any outliers the demand and inter-demand quotient is increased with 2 percent plus 0.406. Therefore the largest deviance are between demand rate and the demand, inter demand quotient when the mean demand is low and the inter demand periods are high.

The bias of ModCr can be studied in the following example in Table 4.33. Suppose a demand series consist of two demand occasion with a total demand of 10. The demand occurs over the period of six days. The mean demand for each day is 1.67. The mean can be calculated by the quotient of sum of the demand and sum of inter demand periods or by the quotient of mean demand and mean inter demand periods. In the first example (1) the mean of the demand rate is identical to mean demand per period. In the second example the demand rate is not identical to mean demand per period.

Table 4.33 Difference between demand/period and mean demand rate.

	1			2		
Period	*Demand*	*Inter-demand*	*Demand rate*	*Demand*	*Inter-demand*	*Demand rate*
1						
2						
3	1	3	0,33			
4						
5				1	5	0,2
6	9	3	3	9	1	9
Sum	10	6	2	10	10	9,11
Mean	5	3	*1,67*	5	3	*4,6*
Demand/period		1,67				

A further example of time series can be found in Table 4.34, with two demand occasions over the course of six periods with a demand of ten, shows that the mean demand rate can be above as well as below the mean demand per period.

Table 4.34 Variations of a time series with a mean demand period of 1.67

Variation	Demand	ID	DR	Mean DR	Quotient	Difference
1	1	1	1			
	9	5	1,8	1,4	0,84	-0,267
2	1	2	0,5			
	9	4	2,25	1,375	0,825	-0,292
3	1	3	0,33			
	9	3	3	1,667	1	0,000
4	1	4	0,25			
	9	2	4,5	2,375	1,425	0,708
5	1	5	0,2			
	9	1	9	4,6	2,76	2,933

In variation 1, see Table 4.34, the first demand occurs in the first period which makes the demand rate equal to 1 and the mean demand for variation 1 equal to 1.4. The mean demand is lower than demand/period (1.67). Quotient is the relation between mean demand rate and mean demand per period. The quotient of variation 1 is 0.84. The difference between the mean demand rate and the mean demand per period is -0.267. The differences and the quotients are smaller when the mean demand rates are lower than the actual demand compared to the larger differences and quotients when the mean demand rate are higher than the actual demand. Even if there is a chance of a lower mean demand rate than demand/period the most likely event is the opposite of which the deviation in the statistics from the items, a reasoning that the regression analysis support. The quotient between mean demand and mean inter demand periods is equal to mean demand per period.

In the previous examples the difference between the mean demand rate and the mean demand per period was shown with numbers. A general description of the change between the different means is presented in equation 4.10-4.12. Assume that there is an initial time series where the mean demand rate is equal to mean demand per period. The initial time series consist of a finite number of periods, see equation 4.10-4.12. To this initial time series one demand, D_a, is added. The inter demand period, ID_a, is larger than one for the new demand. If the inter demand period is allowed to approach infinity, the mean demand is not affected more than the new demand is added and the denominator is increased by one, but since both demand and inter demand has originally a denominator the equation is simplified by removing the denominators, see equation 4.13.

However the inter demand periods approaches infinity as ID_a approaches infinity as a result the mean demand per period will approach zero. The mean demand rate will not approach zero. It is only the new quotient that will approach zero as ID_a approaches infinity. Therefore the new demand rate will instead be a version of the initial mean demand rate with slightly lower mean. The decrease of the mean demand rate is dependent on how many time periods that the initial time series consisted of. The larger the number of time periods the less influence on the mean demand rate, see equation 4.14.

$$D_{initial} = D_1 + D_2 + \ldots\ldots + D_n \tag{4.10}$$

$$ID_{initial} = ID_1 + ID_2 + \ldots\ldots + ID_n = n \tag{4.11}$$

$$DR_{initial} = \frac{D_1}{ID_1} + \frac{D_2}{ID_2} + \quad + \frac{D_n}{ID_n} \tag{4.12}$$

$$\frac{D_{mean}}{ID_{mean}} = \frac{D_{initial} + D_a}{ID_{initial} + ID_a} \tag{4.13}$$

$$DR_{mean} = \frac{\left(DR_{initial} + \frac{D_a}{ID_a}\right)}{n+1} \tag{4.14}$$

If instead, one demand occasion is allowed to approach infinity both the mean demand per period and the mean period will approach infinity but at different rates. The condition that the mean demand rate should be equal to the mean demand per period is still valid. The increase, *c*, is added to the quotient between D_a and ID_a. The ratio of the quotient is the same as for the mean demand per period. Then in equation 4.15 the expression can be divided in two quotients; the first one is the mean demand per period and the second one is the marginal change with $n+ID_a$ as the denominator. The expression for the demand rate also has a quotient equal to the mean demand but the marginal change quotient is different. Instead of $n+ID_a$ the marginal change for the mean demand rate is $ID_a(n+1)$ as the denominator, see equation 4.16. The different denominators make the mean demand rate to change at a slower rate than the mean demand per period, see equation 4.17. The slower rate is the same for increase as well as decrease.

$$\begin{aligned} \frac{D_{mean\_c}}{ID_{mean\_c}} &= \frac{D_{initial} + D_a \pm c}{ID_{initial} + ID_a} = \frac{D_{initial} + D_a \pm c}{n + ID_a} \\ &= \frac{D_{initial} + D_a}{n + ID_a} \pm \frac{c}{n + ID_a} \end{aligned} \tag{4.15}$$

$$DR_{mean\_c} = \frac{\left(DR_{initial)} + \dfrac{D_a \pm c}{ID_a} \right)}{n+1}$$
$$= \frac{DR_{initial)} + \dfrac{D_a}{ID_a}}{n+1} \pm \frac{c}{ID_a \cdot (n+1)} \tag{4.16}$$

$$\frac{1}{ID_a \cdot (n+1)} < \frac{1}{ID_a + n} \tag{4.17}$$

Just because the marginal change have equal absolute values it is not an assurance of equal marginal change even if a time series fluctuates around a mean. The distribution of the fluctuation can be skewed which results in different means for demand per period and demand rate. In Figure 4.14 the quotient between the mean demand rate and the mean inter-demand period shows the deviation between the two means. As stated previously, the deviation increases as the number of demand occasion decreases in time series of equal length.

Even if the marginal change of mean shows the possibility of a mean demand rate lower than the mean demand per period, none of the items has a mean demand rate that is lower. The variations between the nominators and denominators of the quotients in a demand rate series varies in a less static and more complex manner. A simple variation is between two integers that have the same probability to occur as a nominator or as a denominator. Then four quotients can be formed out of the two integers. Two of those quotients are equal to one and are omitted from equation 4.18. The mean of the remaining two quotients is always larger than one assumed that one integer is larger than the other. This also applies if the two quotients equal to one are added. This sub-series always form a sub-mean that will increase the mean demand rate compared to the mean demand per period. In equation 4.19 the mean for same variation as mean demand per period will not affect the overall mean at all

$$\left(\frac{d}{e}+\frac{e}{d}\right)\Big/2=\frac{d^2+e^2}{2ed} \qquad d<e$$
$$(e-d)^2>1 \tag{4.18}$$

$$\frac{d+e}{e+d}=1 \tag{4.19}$$

Teunter and Sani (2009) state that when a demand occurs less than 1 out of 3 periods the bias of ModCr is more than 50%, this is comparable to the values of the quotients of deviation and the percentage of demand occasions in Figure 4.13. There is an item in the figure that has a demand occasion value of approximately one third and the deviation quotient is approximately 1.5 which is a deviation of 50% between the mean demand rate and the mean demand per period. As the percentage of demand occasion approaches 1.0 (100%) the deviation decreases. A likely explanation is that mean demand rate and the mean demand per period are the same when a demand occurs in every period.

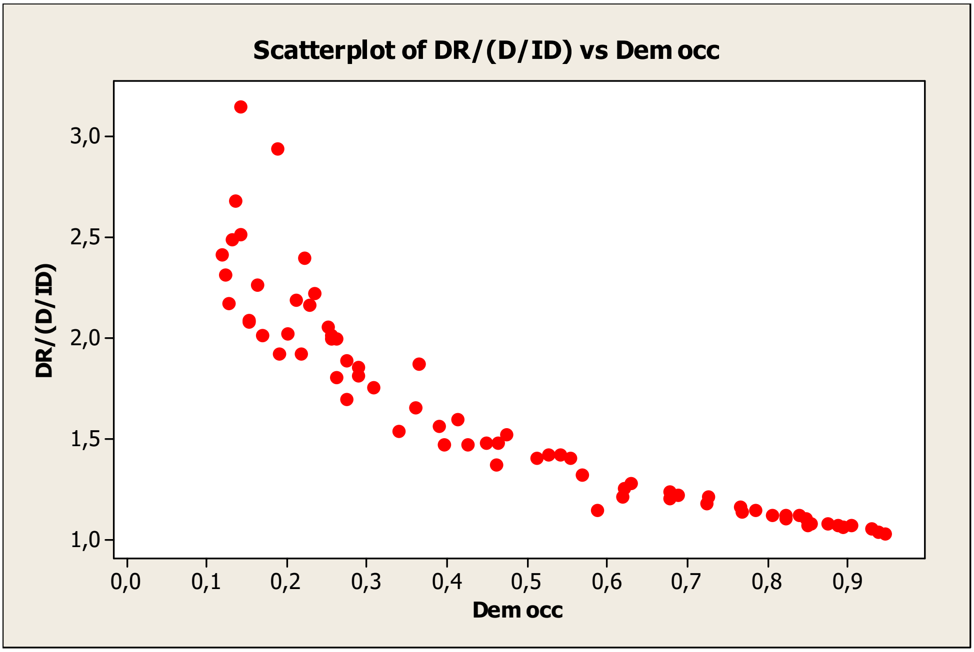


Figure 4.13 Scatterplot of Demand rate/(Demand/Inter demand periods) and the proportion demand occasions of the number periods in percent

Based on Figure 4.13 the deviation between the demand rate and the quotient of mean demand inter-demand period will approach infinity. However, the spread in curvature in Figure 4.13 increases for lower values of demand occasions and therefore diminishes such an assumption. Furthermore, the demand is also influential for the outcome of the demand rate. Figure 4.14 displays the fact that the proportion demand occasions alone will not explain the devation. The highest devations, in the proximity of 3, have two items with a low number of demand occasions but not the lowest number of demand occasions.

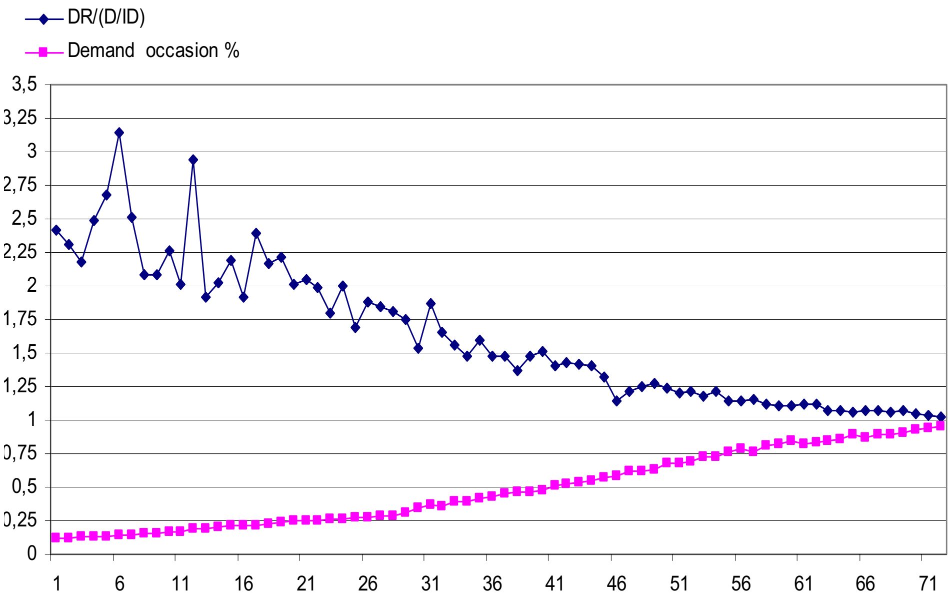


Figure 4.14 Demand rate/(Demand/Inter demand periods) and the percentage of demand occasions for all items. As the percentage of demand occasions is approaching 100% (1.0) the same is also valid for DR/(D/ID).

In Figure 4.15, the MACs values are plotted for every item and their respective demand, inter demand period and demand rate. The CV values have a similar appearance as MACs but the lines are not equally close to each other. Since MACs takes the sequence of the time series in account it is used here instead of CV that do not consider the order of the sequence. MACs is the mean absolute change in relation to the mean of the series. As the inter demand period decreases and approaches 1.00 the values and lines of demand and demand rate become more comparable, which is expected as the inter demand approaches 1.00. A similar pattern can be found as early as from item 8. The values are not as close as for a higher number of demand occasions, but the lines shows evidence of a parallel change.

The main problem with ModCr is that it tries to forecast the new mean demand per period based on the mean demand rate that usually has a higher mean. The regression analysis with demand and inter demand periods as predictor indicates that even if ModCr can perform unbiased forecasts of the demand rate, what the method forecasts is not the mean demand per period but the demand rate which is something else than demand per period. This deviation becomes in practice equal to a biased forecasting method. The one exception is when there is a demand in every period which makes the sum of inter demand periods equal to the number of demand rates.

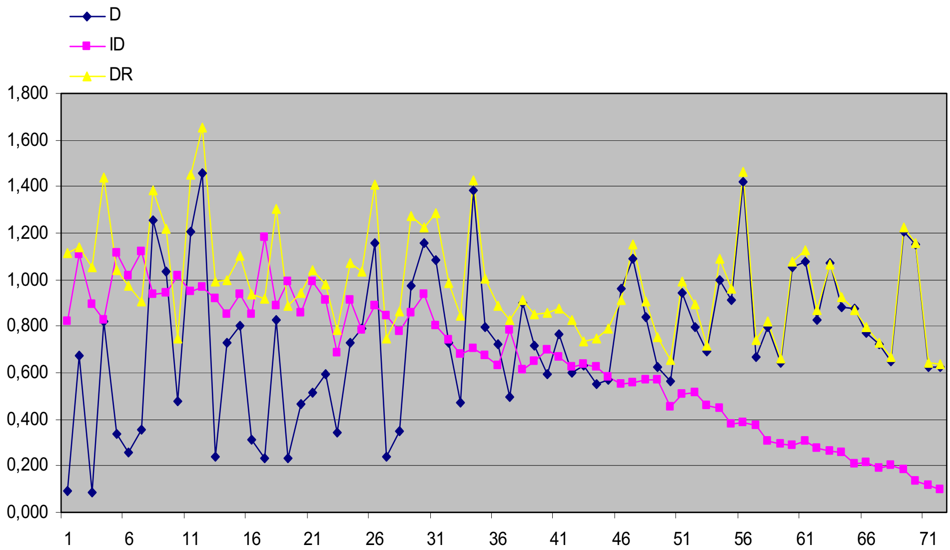


Figure 4.15 MACs values for all items and the items respective demand rate, demand and inter-demand interval.

4.12 SES

The Max-Min quotients of SES are partly similar to ModCr regarding the quotients of the variance measures. SES has, after ModCr, the highest spread, see Table 4.35. The differences concerning MAD quotients are much smaller then the MSE quotients, where SES and ModCr form the high spread subgroup and Croston and SyBo form the low spread subgroup. The distinction of subgroups coincides with how the different method works. SES and ModCr makes one forecast instead of two as Croston and SyBo do. The quotients of SES are similar for MSE and MAD regardless of start value.

It is therefore of importance to find a suitable smoothing constant in order to reduce the size of safety stock that is affected by the forecast error. If MSE is used as a base for safety stock, the safety stock would probably be larger compared to a safety stock based on MAD, because MSE has a larger spread.

SES has a better MSE and MAD performance than Croston but is more sensitive to the value of the smoothing constant.

Unlike the variance measures the Max-Min quotient of NOSp varies between the start values. The highest spread, 0.874, has +25-s due to the high start value in combination with the smoothing constants. The higher start value reduces NOSp which is affected by the smoothing constant, the higher smoothing constant the higher NOSp. This is expected since the SES decreases the forecast for every period without a demand and the decrease is faster when using a higher smoother constant. CFE and PIS have large spreads. The difference between maximum and minimum is not the largest of the forecasting methods, but the median is usually closer to zero than for any of the other methods and therefore the spread becomes larger. CFE_{max} has a mean quotient that decreases when start value increases. With a higher start value CFE_{max} for the different smoothing constants will be more similar and the spread decreases.

Table 4.35 SES Max-Min quotients

	-25-s	*mean-s*	*+25-s*
MSE	0,139	0,140	0,139
MAD	0,063	0,059	0,056
sMAPE	0,044	0,045	0,046
MADn	0,130	0,133	0,138
MSEn	0,263	0,274	0,296
NOSp	0,184	0,301	0,847
CFE	2,483	177,91	-6,459
PIS	-4,919	-5,128	6,128
CFE min	-1,479	-2,162	-2,684
CFE max	1,058	0,645	0,456

The way the start values influence the variance errors is similar and dissimilar to Croston and ModCr; similar because MAD increases as the start value increases, dissimilar because MSE decreases instead of increases. The exception is 0.025 where mean-s instead is significant, see Table 4.36. Mean-s results in a lower MSE. However, the quantity of the items decreasing error trend is less than the quantity of the items with increases trend for Croston and ModCr.

Table 4.36 Number of items with lowest error for a certain scenario concerning MAD and MSE for SES.

	MAD			MSE		
	Increasing	*Decreasing*	*mean-s*	*Increasing*	*Decreasing*	*mean-s*
0.025	69	1	1	18	24	29
0.05	66	2	3	27	29	15
0.075	66	2	3	25	29	17
0.10	65	3	3	24	34	13
0.15	63	5	3	24	34	13
0.20	61	7	3	24	35	12
0.25	61	8	2	23	34	14
0.30	61	10	0	26	34	11

The PCA:s of SES show that two components can account for approximately 83% of the variability concerning -25-s and +25-s. A third component is needed to increase the degree of explanation from 73.5% to 83.7% for mean-s. The third component has an eigenvalue of 1.025 and accounts for 10.3%. Mean-s differs from the other two start values because CFE is hardly present in the first two components but has the highest loading of the errors for the third component, see Figure 4.16 and Figure 4.17.

The third component is the bias component while the second component is a mixture of PIS (bias) and sMAPE. The second component represents 14.3% of the variability. The first component accounts for 59.1% and is a mixture between the variance errors and the max and min errors of CFE. For -25-s and +25-s the first component is also a mixture of variance errors and the max and min errors of CFE, but with the addition of PIS.

The first components represent 67.3% or 68.1%. The second component for -25-s and +25-s are a mixture of bias (NOSp) and sMAPE. The second component accounts for 14.8% or 15.8%. The relation between CFE and PIS in the loading plots indicates a weaker correlation than for any of the other methods. PIS usually has lower values, closer to zero, than PIS has for any of the other methods. Even if the size, in terms of difference, of the variation of CFE and PIS are approximately the same, the variation around zero can distort the correlation. No study of the quotients between CFE and CFE_{max} is necessary and will not reveal any eventual bias.Since the loadings of CFE_{max} and CFE have different loadings and for +25-s, different directions.

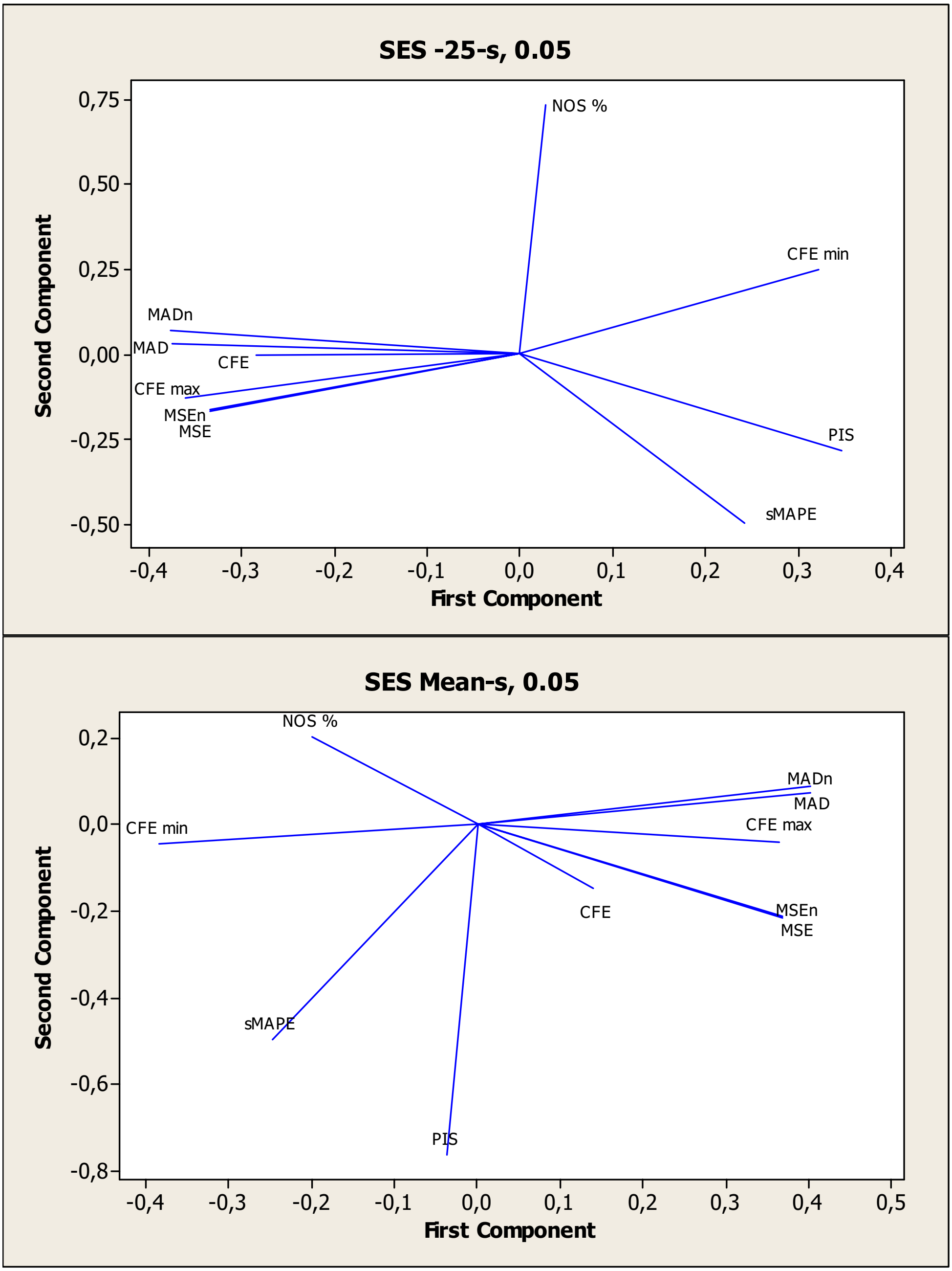


Figure 4.16 Loading plot of SES -25-s and mean-s both with 10 forecast errors.

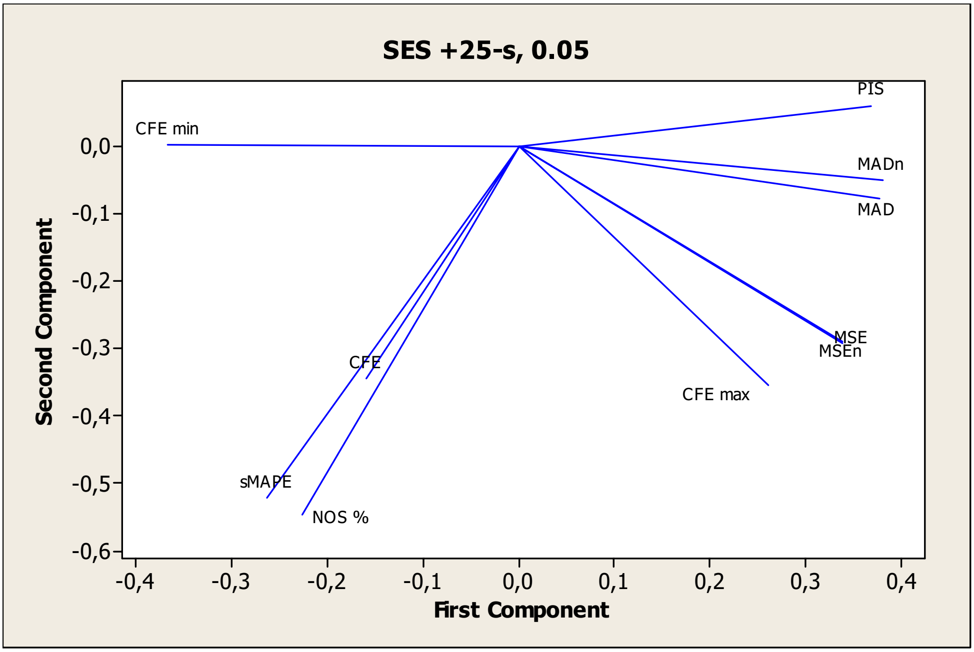


Figure 4.17 Loading plot of SES +25-s with 10 forecast errors.

4.13 SyBo

The Max-Min quotients of SyBo have some similarities to the Max-Min quotients of Croston. SyBo has the lowest spread values concerning the variance measures including MADn and MSEn, see Table 4.37. The variance quotients are slightly lower than Croston's variance quotients. This means that SyBo is not as sensitive to a demand smoothing constant that is not optimal from the lowest variance measure point of view. MSE has a mean of 0.044 for both mean-s and +25-s. The difference between the largest and smallest error expressed in number of medians is 4.4%. To find the demand smoothing constant that has the lowest error in conjunction with a satisfactory bias performance is not as important for SyBo compared to SES and ModCr since the spread is lower, as long as SyBo is reasonable bias free. Therefore less effort can be taken, in order to set up the forecasting before using the method for forecasting. The quotients are stable for the different start values.

The spread of the bias quotients are larger for CFE and CFE_{min} than for CFE_{max}. Of the four methods, SyBo has the lowest spread concerning CFE_{max} especially for mean-s, 0.188. The spread is only to some extent dependent on the smoothing constant because the spread is lower than for any other method. Still the quotient is approximately three times larger than the MSE quotient. PIS has a low spread for +25-s and mean-s compared to -25-s. With the lowest start value, the lowest smoothing constant will slowly adapt to a more accurate mean while the highest smoothing constant will make the same adaptation faster. The different adaptation is a cause of the larger spread. Along with SES, SyBo has the least spread of NOSp.

Table 4.37 SyBo Max-Min quotients

	-25-s	mean-s	+25-s
MSE	0,042	0,044	0,044
MAD	0,047	0,032	0,038
sMAPE	0,020	0,020	0,024
MADn	0,068	0,046	0,052
MSEn	0,154	0,116	0,141
NOSp	0,334	0,241	0,789
CFE	1,334	-0,996	-0,759
PIS	-20,782	-1,886	-1,546
CFE min	-1,243	-0,991	-1,692
CFE max	0,586	0,188	0,464

The start values influence on variance errors is comparable to SES. As for every other method MAD increases as the start value increases. Like SES, SyBo has a significant mean concerning mean and smoothing constant 0.025, see Table 4.38. With mean-s instead of the other two start values the method performs better from a MSE perspective. Similar to SES but dissimilar to Croston and ModCr, the significant part of MSE is decreasing as the start value increases. The quantity of decreasing items is lower than the increasing quantity of increasing items for Croston and ModCr, which is comparable to SES.

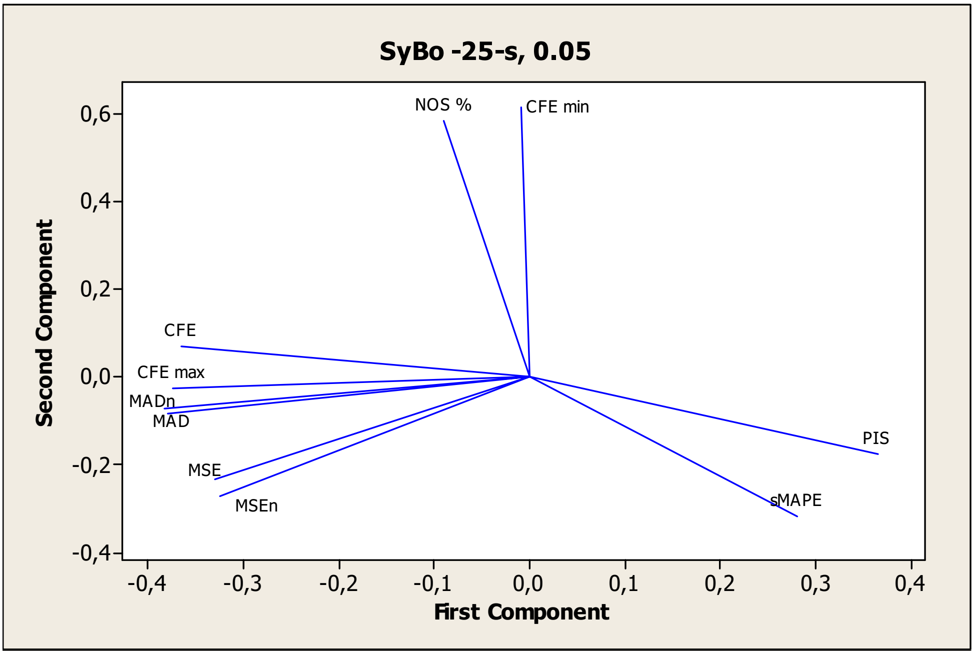
Even if MSE can not detect a bias, covaries the significant increasing or decreasing error with the method bias tendency. Croston and ModCr have lower MSE with low start values than with higher start values. SES and SyBo have higher MSE when the start value is high.

Table 4.38 Number of items with lowest error for a certain scenario concerning MAD and MSE for SyBo.

	MAD			MSE		
	Increasing	*Decreasing*	*mean-s*	*Increasing*	*Decreasing*	*mean-s*
0.025	67	0	5	18	19	35
0.05	64	1	7	25	24	23
0.075	64	4	4	25	30	17
0.10	64	4	4	27	33	12
0.15	63	6	3	28	35	9
0.20	62	7	3	28	33	11
0.25	61	9	2	27	33	12
0.30	59	10	3	25	33	14

The variability of the ten errors requires two components if one settles for an accountability of 84.9-88.8%. The third component has an eigenvalue of 0.827 when the lowest start value is used and accounts for 8.2%. The rest of the components have lower eigenvalues and therefore a higher amount of noise. The explained variability of the two components is in the same region as with three components for the other methods; see Figure 4.18 and Figure 4.19.

The first component accounts for 66.6-67.4% and is a mixture of errors, where every error except CFE_{min} and NOSp have loadings between 0.28-0.38. CFE_{min} and NOSp dominate the second component that accounts for 18.3-21.4%. While the second component can be referred to as a bias component, the first component lacks a distinct identity more or less the same as the case with three errors, see MSE correlation part in 4.1.1 Correlation between MSE, MAD and CFE. NOSp decreases its loading of the second component as the start value increases due to its increasing covariation with the errors of the first component as the percentage of shortages decreases with a higher start value. The loadings and angles of CFE and CFE_{min} are similar which can indicate an underestimating bias.



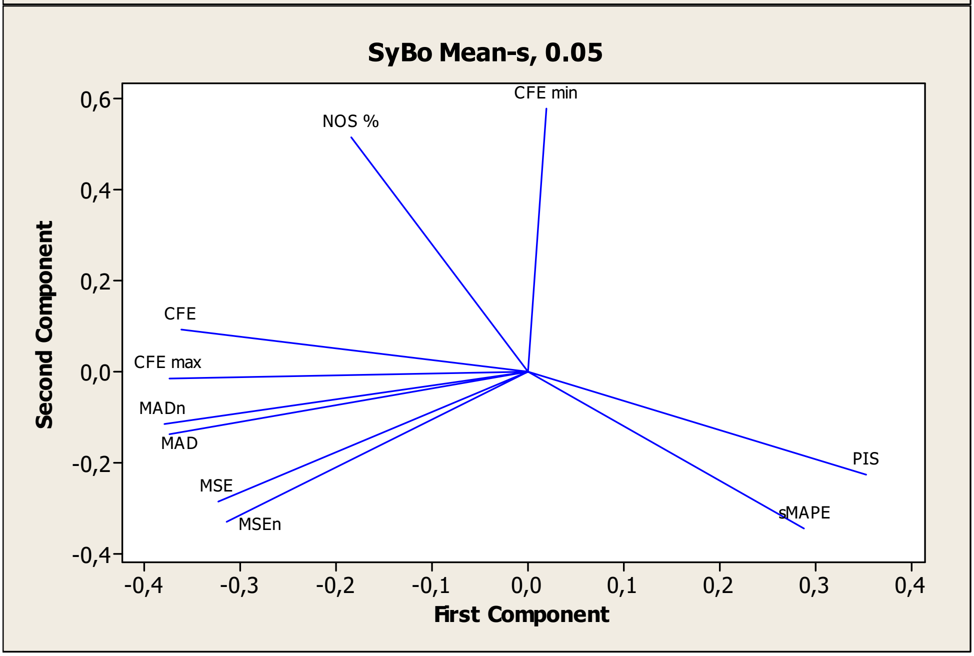


Figure 4.18 Loading plot of SyBo -25-s and mean-s both with 10 forecast errors.

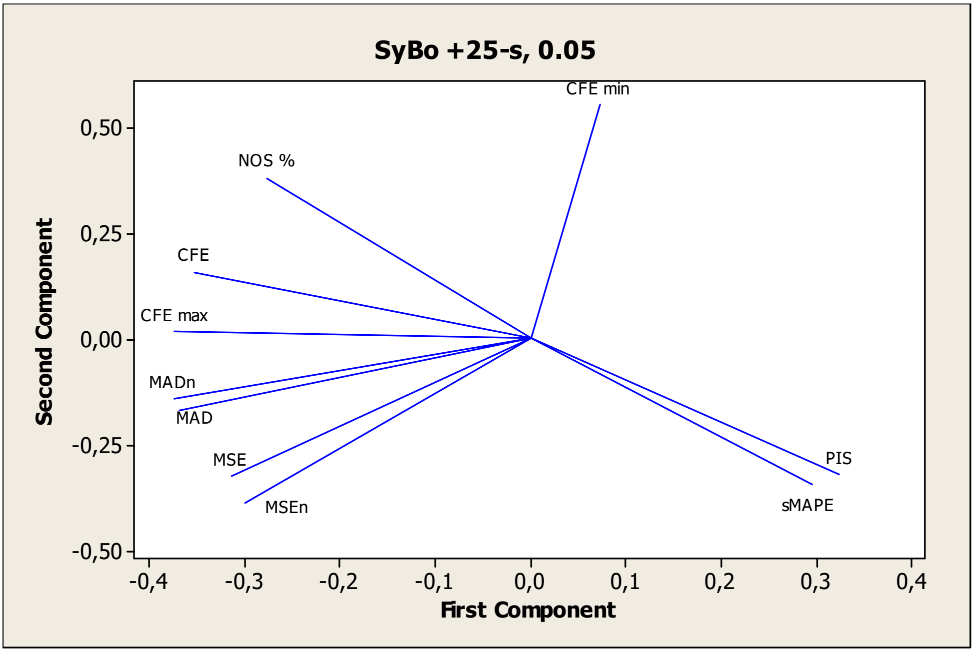


Figure 4.19 Loading plot of SyBo +25-s with 10 forecast errors.

The loading plots of the PCA indicates a possible bias which cannot be confirmed by the quotients between CFE_{max} and CFE, see Table 4.39. The mean varies and are on a few occasions in the proximity of 1.0. The values of the medians are closer to 1.0, especially for +25-s. The values of the standard deviations are large compared to the means and so are also the min and max values. If an overall systematic bias exists the standard deviation should be lower instead of the higher values that indicate a large variation between the items. This variation is also traceable to the values of max and min. For more loading plots, see Appendix – PCA.

Table 4.39 CFE$_{max}$/CFE Quotients of SyBo

				-25-s				
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	6,67	3,88	-0,69	2,05	4,95	2,37	2,03	5,15
Median	1,26	1,38	1,27	1,23	1,23	1,21	1,20	1,18
Std	44,83	10,79	46,80	7,34	19,02	6,03	5,53	23,78
Min	-8,66	-2,03	-353,11	-26,78	-28,28	-15,13	-17,35	-21,46
Max	381,60	78,07	163,49	34,58	123,23	30,15	27,30	189,76
				mean-s				
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	1,83	0,43	2,52	22,15	2,54	1,72	2,75	0,79
Median	1,17	1,14	1,15	1,16	1,13	1,13	1,11	1,09
Std	14,64	12,63	10,87	168,37	14,66	8,99	16,76	4,42
Min	-17,75	-63,19	-14,23	-25,84	-37,60	-32,16	-15,22	-17,26
Max	118,32	55,94	88,25	1426,54	104,35	55,61	129,45	15,65
				+25-s				
	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
Mean	-0,16	3,26	-0,25	-4,31	-0,17	-3,63	2,01	1,31
Median	-0,14	1,00	1,01	1,02	1,05	1,05	1,08	1,08
Std	3,29	28,11	4,40	49,63	11,05	19,15	14,96	4,61
Min	-23,65	-49,74	-27,49	-346,05	-77,96	-132,96	-59,11	-7,68
Max	4,20	231,51	10,16	183,67	34,78	14,40	106,06	26,62

SyBo is not without bias, but it is not as biased as ModCr or Croston. The method has a tendency to underestimate the demand for a majority of the items. In Table 4.40 is a summary of the number of items that have positive PIS values. Approximately 25% of the total 1728 forecasts have a positive number which means that 75% has a negative number, an indication of underestimation over time.

Table 4.40 Summary of number of items with positive PIS values for SyBo. Maximum value is 72

	0,025	*0,05*	*0,075*	*0,10*	*0,15*	*0,20*	*0,25*	*0,30*
-25-s	2	2	2	3	7	10	11	13
mean-s	17	19	20	19	19	17	16	16
+25-s	50	40	34	31	27	25	22	21

If one forecast has a PIS value of -1, it is hardly an evidence of bias, but when a majority of the forecasts have negative values it is another matter. ModCr has 8 forecasts out of 1728 with negative PIS. The sign of PIS seems to be linked to the percentage of demand occasions. The tendency of a negative PIS increases as the percentage of demand occasions increases. SyBo has been proven to have an underestimating bias in a study of Teunter and Sani (2009). When a demand series is short and the mean is fluctuating, a forecasting method, that in the long run is unbiased, can show evidence of bias, but it is not likely the reason for the underestimation. The tendency of SyBo's underestimation results in a method that needs the safety stock in order to avoid shortages.

4.14 Comparison of the Forecasting Methods with CFE and PIS

To demonstrate the differences among the four methods an example with 20 forecasting periods has been constructed. The sum of the demand are 22 and there are 8 demand occasions, see Figure 4.20. The smoothing constants are 0.30 to make the characters of the different methods more traceable. The start value is 1 for every forecasting method.

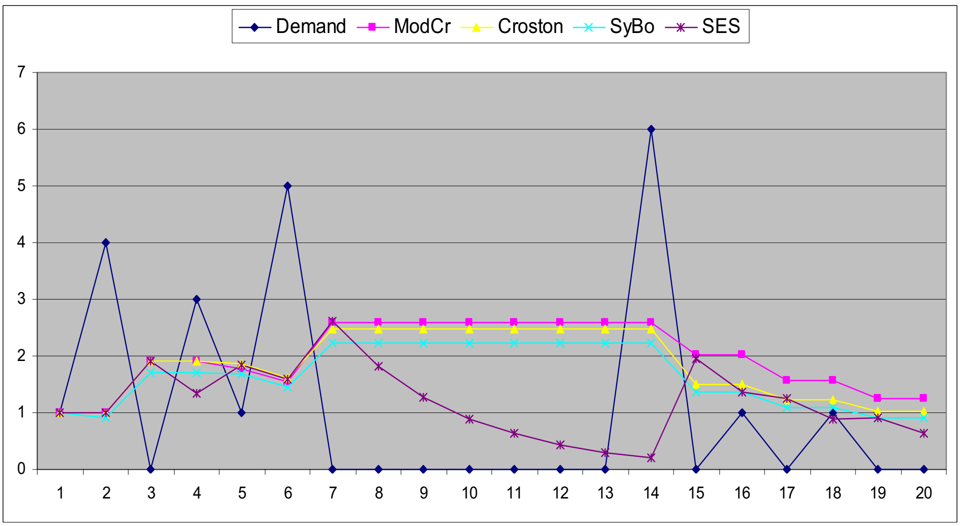


Figure 4.20 Demand and the forecasts of the four methods.

SES differs from the rest due to the update in every period. The other methods are more or less parallel. The differences are the amount of the forecasts. ModCr do generally have a higher forecast than the rest. Croston has forecast that is in between ModCr and SES.

SyBo has the lowest forecast among the three variants of the Croston method, which is also noticeable in the CFE performance, see Figure 4.21. SyBo, as ModCr and Croston, overestimates the demand but has the lowest overestimation of the three. The CFE of SES has not the clear tendency of the other methods. SES goes from a low underestimation to low overestimation compared to the other methods. The demand of 6 in period 14 does affect the CFE values for all methods and that demand makes the CFE of SES to change the sign as well as the momentary bias. SES goes from underestimation to overestimation and ends with a CFE, which is a sign of a low overestimation. All four methods are overestimating the demand according to CFE.

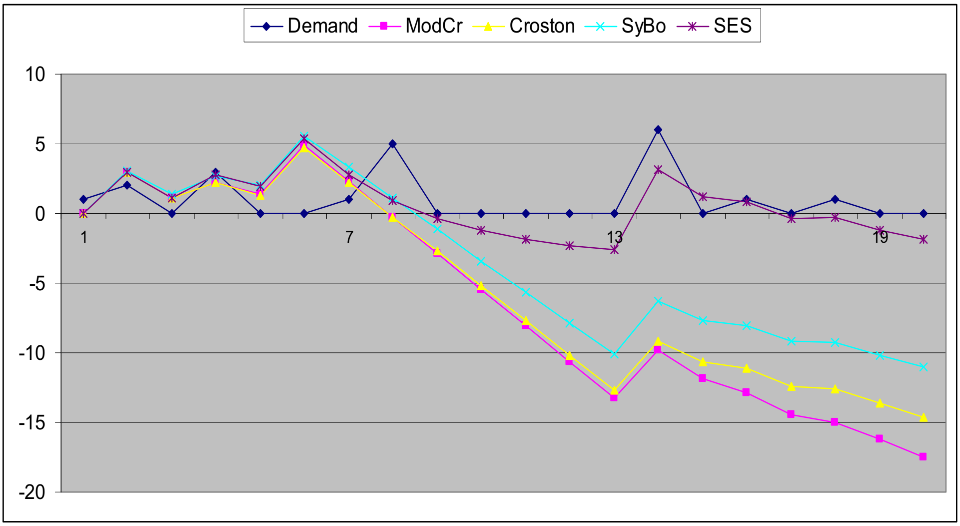


Figure 4.21 Demand and CFE performances of the four methods.

The bias tendencies of Croston, ModCr and SyBo are confirmed with PIS, see Figure 4.22. SyBo still has the lowest overestimation of the three methods. ModCr is the method that has the highest degree of overestimation followed by Croston. The demand in period 14 does not affect the values of PIS as it does concerning CFE. There is no transient response in the PIS curves compared to the curves of CFE. SES is according to its PIS value not mildly overestimating the demand as CFE indicated, but rather underestimating the demand. The demand of 6 in the fourteenth period does not have the same effect on the PIS performance of SES compared to the CFE performance of SES. The update of the forecast in every period that eventually will approach zero makes SES an overall underestimating method.

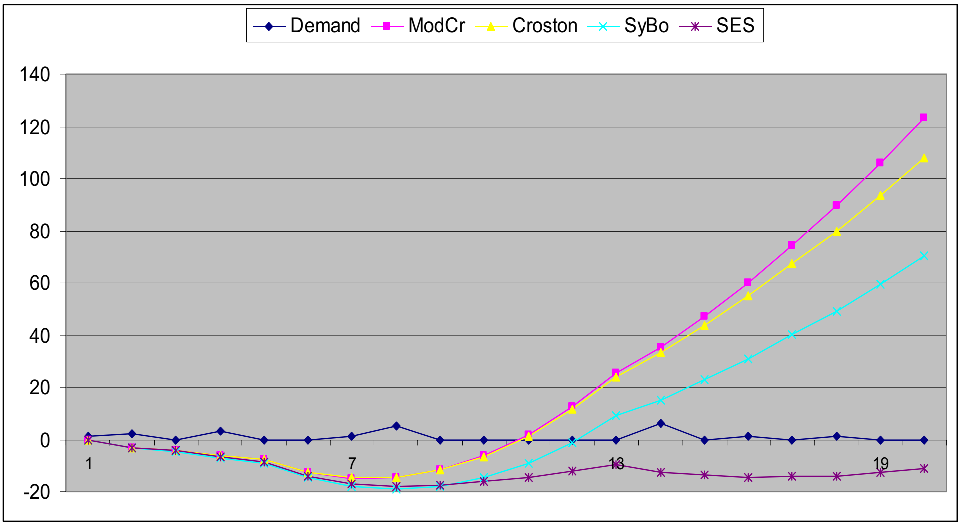


Figure 4.22   Demand and PIS performances of the four methods.

5 DISCUSSION

The fifth and final chapter discusses and summarises the findings of the experiments. Suitable forecast errors are identified among the variance errors (MAD, MSE, MADn and MSEn), bias errors (CFE, PIS, NOSp, CFE_{min} and CFE_{max}) and the symmetric error sMAPE. Suitable forecasting methods for forecasting intermittent demand are identified among the four evaluated methods, Croston, ModCr, SES and SyBo. Further the error dimensions in relation to the forecasting methods are discussed as well as the validity of the thesis and future research. The research questions are answered in the summary.

5.1 The Choice of Forecast Method and Forecast Errors

Which one of the tested methods is most suitable for intermittent demand? If the number of first places in Table 5.1 should serve as guide for the four errors; MSE, MAD, sMAPE and CFE, no method stands out as the single most suitable method. However one method is not represented in the table, Croston. In previous research by Syntetos and Boylan, (2005b) Croston should be an appropriate choice when the demand is considered smooth, which 18 of the 72 items in this study have. It might not be a sufficient number of items that have a smooth demand in order for Croston to perform well, which might explain the methods limited success in this study. The performance of Croston when MSE is considered is not far from the best methods when the relative quotients are examined.

Table 5.1 Summary of the method or methods with the highest number of first places for MSE, MAD, sMAPE and CFE

	Method with lowest error	Comments
MSE	SES and SyBo	SES better for 0.025 and 0.05
MAD	SyBo	
sMAPE	ModCr	
CFE	SES	

5.1.1 Forecast Errors

MSE

Another relevant question is; are the errors presented in the Table 5.1 reliable in an intermittent context? The answer is both yes and no. MSE do not have any collapsing behaviour in the form of steadily decreasing error with an increasing smoothing constant with the exception of SyBo. SyBo is the only method that has two items where the errors decrease when the smoothing constants increase, but with just 2 of 72 items it is not significant with a p-value of 0.01 when the binomial situation is assumed. Apart from this, only items with increasing errors as the smoothing constants increase could be found in sufficient numbers (8 or more) to form subgroups (here: items with increasing errors as the smoothing constant is increased). For the MSE subgroups, the significant variables are CV and MACs for demand and demand rate regardless of method. This makes sense because when the variation increases, knowledge of the mean is more valuable than a naïve forecast in order to increase the accuracy. The sizes of the mean are not significant. MSE is therefore considered a reliable measure.

MAD

The other measure of variance, MAD, had collapsing behaviour for every method with the exception of ModCr. SES had six items and was therefore not examined further. For Croston and SyBo 9 items decreased MAD when the smoothing constant was increased. The significant variables are identical; every descriptive statistic for demand should be low as well as the mean and standard deviation of demand rate while the mean and MACs of the inter-demand periods should be as high as possible. Under these circumstances MAD will be more distorted and a higher inter-demand will make MAD less reliable.

A higher inter-demand interval is what partly defines intermittent demand. From -25-s up to +25-s the only methods that have first places are the two methods that have a tendency to underestimate the demand, SES and SyBo. MAD is therefore not considered a measure that is suitable to use when the demand is intermittent which agrees with the findings of Teunter and Duncan (2009) who states that MAD favours underestimating methods which MSE also does but not to the same degree.

Furthermore the PCA loading plots of the errors revealed that MAD and MSE have similar loadings and therefore the two measures have approximately the same variability information, which do not make it necessary to use MAD. The higher sensitivity of outliers, larger errors, for MSE did not make any difference. The cut-off value of six standard deviations for each item removed the most extreme outliers from the used demand data.

sMAPE

sMAPE is the opposite measure compared to MAD. While MAD tends to benefit underestimating forecasting methods, sMAPE benefits the overestimating forecasting methods especially when the percentage of demand occasions is low. When the demand is high the underestimating forecasting methods usually have the lowest error. It is not what can be defined as a reliable measure for intermittent demand.

MADn and MSEn

MADn and MSEn were an attempt to decrease the tendency of zero forecasts when optimising in Excel. It proved to be the opposite. How bad MAD might be when optimising it does not favour a single forecast method as MADn and MSEn do with SES. The use of MADn and MSEn is not to be recommended.

CFE, PIS and NOSp

The bias measures (CFE, PIS and NOSp) work better when the combination of them are used instead of the use of a single measure. CFE can conceal the bias tendency when a time in point is considered. If the CFE value is low in absolute terms the sign do not reveal any bias information. A positive CFE (underestimating) might just be a random figure for a method that is overestimating the demand when the other measures are checked. If CFE is calculated with a limited number of observations, such as real data, it is questionable if CFE always offer reliable information without confirmation from other measures. According to CFE, SES performed best when it could not fulfil a single of the demand occasions which was also the fact for SyBo. The lowest CFE occurred when none of the demand occasions were met by the forecast. The low CFE is the result of fulfilling the demand afterwards the demand has occured which is not traceable by CFE.

The fast transient response, that makes CFE more sensitive to conceal underestimating tendencies, is slower for PIS and therefore some random high demand at the end of an evaluation period will not affect PIS as much. Therefore CFE_{min} and CFE_{max} should be replaced by PIS_{min} and PIS_{max}.

However, even if PIS is more stable than CFE for Croston, ModCr and SyBo, it has the same problems as CFE with SES. SES can have the best performance when it could not fulfil any of the demand occasions. This makes NOSp a relevant measure since it is not sensitive for this type of collapsing. But to only use NOSp means that the majority of the information of CFE or PIS is lost. If two methods has the same amount of unmet demand occasions but one method has a lower PIS, it suggests that that method is more efficient. The sensitivity with SES are likely to decrease when the update of SES is done less frequent. With the frequent update, SES can meet demand afterwards which is not reflected in the value of PIS when SES has a higher smoothing constant.

In the parts of the thesis, when the biases of methods are discussed, the bias is more or less prominent depending on the start value. Just because that the bias is indicated for some of the start values does not imply that a change of the start value will reduce or make the bias vanish. But under certain circumstances a start value that is the opposite of the bias, for a certain forecasting method, will mask the bias. If the start value is to low in comparison to the mean-s and the forecasting method tends to overestimate the demand, then the method will still overestimate the demand in the long run. How long it takes depends on the smoothing value and the demand. If the first part is the first nine months and the second half is the nine last months, then the second part of the forecasting series will probably be overestimated but since the start value was low enough for an initial underestimation this will go on unnoticed. Monitoring the tracking signal would take care of this matter, although it assumes a distribution of known proportions. An alternative is to monitor the development of the errors during the whole time series instead of just the last point.

Suitable Errors

The chosen errors based on the previous discussion are; MSE, CFE, PIS and NOSp. MSE as the sole variance measure because MAD can distort under certain circumstances and also the variability in the PCA are similar for MSE and MAD. CFE, PIS and NOSp as bias measures because the forecast errors measures to a certain extent different dimension of the bias. CFE summarises the errors, PIS integrates the errors and NOSp detect the number of shortages. In Table 5.2 is a summary.

Table 5.2 Recommendation regarding suitable errors for intermittent demand.

	Recommendation
MSE	Sufficient stable
MAD	Not suitable – Distorts under certain circumstances
sMAPE	Not suitable – Distorts under certain circumstances
MADn	Not suitable – Distorts under certain circumstances
MSEn	Not suitable – Distorts under certain circumstances
CFE	Suitable in combination with other bias errors
CFEmin	If used it should be replaced PISmin
CFEmax	If used it should be replaced PISmax
PIS	Suitable in combination with other bias errors
NOSp	Suitable in combination with other bias errors

5.1.2 Forecasting Methods

When a method will be chosen the costs that a forecast method causes should be considered. But since no cost information is available for the data set in this study from where the items where chosen, the decisions are based on the forecast errors. In the experiments only one-period-ahead forecasts have been used. If it is possible to have a lead time of one period the most suitable method is SES due to its low variance and it is the least biased method, at least when the lowest smoothing constants are concerned. It is not an ultimate solution since it is one of the most sensitive methods of the four regarding the choice of the smoothing constant. When naïve start values are used, SES is still good but not as good as when the other start values are considered. Even with a naïve start value the smoothing constant should be low. For some items the other methods have lower errors than SES. Therefore another method could be a better alternative than SES. However no analysis has been done of the forecasts and the items where the combination of variance and bias, MSE and CFE/PIS/NOSp, are evaluated.

If forecasts beyond one-period-ahead forecasts are needed it becomes more complicated to find a suitable method or methods. Considering the variance; a low error would reduce the safety stock as long as the safety stock is based on the error variance. A low error means a lower safety stock. For MSE it is SES and SyBo that has lowest errors. SES is better than SyBo when the smoothing constants are low.

However if relative quotients are considered, the differences between the methods when they performed best was not large. If the Max-Min quotients also are considered, Croston and SyBo have the smallest variation between the best and the worst case. The use of separate demand and inter-demand forecasts results in a more stable variance than for ModCr and SES. Both methods use one forecast each. The quotient is lower for SyBo than for Croston, 0.044 compared to 0.053. The lower the number is the less difference between the best or the worst forecast error in the relation to the median forecast error of a certain method. The quotient is an internal measure of a forecast method concerning the error stability.

If only the lower half of numbered value of the items is taken into account, SyBo is best from a stability perspective. But the value of the Max-Min quotient do not reveal any information how the error change when the smoothing constant changes. If the maximum or minimum value is an extreme point the quotient is not representative for the error change when the smoothing constant is changed. Therefore the curvature of the eight quotients is of interest. The quotient is formed between an error for a certain smoothing constant and the median of the eight errors. The start value is mean-s. For each smoothing constant and forecasting method, a mean quotient is calculated based on the 72 items.

In Figure 5.1 the four methods are plotted. The points in the lines represent the value of the quotients and therefore a lower value does not mean that a method has a lower MSE than the other methods. As with the Max-Min quotients SES and ModCr have the largest range between the lowest and highest error and the development of the error increases faster as the smoothing constants becomes larger. SyBo and Croston have a smaller difference between the smoothing constants and the changes between the smoothing constants are more constant and linear. A smoothing constant with a value between the used smoothing constants, for example 0.06, might be higher or lower than its neighbours, 0.05

and 0.075, but it is the tendency of the used smoothing constants that is important.

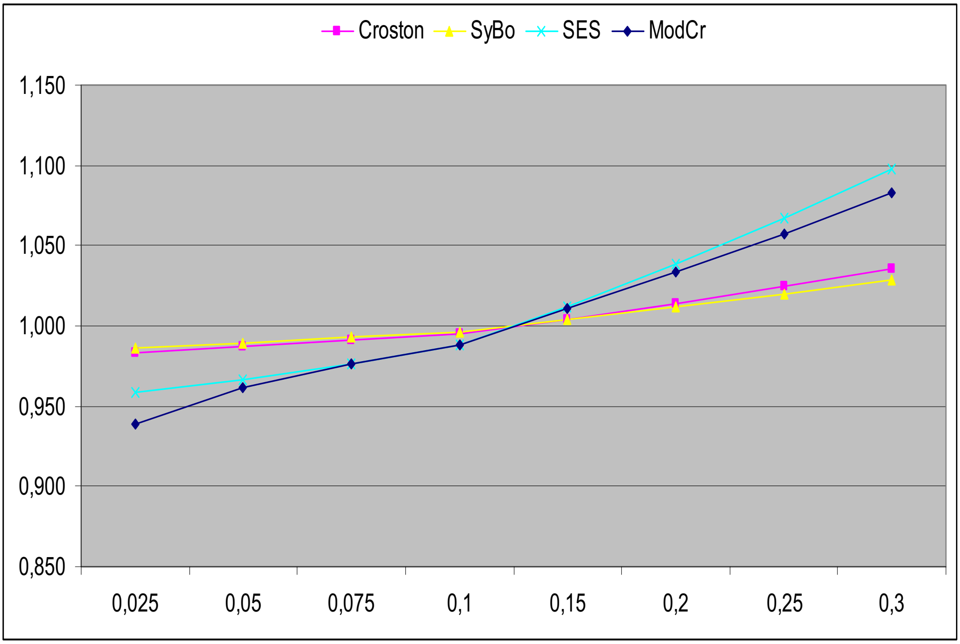


Figure 5.1 The mean of the quotient between error and median error for MSE. The lines are used for ease of reading not to represent the values between the smoothing constant. The smoothing constants have the same distance from each other regardless of distance.

If the most suitable smoothing constant may not be the chosen one, it does not affect the size of the error to the same degree for SyBo and Croston. Another advantage with a more stable error size is that the different smoothing constants are possible to use according to how sensitive to changes one wishes the forecast method to be without increasing the errors. With Croston or SyBo values larger than 0.2 are possible to use without a large increase of MSE, while for SES values larger than 0.1 increases the error rapidly.

None of the tested methods is free of bias. ModCr and Croston are overestimating the demand while SES and SyBo underestimate the demand. SES is the method with the least bias regardless of which of the three bias errors that has been used; CFE, PIS or NOSp. But since the lead times have been considered is 1 period. It is questionable if SES continues to have the same bias performance when the lead time is larger than 1. When the lead time increases it is not possible to use the one-period-ahead forecast that has been used throughout this evaluation, which forecast should be used? As Croston (1972) discussed, to use the forecast right after a demand will lead to overestimating the demand. If instead a mean of every forecast generated since the previous demand is used, the bias will decrease provided that there are enough forecasted period since the last demand. If it is one period then SES will be exposed to the same kind of bias that ModCr has. The other methods update the forecast only after a demand has occurred and therefore they do not have the similar problem with the choice of which forecast to use. ModCr has the most bias of all methods. Croston is not equally biased and has a stable MSE performance. SyBo has also a stable MSE performance but has the opposite bias.

What is better, overestimation or underestimation? It depends on the situation. Lee and Everett (1986) came to the conclusion that in a manufacturing situation an overestimation is preferred due to fewer setups and thereby reducing the cost. When stock is considered an underestimation might be better than an overestimation if the safety stock is based on the variance performance. MSE is a quadratic procedure that does not take the sign of the error into account. If a overestimating forecast method is used it is more likely that a greater part of the variance comes from the overestimating. If the variance is used to calculate the safety stock it increases the stock and the forecast is already overestimating the demand. This can be compared to an underestimating method where the variance to a greater part comes from underestimation. The underestimation plus the security stock do have a lower stocking cost compared to the overestimation.

Based on the previous discussion, if one method has to be chosen it is SyBo. It is not the best method for every situation, in particular when the inter-demand periods are shorter. When the percentage of demand occasions is larger than 55% SyBo has usually a NOSp of 90% or larger. When the inter-demand periods are shorter, i.e. when the percentage of demand occasions is larger, then other methods are better. Teunter and Sani (2009) found that SyBo had a slightly worse bias performance but a better MSE performance than the method of Syntetos. A reason why Syntetos and Boylan launched SyBo instead of the method of Syntetos may be the better inventory performance of SyBo that has the most stable variance of all the tested methods.

Croston is the method with the second best variance stability but has been rather anonymous throughout the experiments. One reason for that is what type of items that are suitable for Croston. Boylan and Johnston (1996) reported that SES is not suitable when the mean inter-demand is larger than 1.25 (demand occasions in 80% of the forecasting periods). Syntetos and Boylan (2005) calculated the theoretical cut-off value when SyBo is better than Croston. The adjusted theoretical cut-off value by Kostenko and Hyndman (2005) is 4/3 for the mean of the inter-demand period. The MSE is lower for SES with low smoothing constants and for higher smoothing constants Croston is better, but the bias is larger than for SES. Croston has a bias problem that is not always detectable with CFE but with PIS. Whether Croston or SES is better has not been established.

If the demand is stable SES with a low smoothing constant can work better than Croston which is the other choice instead of SyBo. That SES can work in an intermittent environment has been documented by Eaves (2002) that used smoothing constants from 0.01 to 0.1. The result of the smoothing constants 0.01-0.1 that was used in this experiment for SES is in agreement with Eaves findings. But with a smoothing value of approximately 0.1 the variance begins to increase more rapidly than for any other method. Teunter and Duncan (2009) states that a smoothing constant in the size of 0.1-0.2 has a minor influence; this can not be confirmed for SES.

The counterpart of SES concerning intermittent demand is ModCr. The idea behind ModCr is more appealing than its performance. ModCr proved to be the method with the largest bias. The bias is increasing the more intermittent the demand becomes which correlates to the relationship between the mean demand rate and the quotient of the mean demand and the mean inter-demand. In the present version of ModCr it is the least suitable method for intermittent demand even without the bias problem. There are signs for the need of low smoothing constants; the variance errors are usually higher, the Max-Min quotient has a high value (compared to Croston and SyBo) and in Figure 5.1 the slope of the line is steeper compared to SyBo and Croston. In order for ModCr to work with just one forecast the demand rate needs to be redefined.

5.1.3 Error Dimensions

Mathews and Diamantopoulos (1994) claimed that it was possible to find the ‘true’ underlying dimensions regarding a number of forecast errors and that the dimensions where not dependent on the forecasting methods. They considered the number of error dimension to be less than the numbers of errors since the calculation of the error are partly similar. The definition of forecast errors in their article also included what is considered descriptive statistics in this thesis. Their claim can not be confirmed in this thesis. There are differences among the different forecasting methods. There are differences among different start values for the same forecasting methods. There are differences among the smoothing constants for the same forecasting method and start value. But in general the relationships are more stable within a method than between methods.

The result of the PCA is also affected by the number of errors used. The PCA with 10 errors tend to have a more varied relationship among the errors than the PCA with three errors. The stability of the 'true' number of dimensions is also dependent of the errors that are used as well as the number of errors included in the PCA. In Table 5.3 is a summary of the stability among the forecasting methods based on the PCA with the three errors (MAD, MSE and CFE). ModCr and SyBo have the most stable relationship that change only to a minor degree when the start value and/or smoothing constants are changed. Croston has a relationship that differs among the errors when the start value is -25-s or +25-s. The relationship also varies depending on the smoothing constant where 0.025 is very different. The method with the most unstable relationship is SES the relationships among the errors are affected to a higher degree. SES is the only method that updates the forecast in every time period.

Table 5.3 Stability of the error relationship among the forecasting methods

	Start value	*Smoothing constant*
Croston	-25-s not stable	Especially 0.025 -25-s
ModCr	Stable	Stable
SES	Not stable	Not stable
SyBo	Stable	Stable

In the study of Mathews and Diamantopoulos (1994) the first factor (the counterpart of component in PCA) constitutes of percentage errors, a ratio error and the CV (coefficient of variance), the second factor constitutes of variance errors and the third factor is a mixture of bias errors (CFE and tracking signal). Since the variables in the first factor never were a part of the PCA it not possible to draw any conclusions regarding the amount of variability the PCA could have explained in the experiments.

However if one regards the sequence of similar components and factors there are similarities; most of the variability across the items comes from variance measures (MAD, MSE) while the bias measures accounts for less of the variability. The exceptions for the four methods from variance component first and bias component second, are for SyBo and the other methods when they underestimate the demand. When the forecast is overestimating the demand the correlation between variance and bias errors are much smaller than in an underestimating situation, where the correlation between variance and bias is larger.

Besides the different appearance of the PCA depending of forecasting methods, start values, smoothing constants and the type and number of the included forecast errors, the type of items that is forecasted is also influential. In Figure 5.2 is a loading plot of SyBo when only the appropriate items, for the method, have been used according to Syntetos and Boylan (2005b). Instead of including the descriptive statistics as Mathew and Diamantopoulos (1994) did. It would be more informative to test PCA on subgroups of the data set to trace differences, if the data is possible to divide into sufficient large subgroups.

The relationships between different errors vary depending on the forecast method, start value and smoothing constant. The relationships are also present between variance measures and bias measures. It is therefore not possible to use CFE as a scale independent measure without examine the relationship with the scale dependent measures or descriptive statistics. If it is scale independent the correlation should be low. However both CFE and PIS are correlated with MSE even without the outliers. If only the items that Syntetos and Boylan (2005b) classified as intermittent the correlation for SyBo between MSE and CFE or PIS decreases while the correlation increases for ModCr. The correlation between the number of occasions, sum of demand and CFE or PIS is also present in various degrees for the different methods.

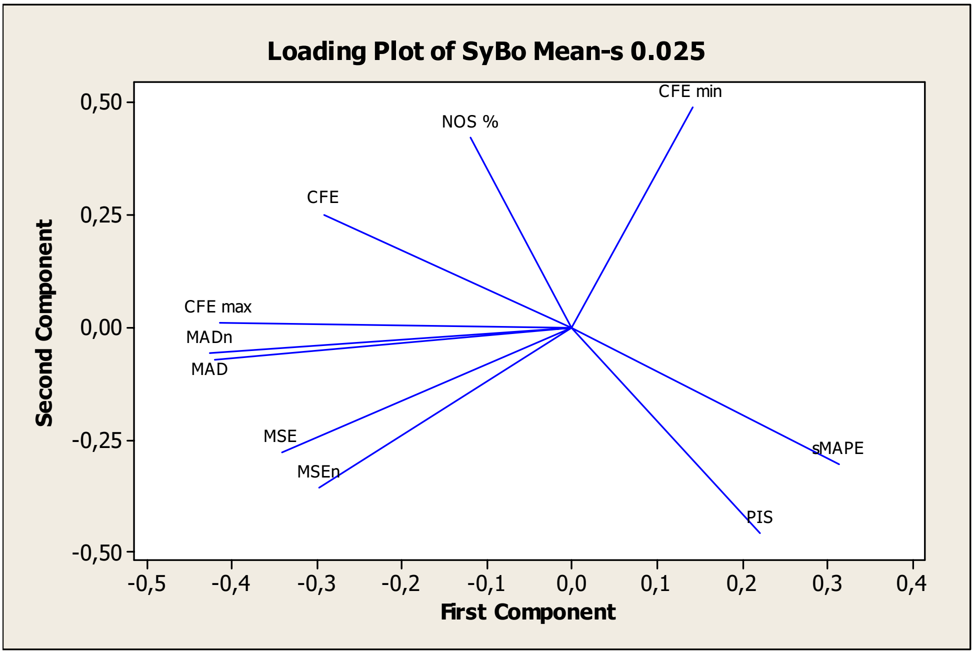


Figure 5.2 Loading plot of the 10 errors where only items that are considered appropriate for the method without the outliers that was identified when the PCA was based initially based on every item.

5.2 Validity

The analysis and conclusions are based on 72 items. Are 72 items enough? One advantage is that it is easier to have an overview and study an individual forecast to increase the understanding of what might cause certain behaviour for a method or an error. Earlier findings like the bias of Croston (Syntetos and Boylan, 2001) or the bias of ModCr (Syntetos and Boylan, 2007; Teunter and Sani, 2009) can be confirmed with the data material. However the validity could have been increased with another type of data for example simulation, especially when the new errors are concerned. The findings that disagree with earlier research are still relevant.

The claim from Mathews and Diamantopoulos (1994) that the 'true' underlying dimensions for a numerous forecast errors and descriptive statistics are the same regardless of forecast method can not be confirmed. Does that mean that the data for this thesis is not reliable? No, it has more to do with the different situations. They based their assumption on partly different errors, no intermittent demand and two forecast methods. A totally different situation compared to the experiments in this thesis.

However the limited number of items does affect the resolution of the conclusions. The logistic regression is based of mainly mean demands lower than 10. A higher number or at least more items where the demand is both intermittent and has a higher mean demand would have been high on the wish list. The creation of subgroups to further investigate the possibility of different characteristics for the different types of errors is limited.

The data has two subgroups when the classification of Syntetos and Boylan (2005b) is used. Two out of four possible groups are used and the size of the groups are 18 (smooth) and 54 (intermittent). To a certain degree the analysis could have been improved if the subgroups where large enough. The PCA of SyBo is not exactly the same when only the non-outliers from the subgroup intermittent are included, see Figure 5.2. CFE is not as close to CFE_{max} compared to the PCA of every non-outlier item, but the similarities are larger than the dissimilarities.

One might argue that eighteen months of data is not enough to carry out the experiments and evaluations. Perhaps, if this had been a simulation study but this is not. Just because it makes statistically sense to use as many periods as possible in an analysis it is not always the best approach from a manager's point of view or even possible. Over the years a product may have altered demand pattern (competitors, change of phase in the products lifecycle, new technology) and only the most recent years, or even just the most recent year; are relevant to the current situation. Another item may have a lifespan much shorter than a simulation. The item with lowest number of demand occasions needs 357 years to reach 10 000 demand occasions. Travelling back 357 years in time from the year 2009 results in the year 1652, that year Isaac Newton was nine years old.

Regarding the theoretical framework, the work of different authors that are published in different journals has been used. However in the field of intermittent demand forecasting, forecast accuracy and forecast evaluation some authors are more frequent than others among the articles to choose from. The choice has been made from what is considered relevant in an article for this thesis not the reputation of the authors.

The different methods has when possible been used to partly portray the same phenomenon from different angles. To use number of first placements and relative quotients is a way of looking at the forecasting methods from different angles. These two methods are examples of the external comparison methods while Max-Min quotients are more of an internal comparison. Probably the least triangulation of the methods is the combination of PCA and correlation since the two methods are both measures of linear dependence. But from an interpretation perspective they offer partly different information. Whether the PCA uses some kind of leave-one-out procedure or cross correlation that decreases the risk of over-fitting is not known. Since the correlation decreased when the influential outliers were removed all PCA and correlations were made without the outliers to decrease the chance of over-interpret the results.

A limitation is the use of only one smoothing constant for the inter-demand period. Since the smoothing constant of inter-demand is fairly high compared to the used smoothing constant for demand and demand rate, an alternative could help to reveal how much increased precision SyBo and Croston could have. It is most likely that SES has benefited on the high smoothing constant for inter-demand.

Another factor that might have been in favour of SES is the generally low mean; 60 items have a mean demand less than 5 while 5 items have a mean demand larger than 10. The identified outliers in the scatterplots were the items with mean demand larger than 10. By removing the 5 outliers the correlations between the errors decreased which was also the case for the correlation between the descriptive statistics.

The logistic regression was used with the outliers present. If the outliers are removed, the significant variables generally become stronger and previously non-significant variables become significant, the opposite of the situation concerning the correlation and PCA. The variables that became significant was ‘borderline significant’ when the outliers where included. In Table 5.4 is a summary of the additional significant variables. The most common new variables are the mean and standard deviation.

Table 5.4 Summary of additional significant variables for the logistic regression without the identified outliers from the PCA.

MSE	*Type*	*Quantity*	*DO%*	*Demand*	*Inter-demand*	*Demand rate*
ModCr	Increase	54		CV -High		
SES	Increase	59			MACs - High	

MAD	*Type*	*Quantity*	*DO%*	*Demand*	*Inter-demand*	*Demand rate*
ModCr	Increase	43		CV -High MACs - High		
Croston	Increase	27		Mean - High Std - High		Mean - High Std - High
SyBo	Increase	28		Std - High		Mean - High Std - High
SES	Increase	36				Std- High

MSEn	*Type*	*Quantity*	*DO%*	*Demand*	*Inter-demand*	*Demand rate*
ModCr	Increase	43		Std - High		
Croston	Increase	29				Mean - High Std - High
SyBo	Increase	31	High	Std - High		Mean - High Std - High CV -High
SES	Decrease	31	Low	Mean - Low		Mean - Low
SES	Increase	15		Mean - High		Mean - High

MADn	*Type*	*Quantity*	*DO%*	*Demand*	*Inter-demand*	*Demand rate*
Croston	Increase	29				Std - High
SyBo	Increase	33				Std - High
SES	Decrease	34			CV - High	
SES	Increase	15		Mean - High		Mean - High

There are also significant variables that are no longer significant when the outliers are removed. There are two variables for SES when MADn is the error. The summary is in Table 5.5.

Table 5.5 Summary of no longer significant variables for the logistic regression without the identified outliers from the PCA

MADn	*Type*	*Quantity*	*DO%*	*Demand*	*Inter-demand*	*Demand rate*
SES	Decrease	34		Std -High		
SES	Increase	15			CV - High	

5.3 Summary

In this summary the research questions (in italics) from chapter one are answered.

How is bias measures affected by the short series that real demand usually are?

For individual series, the value of the last period for CFE can be better than what is representative during the whole forecast period and a method can be considered nonbiased despite a bias. The problem with CFE and bias is when the CFE value is low. PIS and NOSp are valuable to prevent misinterpretations of CFE. If there are several thousand items with similar descriptive statistics that points in one direction concerning a CFE tendency, then numbers makes CFE more trustworthy.

To what degree is the relationship between errors for a forecasting method unique compared to other forecasting methods?

The bias of the tested methods influences to a high degree the relationship among the tested errors. ModCr and Croston are both overestimating methods and there respective PCA are more similar than SES and SyBo. The combination of start value and smoothing constant could also affect this.

Are there common error dimensions among the forecasting methods and how are the error dimensions structured for the forecasting methods?

The common structure among all the tested methods is not equal. SyBo is the exception this is due to the underestimating bias. The separation of a variance dimension and a bias dimension is lacking. The other methods have a variance component that explains the most of the variability and a bias component in common. However the relationships between the errors vary.

What dimension should a forecast error cover? Is it possible to have all the dimensions covered in one error so the evaluation of forecasting methods can be done by using one measure alone? If so which measure?

There is no overall best measure. There are two main reasons for this; the dimension of an error and the distortion of an error in combination with a forecasting method. To evaluate an error there are at least four dimension that are of interest (in no particular order); variance, bias, service and stock consequences, and cost. The traditional measures used in practice are either variance or bias measures and they are not equal to service and stock implications or cost and vice versa. The second reason is that under certain circumstances the errors did favour a certain type of bias. MAD gave lower error for an underestimating forecast method. Even if MSE did not do that the combination of certain characteristics of a time series and forecast method may cause a MSE performance that is not trustworthy. This applies to other types of forecast errors as well.

How robust are the forecasting methods considering different errors smoothing constants and start values?

The split forecasting methods, SyBo and Croston proved to be the most stable methods that were least sensitive for the choice of smoothing constant. If the 'right' start value is hard to predict from the start, the methods can with a slightly higher smoothing constant find the mean compared to the SES and ModCr that needs low values of the smoothing constants. This implies that SES can be used when the mean is known, but if the mean is known, why forecast?

5.4 Main Contributions

The main contributions can be found in two areas, forecast errors and the evaluation methods.

PIS allows a less transient sensitive measurement of the bias and are therefore usually a more dependent measure of bias as the aspect of when the forecast error arises is also measured, especially when the time series is limited and there are only a few time series to evaluate. Even if the name implies that it is a measure related to an inventory situation the measure can be used in other situations when inventory is not the issue.

An extension to PIS is to measure the mean stock for a forecasting method; mean forecasted stock. This might be more valuable in practice since the measure has the same dimension as the real mean stock. The mean forecasted stock is the quotient between PIS in period *t* and the number of periods *t*, see equation 5.1. This measure can be of greater value to the practitioner then both CFE and PIS since it has the same dimension as the measure of stock has in an inventory control situation.

$$\text{Mean forecasted stock}=\frac{PIS_t}{t} \tag{5.1}$$

The example in 4.14 (Comparison between the Forecasting Methods with CFE and PIS) with the 'mean forecasted stock' instead of PIS the interpretations concerning the different forecasting methods are still the same, see Figure 5.3. However the dimension differs. Just like other types of forecast errors it is not necessary to measure PIS from the start period, it just as possible to use a certain number of the latest periods.

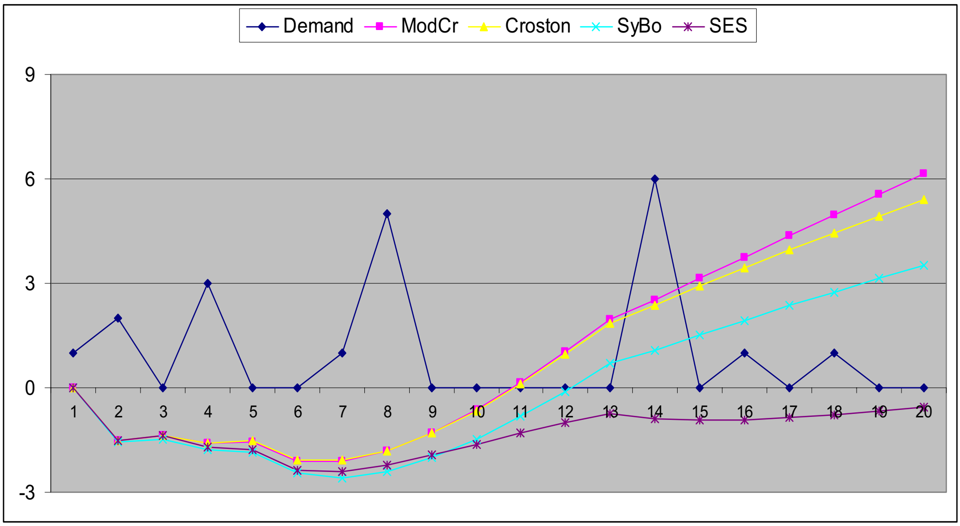


Figure 5.3 Demand and 'Mean forecasted stock' of the four methods based on the example in 4.14 Comparison between the Forecasting Methods with CFE and PIS.

When the adjustment of a forecast method is considered the focus generally lies in finding the setting for the smoothing parameter that delivers the highest accuracy. However in a practical situation it is far from certain that an optimisation takes place, the default setting in the software may be used or just a few alternatives for the smoothing constant. In that type of situation the stability of a forecasting method becomes very important and therefore it is important to evaluate a forecasting methods error performance in relation to the sensitivity of the proper smoothing constant. The Max-Min quotient is such a method along with the version where every quotient between the forecast errors for a certain smoothing constant and the median of the eight smoothing constants for an item was considered, see the text relevant to Figure 5.1.

The use of multivariate methods can improve the understanding of how a forecast error responds to a certain forecasting method. When a forecast error is measured it is not the forecast method that is measured but the combination of the forecast method and the responds of forecast error. The logistic regression is a valuable tool to detect situations where the forecast errors is not suitable for a certain type of forecasting method and/or a certain situation caused by the appearance of the time series.

5.5 Suggestions for Future Research

When an evaluation is done the focus are generally on the forecasting methods but this thesis shows that an evaluation based a certain error or errors can be of limited use since the knowledge of the methods is based on errors that might be distorted, therefore research that can identify when a certain error can be used and when it can not be used is relevant. In this thesis logistic regression has been used when the errors are increasing or decreasing in relation to the changes of the smoothing constants. Teunter and Duncan (2009) used a zero demand to find out how different errors behave.

Of the additional measures, PIS is the measure that provided the most information of the additional forecast errors since it adds another dimension, time, to the measure of error. The sum of PIS indicates if the forecasting method is biased if the value differs significantly from zero, but what is significantly from zero translated in numbers? A further research could be to examine the relationship between PIS and the tracking signal. Also how the error distributions are related to PIS is a possible future research.

The present definition of demand rate makes ModCr biased and the bias increases when the inter-demand interval increases. If the main idea behind ModCr, a single forecast, should remain unaltered; how should the demand rate be defined so that the overestimating bias property will be reduced?

The choice of forecasting method should not only be based on forecast errors but also on the consequences for the organisation the chosen method or methods have. The choice should reflect the organisation’s strategies. However it is easer to write such an aim than to implement one in practice. In logistics there is a tendency to always minimise the cost instead of increasing the profit. Sometimes the minimisation creates a sub-optimisation, either in the own organisation or in the supply chain. To avoid this scenario a relevant question is; what are the relevant parameters when an organisation’s strategies also should be reflected?

References

Adler, K., (2003) *Världens mått: Berättelsen om hur metersystemet förändrade världen* (Original title: *The Measure of All Things: The Seven-Year Odyssey and Hidden Error That Transformed the World*). Norstedts Förlag, Stockholm.

Armstrong, J.S., Collopy, F., (1992) Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8, 69-80

Bayraktara, E., Lenny Kohb, S.C., Gunasekaran, A., Sari, K., Tatoglu, E., The role of forecasting on bullwhip effect for E-SCM applications. *International Journal of Production Economics*, 113, 193-204

Boylan, J.E., Syntetos, A.A,. The accuracy of a Modified Croston procedure. *International Journal of Production Economics*, 107, 511-517

Brown, R. G., (1959) *Statistical forecasting for inventory control*. McGraw-Hill, New York.

Carbone, R., Armstrong J.S., (1982) Evaluation of Extrapolative Forecasting Methods: Results of a Survey of Academicians and Practitioners. *Journal of Forecasting*, 1, 215-217

Chatfield, C., (1988) Apples, Oranges and Mean Square Error. *International Journal of Forecasting*, 17, 515-518

Chatfield, C., (2001) Is ‘bigger’ necessarily ‘better’? *International Journal of Forecasting*,17, 537–584

Clements, M.P., Hendry, D.F. Explaining the results of the M3 forecasting competition. , *International Journal of Forecasting*, 17, 537-584

Croston, J.D., (1972) Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23, 289-303

Dekker, M., van Donselaar, K., Ouwehand, P., (2004), *How to use aggregation and combined forecasting to improve seasonal demand forecasts,* International journal of production economics, No 90 (2004) 151-167.

Denzin N.K., Lincoln Y.S., (2000) *Handbook of Qualitative Research, Second Edition*. Thousand Oaks, CA.

Eriksson, L., Wiedersheim-Paul, F., (2001) *Att utreda, forska och rapportera*, Liber ekonomi, Malmö.

Eaves, A.H.C, Kingsman, B.G., (2004) Forecasting for the Ordering and Stock-Holding of Spare Parts. The Journal of the Operational Research Society, 55, 431-437

Ejvegård, R., (1996) *Vetenskaplig metod*. Studentlitteratur, Lund.

Fildes, R., Bretschneider, S., Collopy, F., Lawrence M., (2003) Researching Sales Forecasting Practice – Commentaries and authors' response on ''Conducting a Sales Forecasting Audit'' by M.A. Moon, J.T. Mentzer & C.D. Smith. *International Journal of Forecasting*, 19, 27–42

Fildes, R., Goodwin, P., Lawrence, M., Nikolopoulos K., (2009) Effective forecasting and judgemental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25, 3-23

Fildes, R., Hibon M., Makridakis, S., Meade, N., (1998) Generalising about univariate forecasting methods: further empirical evidence. *International Journal of Forecasting*, 14, 339-358

Fildes, R., Makridakis, S., (1988) Forecasting and loss functions. *International Journal of Forecasting*, 4, 545-550

Fildes, R., Makridakis, S., (1995) The Impact of Empirical Accuracy Studies on Time Series Analysis and Forecasting. *International Statistical Review*, Vol. 63, No. 3, 289-308

Forsman, B., (1997) *Forskningsetik – en introduktion*. Studentlitteratur, Lund.

Gardner, E. S., (2006) Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, Volume 22, Issue 4, 637-666

Goodwin, P., Lawton, R., (1999) On the asymmetry of the symmetric MAPE. *International Journal of Forecasting*, 15, 405-408

Gorard, S., (2003) *Quantitative Methods in Social Science*. Continuum, London.

Hamilton, L.C., *Regression with Graphics – A Second Course in Applied Statistics*. Duxbury Press, Belmont, CA.

Hellevik, O. (1984) *Forskningsmetoder i sociologi och statsvetenskap*. Natur och Kultur, Borås.

Holme, I.M., Solvang, B.K, (1997) *Forskningsmetodik – Om kvalitativa och kvantitativa metoder*. Studentlitteratur, Lund.

Hyndman, R.J., Koehler A.B., (2006) Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688

Johnsson, D., (1998), *Applied multivariate methods for data analysts,* Duxbury Press, Pacific Groove.

Johnston, F.R., Boylan, J.E., (1996) Forecasting for Items with Intermittent Demand. *The Journal of the Operational Research Society*, 47, 113-121

Johnston, F.R., Boylan, J.E., Shale, E.A., (2003). An examination of the size of orders from customers, their characterization and the implications for inventory control of slowmoving items. *Journal of the Operational Research Society,* 54,833–837.

Koehler, A.B. (2001) The asymmetry of the sAPE measure and other comments on the M3-Competition. *International Journal of Forecasting*, 17, 537-584

Kostenko, A., V., Hyndman R.J., Viewpoints (2006) A note on the categorization of demand patterns. *Journal of Operational Research Society*, 57, 1256-1257

Körner, S., Ek, L., Berg, S., (1984) *Deskriptiv statistik*. Studentlitteratur, Lund.

Lekvall, P., Wahlbin, C., (2003), *Information för marknadsföringsbeslut,* IHM förlag, Göteborg.

Lee, H.L., Padmanabhan, V., Whang, S., (1997) Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, Vol. 43, No. 4, 546-558

Lee, T.S., Everett, E.A. Jr., (1986) Forecasting Error Evaluation in Material Requirements Planning (MRP) Production-Inventory Systems. *Management Science*, 32, 1186-1205

Levén E., Segerstedt A., (2004) Inventory control with a modified Croston procedure and Erlang distribution. *International Journal of Production Economics*, 90, 361-367.

Lundahl, U., Skärvad P-H., (1992) *Utredningmetodik för samhällsvetare och ekonomer*. Studentlitteratur, Lund.

Mathews, B.P., Diamantopoulos, A., (1994) Towards a Taxonomy of Forecast Error Measures – A Factor-comparative Investigation of Forecast Error Dimensions. *Journal of Forecasting*, Vol.13, 409-416

Makridakis, S., (1986) The art and science of forecasting An assessment and future directions. *International Journal of Forecasting*, Volume 2, Issue 1, 1986, 15-39

Makridakis, S., (1993) Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9, 527-529

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R., (1982) The accuracy of extrapolative (timeseries) methods: The results of a forecasting competition. *Journal of Forecasting*, 1, 111-153.

Makridakis, S., Chatfield, C., Hibon M., Lawrence, M., Mills, T., Ord, K., Simmons, L., (1993) The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, Volume 9, Issue 1, 5-22

Makridakis, S., Hibon, M., (1991) Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy. *International Journal of Forecasting*, 7, 317-330

Makridakis, S., Hibon, M., (2000) The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, Volume 16, Issue 4, 451-476

Makridakis, S., Hibon, M., (2001) Response to the commentaries on 'The M3-Competition: results, conclusions and implications'. International Journal of Forecasting, 17, 537 –584

Makridakis, S., Hibon, M., Moser, C., (1979) Accuracy of Forecasting: An Empirical Investigation. *Journal of Royal Statistical Society. Series A*, Vol.142, 97-145

Makridakis, S., Wheelwright, S.C., (1989) *Forcasting Methods for Management.,* John Wiley & Sons, New York.

Makridakis, S., Wheelwright, S.C., Hyndman, R.J., (1998) *Forecasting, Methods and application 3rd edition.* John Wiley & Sons, Hoboken NJ.

Merriam, S.B., (1994) *Fallstudien som forskningsmetod*, Studentlitteratur, Lund.

Montgomery, D.C., (2005) *Introduction to Statistical Quality Control 5th edition,* Wiley & Sons. Hoboken, NJ.

Montgomery, D.C., Johnson, L.A., Gardiner, J.S. (1990) *Forecasting and time series analysis*. McGraw-Hill, New York.

Moon, M.A., Mentzer, J.T., Smith, C.D., (2003) Conducting a sales forecasting audit. *International Journal of Forecasting*, 19, 5-25

Ord, K., (2001) Commentaries on the M3-Competition An introduction, some comments and a scorecard. *International Journal of Forecasting*, 17, 537-584

Patel, R., Davidsson, B., (2003), *Forskningsmetodikens grunder – att planera, genomföra och rapportera en undersökning,* Studentlitteratur, Lund.

Rao, A.V., A comment on: Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, volume 24, 639-640

Rasmussen, R., (2004) On time series data and optimal parameters. *Omega*, 32, 111-120.

Remenyi, D., Williams, B., Money, A., Swartz E., (1998) *Doing Research in Business and Management*. Sage publications, London.

Schultz, C.R., (1987) Forecasting and Inventory Control for Sporadic Demand under Periodic Review. *The Journal of the Operational Research Society*, 38, 453-458

Sani, B. Kingsman, B.G., (1997) Selecting the Best Periodic Inventory Control and Demand Forecasting Methods for Low Demand Items. The Journal of the Operational Research Society, 48, 700-713

Segerstedt, A., (2000) *Forecasting slow-moving items and ordinary items - A modification of Croston's idea,* Working Paper, Industrial logistics, Luleå University of Technology,.

Shah, C., (1997) Model selection in univariate time series forecasting using discriminant analysis. *International Journal of Forecasting*, 13, 489-500.

Silver, E.A., (1981) Operations research in inventory management: A review and critique. *Operations Research*, 29, 628-645

Silver, E. A., Pyke, D. F. and R. Peterson, 1998. *Inventory Management and Production Planning (3rd Ed.)*. John Wiley & Sons, New York.

Snyder, R., (2002) Forecasting sales of slow and fast moving inventories. *European journal of operational research*, 140,684-699

Sohn, S.Y., Lim, M., (2008) The effect of forecasting and information sharing in SCM for multi-generation products. *European Journal of Operational Research,* 186, 276–287

Sprent, P., Smeeton, N.C., (2001) *Applied Nonparametric Statistical Methods, Third edition.* Chapman & Hall/CRC, Boca Raton, Florida.

Sucky, E., (2009) The bullwhip effect in supply chains – An overestimated problem? *Journal of Production Economics*, 118, 311–322

Syntetos, A.A., Boylan, J.E., (2001) On the bias of intermittent demand estimates. *International Journal of Production Economics,* 71, 457–466.

Syntetos, A.A., Boylan, J.E., (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21, 303–314.

Syntetos, A.A., Boylan, J.E., (2005b) On categorization of demand patterns. *Journal of Operational Research Society*, 56, 495-503

Syntetos, A.A., Boylan, J.E., (2006) On the stock control performance of intermittent demand estimators, *International Journal of Production Economics*, 103, 36-47.

Tashman, L., (2001) The M3-Competition and forecasting software. *International Journal of Forecasting*, 17, 537-584

Tersine, R.J., Principles of Inventory and Materials Management. Elsevier Science Publishing Co, New York.

Teunter, R.H., Duncan, L., (2009) Forecasting intermittent demand: a comparative study. *Journal of Operational Research Society*, 60, 321-329

Teunter, R., Sani, B., (2009) On the bias of Croston's forecasting method. *European Journal of Operational Research*, 194, 177–183

Vokurka, R.J., Flores, B.E., Pearce, S.L., Automatic feature identification and graphical support in rule-based forecasting: a comparison. *International Journal of Forecasting*, 12, 495-512

Wallström, P., (2006) *Prognoser vid Ahlsell*. Master thesis 2006:188, ISSN 1402-1617 / ISRN: LTU-EX-06/188-SE

Weisberg, S., (2005) *Applied Linear Regression, Third Edition*. John Wiley & Sons, Hoboken, NJ.

Willemain, T.R., Smart, C.N., Schwartz H.F., (2004) A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting, 20,* 375-387

Willemain, T.R., Smart, C.N., Shockor, J.H., DeSautels, P.A., (1994) Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting, 10,* 529-538

Yin, R.K., (1994), *Case study research: design and methods,* Sage publications, Thousand Oaks, CA.

Yokum, J.T., Armstrong, J.S., (1995) Beyond accuracy: Comparisons of criteria used to select forecasting methods. *International Journal of Forecasting*, 11, 591-597

Zhao, X., Xie, J., Leung, J., (2002) The impact of forecasting model selection on the value of information sharing in a supply chain. *European Journal of Operational Research*, 142, 321–344

Internet Sources

Famous Forecasting Quotes - School of Engineering, Computing and Mathematics [SECaM]. Exeter University. (6th April 2009):
(http://www1.secam.ex.ac.uk/famous-forecasting-quotes.dhtml)

Demand, Items 1-39

	No DO	Mean	Median	Std	CV	MACs
Item 1	42	1,05	1	0,22	0,21	0,09
Item 2	46	1,91	1	1,91	1,00	0,67
Item 3	47	1,04	1	0,20	0,20	0,08
Item 4	50	1,74	1	1,87	1,08	0,82
Item 5	51	1,31	1	0,55	0,42	0,33
Item 6	53	1,19	1	0,44	0,37	0,26
Item 7	54	1,28	1	0,60	0,47	0,35
Item 8	58	3,14	1	5,66	1,80	1,25
Item 9	58	5,79	4	5,57	0,96	1,03
Item 10	63	1,97	2	0,90	0,46	0,48
Item 11	63	5,54	2	8,23	1,49	1,21
Item 12	72	6,17	1	11,86	1,92	1,46
Item 13	73	1,16	1	0,47	0,41	0,24
Item 14	75	1,71	1	1,91	1,12	0,73
Item 15	80	2,39	1	2,09	0,88	0,80
Item 16	82	1,26	1	0,49	0,39	0,31
Item 17	84	1,14	1	0,35	0,31	0,23
Item 18	87	1,77	1	2,19	1,24	0,83
Item 19	89	1,17	1	0,43	0,37	0,23
Item 20	96	1,36	1	0,86	0,63	0,46
Item 21	96	1,43	1	1,01	0,71	0,52
Item 22	98	1,59	1	1,05	0,66	0,60
Item 23	100	1,24	1	0,64	0,51	0,34
Item 24	100	1,81	1	1,86	1,03	0,73
Item 25	104	3,12	2	2,90	0,93	0,79
Item 26	105	4,86	2	5,75	1,18	1,16
Item 27	110	1,18	1	0,45	0,38	0,24
Item 28	111	1,25	1	0,56	0,45	0,35
Item 29	119	2,61	1	3,49	1,34	0,97
Item 30	130	3,21	1	5,90	1,84	1,16
Item 31	139	3,68	2	5,85	1,59	1,08
Item 32	137	2,26	2	1,98	0,87	0,73
Item 33	152	1,47	1	0,79	0,54	0,47
Item 34	154	4,66	2	10,13	2,17	1,39
Item 35	161	2,40	2	2,33	0,97	0,80
Item 36	166	2,90	2	2,27	0,78	0,72
Item 37	175	1,46	1	0,83	0,57	0,50
Item 38	177	3,86	2	3,42	0,89	0,90
Item 39	181	2,25	2	1,70	0,76	0,71

Demand, Items 40-72

	No DO	Mean	Median	Std	CV	MACs
Item 40	187	16,47	10	11,66	0,71	0,59
Item 41	199	21,31	10	20,80	0,98	0,77
Item 42	202	1,82	1	1,73	0,95	0,60
Item 43	208	1,88	1	1,35	0,72	0,63
Item 44	212	1,72	1	1,12	0,65	0,55
Item 45	218	1,68	1	0,96	0,57	0,57
Item 46	230	2,56	1	3,69	1,44	0,96
Item 47	239	3,20	2	3,79	1,18	1,09
Item 48	248	2,25	2	2,27	1,01	0,84
Item 49	248	1,93	1	1,37	0,71	0,63
Item 50	262	1,83	2	1,04	0,57	0,56
Item 51	263	4,00	2	5,43	1,36	0,94
Item 52	272	2,46	2	2,35	0,96	0,80
Item 53	279	23,98	20	16,32	0,68	0,69
Item 54	289	3,65	2	4,35	1,19	1,00
Item 55	299	3,08	2	2,88	0,94	0,91
Item 56	305	7,05	2	12,09	1,72	1,42
Item 57	308	2,56	2	1,73	0,68	0,67
Item 58	316	2,79	2	2,32	0,83	0,79
Item 59	320	2,47	2	1,56	0,63	0,64
Item 60	331	5,63	3	7,30	1,29	1,05
Item 61	334	41,66	20	46,26	1,11	1,08
Item 62	336	3,83	3	3,20	0,84	0,83
Item 63	339	4,64	3	6,63	1,43	1,07
Item 64	348	4,29	3	4,55	1,06	0,88
Item 65	357	4,50	3	4,07	0,91	0,88
Item 66	355	3,72	3	3,40	0,91	0,77
Item 67	364	7,04	6	5,91	0,84	0,72
Item 68	368	3,87	3	2,55	0,66	0,65
Item 69	371	13,23	6	17,61	1,33	1,21
Item 70	380	7,18	4	12,29	1,71	1,15
Item 71	387	3,83	3	2,39	0,62	0,63
Item 72	391	4,09	4	2,52	0,62	0,63

The descriptive statistics are based on every demand occasion including the outlier demand occasions.

Inter-demand interval, Items 1-39

	Mean	Median	Std	CV	MACs
Item 1	8,50	7	6,49	0,76	0,82
Item 2	8,22	5	9,65	1,17	1,11
Item 3	7,98	6	6,26	0,78	0,89
Item 4	7,70	6	5,82	0,76	0,83
Item 5	7,45	4	7,09	0,95	1,12
Item 6	7,06	4	8,20	1,16	1,02
Item 7	7,07	4,5	10,83	1,53	1,12
Item 8	6,57	5	6,32	0,96	0,93
Item 9	6,62	4,5	7,31	1,10	0,94
Item 10	6,21	4	7,16	1,15	1,01
Item 11	5,95	4	6,40	1,08	0,95
Item 12	5,32	2	12,11	2,28	0,97
Item 13	5,29	4	4,35	0,82	0,92
Item 14	4,99	3	4,71	0,94	0,85
Item 15	4,78	3	4,77	1,00	0,94
Item 16	4,60	4	3,54	0,77	0,85
Item 17	4,55	3	5,78	1,27	1,18
Item 18	4,40	3	4,01	0,91	0,88
Item 19	4,29	3	4,53	1,05	0,99
Item 20	3,92	2	3,97	1,01	0,86
Item 21	3,98	2,5	3,99	1,00	0,99
Item 22	3,92	2	3,72	0,95	0,91
Item 23	3,84	3	2,91	0,76	0,69
Item 24	3,83	3	4,54	1,18	0,92
Item 25	3,66	2,5	3,06	0,84	0,78
Item 26	3,66	3	3,36	0,92	0,89
Item 27	3,48	2	3,33	0,96	0,85
Item 28	3,46	2	3,01	0,87	0,78
Item 29	3,25	2	2,83	0,87	0,86
Item 30	2,95	2	2,69	0,91	0,94
Item 31	2,75	2	2,41	0,88	0,80
Item 32	2,79	2	2,13	0,77	0,74
Item 33	2,57	2	1,91	0,74	0,68
Item 34	2,53	2	1,95	0,77	0,70
Item 35	2,42	2	2,10	0,87	0,67
Item 36	2,35	2	1,77	0,75	0,63
Item 37	2,23	2	1,74	0,78	0,78
Item 38	2,17	2	1,58	0,73	0,61
Item 39	2,17	1	1,84	0,85	0,65

Inter-demand interval, Items 40-72

	Mean	Median	Std	CV	MACs
Item 40	2,12	1	1,65	0,78	0,70
Item 41	1,95	1	1,56	0,80	0,67
Item 42	1,90	1	1,30	0,69	0,63
Item 43	1,85	1	1,69	0,91	0,63
Item 44	1,81	1	1,48	0,82	0,63
Item 45	1,76	1	1,18	0,67	0,58
Item 46	1,70	1	1,04	0,61	0,55
Item 47	1,62	1	1,06	0,66	0,56
Item 48	1,61	1	1,07	0,66	0,57
Item 49	1,59	1	1,16	0,73	0,57
Item 50	1,48	1	0,84	0,57	0,45
Item 51	1,48	1	0,93	0,63	0,51
Item 52	1,46	1	0,87	0,60	0,52
Item 53	1,38	1	0,81	0,59	0,46
Item 54	1,38	1	0,73	0,53	0,45
Item 55	1,30	1	0,64	0,49	0,38
Item 56	1,28	1	0,60	0,47	0,39
Item 57	1,31	1	0,68	0,52	0,37
Item 58	1,24	1	0,66	0,53	0,31
Item 59	1,22	1	0,52	0,43	0,29
Item 60	1,18	1	0,64	0,54	0,29
Item 61	1,22	1	0,50	0,41	0,31
Item 62	1,19	1	0,52	0,44	0,28
Item 63	1,18	1	0,52	0,44	0,26
Item 64	1,17	1	0,45	0,38	0,26
Item 65	1,12	1	0,55	0,49	0,21
Item 66	1,14	1	0,46	0,40	0,21
Item 67	1,13	1	0,39	0,35	0,19
Item 68	1,12	1	0,36	0,33	0,20
Item 69	1,11	1	0,39	0,35	0,18
Item 70	1,08	1	0,43	0,40	0,14
Item 71	1,07	1	0,26	0,24	0,12
Item 72	1,06	1	0,24	0,23	0,10

Demand Rate, Items 1-39

	Mean	Median	Std	CV	MACs
Item 1	0,30	0,14	0,39	2,70	1,12
Item 2	0,54	0,42	0,57	1,36	1,14
Item 3	0,28	0,17	0,35	2,10	1,05
Item 4	0,56	0,19	0,95	4,88	1,44
Item 5	0,47	0,25	0,51	2,03	1,04
Item 6	0,53	0,25	0,58	2,33	0,98
Item 7	0,45	0,33	0,51	1,53	0,91
Item 8	1,00	0,33	1,83	5,48	1,38
Item 9	1,82	0,95	2,82	2,96	1,22
Item 10	0,72	0,5	0,66	1,32	0,74
Item 11	1,87	0,67	3,95	5,93	1,45
Item 12	3,40	1	8,86	8,86	1,65
Item 13	0,42	0,33	0,38	1,13	0,99
Item 14	0,69	0,33	0,86	2,58	1,00
Item 15	1,10	0,5	1,43	2,86	1,10
Item 16	0,52	0,33	0,51	1,54	0,94
Item 17	0,60	0,33	0,52	1,56	0,92
Item 18	0,87	0,33	1,65	4,94	1,30
Item 19	0,60	0,5	0,54	1,09	0,89
Item 20	0,70	0,5	0,76	1,51	0,94
Item 21	0,74	0,5	0,72	1,44	1,04
Item 22	0,81	0,5	0,92	1,83	0,98
Item 23	0,58	0,33	0,63	1,90	0,79
Item 24	0,94	0,5	1,34	2,68	1,07
Item 25	1,44	0,9	1,72	1,91	1,03
Item 26	2,50	1	4,12	4,12	1,41
Item 27	0,63	0,5	0,54	1,07	0,75
Item 28	0,66	0,5	0,57	1,14	0,86
Item 29	1,40	0,5	2,31	4,62	1,28
Item 30	1,68	1	3,41	3,41	1,22
Item 31	2,50	1	5,41	5,41	1,28
Item 32	1,34	1	1,67	1,67	0,99
Item 33	0,89	0,67	0,79	1,19	0,84
Item 34	2,72	1	6,61	6,61	1,43
Item 35	1,59	1	1,84	1,84	1,00
Item 36	1,82	1,33	1,87	1,40	0,89
Item 37	0,97	1	0,72	0,72	0,82
Item 38	2,43	2	2,56	1,28	0,92
Item 39	1,54	1	1,45	1,45	0,85

Demand Rate, Items 40-72

	Mean	Median	Std	CV	MACs
Item 40	11,79	10	11,76	1,18	0,86
Item 41	15,30	10	17,62	1,76	0,88
Item 42	1,36	1	1,77	1,77	0,83
Item 43	1,44	1	1,25	1,25	0,73
Item 44	1,33	1	1,07	1,07	0,75
Item 45	1,26	1	0,93	0,93	0,79
Item 46	1,72	1	2,31	2,31	0,91
Item 47	2,40	1	3,05	3,05	1,15
Item 48	1,75	1	1,68	1,68	0,91
Item 49	1,55	1	1,25	1,25	0,76
Item 50	1,53	1	1,06	1,06	0,66
Item 51	3,24	2	4,49	2,24	0,99
Item 52	2,06	1	2,29	2,29	0,90
Item 53	20,42	20	14,95	0,75	0,72
Item 54	3,21	2	4,14	2,07	1,09
Item 55	2,69	2	2,81	1,41	0,96
Item 56	6,33	2	11,56	5,78	1,46
Item 57	2,27	2	1,77	0,89	0,74
Item 58	2,53	2	2,20	1,10	0,82
Item 59	2,24	2	1,55	0,78	0,66
Item 60	5,27	3	7,09	2,36	1,08
Item 61	38,45	20	46,02	2,30	1,13
Item 62	3,59	2,75	3,21	1,17	0,87
Item 63	4,21	2,33	5,88	2,52	1,07
Item 64	3,94	3	4,21	1,40	0,92
Item 65	4,26	3	3,92	1,31	0,87
Item 66	3,50	3	3,35	1,12	0,80
Item 67	6,69	6	5,81	0,97	0,73
Item 68	3,68	3	2,52	0,84	0,67
Item 69	12,82	5	17,57	3,51	1,22
Item 70	7,02	4	12,27	3,07	1,16
Item 71	3,71	3	2,40	0,80	0,64
Item 72	3,98	3	2,52	0,84	0,64

Relative Error Quotients

MSE

Table 1 The mean and median for the MSE quotients with -25-s as start value.

	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	1,038	1,063	1,077	1,087	1,103	1,118	1,134	1,149
ModCr/SyBo	1,037	1,065	1,081	1,092	1,110	1,128	1,145	1,163
ModCr/SES	1,044	1,066	1,073	1,076	1,076	1,073	1,069	1,064
Croston/SyBo	0,999	1,002	1,003	1,004	1,007	1,008	1,010	1,012
Croston/SES	1,006	1,002	0,997	0,990	0,976	0,960	0,944	0,928
SyBo/SES	1,007	1,000	0,993	0,986	0,969	0,952	0,934	0,916
				Median				
	0,025	0,050	0,075	0,100	0,150	0,200	0,250	0,300
ModCr/Croston	1,018	1,031	1,045	1,055	1,065	1,076	1,090	1,097
ModCr/SyBo	1,017	1,033	1,050	1,061	1,074	1,082	1,096	1,104
ModCr/SES	1,021	1,033	1,036	1,040	1,037	1,034	1,027	1,024
Croston/SyBo	1,000	1,002	1,004	1,005	1,007	1,009	1,010	1,011
Croston/SES	1,006	1,002	0,998	0,994	0,979	0,962	0,945	0,925
SyBo/SES	1,007	1,000	0,994	0,989	0,973	0,954	0,936	0,915

Table 2 The mean and median for the MSE quotients with +25-s as start value.

	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	1,046	1,069	1,081	1,090	1,106	1,120	1,135	1,151
ModCr/SyBo	1,056	1,077	1,089	1,098	1,115	1,131	1,148	1,165
ModCr/SES	1,063	1,077	1,082	1,082	1,080	1,076	1,071	1,066
Croston/SyBo	1,009	1,007	1,007	1,007	1,008	1,010	1,011	1,013
Croston/SES	1,016	1,007	1,000	0,992	0,977	0,961	0,945	0,928
SyBo/SES	1,007	1,000	0,993	0,986	0,969	0,952	0,934	0,916
				Median				
	0,025	0,050	0,075	0,100	0,150	0,200	0,250	0,300
ModCr/Croston	1,028	1,037	1,052	1,059	1,069	1,079	1,093	1,100
ModCr/SyBo	1,038	1,041	1,057	1,066	1,077	1,085	1,100	1,107
ModCr/SES	1,041	1,044	1,046	1,041	1,035	1,028	1,025	0,000
Croston/SyBo	1,008	1,007	1,006	1,007	1,009	1,010	1,011	1,012
Croston/SES	1,012	1,006	1,001	0,996	0,982	0,965	0,946	0,926
SyBo/SES	1,005	1,000	0,994	0,989	0,972	0,954	0,935	0,914

MAD

Table 3 *The mean and median for the MAD quotients with -25-s as start value.*

	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	1,122	1,156	1,170	1,178	1,185	1,190	1,194	1,197
ModCr/SyBo	1,161	1,198	1,214	1,222	1,231	1,236	1,241	1,245
ModCr/SES	1,132	1,176	1,193	1,202	1,208	1,211	1,212	1,212
Croston/SyBo	1,034	1,036	1,037	1,037	1,038	1,038	1,039	1,039
Croston/SES	1,009	1,016	1,018	1,019	1,017	1,015	1,012	1,009
SyBo/SES	0,976	0,981	0,983	0,982	0,980	0,977	0,974	0,971
				Median				
	0,025	0,050	0,075	0,100	0,150	0,200	0,250	0,300
ModCr/Croston	1,107	1,136	1,126	1,122	1,130	1,136	1,141	1,146
ModCr/SyBo	1,135	1,165	1,156	1,151	1,160	1,167	1,175	1,183
ModCr/SES	1,118	1,152	1,150	1,144	1,138	1,139	1,135	1,127
Croston/SyBo	1,037	1,039	1,039	1,039	1,040	1,040	1,040	1,040
Croston/SES	1,008	1,013	1,013	1,010	1,006	1,002	0,999	0,997
SyBo/SES	0,973	0,984	0,982	0,980	0,976	0,972	0,969	0,964

Table 4 *The mean and median for the MAD quotients with +25-s as start value.*

	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	1,113	1,150	1,166	1,174	1,183	1,188	1,192	1,196
ModCr/SyBo	1,159	1,196	1,212	1,220	1,229	1,235	1,240	1,244
ModCr/SES	1,168	1,199	1,210	1,214	1,217	1,217	1,217	1,216
Croston/SyBo	1,041	1,040	1,039	1,039	1,039	1,039	1,039	1,039
Croston/SES	1,048	1,040	1,035	1,031	1,026	1,021	1,017	1,013
SyBo/SES	1,007	1,000	0,996	0,993	0,987	0,983	0,979	0,975
				Median				
	0,025	0,050	0,075	0,100	0,150	0,200	0,250	0,300
ModCr/Croston	1,104	1,134	1,127	1,124	1,130	1,136	1,141	1,147
ModCr/SyBo	1,146	1,167	1,157	1,152	1,160	1,167	1,175	1,182
ModCr/SES	1,164	1,162	1,153	1,146	1,142	1,142	1,137	1,130
Croston/SyBo	1,043	1,041	1,041	1,041	1,041	1,041	1,041	1,041
Croston/SES	1,035	1,028	1,021	1,017	1,011	1,006	1,004	0,999
SyBo/SES	0,999	0,993	0,992	0,985	0,981	0,976	0,971	0,965

sMAPE

Table 5 The mean and median for the sMAPE quotients with -25-s as start value.

	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	0,983	0,980	0,979	0,978	0,978	0,978	0,979	0,979
ModCr/SyBo	0,979	0,976	0,975	0,974	0,974	0,974	0,974	0,974
ModCr/SES	0,984	0,977	0,973	0,969	0,965	0,961	0,957	0,954
Croston/SyBo	0,996	0,996	0,996	0,996	0,995	0,995	0,995	0,995
Croston/SES	1,002	0,997	0,994	0,991	0,986	0,982	0,978	0,974
SyBo/SES	1,006	1,001	0,998	0,995	0,991	0,987	0,983	0,979
				Median				
	0,025	0,05	0,075	0,1	0,15	0,2	0,25	0,3
ModCr/Croston	0,983	0,982	0,981	0,980	0,979	0,979	0,979	0,980
ModCr/SyBo	0,977	0,976	0,976	0,976	0,975	0,975	0,975	0,975
ModCr/SES	0,985	0,977	0,972	0,969	0,961	0,957	0,954	0,949
Croston/SyBo	0,994	0,994	0,994	0,993	0,993	0,994	0,993	0,993
Croston/SES	1,001	0,998	0,994	0,993	0,988	0,984	0,981	0,976
SyBo/SES	1,006	1,003	1,000	0,998	0,992	0,987	0,983	0,980

Table 6 The mean and median for the sMAPE quotients with +25-s as start value.

	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	0,986	0,982	0,980	0,979	0,978	0,978	0,979	0,979
ModCr/SyBo	0,984	0,979	0,977	0,976	0,974	0,974	0,974	0,974
ModCr/SES	0,983	0,976	0,972	0,969	0,964	0,960	0,957	0,953
Croston/SyBo	0,998	0,997	0,996	0,996	0,996	0,995	0,995	0,995
Croston/SES	0,997	0,994	0,992	0,989	0,985	0,981	0,977	0,974
SyBo/SES	0,999	0,997	0,995	0,993	0,989	0,986	0,982	0,979
				Median				
	0,025	0,05	0,075	0,1	0,15	0,2	0,25	0,3
ModCr/Croston	0,990	0,986	0,982	0,980	0,979	0,979	0,979	0,980
ModCr/SyBo	0,985	0,979	0,977	0,976	0,975	0,975	0,975	0,975
ModCr/SES	0,985	0,977	0,971	0,968	0,961	0,956	0,953	0,949
Croston/SyBo	0,996	0,995	0,995	0,995	0,994	0,994	0,993	0,993
Croston/SES	0,996	0,993	0,991	0,989	0,986	0,983	0,981	0,975
SyBo/SES	0,997	0,997	0,997	0,996	0,991	0,986	0,982	0,979

MAD_n

Table 7 The mean and median for the MAD_n quotients with -25-s as start value.

	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	1,201	1,273	1,303	1,319	1,334	1,342	1,349	1,353
ModCr/SyBo	1,260	1,341	1,375	1,393	1,410	1,419	1,426	1,431
ModCr/SES	1,249	1,366	1,428	1,471	1,540	1,600	1,655	1,709
Croston/SyBo	1,048	1,052	1,053	1,054	1,055	1,056	1,056	1,056
Croston/SES	1,037	1,064	1,083	1,099	1,129	1,159	1,187	1,214
SyBo/SES	0,989	1,012	1,028	1,042	1,070	1,097	1,123	1,148
				Median				
	0,025	0,05	0,075	0,1	0,15	0,2	0,25	0,3
ModCr/Croston	1,171	1,217	1,226	1,236	1,242	1,247	1,255	1,262
ModCr/SyBo	1,220	1,277	1,291	1,295	1,305	1,310	1,319	1,327
ModCr/SES	1,213	1,284	1,317	1,342	1,373	1,406	1,423	1,431
Croston/SyBo	1,051	1,055	1,056	1,057	1,056	1,056	1,057	1,056
Croston/SES	1,027	1,047	1,058	1,070	1,089	1,103	1,107	1,107
SyBo/SES	0,983	1,002	1,010	1,016	1,038	1,050	1,057	1,057

Table 8 The mean and median for the MAD_n quotients with +25-s as start value.

	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	1,188	1,262	1,295	1,312	1,329	1,338	1,345	1,351
ModCr/SyBo	1,265	1,340	1,373	1,391	1,408	1,417	1,424	1,430
ModCr/SES	1,319	1,411	1,461	1,497	1,560	1,617	1,671	1,723
Croston/SyBo	1,064	1,060	1,059	1,058	1,057	1,057	1,057	1,057
Croston/SES	1,102	1,104	1,110	1,121	1,146	1,173	1,199	1,224
SyBo/SES	1,036	1,041	1,048	1,059	1,083	1,108	1,133	1,158
				Median				
	0,025	0,05	0,075	0,1	0,15	0,2	0,25	0,3
ModCr/Croston	1,174	1,219	1,229	1,246	1,257	1,259	1,262	1,263
ModCr/SyBo	1,245	1,285	1,301	1,306	1,324	1,323	1,328	1,333
ModCr/SES	1,286	1,328	1,350	1,363	1,388	1,414	1,424	1,432
Croston/SyBo	1,062	1,060	1,059	1,059	1,059	1,058	1,058	1,057
Croston/SES	1,076	1,068	1,076	1,079	1,092	1,106	1,114	1,120
SyBo/SES	1,012	1,012	1,020	1,020	1,041	1,054	1,064	1,070

MSE$_n$

Table 9 The mean and median for the MSE$_n$ quotients with -25-s as start value.

	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	1,562	1,787	1,882	1,936	2,009	2,073	2,138	2,207
ModCr/SyBo	1,730	2,005	2,124	2,192	2,283	2,362	2,441	2,524
ModCr/SES	1,804	2,329	2,658	2,918	3,368	3,789	4,208	4,639
Croston/SyBo	1,086	1,095	1,099	1,102	1,106	1,109	1,111	1,114
Croston/SES	1,116	1,207	1,278	1,340	1,451	1,545	1,629	1,704
SyBo/SES	1,024	1,095	1,152	1,202	1,291	1,368	1,435	1,495
				Median				
	0,025	0,05	0,075	0,1	0,15	0,2	0,25	0,3
ModCr/Croston	1,257	1,353	1,458	1,479	1,496	1,545	1,594	1,630
ModCr/SyBo	1,357	1,461	1,612	1,652	1,695	1,717	1,753	1,787
ModCr/SES	1,364	1,627	1,782	1,874	2,025	2,095	2,167	2,285
Croston/SyBo	1,081	1,095	1,097	1,097	1,096	1,097	1,103	1,107
Croston/SES	1,072	1,131	1,161	1,188	1,216	1,248	1,284	1,302
SyBo/SES	1,001	1,044	1,071	1,093	1,114	1,140	1,173	1,185

Table 10 The mean and median for the MSE$_n$ quotients with +25-s as start value.

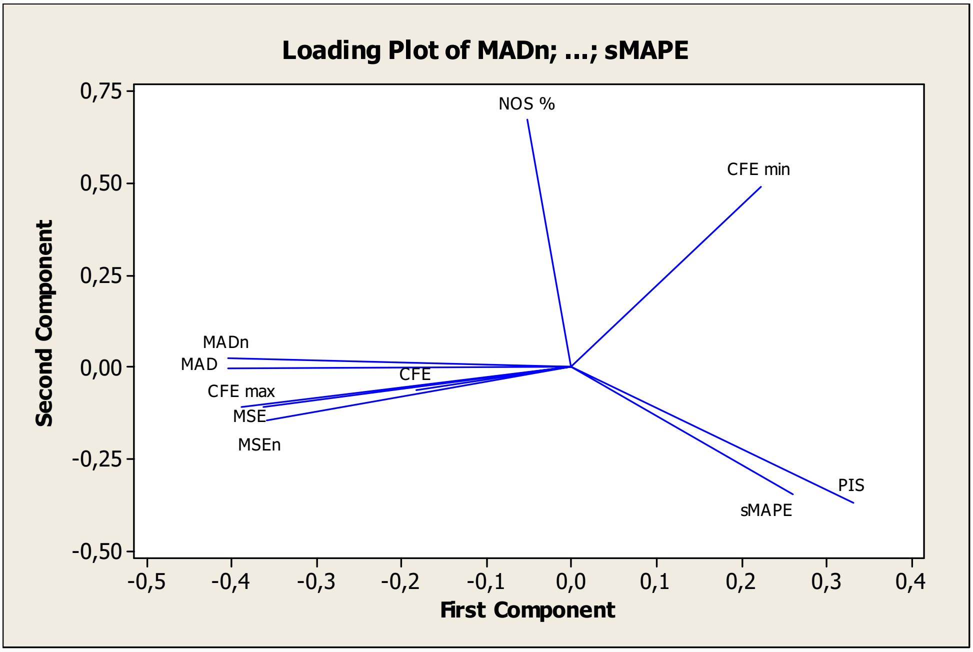
	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	1,451	1,676	1,787	1,855	1,947	2,023	2,096	2,171
ModCr/SyBo	1,655	1,913	2,039	2,117	2,223	2,311	2,398	2,486
ModCr/SES	2,016	2,484	2,787	3,035	3,477	3,895	4,313	4,742
Croston/SyBo	1,120	1,114	1,112	1,111	1,111	1,112	1,114	1,116
Croston/SES	1,305	1,340	1,385	1,434	1,530	1,618	1,696	1,767
SyBo/SES	1,154	1,189	1,229	1,271	1,352	1,424	1,488	1,546
				Median				
	0,025	0,05	0,075	0,1	0,15	0,2	0,25	0,3
ModCr/Croston	1,263	1,378	1,441	1,491	1,494	1,542	1,592	1,628
ModCr/SyBo	1,404	1,506	1,622	1,659	1,685	1,717	1,749	1,784
ModCr/SES	1,571	1,739	1,817	1,915	2,042	2,103	2,175	2,307
Croston/SyBo	1,118	1,111	1,109	1,109	1,103	1,102	1,104	1,109
Croston/SES	1,149	1,183	1,212	1,235	1,240	1,264	1,287	1,306
SyBo/SES	1,043	1,067	1,097	1,111	1,123	1,150	1,184	1,196

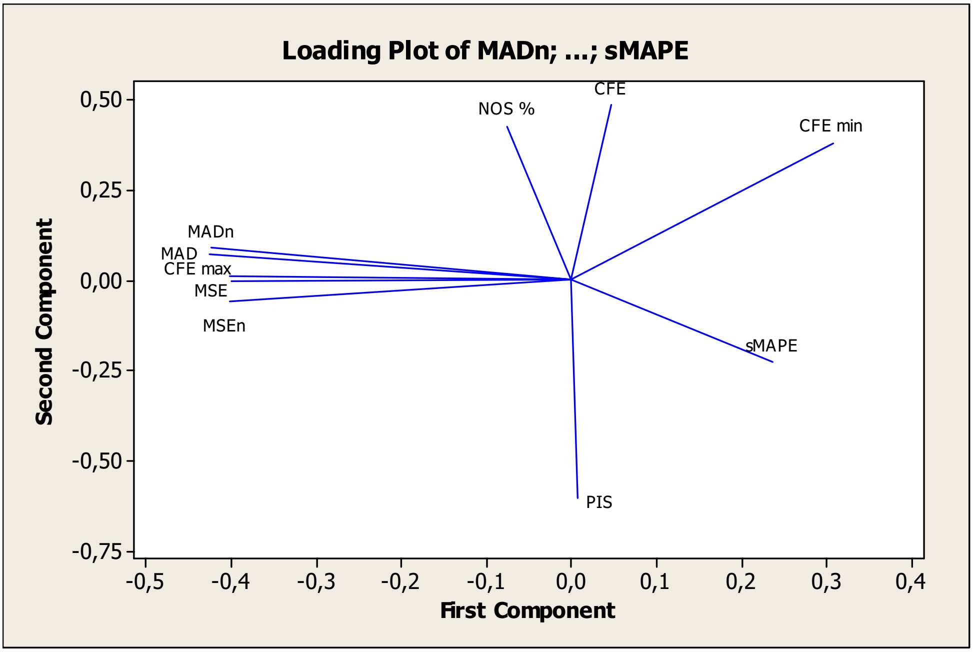
The evaluation between the different start values have been done by using the quotients between the different start values for the same smoothing constant. A value close to 1.00 means that the error is affected by the start value to a minor degree, see Table 11.

Table 11 Quotients between mean-s and+25-s for MSE_n.

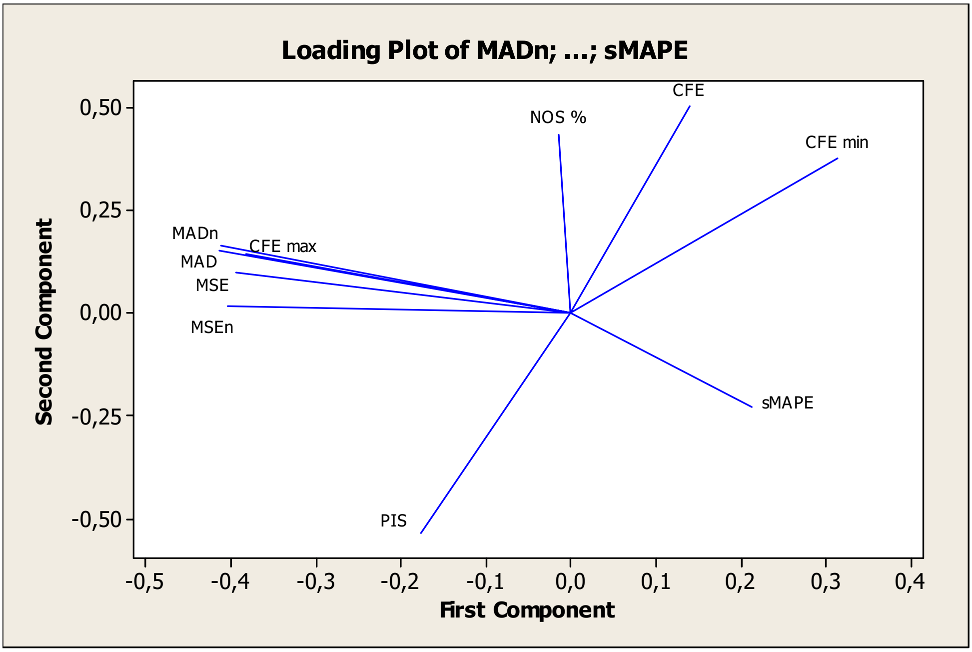
	0,025	0,05	0,075	0,10	0,15	0,20	0,25	0,30
ModCr/Croston	0,996	0,997	0,998	0,998	0,999	0,999	0,999	0,999
ModCr/SyBo	0,999	0,999	0,999	0,999	0,999	1,000	1,000	1,000
ModCr/SES	1,016	1,010	1,007	1,005	1,004	1,003	1,002	1,002
Croston/SyBo	1,003	1,002	1,001	1,001	1,001	1,000	1,000	1,000
Croston/SES	1,019	1,011	1,008	1,006	1,004	1,003	1,003	1,002
SyBo/SES	1,016	1,010	1,007	1,005	1,004	1,003	1,002	1,002
				Median				
	0,025	0,05	0,075	0,1	0,15	0,2	0,25	0,3
ModCr/Croston	1,003	0,999	1,000	1,001	1,000	1,000	1,000	1,000
ModCr/SyBo	1,001	1,001	1,001	1,000	1,000	1,000	1,000	1,000
ModCr/SES	1,020	1,006	1,001	1,001	1,001	1,001	1,001	1,001
Croston/SyBo	1,003	1,001	1,001	1,001	1,001	1,000	1,000	1,000
Croston/SES	1,013	1,007	1,005	1,004	1,004	1,002	1,002	1,001
SyBo/SES	1,016	1,005	1,003	1,003	1,002	1,002	1,001	1,001

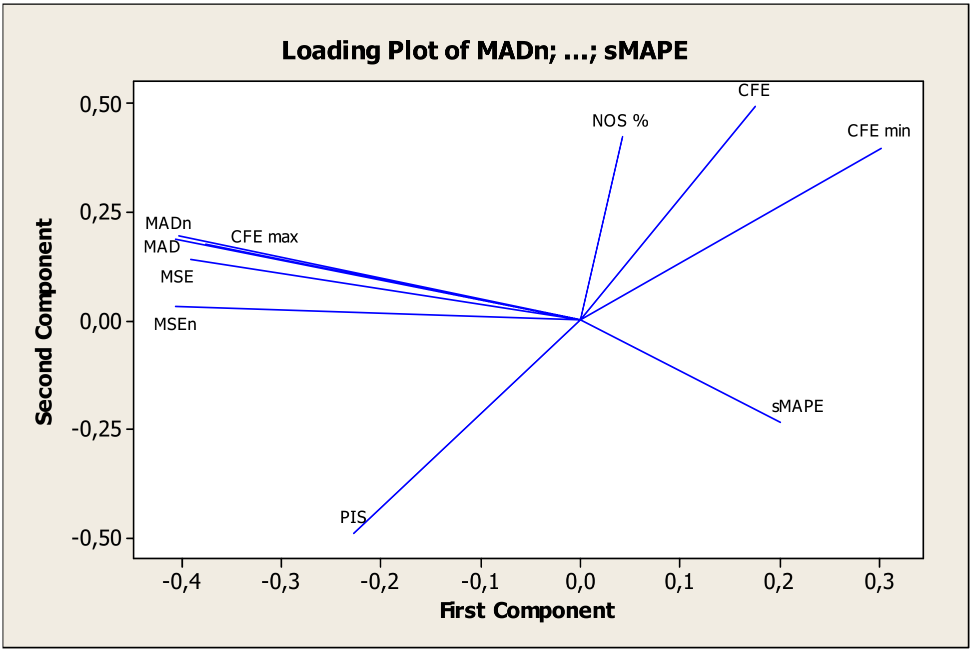
Croston 0.025 (upper), 0.075 (Lower) -25-4-s



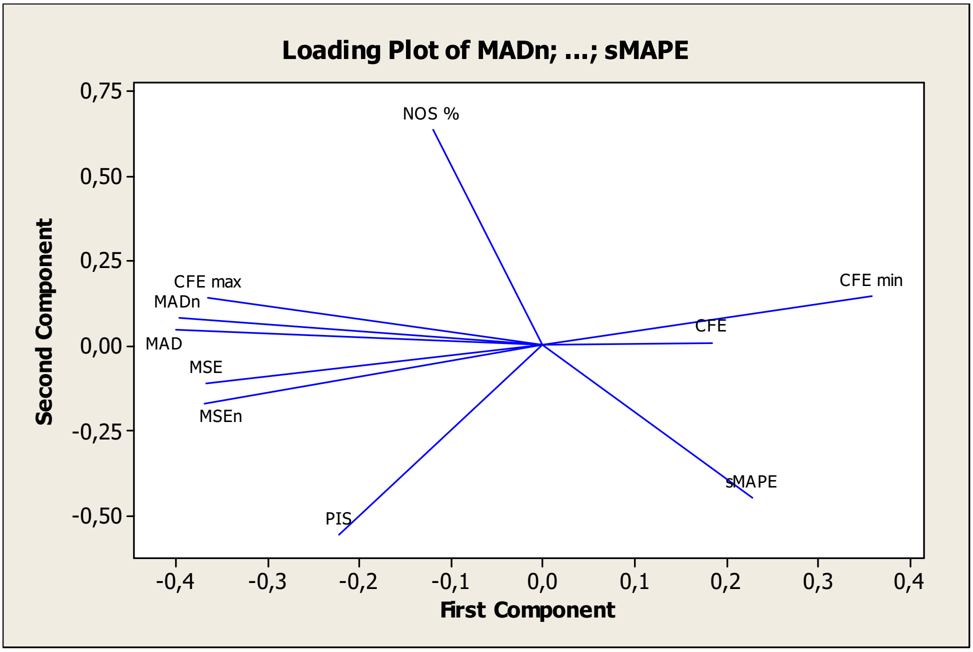


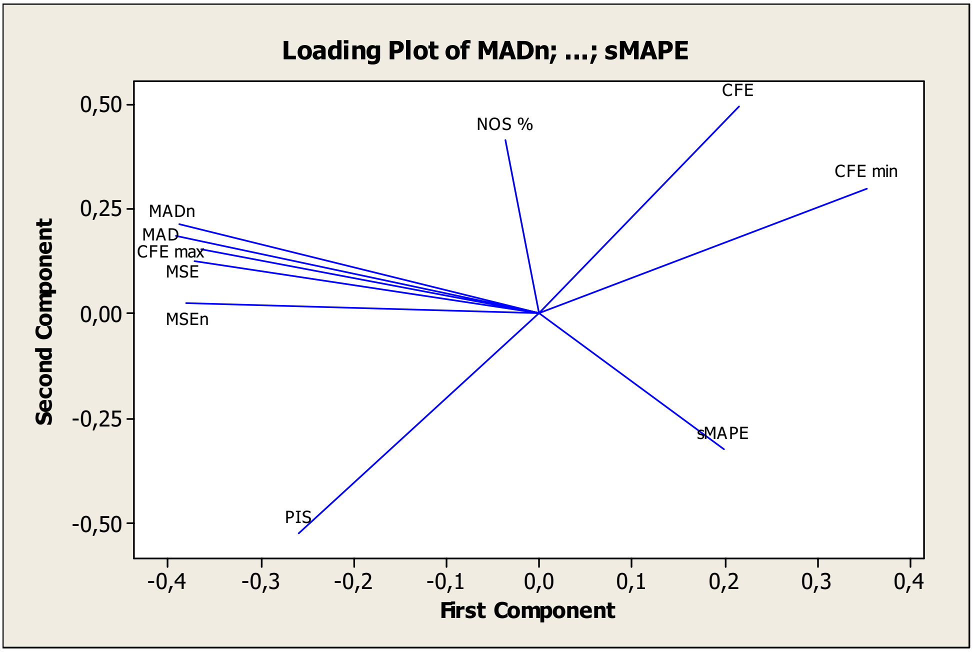
Croston 0.15 (upper), 0.25 (Lower) -25-s



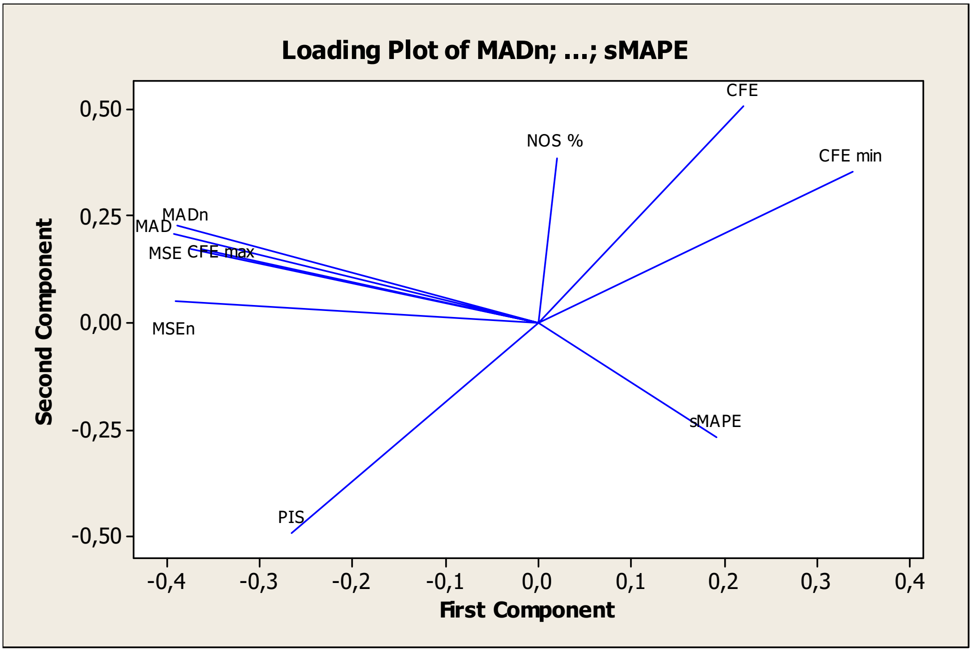


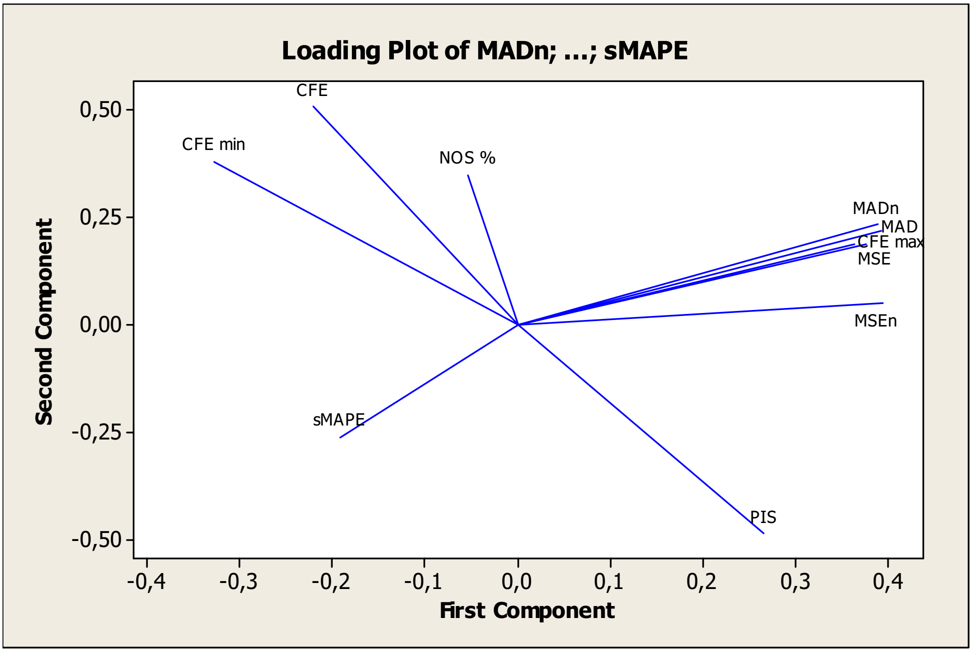
Croston 0.025 (upper), 0.075 (Lower) mean-s



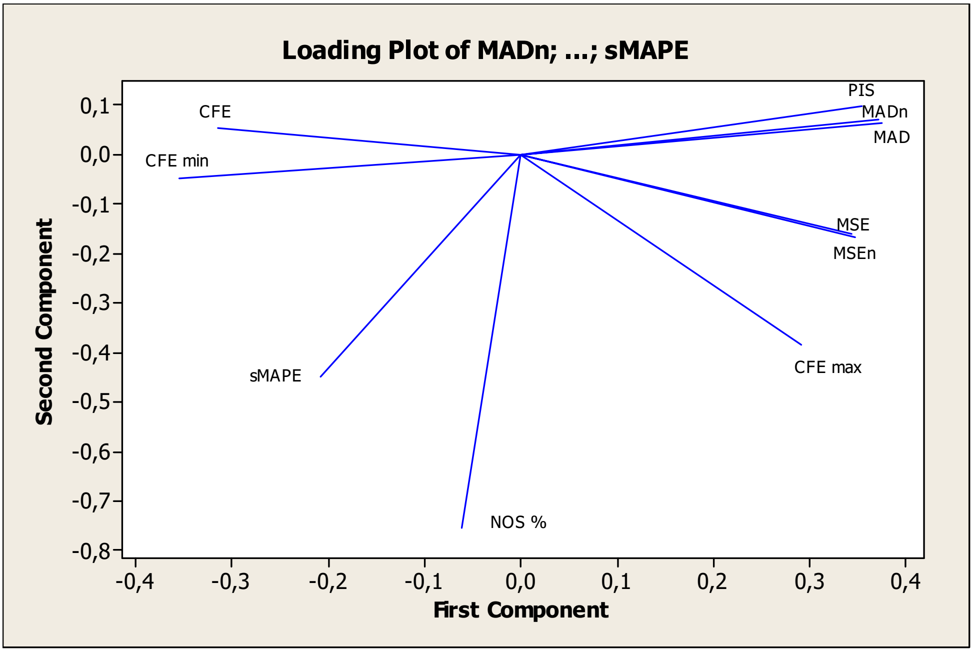


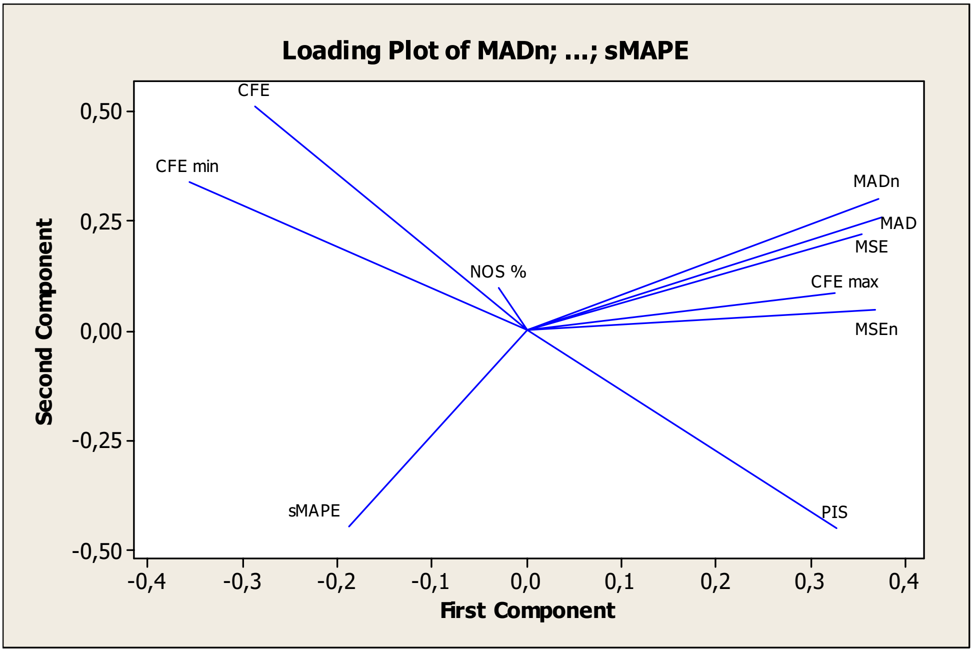
Croston 0.15 (upper), 0.25 (Lower) mean-s



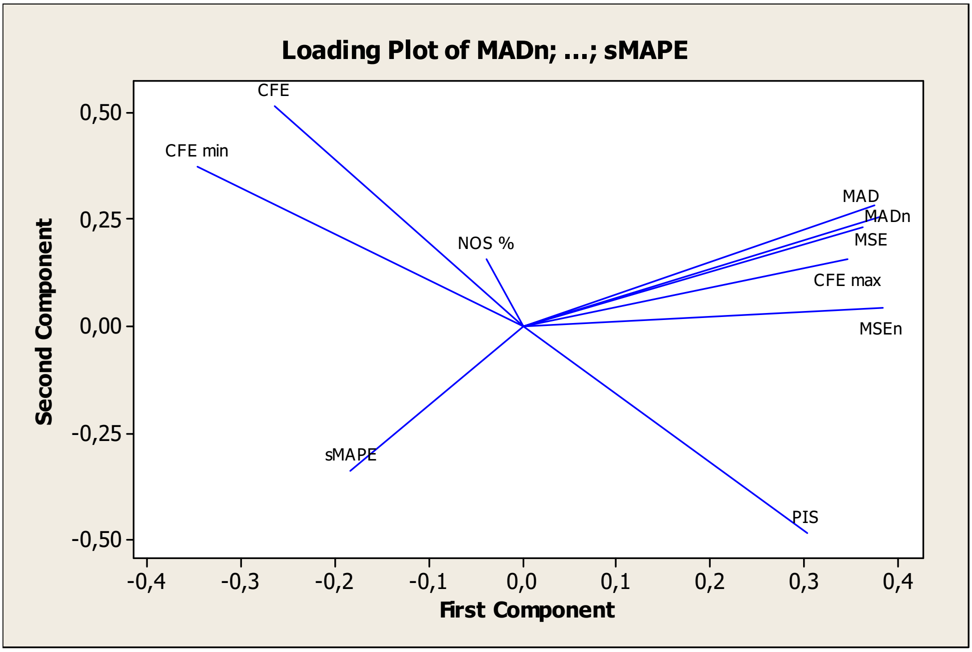


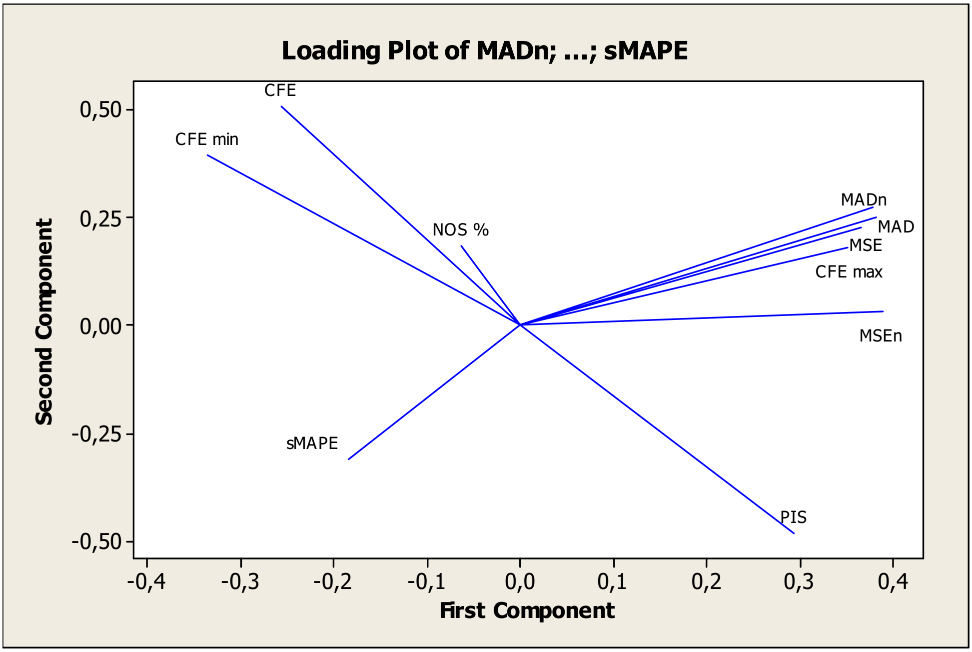
Croston 0.025 (upper), 0.075 (Lower) +25-s



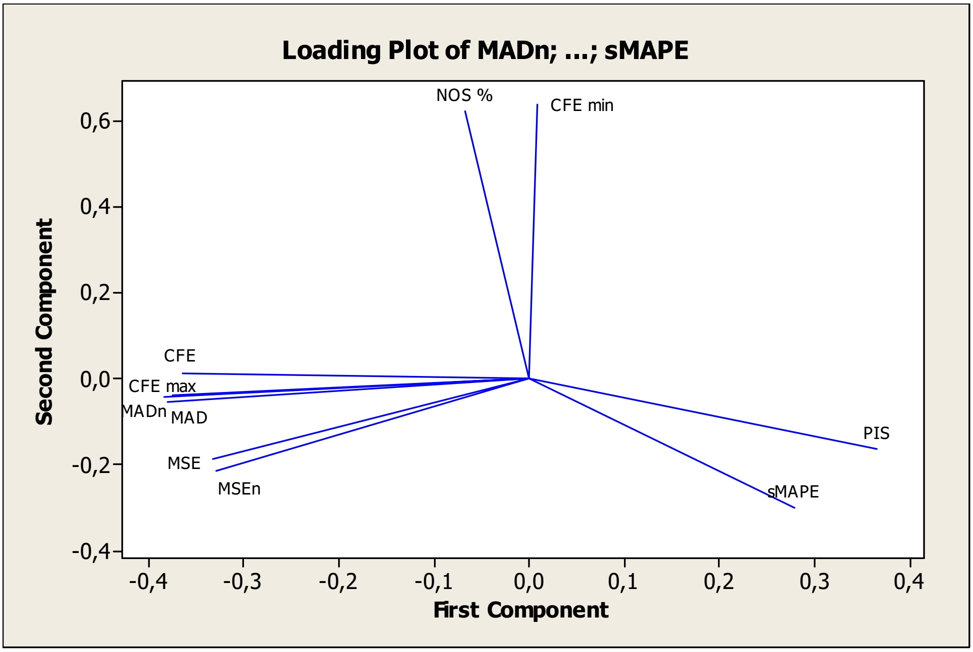


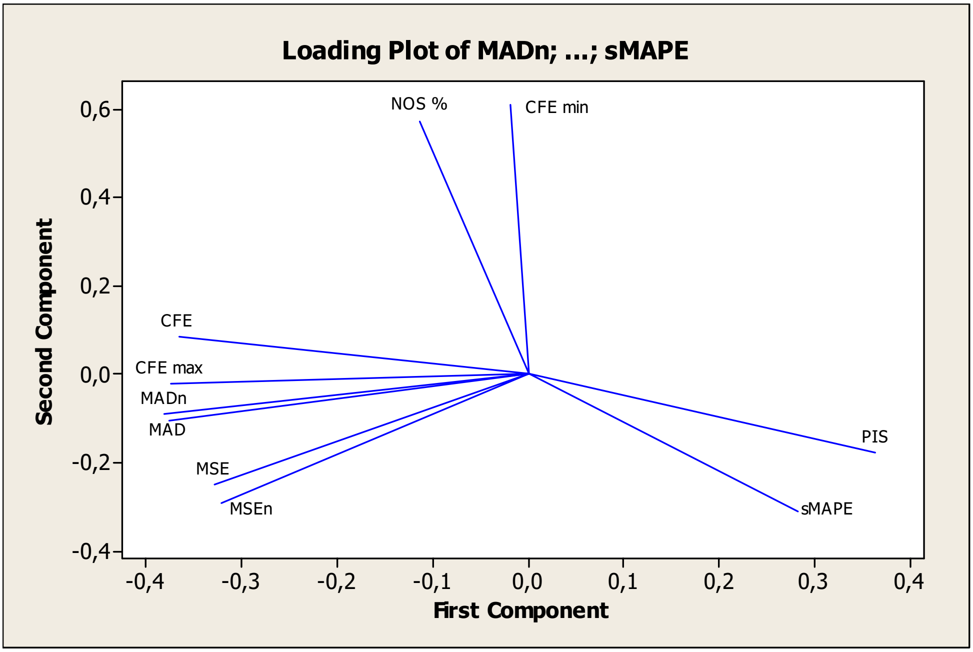
Croston 0.15 (upper), 0.25 (Lower) +25-s



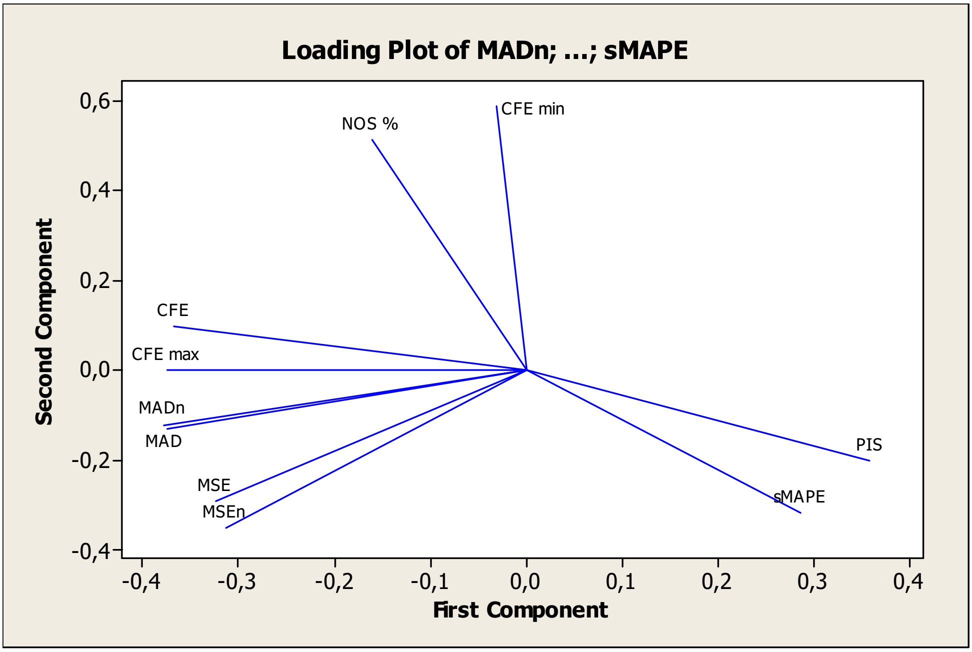


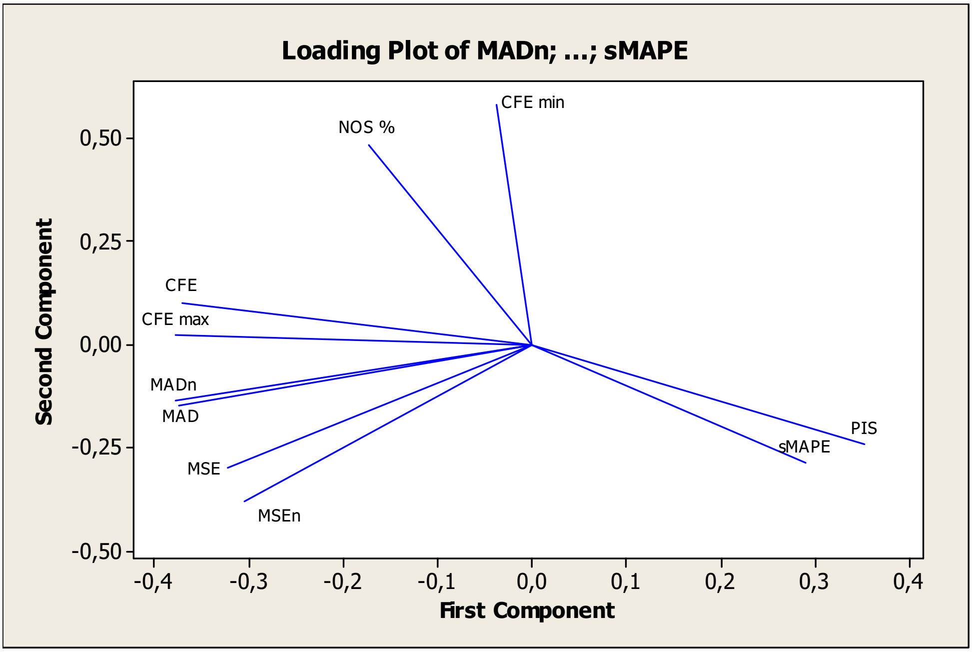
SyBo 0.025 (upper), 0.075 (Lower) -25-s



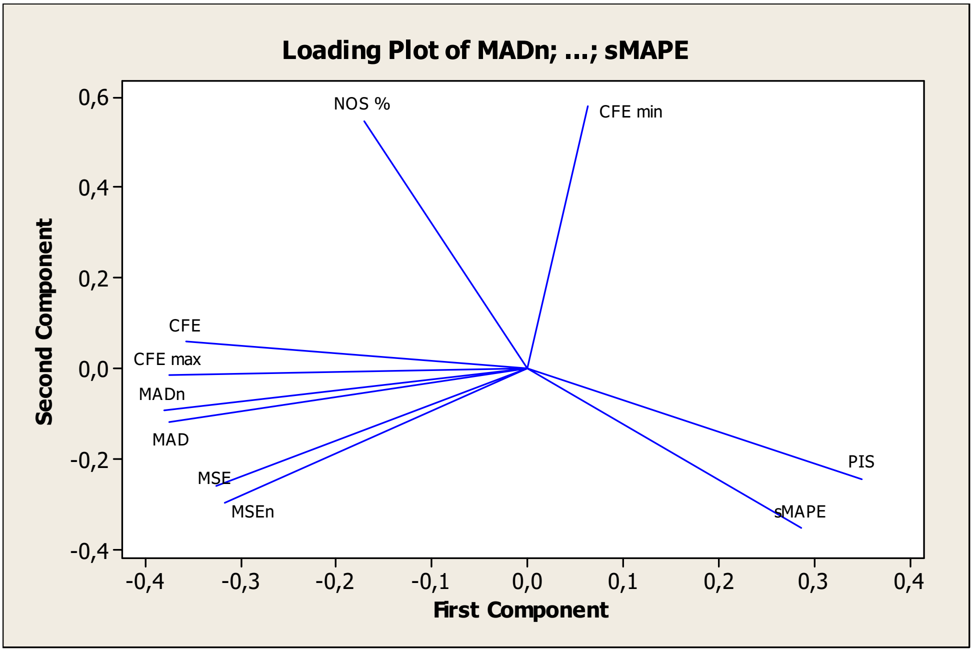


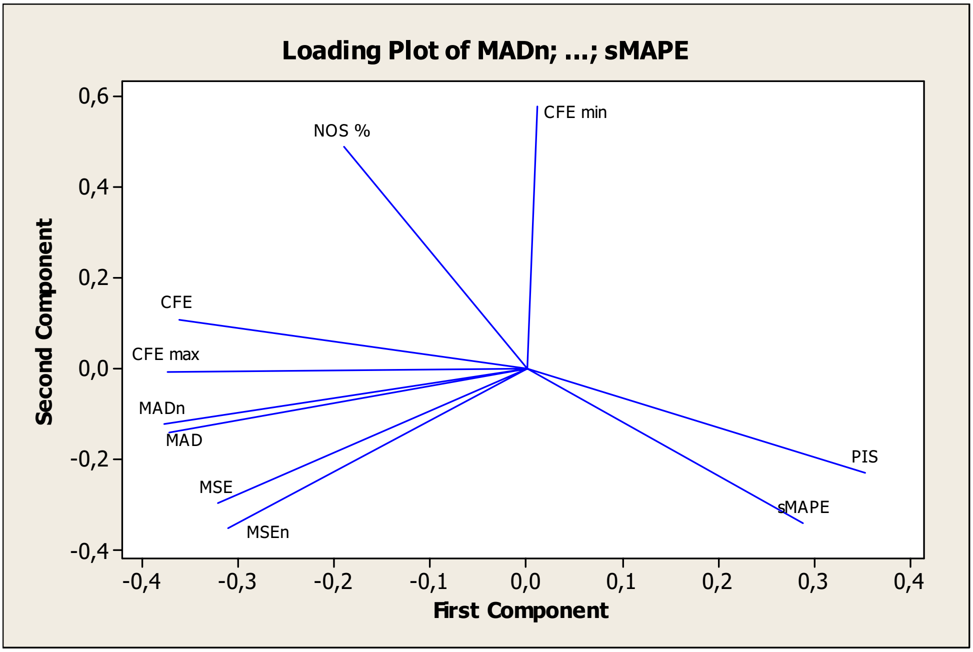
SyBo 0.15 (upper), 0.25 (Lower) -25-s



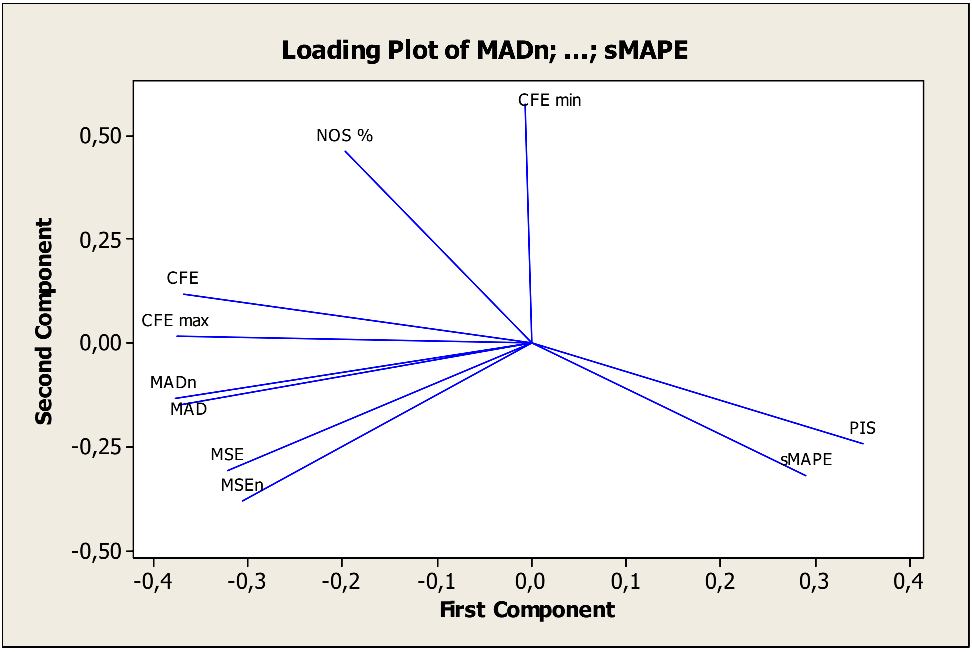


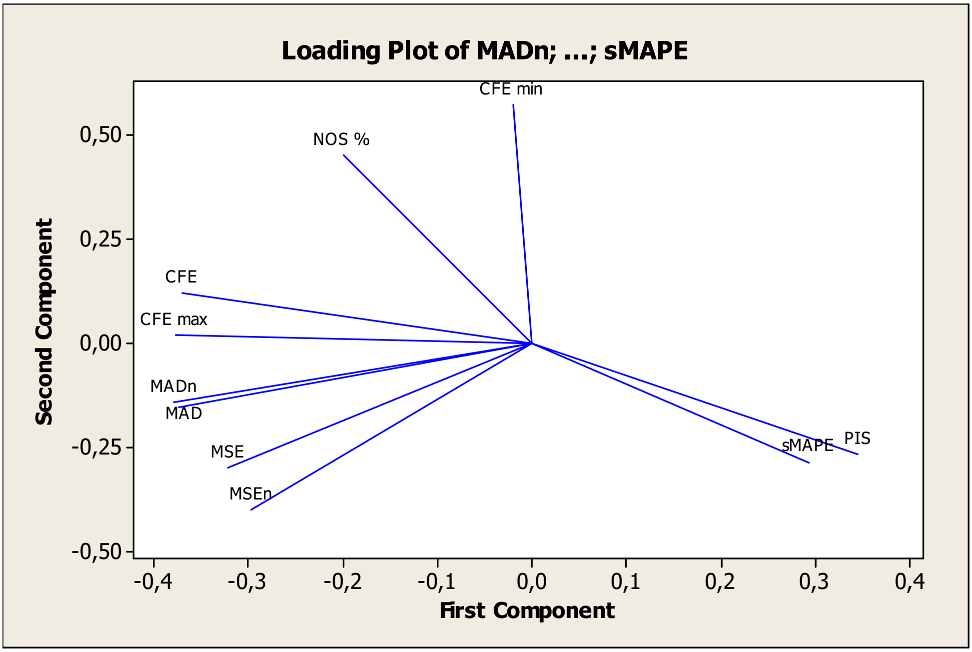
SyBo 0.025 (upper), 0.075 (Lower) mean-s



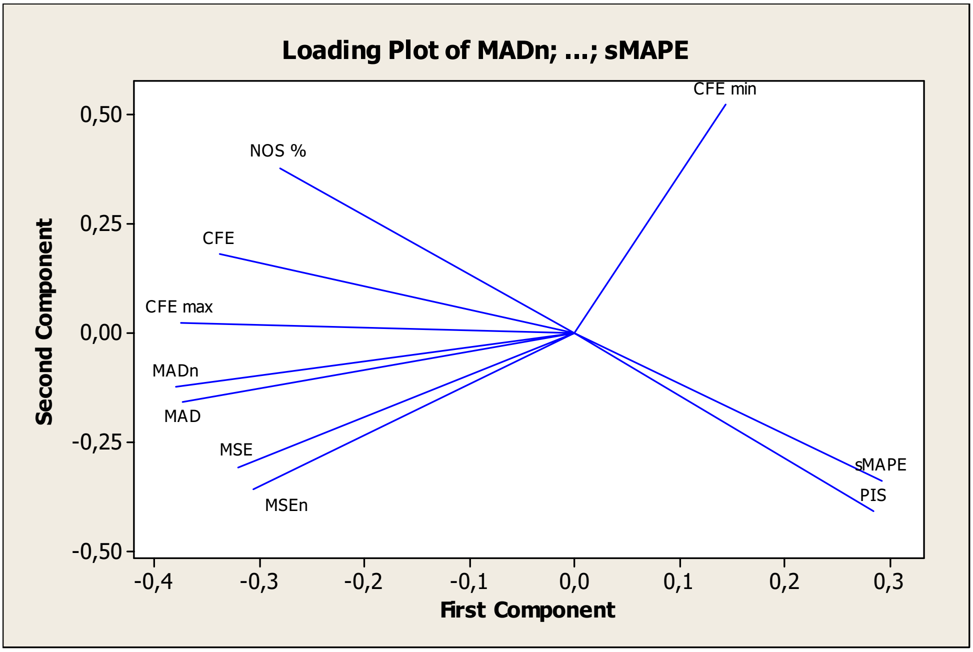


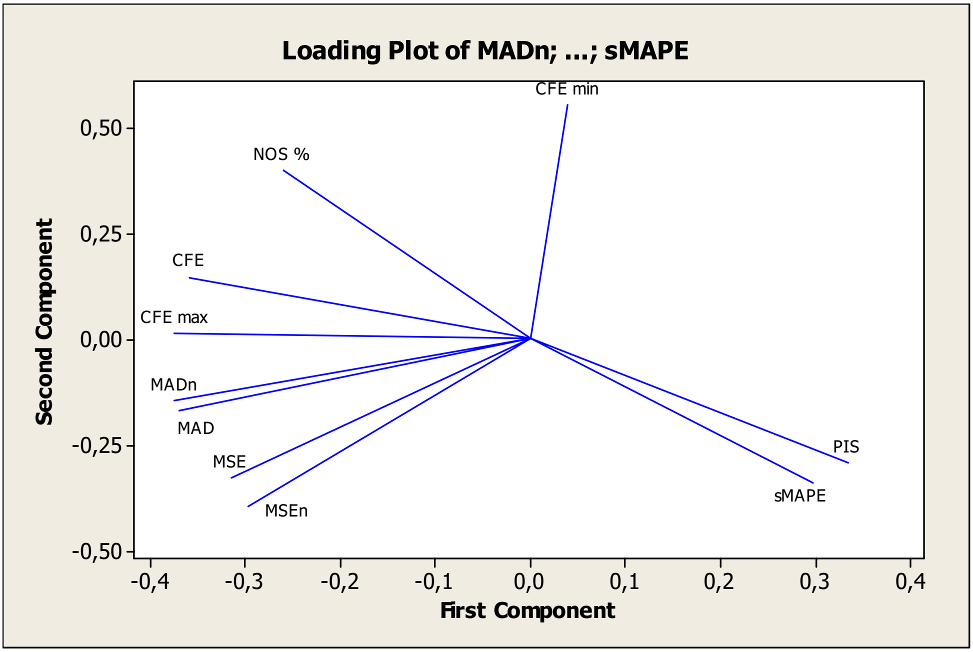
SyBo 0.15 (upper), 0.25 (Lower) mean-s





SyBo 0.025 (upper), 0.075 (Lower) +25-s





SyBo 0.15 (upper), 0.25 (Lower) +25-s