OPTIMISING TIME SERIES FORECASTS THROUGH LINEAR PROGRAMMING

Apostolos Panagiotopoulos, MSc

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

December 2011

To the memory of my beloved aunt Fotini Schiza and my fraternal friend Georgios Brillakis

ABSTRACT

This study explores the usage of linear programming (LP) as a tool to optimise the parameters of time series forecasting models. LP is the most well-known tool in the field of operational research and it has been used for a wide range of optimisation problems. Nonetheless, there are very few applications in forecasting and all of them are limited to causal modelling. The rationale behind this study is that time series forecasting problems can be treated as optimisation problems, where the objective is to minimise the forecasting error.

The research topic is very interesting from a theoretical and mathematical prospective. LP is a very strong tool but simple to use; hence, an LP-based approach will give to forecasters the opportunity to do accurate forecasts quickly and easily. In addition, the flexibility of LP can help analysts to deal with situations that other methods cannot deal with.

The study consists of five parts where the parameters of forecasting models are estimated by using LP to minimise one or more accuracy (error) indices (sum of absolute deviations – SAD, sum of absolute percentage errors – SAPE, maximum absolute deviation – MaxAD, absolute differences between deviations – ADBD and absolute differences between percentage deviations – ADBPD). In order to test the accuracy of the approaches two samples of series from the M3 competition are used and the results are compared with traditional techniques that are found in the literature.

In the first part simple LP is used to estimate the parameters of autoregressive based forecasting models by minimising one error index and they are compared with the method of the ordinary least squares (OLS minimises the sum of squared errors, SSE). The experiments show that the decision maker has to choose the best optimisation objective according to the characteristic of the series. In the second part, goal programming (GP) formulations are applied to similar models by minimising a combination of two accuracy indices. The experiments show that goal programming improves the performance of the single objective approaches.

In the third part, several constraints to the initial simple LP and GP formulations are added to improve their performance on series with high randomness and their accuracy is compared with techniques that perform well on these series. The additional constraints improve the results and outperform all the other techniques. In the fourth part, simple LP and GP are used to combine forecasts. Eight simple individual techniques are combined and LP is compared with five traditional combination methods. The LP combinations outperform the other methods according to several performance indices. Finally, LP is used to estimate the parameters of autoregressive based models with optimisation objectives to minimise forecasting cost and it is compared them with the OLS. The experiments show that LP approaches perform better in terms of cost.

The research shows that LP is a very useful tool that can be used to make accurate time series forecasts, which can outperform the traditional approaches that are found in forecasting literature and in practise.

ACKNOWLEDGMENTS

There are several people I wish to thank for their support for the period of my doctoral studies. I would like to distinguish my supervisor Dr Luc Muyldermans and my uncle and godfather Apostolos Schizas. I am very appreciative to them because without the contribution of Luc this effort would have never been completed and without the contribution of Apostolos it would have never begun. Closing, I would like to express my gratefulness to my partner Georgia Latsi for being so supportive during the last year.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

Linear programming (LP) is the most well-known tool in the field of operational research. Since the formulation of the first LP problems by Leonid Kantorovich in 1939 (Kantorovich, 1960) and the development of the simplex algorithm by George B. Dantzig in 1947 (Dantzig, 2002), linear programming has been used in a wide range of optimisation problems that are found in business and management, such as transportation routing, project planning, production planning, supply chain management and portfolio optimisation.

Forecasting is a vital managerial activity as it is the first stage of every planning procedure. This research focuses on applying linear programming to solve forecasting problems. The main idea behind the study is that forecasting problems can be treated as optimisation problems, where the objective is to minimise the forecasting error.

1.1 RESEARCH SCOPE AND OBJECTIVES

The aim of the study is to examine the application of linear programming as a tool to estimate the parameters of time series forecasting models. The research compares the performance of the LP-based approaches with the traditional statistical tools that are found in the literature, specifies the advantages and the disadvantages of the former and concludes the cases, where they should be preferred.

The rationale for this topic is that it is very interesting from a theoretical and mathematical prospective. The topic rather focuses on the advancement of scientific knowledge about forecasting and it is not an actual forecasting application for business. However, LP is a very strong tool but simple to understand and to use; thus, if LP-based approaches are shown to be more or as accurate as the traditional methods it will give to forecasters the opportunity to do accurate forecasts quickly and easily.

The purpose of the research is explorative, because it tests the implementation of an idea that already exists. There is important research on the use of linear programming and optimisation as a tool for solving statistical problems, with a wide area of applications (e.g. descriminant analysis). However, applications of the former for optimising forecasts have not been investigated in detail and they are limited to causal forecasting applications (Trapp, 1986, Soliman et al. 1997). In addition, the performance of LP-based approaches has neither been tested nor compared with traditional methods. Linear programming is a flexible tool that can exceed many limitations of the latter. This research aims to test the performance of LP-based approaches as an alternative and also to use it for solving problems that traditional statistical tools cannot deal with. The objectives of the study are to answer the following five questions and as a result to develop insights into the main aim:

I. How does linear programming perform in estimating the parameters of autoregressive based forecasting models? The traditional tool for the estimation of an autoregressive equation is the ordinary least squares method, which minimises the Sum of Squared Errors (*SSE*). LP gives the opportunity to minimise different accuracy indices, such as the Sum of Absolute Errors (*SAD*) and the Sum of Absolute Percentage Errors (*SAPE*). Thus, an LP approach will show how other indices perform compared with the SSE and in what situations they should be preferred.

II. "*The performance of a technique may differ according to different accuracy measures*" (Makridakis et al. 1984). Traditional tools aim to optimise one accuracy index (e.g. the least squares method minimises only the SSE). LP, in contrast with other techniques, can be applied for multi-objective optimisation (e.g. goal programming). Can linear goal programming be used to minimise two or more accuracy indices (e.g. SAD, Maximum Absolute Deviation - *MaxAD*) instead of only one? The study will show its performance and compare multi-objective and single objective LP optimisation methods.

III. One of the outcomes of past research (e.g Makridakis et. al, 1984) was that simple techniques, like moving average and exponential smoothing outperform more sophisticated techniques in series with high randomness. Can the flexibility of the linear programming approaches be exploited to improve the performance on series with high randomness? If yes, how does it perform compared with simpler techniques?

IV. Linear programming was suggested as a combination forecasting technique (Reeves and Lawrence 1982). Nevertheless, comparison with other methods to develop a good combination of forecasts (e.g. *simple average, inverse proportion*) is not available in the literature. How does linear programming perform as a tool for combining forecasts? LP guarantees the optimal combination between all the available forecasts, according to a preselected optimisation criterion (e.g. SAD). The study will show how LP models for combined forecasting perform compared with individual as well as traditional combination methods.

V. “*Situations where the cost of overestimation differs from this of underestimation are very common*” (Newbold and Bos, 1994). Can linear programming be used to minimise the total forecasting cost (instead of error) in case the costs of overestimation and underestimation are different? The results will demonstrate how a cost minimisation model performs compared with the more traditional accuracy optimisation models and a sensitivity analysis will show in which cases (differences between underestimation and overestimation cost) the results are significantly different.

The first, the second and the fourth questions aim to test the applicability of linear programming for estimating the parameters of forecasting models, while the other two focus on exploiting the flexibility of linear programming to overcome some of the limitations of traditional statistical methods.

The study belongs in the scientific field of operational research/management science. Reisman and Kirschnick (1994, 1995) classify studies in this field in three categories, according to the research strategy and aim:

a. The first category includes studies of “meta-research” and research on the philosophy and history of OR.

b. Second is the “untested theory” that includes studies, which focus on theoretical OR topics, for example research on new OR tools, and are not real-world applications.

c. Third are the real-world applications that deal with real-world problems.

This study belongs to the second category, since it is neither research on the philosophy and history of OR, nor a real world managerial application. It is a pure theoretical OR study that focuses on a new LP-based methodology for estimating the parameters of well-known forecasting models.

1.2 RESEARCH OUTLINE

The thesis’ outline is as follows: There is a review of the related literature of the area, then the methodology of the study follows; I continue with the mathematical models and the results of the experiments and I finish with several conclusions. The structure of the thesis is the following.

Chapter 2 presents what is known in the field so far and it consists of seven parts. First is an introduction to forecasting, followed by a general review of the types of forecasting techniques. The third part talks about the forecasting error. As the study is on time series forecasting the fourth part is more specific on time series analysis and forecasting techniques. Part five focuses on the field of combining forecasts. Part six is a review of the mathematical programming approaches for forecasting. The last part talks about seven forecasting competitions (M Competitions and NN Competitions) with objective to investigate how different techniques differ from each other and how forecasters can be able to make practical choices (Makridakis et al. 1984) and the future of forecasting research.

The methodology of the research can be found in Chapter 3. First is the development where I talk about the linear programming formulations. Second is the data, where I talk about the selection of the series for testing the techniques, the statistical analysis of them and their decomposition. The last part is the methodology of testing the forecasts, the comparison with traditional techniques found in the literature and the evaluation of the whole process.

The next five chapters aim to answer the five research questions respectively. In Chapter 4 simple LP is used to develop and optimise autoregressive based forecasting models. I estimate the coefficients of simple autoregressive models (AR) and autoregressive models with additive seasonality (ARS) by minimising SAD, SAPE, MaxAD, the Absolute Difference between Deviations (*ADBD*), and the Absolute Percentage Difference between Deviations (*ADBPD*). The accuracy of the LP based approaches is compared with the Ordinary Least Squares (*OLS*) method (minimising the Sum of Squared Errors). The study is mainly focused on this specific ARIMA (d,0,0) models due to the limitations of the LP. Only autoregressive models can be formulated as linear programs and the above minimisation objectives can follow a linear structure.

In Chapter 5 Linear Goal Programming formulations are applied to estimate the parameters of the same models. MinSum and MinMax pre-emptive and weighted goal programming is used (the latter is a relaxation of the former) to minimise both the SAD and the MaxAD. The results are compared with the OLS and the single objective approaches from Chapter 4.

Chapter 6 presents how the flexibility of LP can be exploited to improve the accuracy of autoregressive based forecasts on time series with high level variability and low predictability. I use all the simple LP and weighted goal programming models from Chapters 4 and 5 and I run experiments with additional constraints on a data set of series with high variability. The accuracy of the new approaches is compared with five traditional techniques, where the literature shows that they perform well in these cases.

In Chapter 7 I explore the use of LP as a tool to combine forecasts. I use simple LP and weighted goal programming. The former estimates the weights of several models by minimising the SAD, the SAPE and the MaxAD, whereas the latter minimises both the SAD and the MaxAD. The models combine eight individual forecasting techniques and we compare their accuracy with five other traditional combination methods.

Finally, in Chapter 8 I explore cases where the cost of the underestimation forecasting error differs from the overestimation cost. I apply simple LP methods to minimise the forecasting cost, instead of the error, or I use the simple LP methods from Chapter 4 adding the cost relationship of the underestimation and overestimation errors as a constraint. Experiments for five different cost relations are run and the approaches are compared with the OLS in terms of accuracy and cost. The approaches are limited on cases where the cost is a linear function to the forecasting error due to the LP limitations. The thesis finishes with several conclusions and recommendations for further research.

2 CURRENT RESEARCH STATUS

This chapter is a discussion of what is already known in the field. There is a general introduction to forecasting, where the distinctions between qualitative and quantitative forecasting and between causal and time series methods are presented. An analysis on the forecasting error (measurement methods and cost) follows. The next section focuses on time series analysis. There is a discussion about time series decomposition, the most common time series forecasting techniques and the Box-Jenkins methodology. Furthermore, there is a review of the area of combined forecasting. Subsequently, the chapter focuses on mathematical programming applications for forecasting. I present what has been done so far and I identify the research gaps that this study aims to cover. The chapter closes with a review of the forecasting competitions and the future of forecasting research.

2.1 INTRODUCTION TO FORECASTING

According to Armstrong (2001) forecasting is defined as the prediction of an actual value in a future time period. Makridakis et al. (1998) state that forecasting supplies information of what may occur in the future. Thus, it is used to estimate when an event is probable to happen so that proper action can be taken.

Forecasting in business practice is the basis every planning process; hence, it affects decisions and activities throughout an organisation. Examples of using forecasts in different areas of business practice are:

Accounting: Estimation of new product cost and cash flow management.

Finance: Time and amount of funding needs, budgeting, investment selection, credit scoring, credit risk management.

Human resources: recruitment needs, layoff planning.

Marketing: Pricing, placing and promotion, market entrance, competition strategies, direct marketing.

Operations: Inventory planning, capacity planning, supply-chain planning, work scheduling, production planning.

Information systems: systems revision.

R&D and design: New products and services introduction, technological progress.

Strategic management: Competition, economic conditions, new markets, goals planning.

According to Stevenson (2005) there are two applications of business forecasting. The first is to help the decision maker to *plan the system* and the second to *plan the use of the system*. Planning the system normally involves long-range plans such as product design, facilities layout, procurement of new equipment and location. On the other hand, planning the use of the system has to do with short and intermediate-range planning, such as inventory and workforce planning, work scheduling and budgeting.

In order to develop a forecast, the decision maker has to follow several steps. The number of steps varies, but, most researchers (e.g. Armstrong, 2001, Stevenson, 2005) agree on the following six:

1. *Determination of the purpose of the forecast*: That is the use and the objectives of the forecast. This will indicate the necessary accuracy level, the amount of resources that should be committed (people, computer time, money) and the cost of the forecasting error.

2. *Specify the time horizon*: A forecast may be long-range, intermediate-range or short-range, according to the forecasting purpose.

3. *Method selection*.

4. *Data gathering and analysis*: The data sources may be internal records (e.g. sales, demand, costs, stock control data, accounting data), external records (e.g. online data, government sources, periodicals and journals). Some data may not be available.

5. *Make the forecast*.

6. *Monitor the forecast*: Monitoring determines the performance of the forecast. If the forecast is not satisfactory, the decision maker has to re-examine the method, the data, the time horizon or even the purpose of the forecast. Then, (s)he has to start the process again from the corresponding step.

It is clear that forecasting is the starting point for various business decisions. The better an organisation's forecasts are, the more it is ready to utilize potential prospects and decrease prospective risks. Thus, forecasters should be very keen in selecting the most appropriate techniques and maintain their information sources up to date in order to keep the accuracy of their forecasts high.

2.2 FORECASTING TECHNIQUES

Forecasting techniques are classified in two categories: *quantitative* and *qualitative* (Makridakis et al., 1998 and Armstrong, 2001). They can also be found in the literature as *objective* and *subjective*, respectively (Nahmias, 2005). According to Makridakis et al. (1998) quantitative forecasting can be applied under three conditions:

1. Quantitative information availability about the past
2. This information can be expressed in numerical data.
3. Assumption of continuity, which is the statement that characteristics of the past patterns will continue in the future.

On the other hand, qualitative forecasting is applied in case of lack of quantitative information, but sufficient qualitative knowledge and experience exists. Finally, when neither quantitative information nor qualitative knowledge is available a satisfactory forecast cannot be performed.

Both quantitative and qualitative techniques differ extensively in accuracy, cost and complexity. Qualitative techniques, in general, are applied for longer term forecasting. Nonetheless, it is common for both methods to be combined. In practice, Sanders and Manrodt (2003) found significant differences in accuracy between companies that focus on only one of the above methods: organisations focusing on quantitative techniques tend to obtain better forecasts. However, the authors conclude that judgmental forecasting focused firms operate in more uncertain environments, which may explain the higher forecasting error.

2.2.1 Quantitative forecasting

Quantitative techniques are divided in two categories: *explanatory* (*causal*) models and *time series* models. The first category investigates the cause and effect relationship between the forecasted variable and one or more independent variables. Time series models predict the future value of a variable based upon its past values without attempting to estimate the external factors that affect this behaviour.

Specifically, explanatory forecasting is based on models in which the predicted value is related to various explanatory variables based on a specified theory (Armstrong, 2001). "*The purpose of explanatory models is to discover the form of the relationship and use it to forecast future values of the forecast variable*" (Makridakis et al., 1998).

The most common causal forecasting techniques are variations of linear and non-linear (e.g. logistic) simple and multiple regression models, where the dependent variable is the forecasted value and the independent variables are exogenous to this value. If Y is the dependent variable and $X_1, X_2, X_3, \ldots\ldots, X_N$ are the N independent variables, then:

$$Y = f\left(X_1, X_2, X_3, \ldots., X_N\right) \tag{2. 1}$$

Econometric models are defined as a special category of regression models in which the relationship between dependent and independent variables is linear. The most common ways for the estimation of the parameters of regression based models are the least squares and the maximum likelihood method (Makridakis et al., 1998). Nevertheless, in case of more complicated non-linear relationships, more sophisticated estimation techniques can be used, like Bayesian networks or artificial neural networks.

On the other hand, time series forecasting is set from the theory that the history of incidences over time can be used to forecast the future. Thus, time series forecasting techniques are based on the concept of recognising a pattern that exists in a series. This study is focused on time series forecasting; hence, an extended review on time series analysis and techniques will follow.

2.2.2 Qualitative Forecasting

As it was mentioned above, qualitative techniques do not require numerical information, but their outcomes are based on the judgment and accumulated knowledge of "*specially trained people*" (Makridakis et al., 1998). Even if the forecasting research and practice has proved that quantitative forecasting is more accurate, qualitative forecasting is widely applied in business practice, especially in situations where no past information is available, or it cannot be quantified. The most common qualitative methods are presented in the following table.

Table 2. 1 Qualitative forecasting techniques

Technique	Description
Grass Roots	Forecasters gather information from the executives and personnel (e.g. workers) who are at the lowest place of the hierarchy and usually closer to the forecasting problem. They use that information as a basis for judgmental forecasting.
Market Research	It is mainly used for long term market forecasting. The input is collected data from many ways, such as surveys, interviews and salesmen opinions.
Panel Consensus	Free open discussion of an idea at meetings. All participants have the right to express their ideas about the future (Galbraith et al. 2010).
Historical Analogy	It is based on finding analogies with similar situations of the past and identifying historical patterns (Dortmans and Eiffe, 2004).
Delphi Method	Group of experts responds to a questionnaire individually. Then a mediator gathers the results and formulates a new questionnaire that is resubmitted to the same group and the process is repeated. The repetition goes on until a forecast emerges (e.g. Kaynak et al., 1994, Lilja et al., 2011, Liu et al., 2010).
Sales Force Composite	Sales executives forecast according to their daily interaction with customers (Peterson, 1989 and 1993).
Unaided Judgment	This is a fast and inexpensive method, where a team of experts predict the result of current situations not aided by a formal forecasting technique and based only on their experience and possible data availability (Green, 2002). It has been proved very useful in cases where the expert has got good feedback about her/his forecasting accuracy. It is widely applied in the area of betting on sports.
Customer Surveys	They are usually used to signal preferences and opinions about new products and services.

Cross – Impact Analysis	Forecasters submit their opinions about what is likely to influence the area of interest. It is common to be used in combination with Delphi method (Banuls and Turoff, 2011).
Scenario Writing	This technique is widely used for long term planning and strategic analysis. It is based on developing the most possible and probable scenarios about the future (e.g. Bunn and Salo, 1993, Kanama, 2010).
Economic Indicators	These are tracked across a time series. The economic description of the behaviour of the series identifies the situation and helps experts to develop judgmental forecasts (e.g. Fite et al., 2002, Ozyildirim et al. 2010).

Source: Armstrong (2001), Chase et al. (2006), Nahmias (2005) and Newbold and Bos (1994)

2.2.3 Other forecasting techniques

Academic research and business practice have produced several forecasting methods that are not classified according to the traditional quantitative – qualitative and time series – causal clustering. The majority of these techniques tend to follow a mixed quantitative/qualitative methodology and they aim to “balance data and judgment” (Bunn, 1996), without this being the rule. The most common are the following.

Simulation: Simulation is common to be used when an analyst is asked to forecast the behaviour of a complex system over time. Simulation programs are designed to reflect the key aspects of a real situation (Pidd, 1998). The simulation method that is used depends on the characteristics of the system and the data availability. It usually combines both quantitative and qualitative elements and the balance between them differs according to the specific simulation method that is used. In business practice we can find applications of *Monte Carlo* (Pflaumer, 1988, Billio and Casarin, 2010), *Discrete Event* (Cheng and Duran, 2004), *System Dynamics* (Higuchi and Troutt, 2004, Wu et al., 2010), *Role Playing* (Green, 2002 and 2005) simulation and others. By running the simulation program under different starting conditions, a forecast for different situations is created (Nahmias, 2005).

Focus forecasting: The method is a rule based forecasting technique, where the analyst creates a simulation program under these rules. The program uses past data to measure how well the issued rules are performed (Chase et al., 2006).

Technical analysis: This method is also known as *Chartism* (Lo et al., 2000) and it has been part of the business and financial forecasting practice for many decades. Nevertheless, most academics recognise it as a highly subjective method and it does not receive the same acceptance as the traditional forecasting approaches. The theory behind technical analysis is that the recognition of a time series pattern can be achieved by looking how the time series charts have changed in the past (Kirkpatrick and Dahlquist, 2010). This will lead to predictions of future changes (Holden et al., 1990).

Game theory: While this technique is a fundamental tool for supporting strategic decisions under conflict, many researchers (e.g. Green, 2002, Goodwin, 2002, Bolton, 2002) have investigated its usage for making forecasts. This idea is also supported by Dixit and Skeath (1999), who state that the second use of game theory is in prediction. When decision makers have to deal with multiple interacting decisions, game theory can be used to predict the undertaken actions together with their results. In practice, game theory's usage for forecasting is very common. An example is this of Decisions Insights Inc. (a consultancy corporation in New York). They state on their website that they develop game theory models to forecast events that affect the business activity (www.diiusa.com).

Rule based forecasting: This is an expert systems application for prediction and it is the most characteristic example of an approach that incorporates judgment into the extrapolation process (Collopy and Armstrong, 1992, Armstrong, 2001). The forecaster develops an expert system that uses the experts' judgements as the rules to identify the quantitative forecasting technique that fits best on a time series.

Conjoint analysis: Conjoint analysis is characterised as a set of techniques for measuring buyers' tradeoffs among multi-attribute products and services (Green and Srinivasan, 1990, Halme and Kallio, 2011). Regression-like analyses are then used to predict the most desirable design (Armstrong, 2001).

Forecasting support systems (FSS)**:** FFS are decision support systems focused on forecasting decisions and consist of a combination of qualitative and quantitative forecasting. According to Armstrong, (2001) a FSS "*allows the analyst to easily access, organise and analyse a variety of information. It might also enable the analyst to incorporate judgment and monitor forecast accuracy*". FSS have found a wide area of application. They are very common in manufacturing and retail as part of an ERP system (Fildes et al., 2006, van Bruggen et al. 2010) but they are not rare in services (Croce and Wober, 2011). The importance of FSS is that managers can add non-time series information (especially event information) to their forecasts to increase their accuracy (Webby et al. 2005).

Armstrong (2001) presents a chart with the most common forecasting techniques, in which relations and interactions between them are indicated:

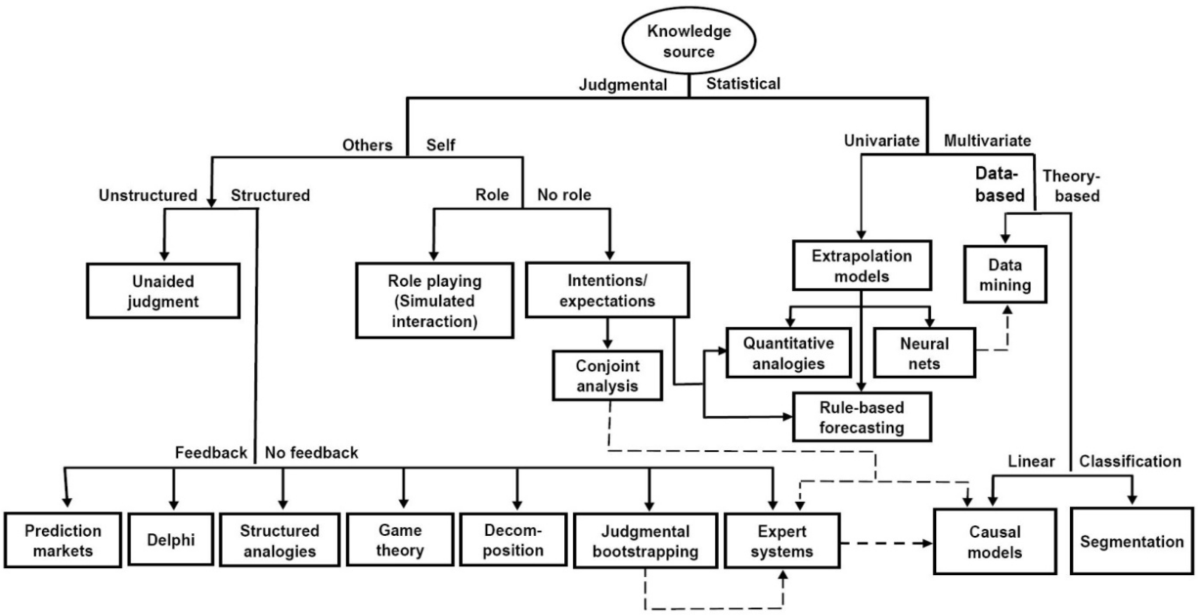

Figure 2. 1 The most common forecasting techniques and their interactions

Source: Principles of Forecasting website, Armstrong (2001)

2.2.4 Judgmental adjustments of quantitative forecasts

The above examples indicate that qualitative forecasting is rather supplementary than alternative to quantitative forecasting. In business practice it is quite common to judgementally adjust statistical based forecasts. The study of Sanders and Manrodt (1994) shows that about 45% of 96 US companies always judgmentally adjust quantitative forecasts, while only 9% never do. There is a large conversation about if judgmental adjustments improve quantitative forecasts. The survey of Fields and Goodwin (2007) concludes that judgmental adjustments tend to decrease the accuracy of statistical forecasts. Forecasters in practice rely a lot on judgment and use statistical forecasts inefficiently. Moreover, forecasts are adjusted by senior managers usually with no discussion and due to political motivation. In addition, they state that about half of respondents of their survey did not examine if their judgmental adjustments improved accuracy and almost a third did not record the cause for these adjustments.

The current research has underlined two main reasons why judgmental adjustments may harm forecasting accuracy. The first is that forecasters often make unnecessary adjustments to statistical forecasts and use statistical forecasts inefficiently (Lawrence et al., 2006). In order to avoid unnecessary adjustments, Goodwin (2000) has tested and suggested three simple methods to improve the use of statistical forecasts in business practice: "*(a) making the statistical forecast the default and requiring to make an explicit request to change this forecast, (b) requiring the judge to record a reason for changing the statistical forecast and (c) eliciting adjustments to the statistical forecast, rather than revised forecasts.*" The study shows that the first two methods significantly improve the use and accuracy of statistical forecasts, while in the third the improvement is rather small.

According to Eroglu and Croxton (2010) the second reason is that judgmental adjustments may introduce three types of bias: 1) optimism bias, 2) anchoring bias, and 3) overreaction bias. These biases are positively or negatively affected by the forecaster's personality (conscientiousness, openness to experience, neuroticism and extraversion), motivational orientation (seeking of compensation, recognition, enjoyment and/or challenge) and work locus of control (internal or external). These types of bias are the reason why forecasters tend to see false patterns in random movements (Goodwin and Fildes, 1999).

Forecasting practice shows that if the qualitative adjustment is necessary and not biased, then it marginally improves the accuracy of the statistical forecast. Fildes et al. (2009) suggest that the most reliable method for adjustment is bootstrapping. There are three well known bootstrapping methods:

- **Blattberg – Hoch** (50-50)**:** This is heuristic method where the adjusted forecast consists of 50% the statistical forecast and 50% the qualitative forecast (Blattberg and Hoch, 1990).
- **Judgmental bootstrapping:** Where the decision maker selects the optimal combination between the statistical forecast and the adjustment.
- **Error bootstrapping:** This is a more sophisticated technique, which models the relationship between the judgmental forecast and the statistical forecasts (Fildes et al. 2009).

Nonetheless, Fildes et al. (2009) state that if the judgment is biased, bootstrapping cannot be optimum.

As we can see, the practice shows that qualitative adjustment usually decreases the forecasting accuracy; however, if it is performed properly it may improve the performance of a statistical forecast, especially when new information is available, which is not already reflected in the pattern of the time series. Nonetheless, the decision maker should be sure that the statistical forecast is utilised, the adjustment is necessary and the judgment is not biased, in order to avoid harming the performance of the statistical approach.

2.3 FORECASTING ERROR

The accuracy level of a forecast is vital for an organisation. An analyst must not only make a good forecast, but also know what the expected error is and how flexible the operating system should be in order to meet the expected differences between forecast and reality.

2.3.1 Measuring forecasting error

The forecasting accuracy should be tested according to different perspectives. First is the *goodness of fit*, which shows how well the model is able to reproduce the actual known data. On the other hand, the *out of sample* perspective shows the predictive accuracy to unknown data. In order to measure the out of sample accuracy, the full amount of data is separated into a *training* and *test* set. The training set is used for the estimate the parameters of the forecasting model. Firstly the model is formulated, then the data of the training set are initialised and the parameters of the model are optimised by the most appropriate method (depending on the model) and according to the values of the data. Then, the model is ready to generate forecasts for the test set data. The out of sample forecast accuracy is then determined by comparing the forecasts with the actual data, which have not been used for the model development (Makridakis et al., 1998).

The forecasting error can be calculated as:

$$e_t = Y_t - F_t \tag{2. 2}$$

With e_t is the forecasting error, Y_t the actual value and F_t the forecast for period t.

Hyndman and Koehler (2006) classify five types of statistical indices that measures forecasting accuracy. These are:

***Scale dependent indices*:** They are useful for comparing the accuracy of different forecasting techniques on the same data set, but useless for comparison of different data sets or sets with different scales. These are:

Mean error: $$ME = \frac{1}{n}\sum_{t=1}^{n} e_t \tag{2. 3}$$

Mean error is mainly used to find if the forecast is biased. If mean error is zero, the forecast is unbiased because the total underestimation error is equal to the total overestimation error. If the mean error is positive, there is underestimation bias because the forecasts tend to be smaller than the actual values **(2. 2)**. On the other hand, if it is negative, there is overestimation bias.

Mean squared error: $$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^{\,2} \tag{2. 4}$$

Root mean squared error: $$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^{\,2}} \tag{2. 5}$$

Mean absolute error: $$MAE = \frac{1}{n}\sum_{t=1}^{n} \left|e_t\right| \tag{2. 6}$$

Median absolute error: $$MdAE = median\left|e_t\right| \tag{2. 7}$$

***Percentage errors*:** They are scale-independent and they can be applied for comparing different series:

Mean absolute percentage error: $$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{e_t}{Y_t}\times 100\right| \tag{2. 8}$$

Median absolute percentage error: $$MdAPE = median\left|\frac{e_t}{Y_t}\times 100\right| \tag{2. 9}$$

Root mean squared percentage error: $$RMSPE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{e_t}{Y_t}\times 100\right)^2} \quad \textbf{(2. 10)}$$

Root median squared percentage error: $$MdAPE = \sqrt{median\left(\frac{e_t}{Y_t}\times 100\right)^2} \quad \textbf{(2. 11)}$$

Despite their widespread use, percentage errors have several disadvantages. One disadvantage is that they are infinity for $Y_t = 0$ and they have an extremely skewed distribution when any value of Y_t is close to zero (Hyndman and Koehler, 2006). In adition, According to many authors stated that the biggest disadvantage of percentage errors is that they are asymmetric. Makridakis (1993) stated that "*equal errors above the actual value result in a greater MAPE (or MdAPE) than those below the actual value*". Makridakis presented the asymmetry of percentage errors with the following example: for Y_t = 100 and F_t = 150, the absolute e_t = 50 and the absolute percentage error is 50%, while for Y_t = 150 and F_t = 100 the absolute e_t will still be 50, but the absolute percentage error will be 33.33%. In addition, Armstrong and Collopy (1992) argued that "*the MAPE puts a heavier penalty on forecasts that exceed the actual than those that are less than the actual.*"In case of underestimation, the maximum possible MAPE is 100%, whereas, in case of overestimation it can be infinity.

Symmetric errors: These indices are suggested to overcome the disadvantages of the percentage errors:

Symmetric mean absolute percentage error: $$sMAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{(Y_t + F_t)}\times 200\right| \quad \textbf{(2. 12)}$$

Symmetric median absolute percentage error: $$sMdAPE = median\left|\frac{e_t}{(Y_t + F_t)}\times 200\right| \quad \textbf{(2. 13)}$$

Indeed, the symmetric absolute percentage error of the above example will be 40% for both cases. However, Goodwin and Lawton (1999) underline three main problems of these measurements:

1. There is a new type of asymmetry between the positive and negative errors. For example, if Y_t = 100 and e_t = 10 the symmetric absolute percentage error will be 9.52%, but if e_t = - 10,

the symmetric absolute percentage error will be 10.53%. However, in both cases the simple absolute percentage error will be 10%.

2. If the forecasts and actual values are of opposite sign, the symmetric MAPE will be very large. Especially, if the absolute values of the forecast and the actual value are equal, but they are of opposite signs, the symmetric MAPE is undefined.

3. If $| e_t | > 2| Y_t |$ then the $| e_t |$ will be reverse proportionate to the symmetric MAPE of period t.

For the above reasons, Goodwin and Lawton (1999) support that the use of symmetric percentage errors should be avoided in favor the simple percentage errors.

Both simple and symmetric percentage errors have several advantages and disadvantages; hence, they should be selected as accuracy measures according to the characteristics of the forecasting problem. If the forecasting error is relatively small, a simple percentage error measure should be preferred, because, there is no problem in measuring small errors, and symmetric errors tend to be asymmetric too. On the other hand, if the error is expected to be relatively big, the symmetric percentage errors should be preferred (with the exception when the absolute error is two times bigger than the actual observation, or when the forecast is negative). Nonetheless, there are no benchmarks; thus, it is up to the experience of the forecaster to select the most appropriate measurement.

Relative errors: This is an alternative to the above. If e_t* is the forecast error from a benchmark forecasting technique (usually a simple random walk), then the relative error is e_t/e_t* (Hyndman and Koehler, 2006). The available indices are:

Mean relative absolute percentage error: $$MRAE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{e_t^*}\right| \quad \textbf{(2. 14)}$$

Median relative absolute percentage error: $$MdRAE = median\left|\frac{e_t}{e_t^*}\right| \quad \textbf{(2. 15)}$$

The relative errors overcome the disadvantages of the percentage errors. Nevertheless, their main disadvantage is that they tend to be infinite if e_t* is close to 0.

Scaled errors: Hyndman and Koehler (2006) state that scaled error indices are widely applicable and are always defined and finite, in contrast with the relative errors. The proposed indices are:

Mean absolute scaled error: $$MASE = \frac{1}{n}\sum_{t=1}^{n}|q_t| \qquad \textbf{(2. 16)}$$

where: $$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{t=2}^{n}|Y_t - Y_{t-1}|} \qquad \textbf{(2. 17)}$$

Theil's U-statistic: $$U = \sqrt{\frac{\sum_{t=1}^{n}(FPE_t - APE_t)^2}{\sum_{t=1}^{n} APE_t^2}} \qquad \textbf{(2. 18)}$$

where: $$FPE_t = \frac{F_t - Y_t}{Y_{t-1}} \qquad \textbf{(2. 19)}$$

And: $$APE_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \qquad \textbf{(2. 20)}$$

The explanation of a scaled error index is the following:

- If it is equal to 1 then, the accuracy of the model is the same as with the naïve $F_t = Y_{t-1}$ method
- If it is smaller than 1, then the model being tested gives better results than the naïve method and the smaller the index, the better the model is.
- If it is greater than 1, then naïve method produces better results.

Both the relative and scaled errors are good accuracy measures for comparing forecasts, but they do not compare the error with the actual observation; thus, they do not make clear of how good or bad a forecast is independently. For this reason, they should be considered rather supplementary instead of alternative to percentage errors.

It may be difficult to select the most accurate forecasting method based on several accuracy measures. The reason is that models may perform dissimilar on different evaluation indices. Thus the analyst should specify a cost function before selecting of the most suitable forecasting model (Swanson and White, 1997).

The level of accuracy is usually the main criterion for the selection of the best forecasting method. Nevertheless, Yokum and Armstrong (1995) state that, in addition to the accuracy, there are other criteria that analysts should take into account when they choose the most suitable method. Additional criteria may be interpretation, functionality, flexibility or required data availability. In practice, models have a tendency to do better on some criteria and worse on other. The number and the hierarchy of the selection criteria always depend on the judgment of the analyst.

2.3.2 Cost of forecasting error

The error of a forecast results in a cost for the organisation. The cost of the forecasting error is given by the function:

$$C = C(e) \tag{2. 21}$$

Where *e* if the error in a forecast and *C* the associated cost.

According to Newbold and Bos (1994), the forecasting error equation has the following characteristics:

1. If the error is zero, then the cost is zero; thus: $C(0) = 0$
2. There is a positive relationship between the cost and the absolute value of the error; thus, the greater the error, the greater the associated cost. Hence, for $|e_1| > |e_2|$, $C(e_1) > C(e_2)$
3. The cost of error equation is often symmetric; hence the cost of a positive error is often equal to this of a negative error: $C(e) = C(-e)$

The first two characteristics are always applicable; nevertheless, situations where the cost of overestimation differs from this of underestimation are very common. For example, the cost of undersupply usually differs from cost of the oversupply. From a microeconomic perspective the cost of over supply is often greater; whereas from a marketing perspective uncovered demand tends to cost more than unexploited reserves.

The most common symmetric cost functions are:

I. *Quadratic error cost function*. The error cost is directly proportional to the squared error: $C(e) \sim e^2$

II. *Absolute cost error function*. The error cost is assumed to be proportional to the absolute error: $C(e) \sim |e|$

There are additional factors that affect the cost of errors. Sanders and Graman (2006), in their effort to quantify the cost of forecasting error and the impact in the warehouse, found that forecast bias is significantly more detrimental to cost compared to the standard deviation of forecasts. Standard deviations of forecasts result from poor forecasting, whereas forecast bias is typically managerially introduced.

2.4 TIME SERIES ANALYSIS

There are two types of time series analysis: time series decomposition and forecasting.

2.4.1 Decomposition

A time series pattern can be usually *decomposed* into sub-patterns that represent different elements of the time series. In economic and business series, patterns are usually decomposed in three parts, *trend-cycle*, *seasonality* and *randomness*. The trend-cycle represents long term changes in the level of the series, whereas the seasonality presents periodic variation of regular length (like the variations of the temperature during a year). On the other hand, randomness represents the error or difference between the combined effect of the previous patterns of the series and the actual data (Makridakis et al., 1998). Thus, according to Makridakis et al. (1998), time series are made up as:

Data = pattern + error

= *f* (trend-cycle, seasonality, error)

Decomposition, does not aim directly to forecasting, but to analysing the time series and identifying its characteristics. Its general mathematical representation is:

$$Y_t = f(T_t, S_t, E_t) \qquad \textbf{(2. 22)}$$

Where Y_t is the data value, S_t and T_t are the seasonal and trend sub-patterns and E_t the irregular pattern for time *t*.

The decomposition equation usually follows an *additive* or a *multiplicative* formulation, which are:

a) Additive: $$Y_t = T_t + S_t + E_t \qquad \textbf{(2. 23)}$$

b) Multiplicative: $$Y_t = T_t \times S_t \times E_t \qquad \textbf{(2. 24)}$$

In addition, Newbold and Bos (1994) suggest the *unobserved* decomposition model, in which sub-patterns are not observed. Forecasting practice has shown that this model is very applicable on most time series regardless of their characteristics (Newbold and Bos, 1994): The formulation is the following:

c) Unobserved: $$Y_t = (T_t + E_t) \times S_t \qquad \textbf{(2. 25)}$$

A way to estimate the trend component is by smoothing the series to reduce the random variation. There are several smoothing methods, such as the simple moving average, double moving average, weighted moving average and regression smoothing (Makridakis et al., 1998). For more complicated series, more sophisticated techniques have been developed, such as the *Census Bureau* (X-11, X-12 and X-12-Arima).

In addition, decomposition can also be done graphically; by separating the series into three plots (trend-cycle, seasonal and random plot, Makridakis et al., 1998). Diagrams with most common time series patterns are presented in Figure 2.2.

Decomposition can be used for forecasting, by projecting the separate plots into the future and re-merging them to develop the forecast. The difficulty of the method lies in the accuracy of the components' forecasts (Makridakis et al., 1998).

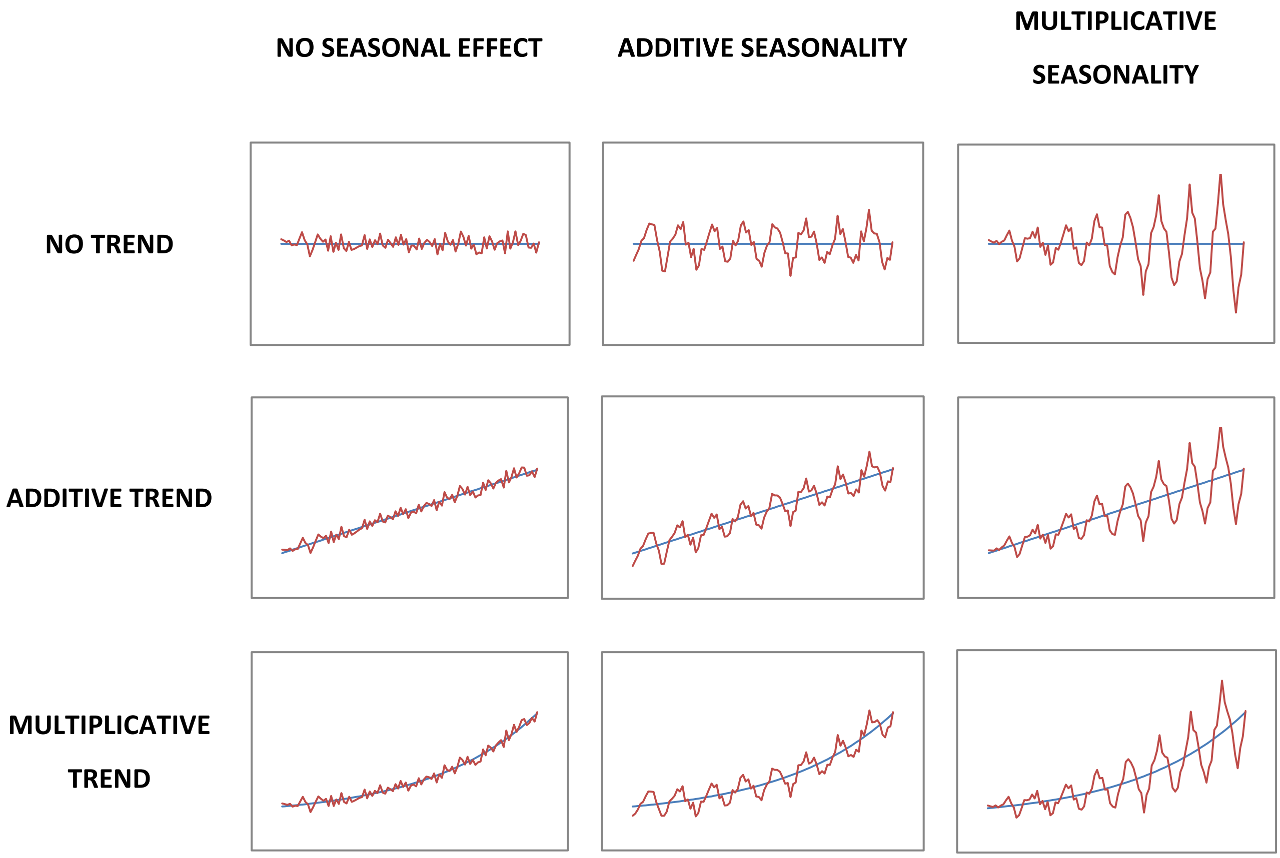

Figure 2. 2 Time series patterns

Decomposition is mainly a method of understanding rather than forecasting a time series. It represents the behaviour of the series, which helps the analyst to understand better the forecast problem. Decomposition is useful as a preliminary step before selecting and applying a forecasting method (Makridakis et al., 1998).

2.4.2 Time series forecasting techniques

Most researchers (Anderson et al. 1998, Armstrong 2001, Hand et al. 2001, Makridakis et al. 1998) classify time series in four categories. These, together with the most common forecasting techniques, are the following:

Simple methods

These are the simplest forecasting techniques, which can be applied for any type of series; however, they do not give very accurate results for series with strong trend or/and seasonal pattern:

Naïve: The simplest, but widely used forecasting approach. The forecast is simply the last value of the time series (Aaker and Jacobson, 1987).

Simple moving average: The forecast of is the average of a number of previous period values (Johnston et al. 1999)

Cumulative moving average (total average)**:** It is similar to the simple moving average; the forecast is the average of all the previous periods. This technique is very applicable for forecasting stationary series (series of data that is generated by a process which is in equilibrium around a constant value and where the variance around the mean remains constant over time, Makridakis et al., 1998).

Weighted moving average: An extension of the simple moving average, where the values of the previous periods are weighted differently (Perry, 2010).

Simple exponential smoothing: The forecast is based on two factors, the last period's forecast and the last period's actual value (Hyndman et al., 2008).

Adaptive response rate exponential smoothing: An extension of simple exponential smoothing where the importance of the last period's forecast and actual value change during the forecasting process (Trigg and Leach, 1967).

Methods for series with trend

Simple forecasting techniques are less effective on series that display a very strong trend. The following techniques can produce more accurate forecasts for series with a strong trend.

Holt's linear method: This is an extension of single exponential smoothing to linear exponential smoothing. In this case, there are two smoothing equations, where the first estimates the level of the series and the second the trend at a specific time (Hyndman et al., 2008).

Damped exponential smoothing: This technique is an extension of Holt's linear method and it is used when the time series trend is not linear, but there is a local slope to a future level of the data (Hyndman et al., 2008).

Regression analysis: Measures the linear or non-linear relationship between the predicted variable (dependent) and the time (independent variable). It is very useful for the estimation of the trend of a time series.

Trend projections: A simple method that identifies the trend of the time series and projects it into the future (Dugdale, 1971).

Methods for series with trend and seasonality

In more complicated time series an additional pattern of seasonality can be observed. In this case, only techniques that consider the seasonality factor can produce accurate forecasts.

Holt-Winters: Winter improved Holt's linear method by adding a third smoothing equation that estimates seasonality. Thus, this technique allows both seasonal and trend influences to be incorporated into the forecast (Hyndman et al., 2008).

Advanced forecasting methods

For more complicated time series, the usage of more sophisticated techniques is required. The most common are the following.

Box-Jenkins: This method was introduced by Box and Jenkins in 1970. It estimates the possible dependencies between the values of the times series from period to period. A more detailed presentation of this method will follow.

Shiskin time series (X-11): This method separates the time series into seasonal, trend and error parts. It is very effective, but it requires a large amount of past data points (at least 36 data points of history).

Data mining: The method uses statistical analysis and machine learning tools on large amounts of data in order to determine patterns of the time series that will aid forecasting (Morales and Wang, 2010, Delen et al., 2011).

Bayesian forecasting: These are forecasting techniques based on Bayesian statistics. In these methods, the forecasts are based on parametric modelling. The parameters of the model are estimated according to the priori probability distribution of the observation of the series. The advantage of Bayesian forecasting is that it presents the probability distribution of the forecast that reflects the uncertainty due to the parameter estimation (Hoogerheide and van Dijk, 2011, Yelland, 2010, Smith and Freeman, 2010, Chen et al. 2011).

Computational intelligence: Instead of statistical methods, quantitative forecasting can be based on computational intelligence tools. These approaches are favourable for forecasting long series with complex, nonlinear patterns. Computational intelligence based techniques are common to be black-box forecasting because the relationship between the time and the values remain hidden from the practitioner. Such methods are artificial neural networks (Wong et al., 2010, Shah and Guez, 2009), fuzzy predictions (Luna and Ballini, 2011), evolutionary and genetic algorithms (Jursa and Rohring, 2008, Venkatesan and Kumar, 2002) or hybrid. According to Simpson (1992), the removal of the undesirable noise (error) of a pattern is one of the most common operations that computational intelligence approaches perform.

2.4.3 The Box-Jenkins methodology for ARIMA models

The main objective of this research is to explore the usage of mathematical programming and linear programming in particular to optimise autoregressive based forecasting models. Thus, this part of the literature review focuses on a more detailed review of ARIMA models (Autoregressive-Integrated-Moving Average). ARIMA models were introduced by George Box and Gwilym Jenkins in the early 70s. This methodology utilises dependencies among values of the series during discrete times. The ARIMA models are combinations of autoregressive, moving average and random walk (integrated) models to produce forecasts for both stationary and non-stationary time series. Thus, the name of the methodology is *Autoregressive* (AR) *Integrated* (I) *Moving Average* (MA) models. The three parts are as follows:

1. Autoregressive models:

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + ... + b_p Y_{t-p} + e_t \qquad \textbf{(2. 26)}$$

This is a regression equation, where the independent variables are *time-lagged* values of the predicted variable Y_t, b_0 is the constant coefficient, b_i ($i \in [1, p]$) are the parameters and e_t is the white noise (error) for period t.

2. Moving averages are described by the following equation:

$$Y_t = c_0 + c_1 e_{t-1} + c_2 e_{t-2} + \ldots + c_q e_{t-q} + e_t \qquad \textbf{(2. 27)}$$

In this case, the independent variables of the regression are the past errors of the forecasts. This equation produces the moving average of the error series e_t for period t, while c_0 is the constant coefficient and c_i ($i \in [1, q]$) are the parameters of the model.

3. Integrated models reduce the difference level of the series that takes place in order to transform a non-stationary series into stationary ones. The difference is defined as the difference between two observations in the series. Thus:

$$Y_t' = Y_t - Y_{t-1} \qquad \textbf{(2. 28)}$$

This equation produces a first-order difference. According to Makridakis et al. (1998), stationarity is usually achieved by taking the first-order difference. Nevertheless, if it is necessary for additional differencing, the second-order difference is:

$$Y_t'' = Y_t' - Y_{t-1}' = Y_t - 2Y_{t-1} + Y_{t-2} \qquad \textbf{(2. 29)}$$

In case of series with seasonality, seasonal stationarity is required. The seasonal difference is the difference between an observed value and the corresponding observation from a previous period. For example, for monthly data with annual seasonality, the first order difference will be:

$$Y_t' = Y_t - Y_{t-12} \qquad \textbf{(2. 30)}$$

For a non-stationary time series the integrated model can be also written as:

$$Y_t - Y_{t-1} = e_t \qquad \textbf{(2. 31)}$$

Where e_t is the *white noise* (Makridakis et al. 1998). This can be rewritten as:

$$Y_t = Y_{t-1} + e_t \qquad \textbf{(2. 32)}$$

This is widely used for non-stationary data and is known as a *random walk* model (Box and Jenkins, 1970).

There are several ways to test the stationarity of a time series. The most common are the plot of the *autocorrelation function* (ACF), the plot of the *partial autocorrelation function* (PACF), the *Ljung – Box* test and *Portmanteau* tests (Makridakis et al., 1998).

According to Newbold and Bos (1994), the ARIMA methodology is limited to time series with the following two characteristics:

1. There is a linear correlation between the forecasts and the actual values of the series.
2. The objective is to develop *efficiently parameterised* models, which are models that present a satisfactory explanation of the characteristics of a time series with the minimum possible parameters.

The general model of the Box-Jenkins methodology is presented as ARIMA (p, d, q), where:

p: order of the AR part (number of the explanatory variables of the autoregressive model).

d: difference order of the Integrated part.

q: order of the MA part (number of coefficients of the moving average model).

The optimal order *p* and *q* for an ARIMA model is estimated with the use of the time plot of ACF and the PACF. For the AR the optimal order *p* is indicated by the lag where the PACF drop to or near to zero. In the same way, for the MA the optimal order *q* is the lag where the ACF drops to or near to zero. The ACF and PACF plots are an indication on the identification of the optimal order of pure AR or MA models. The order of mixed ARMA or ARIMA models is more difficult to identify. Hence, the decision maker should begin with a pure AR or MA model and consider extending it to ARMA or ARIMA.

There may be more than one optimal value for the order m (= $p + q$) of an ARIMA model. The way to select the best alternative is by using the *Akaike's Information Criterion* (AIC, Akaike 1974). If L is the likelihood for a model of order m to be the optimal then:

$$AIC = -2\log L + 2m \qquad \textbf{(2. 33)}$$

The optimal order is the one with the smallest AIC. If the decision maker does not have the means to estimate the actual AIC, an approximation is given as:

$$-2\log L \approx n(1+\log(2\pi)) + n\log\sigma^2 \qquad \textbf{(2. 34)}$$

Hence: $$AIC \approx n(1+\log(2\pi)) + n\log\sigma^2 + 2m \qquad \textbf{(2. 35)}$$

Where n is the number of the observations of the series and σ^2 is the variance of the residuals.

In the literature, there can be found many variations of the AIC, such as the *Bayesian Information Criterion* (BIC, Schwarz, 1978) or the *Final Prediction Error* (FPE, Akaike, 1969). A completed review can found in Konishi and Kitagawa (2008).

The main characteristic of an ARIMA model is that it covers a variety of models. Makridakis et al. (1998) presents a stepwise procedure to assist in the identification of the parameters of the model (p,d,q). After the estimation of the orders, the coefficients of the different parts should be estimated. Makridakis et al. (1998) suggest that the most common methods are these of the least-squares and the maximum likelihood estimation.

Box and Jenkins (1970) have summarised their ARIMA methodology in three phases, using the following diagram:

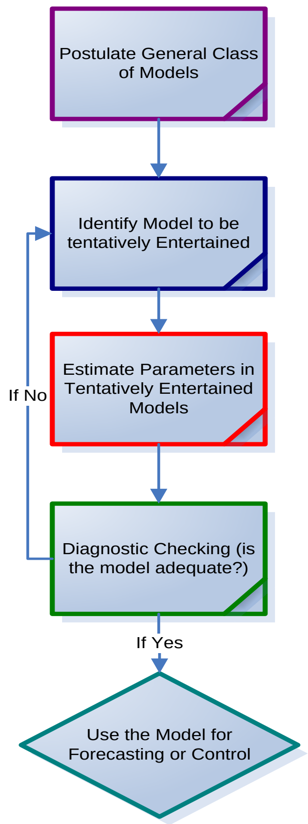

Figure 2. 3 The Box-Jenkins ARIMA methodology

Source: Box-Jenkins (1970)

2.4.4 ARIMA extensions

The general ARIMA model has been the basis for extended, more complicated forecasting models, in order to deal with issues, such as missing data in the time series, and or also considering external information (mixed time series – causal models). Some of the most well-known extensions are the following:

- Autoregressive conditional heteroskedasticity (ARCH, Engle, 1982, 1987) and generalised autoregressive conditional heteroskedasticity (GARCH, Bollerslev. 1986) for financial time series with time-varying volatility.
- Regression with ARIMA errors.
- Dynamic analysis models (Winker, 2006 and Fan and Söderström, 1997) for continuous time representation (e.g. dynamic inventory control systems).

- Auto-regressive auto-regressive moving averages (ARARMA) models (Parzen, 1982).
- Multivariate autoregressive (VARIMA) models.
- Robust trend models (Grambsch and Stahel, 1990).
- State space models.
- Non-linear ARIMA models (e.g. coefficient estimations with artificial neural networks).
- The X – 12 – ARIMA model (Pierce, 1980) for seasonal adjustment that combines the Census X – 11 technique with the ARIMA methodology.

2.5 COMBINED FORECASTING

Combinations of forecasts were introduced by Bates and Granger (1969) and it is a very common way to improve the forecasting accuracy. The forecasts that are combined can be based on different data or different techniques. The main idea of combining forecasts lies in the fact that different forecasting methods contain useful and independent information. According to Armstrong (2001) the areas of expert forecasting and econometric forecasting have proved good evidence about the improvement of forecasting accuracy through combining individual forecasts. Moreover, combining forecasts has been very useful when it is difficult to select the most accurate forecasting method. It has been shown a good way of hedging the risk in situations of very expensive forecasting errors (Armstrong, 2001). Makridakis (1989) states that the accuracy of an individual forecast is sensitive to several factors that may affect the accuracy. Combined forecasting works because it distributes the risk of such errors using several individual techniques. On the other hand, Andrawis et al. (2011) state that the benefit of combining forecasts is the prospect to combine short-term and long-term forecasting.

Combining can be expressed mathematically as follows:

$$F_{ct} = \sum_{i=1}^{k} w_i F_{it} \qquad \textbf{(2. 39)}$$

with

$$\sum_{i=1}^{k} w_i = 1 \qquad \textbf{(2. 40)}$$

and $$0 \le w_i \le 1 \tag{2. 41}$$

Where there are k forecasts that are combined. F_{ct} is the combined forecast at time t, F_{it} is the result of forecast i ($1 \le i \le k$) and w_i is the weight of forecast i.

Researchers (e.g. Newbold and Bos, 1994, Russel and Adam, 1987 and de Menezez et al., 2000) agree that the most common methods to estimate the values of the combing weights are:

Simple average – equal weights case: The simplest way to combine individual forecasts is to assign them equal weights. Hence:

$$F_{ct} = \frac{\sum_{i=1}^{k} F_{it}}{k} \tag{2. 42}$$

thus $$w_i = \frac{1}{k} \tag{2. 43}$$

An alternative to the simple average is the median.

Inversely proportional weights: This is a more sophisticated method that was introduced by Bates and Granger (1969). The forecasts are combined according to their individual performance. Specifically, the weight of a forecast is estimated according to the accuracy of the forecast compared with the sum. Newbold and Bos (1994) state that there are two factors, which should be considered for the estimation of the weights. The first is that the methods that perform better should have higher weights. Secondly the weighting procedure may need to be adapted in order to take into account the possibility that the performances of a forecasting method may change over time. The weights are estimated according to the inverse proportion of an accuracy index of an individual technique, divided by the sum of the inverse proportion of the accuracy index of all the techniques. The formula for assigning the weights using the inverse proportion of the mean squared error is:

$$w_i = \frac{MSE_i^{-1}}{\sum_{i=1}^{k} MSE_i^{-1}} \tag{2. 44}$$

Where MSE_i is the mean squared error of forecast *i*. The inverse proportion to other accuracy indices (mean absolute error and mean absolute percentage error) is formulated in a similar way.

Regression-based weights: When there are two forecasts to be combined, an alternative approach of assigning weights to forecast in to use a simple linear regression model. Suppose that there are two forecasts with weights w_1 and w_2 respectively, then, it is assumed that $w_2 = 1 - w_1$. Thus, the equation of the combination is:

$$Y_t = w_1 F_{1t} + (1 - w_1) F_{2t} + e_{ct} \qquad \textbf{(2. 45)}$$

Where Y_t is the actual value and e_{ct} is the combined forecasting error. The above equation is rewritten as:

$$(Y_t - F_{2t}) = w_1 (F_{1t} - F_{2t}) + e_{ct} \qquad \textbf{(2. 46)}$$

According to Newbold and Bos (1994) the regression based method is not preferable compared with the other two methods.

Weights based on the absolute error: The weights are assigned according to the number of times a technique gives the minimum absolute error in a series. The formula of the combination is:

$$w_i = \frac{\sum_{t=1}^{T} \delta_{AD_{it}, \min AD_{it}}}{\sum_{i=1}^{k} \sum_{t=1}^{T} \delta_{AD_{it}, \min AD_{it}}} \qquad \textbf{(2. 47)}$$

With δ the Kronecker delta and AD_{it} is the absolute error by individual technique *i* at time *t*. Hence:

$$\delta_{AD_{it}, \min AD_{it}} \begin{cases} = 1 \ \ if \ \ AD_{it} = \min\limits_{i} AD_{it} \\ = 0 \ \ otherwise \end{cases} \qquad \textbf{(2. 48)}$$

A further extension is to introduce judgmental forecasts in the forecasting combination. Sanders and Ritzman (1990) proved that the accuracy of quantitative techniques is improved when they are combined with judgmental forecasting. Especially in time series with a high level of unpredictability and uncertainty the introduction of the judgmental element may increase the accuracy of the results.

Russell and Adam (1987) present an empirical evaluation of the performance of various techniques for combined forecasts. They used a data set of 31 randomly selected series from the M competition and ten simple individual techniques. They evaluate the performance of ten different combination methods, including the simple average, the inverse proportion to MAD, MSE, MAPE on selected series, the average of the three, the average of the best five techniques according to their performance on the MSE, the inverse proportion to the MSE for all the techniques, the selective weights according to the performance based on the absolute error and the selective weights according to the performance based on the absolute error with indicator values. The last is an extension of the simple selective weights based on the absolute error method, where it applies hierarchical weights (indicator value) to the individual techniques in reverse analogy to their performance based on the absolute error. The least accurate technique is weighted with 1, the second least with two and so on and at the end each weight is divided are with the sum of the weights. The results show that the inverse proportion, average inverse proportion and the selective weights based on the absolute error with indicator values were the most accurate.

Hibon and Evgeniou (2005), in their study of the accuracy of forecasting combinations, reached the following three conclusions: 1. The accuracy of the best individual forecasting techniques that are tested on the a time series is the same with the best combinations, however, the worst forecasting techniques perform significant worse than the worst combinations. 2. The accuracy of forecasting combinations may drop significantly when more and more individual forecasts are added. 3. Finally, choosing an individual forecast from a set of available methods is more risky than choosing a combination. Thus, a chosen individual method may have significantly worse performance in practice than a chosen combination (according to the authors this is the most important result of the study).

Finally, Kolassa (2011) presented a method, where the optimal combination is selected according to the AIC and BIC. In his study he combined different exponential smoothing forecasts and tested his method in a sample of series from the M and M3 completions. He combined his method with simple combination techniques (average and median) and the best individual technique according to the AIC and BIC. The experiment show that the AIC and BIC based combinations outperform best individual forecasts overall. On the other hand, simple combinations and AIC and BIC based combinations tend to perform similarly, while simple combinations sometimes are outperformed than the best single forecasts. Even if the results of the paper were promising, the comparison was only between the AIC and BIC based combinations and not other more advanced combination techniques (e.g. Inverse proportion).

The main disadvantage of the above combination techniques is that they distribute the weights between all the individual forecasts that are used. The less accurate forecasts will are having smaller weights compared with the more accurate (with only exemption the simple average, where the weights are distributed equally). However, adding less accurate forecasts to a combination may affect the total accuracy of the combination. Hence, the decision maker should select and combine the most accurate forecasts, or the forecasts that are expected to be more accurate, instead of using all the alternatives.

2.6 MATHEMATICAL PROGRAMMING FOR FORECASTING

Linear optimisation has been presented several times as an alternative to the least squares and maximum likelihood methods for estimating the parameters of a linear regression. Nevertheless, while linear programming has been used in many areas were a regression technique is applicable (e.g. *discriminant analysis*), the literature in the area of forecasting is very limited.

2.6.1 Mathematical programming in statistics

The first time when linear optimisation was identified as a tool for estimating the coefficients of a linear regression was by Wagner (1959). In his study he states that while the traditional least squares method produces the Minimum Squared Error of a regression analysis, the simplex method can produce two additional *best-fit* criteria, the Minimum Absolute Deviation (MinAD) and the Minimum Maximum Absolute Deviation (MinMaxAD). A linear program is used in order to calculate the parameters of the regression equation that minimise the MinAD and MinMaxAD respectively; the objective function is one of these best-fit criteria (MinAD or MinMaxD) and the constraints are the linear equations for each of the observations. If x_{ij} is the independent variable, y_i the dependent variable, b_j ($1 \le j \le k$) the coefficient of x_{ij} and e_i the regression error, i ($1 \le i \le k$) is the index of the observation and j ($1 \le j \le n$) the index of the coefficient then the equation of the regression is[1]:

[1] The author has excluded the constant coefficient b_0 to simplify the example.

$$y_i = \sum_{j=1}^{m} x_{ij} b_j + e_i \qquad \textbf{(2. 49)}$$

The linear programming formulation for the estimation of the MinAD of a linear regression is the following:

Objective function:

$$\underset{b_j}{Min} \sum_{i=1}^{k} \left| \sum_{j=1}^{m} x_{ij} b_j - y_i \right| \qquad \textbf{(2. 50)}$$

This can be rewritten as:

$$Min \sum_{i-1}^{k} e_{1i} + \sum_{i=1}^{k} e_{2i} \qquad \textbf{(2. 51)}$$

Where e_{1i} is the *under-estimation* and e_{2i} the *over-estimation* error; thus, the total estimation error is $e_i = e_{1i} – e_{2i}$. According to this, the constraints of the model are:

$$\sum_{j=1}^{m} x_{ij} b_j + e_{1i} - e_{2i} = y_i \qquad \textbf{(2. 52)}$$

For

i =1, 2, ….., k

b_j unrestricted in sign for each j

e_{1i}, e_{2i} non-negative.

In the same way, the formulation for the estimation of the MinMaxD is:

$$\underset{b_j}{Min}\left\{\underset{i}{Max}\left|\sum_{j=1}^{m} x_{ij}b_j - y_i\right|\right\} \quad \textbf{(2. 53)}$$

Which can be rewritten as:

$$Min(e) \quad \textbf{(2. 54)}$$

Were *e* is the maximum absolute deviation. Thus the constraints of the program are:

$$-\sum_{j=1}^{m} x_{ij}b_j + e \geq -y_i \quad \textbf{(2. 55)}$$

$$\sum_{j=1}^{m} x_{ij}b_j + e \geq y_i \quad \textbf{(2. 56)}$$

$i = 1, 2, \ldots.., k$

b_j unrestricted in sign for each j

e, non-negative.

Kiountouzis (1973) uses Monte Carlo simulation to evaluate the use of the MinAD regression in comparison with the least squares method (that minimises SSE). He conducted the simulation using four different error distributions, uniform, normal, Laplace and Cauchy. The experiments demonstrated that the linear programming was a better estimator of the parameters of the linear regression when the errors were given in a Laplace or Cauchy probability distribution. Nevertheless, the study illustrates that it may introduce a small bias even in symmetrical error distributions.

Arthanari and Dodge (1981) present two additional *best-fit* criteria, the Minimum Absolute Difference between Deviations (ADBD) and the Minimum Absolute Difference between Absolute Deviations (ADBAD). Similarly to the MinAD and MinMaxAD models, a linear program is used in order to calculate the parameters of the regression equation that minimises the ADBD and ADBAD coefficient of x_{ij} and e_i the regression error, i $(1 \leq i \leq k)$ is the index of the observation and j

respectively. If x_{ij} is the independent variable, y_i the dependent variable, b_j the ($1 \leq j \leq n$) the index of the coefficient then the linear programming formulation for the estimation of the ADBD of a linear regression is the following:

$$MinADBD$$

Objective function:

$$Min \sum_{i=1, i<l}^{k} |e_i - e_l| \quad \textbf{(2. 57)}$$

for e_{1il} and e_{2il} the absolute positive and negative differences between the errors e_i and e_l, this can be written as:

$$Min \sum_{i=1, i<l}^{k} (e_{1il} + e_{2il}) \quad \textbf{(2. 58)}$$

Subject to the following constraints:

$$e_i - e_l - e_{1il} + e_{2il} = 0 \quad \forall i < l \quad \textbf{(2. 59)}$$

$$\sum_{j=1}^{m} x_{ij} b_j + e_i = y_i \quad \textbf{(2. 60)}$$

For

i =1, 2, ….., k and l [i, k]

e_{1il}, e_{2il} non-negative and e_i and b_j unrestricted in sign.

Similar will be the formulation of the program that minimises the *ADBAD* is as follows.

$$MinADBAD$$

$$Min \sum_{i=1,\, i<l}^{k} \left\| e_i \right| - \left| e_l \right\| \quad \textbf{(2. 61)}$$

This can be written as:

$$Min \sum_{i=1,\, i<l}^{k} \left(e_{1il} + e_{2il} \right) \quad \textbf{(2. 62)}$$

For e_{1il} and e_{2il} the absolute positive and negative differences between the errors e_i and e_l,.

Subject to the following constraints:

$$e_{1i} + e_{2i} - e_{1l} - e_{2l} - e_{1il} + e_{2il} = 0 \quad \forall i < l \quad \textbf{(2. 63)}$$

Where e_{1i} is the *under-estimation* and e_{2i} the *over-estimation* errors of period *i*

And **(2. 52)**

For

i =1, 2, ….., *k* and *l* [*i*, *k*]

e_{1i}, e_{2i} , e_{1il}, e_{2il} non-negative and b_j unrestricted in sign.

One of the most studied applications of linear programming for solving typically statistical problems is the area of classification and specially *discriminant analysis*. Discriminant analysis is an "*approach to conveniently identify significant subsets of individual observations by exploring common and contrasting features*" (Freed and Glover, 1981). Freed and Glover (1981) present how LP can be used to solve a simple two groups classification problem. Similar to regression, the objective function of the discriminant analysis problem is to minimise the sum of the classification errors, where the constraints are the equations of each observation. The program has to specify the parameters of the equations that minimise the objective function.

Further research on the field has shown that LP models can exceed the limitations of traditional statistic tools and, in turn, deal with more complex problems. Examples are Freed and Glover (1981) who develop a Linear Goal Programming model for classifying observations into more than two groups. Moreover, a hybrid LP model that integrates the simple LP with the previous goal programming has been presented in Glover (1990). In addition, non-linear extensions of the LP model, which deal with higher complexity, can found in the literature (e.g. Glen, 2003 and 2006).

In the field of time series analysis, Feigin and Resnik (1992, 1994, 1995, 1997 and 1999) and Feigin et al. (1996) suggested a linear programming based method for estimating the parameters of a stationary autoregressive function with heavy tailed error distributions (e.g. Pareto distribution and Levy distribution). For the autoregressive processes:

$$Y_i = \sum_{j=1}^{m} b_j Y_{i-j} + e_i \qquad \textbf{(2. 63)}$$

Where Y_i is the predicted variable, Y_{i-j} are the explanatory variables, b_j is the coefficient of Y_{i-j}, e_i is the forecasting error, i ($1 \le i \le n$) is the index of the forecasting period and j ($1 \le j \le k$) is the order. The formulation of the linear program is:

$$Max \sum_{j=1}^{m} b_j \qquad \textbf{(2. 65)}$$

Subject to:

$$Y_i \ge \sum_{j=1}^{m} b_j Y_{i-j} \qquad \textbf{(2. 66)}$$

b_j unrestricted in sign.

2.6.2 Mathematical programming for estimating the parameters of forecasting models

While mathematical programming has been shown to be a strong tool for solving statistical problems, the research in the area of forecasting applications is limited. There are a few studies about the usage of optimisation as a supplementary tool for forecasting. Examples are this of Mosheiov and Raven (1997) who apply simple linear programming to estimate the trend of a time series. Dhahri and Chabchoub (2007) have presented a non-linear goal programming model as a tool to estimate the optimal order of ARIMA models.

One of the studies on the estimation of the parameters of a forecasting model using optimisation is this of Segura and Vercher (2001). They examined the effectiveness of using spreadsheets for optimising a Holt-Winters forecasting model. They use the spreadsheet's solver to develop a simple non-linear program that estimates the optimal parameters of the forecasting model. In addition, Bermudez et al. (2006) applied the approach of Segura and Vercher (2001) for demand forecasting. They used the Holt-Winters formulation in both additive and multiplicative seasonal forms.

A second study on optimisation for the estimation of the parameters of a forecasting model comes from Amin and Emrouznejad (2007). In their study they used the theory of *inverse optimization* (Tarantola, 1987, Zhang and Liu, 1996). According to Ahuja and Orlin (2001), an inverse linear programming problem (ILPP) is an alternative solution method for a linear program. Their theory indicates that the ILPP is the formulation of a new linear problem "*the solution of it (which is similar to the original problem and the associated dual solutions) can be used to solve the inverse problem*" (Amin and Emrouznejad, 2007). The relationship between the original problem and the corresponding ILPP is that the decision variables of the first are used as given parameters to the second, and vice versa. The optimal solution of the original problem is the feasible solution that will give zero as optimal solution to the ILPP.

In their study, Amin and Emrouznejad (2007) have applied inverse linear programming to estimate the parameters of a causal forecasting model. Specifically, they apply inverse linear programming to minimise the MAD of a linear regression. The accuracy of their model is validated by its application on a small sample of actual data.

Trapp (1986) uses linear programming to estimate the parameters of a wheat econometric forecasting/decision model. The forecasting part of the model predicts the storage profit and loss, while the decision part considers the alternatives of either to store or not. The model consist of a multiple linear regression where the dependent variable is the rate of return to storing and the independent variables are the available quantity of wheat production and the changes of wheat stocks during a year. The linear program estimates the parameters of this equation that maximise the profits of the two alternative decisions. The results are compared with an ordinary least squares (OLS) based econometric model on a twenty four years period data set. The comparison shows that the LP formulation results in higher profits compared with the OLS formulations, while the latter is superior in terms of accuracy.

Soliman et al. (1997) use linear programming to minimise the absolute error and the OLS method (minimises MSE) for the applications of short term electricity load forecasting. The models consist of a multiple linear regression, a harmonic decomposition and a hybrid application and they are tested on a contaminated data set with 20% gross error load measurements and an uncontaminated data set without error measurements. The results show that if the data are coming from a contaminated source, both techniques perform very similarly; while, if the source is uncontaminated, linear programming formulations were more accurate.

Mohammadi et al. (2006) used non-linear goal programming to estimate the parameters of an ARMA model for river flow forecasting. The formulation is very similar to the linear program for linear regression model. The first objective is to minimise the absolute deviation in the whole series and the second in a specific periods in the series. The results were compared with the maximum likelihood method. The comparison showed that the second approach was more accurate and better in terms of computational time.

2.6.3 Mathematical programming for combining forecasts

Mathematical programming has been used as a method to combine forecasts. The initial linear programming formulation is similar to the linear regression formulation where:

Objective function:

$$Min \sum_{t=1}^{T} \left| F_{ct} - Y_t \right| \qquad \textbf{(2. 67)}$$

This can be rewritten as:

$$Min \sum_{t=1}^{T} e_{1t} + \sum_{t-1}^{T} e_{2t} \qquad \textbf{(2. 67)}$$

Subject to

$$F_{ct} = \sum_{i=1}^{k} w_i F_{it} \qquad \textbf{(2. 68)}$$

$$\sum_{i=1}^{k} w_i F_{it} + e_{1t} - e_{2t} = Y_t \qquad \textbf{(2. 69)}$$

with
$$\sum_{i=1}^{k} w_i = 1 \qquad \textbf{(2. 70)}$$

and
$$0 \leq w_i \leq 1 \qquad \textbf{(2. 71)}$$

for e_{1t}, e_{2t} non negative

were there are k forecasts that are combined. F_{ct} is the combining forecast and Y_t the actual value at time t, F_{it} is the result of forecast i ($1 \leq i \leq k$), w_i is the weight of forecast i and e_{1t} *and* e_{2t} the underestimation and overestimation errors of the combined forecasts respectively.

The first time that linear programming was used to combine forecasts was by Reeves and Lawrence (1982). In their study they developed two multi-objective linear programming models to combine three simple forecasting techniques in a series of thirty periods. The first linear program minimises the sum of absolute errors in the whole series, the sum of the positive errors in the whole series and the sum of the absolute errors in the most recent data points of the series; the second minimises both the sum of the absolute errors and the maximum absolute error in the whole series. The authors present different combinations that come from alternative possible solutions of both models and give the decision maker the opportunity to select the most favorable. The authors state that while their methodology generates efficient combinations, the presented example cannot be used to draw general conclusions about the applicability of the method.

Reeves et al. (1988) apply the first model of the above paper in a case study to combine the forecasts of the earnings of six major corporations. The results show that the combined forecast utilises all the individuals and it outperforms their accuracy in general.

Similar to the above, Zhou et al. (1999) present a multi-criteria multi-constraint (MC^2) linear programming formulation for combining forecasts. The multi-criteria part reflects the difference in importance of different periods in the series, while the multi-constraint part presents the opinions of different forecasting experts. The authors applied their model to forecast the development rate of the telecommunications industry in south China for five years. They used three different techniques and they considered the opinions of three experts of different backgrounds. The analysis shows that the combined forecast outperforms the individual techniques that have been used; however, the results are not compared with different combination methods.

Lam et al. (2001) present two models that minimise the sum of the absolute percentage errors and the maximum absolute percentage error. They apply their model on the data and the techniques used by Reeves and Lawrence (1982) and they compare their results with three other combination techniques, average, variance – covariance (Kang, 1986) and odds matrix (Clemen and Winkler, 1986, Gupta and Wilton, 1987). The results show that the LP models generally outperformed the other three combination techniques.

Reeves and Lawrence (1991) present a multi-objective linear integer programming model and a multi-objective linear programming model, which is a relaxed version of the former. The models minimise the sum of the absolute errors and the number of periods in which the forecasted direction of change is incorrect. The models are tested on the data by Reeves and Lawrence (1982) and an additional data set. The results show that the combined forecast was better than the individuals in terms of accuracy.

Leung et al. (2001) use non-linear preemptive goal programming to combine three individual forecasts according to their performance on investment returns. The objectives are to minimise the mean (expected) return, the variance (risk) and the skewness of the investment respectively. They tested their model on a period of sixty observations of three stock market indices (S&P 500, FTSE 100 and Nikkei 250) and they compare the accuracy with the individual forecasts and four other combination techniques (simple average, inverse proportion to MAD and MSE and weights based on the absolute error performance). The results indicated that the proposed model outperforms the individual forecasts and the other combination techniques, and its superiority becomes clear when the market displays significant volatility and instability.

Finally, Lawrence et al. (2010) have formulate a preemptive GP model for combining forecasts, where the first objective is to minimise the sum of the absolute deviations on the sum on all the series, the second is to minimise the sum of the underestimation deviations and the third to minimise the sum absolute deviations in the last five periods of the series. The authors try to improve the initial formulation with two alternatives. The first is by *fuzzifying* the constraints of the program with give tolerance values (*Fuzzy Approach Using Soft Constraints*, FAUSC). The second is by transforming the preemptive GP model into a weighted additive GP model, where the objective is to optimise a simple additive fuzzy objective function (*Fuzzy Approach Weighted Additive Model*, FAWAM). The objective function consists of multiplying each membership of the fuzzy goal with the weight of importance and adding them together.

The models are tested on the data by Reeves and Lawrence (1991). The analysis shows that the initial preemptive GB model improves the results of the individual forecasts. However, FAUSC and FAWAM significantly improve the results further. The differences between the two are relatively small.

2.6.4 Discussion

As we can see, the research in the field is very limited. However, the studies indicate that an LP based approach may performs better than the counterpart OLS in some cases. Specifically, in the area of statistical analysis, the study of Kiountouzis (1973) shows that LP is a better estimator when the error are follow specific types of probability distributions (Laplace or Cauchy) and Feigin and Resnik (1992) show that LP is a good method to estimate the parameters of an autoregressive model for a stationary series with positive innovations. Nevertheless, these studies are rather in the field of theoretical statistics than forecasting.

In the field of forecasting, Soliman et al. (1997) shows that LP works better than the OLS on their application and Trapp (1986 shows that LP based forecasts have lower cost of forecasting error than the OLS) in his case study. However, both studies are the area of causal forecasting and they are rather real life applications than research in forecasting theory. The limited studies in the field (there is no study in time series forecasting) and the sign that LP may give good results indicate that this is a promising area for further research.

On the other hand, there are many studies in the area of LP for combined forecasting. The studies show that the LP based combinations produce good results, but their performance has never been compared with different combination methods. In addition, LP does not have the main disadvantage of the traditional combination techniques because it does not have to give a weight to every individual technique that is used, but it can easily distribute the combination weights between the most accurate individual forecasts according to the optimisation criterion. Thus, the comparison of LP based combinations with traditional combination methods is also an interesting area for further research.

2.7 THE FORECASTING COMPETITIONS AND THE FUTURE OF FORECASTING RESEARCH

2.7.1 The M-Competitions

The M-Competition was established by Makridakis in 1982 in a paper which studied the post-sample accuracy of several time series forecasting methods (Makridakis et al. 1982). The participants tested the accuracy of 24 methods on 1001 series with various horizons. The objective of the competition was to investigate how different techniques differ from each other and how information can be provided so that forecasters can be able to make practical choices under different circumstances (Makridakis et al. 1984). According to Makridakis et al. (1984) the most significant conclusions of the competition were the following:

- The performance of a technique may differ according to different accuracy measures. Differences among methods were influenced by differences in the type of series used and the length of the forecasting horizon.
- The nature of the series, such as the period (e.g. monthly, yearly) and data types (e.g. financial, demand) affects the forecasting accuracy of different techniques.
- Seasonality, trend and randomness are the main factors that affect accuracy.
- Sophisticated methods do not perform better than simpler techniques in patterns with considerable randomness.
- Both simple and sophisticated methods perform equally in patterns with a strong element of seasonality.
- Combined forecasting reduces the forecasting error significantly.

A major criticism of M-Competition was that the performance of statistical models was tested only, while in real situations forecasters use additional qualitative information to improve forecasts (Makridakis et al., 1993). This criticism resulted in a second M-Competition (the M2-Competition) in 1987 that aimed to investigate forecasting under real situations. Five forecasting experts were required to produce forecasts for a number of companies using a combination of qualitative and quantitative models. In addition, traditional quantitative techniques were applied to the same data. In addition to the conclusions of the first M-Competition, the conclusions of the M2-Competion are summarized as follows:

- Exponential smoothing methods (simple and damped) were the most accurate.
- Experts’ forecasts proved to be poor.
- Combinations of different exponential smoothing methods produce good results, but are not better than individual.
- Combinations of the experts’ forecasts proved better than individual.

In 2000 Makridakis et al. launched a third competition. The aims of the M3-Competition were to clear up the accuracy issues of several forecasting techniques and to extend the results of the previous competition. The extension involved the use of additional techniques (mainly from the area of artificial intelligence), more practitioners and more series (3003). The findings of the M3 confirmed the conclusions of the previous competitions. Particularly:

- Sophisticated techniques are slightly better than simpler techniques for time series with limited variability.
- The performance of a forecasting method differs according to the used accuracy measure.
- The accuracy of combined forecasting is usually as good as or better than the accuracy of the individual methods that were combined as well as than the accuracy of other methods.
- The best method for a series depends on the forecasting horizon.

The three competitions have played a very important role in the forecasting research the last three decades and their results provided as basis for future forecasting research.

In September 2010 Makridakis et al. launched a fourth competition. According to the competition team, the purpose of the M4 – Competition is to further study the accuracy and the utility of several forecasting techniques. For this reason, the number of the series, the categories and the forecasting techniques are increased. The results of the competition have not been published yet.

2.7.2 The NN Competitions

The NN completions are "*replications of the M competitions with an extension towards neural networks (NN) and computational intelligence (CI) methods*" (Crone et al. 2011). The first competition (NN3) took place in 2006 and it was focused on monthly time series forecasting. The practitioners tested forecasting techniques based on neural networks and computational intelligence on a data sample of either 111 or 11 monthly series from the M3 competition. The NN/CI methods were compared with 17 benchmark techniques base on statistics. The results of the competitions drive further conclusions that mainly confirm these of the M competitions. Nonetheless, in contrast with the latter, the analysis shows that more sophisticated techniques based on NN/CI tend to outperform simples forecasting techniques.

The next competition (NN5) took place in 2008 and was focused on extending the NN3 competition on daily time series. The practitioners tested their methods on a data sample of either 111 or 11 daily series from the M3 competition, similarly to the NN3 competition. The results were compared with 12 benchmark statistical techniques and 12 benchmark NN/CI based techniques. The conclusions of this competition have not been published yet.

Finally on 2010 Crone et al. announced a new competition (NNGC1). The aim of this competition is to extent the earlier NN3 and NN5 competitions on a new set of series of multiple frequencies. In this competition the techniques are tested on at least one of six datasets, containing 11 series each. The frequency of the series ranges from yearly to hourly. The results of the competition have not been published yet.

2.7.3 The future of forecasting research

As we can see, forecasting research and practice have improved during the last three decades, and the main reasons are the improved statistical and analytical models and the development of more mature approaches for estimation and evaluation. In addition, the importance of the rising power of computers and the increased information availability are recognised. Their proposals can be summarised in the following ten forecasting research suggestions (De Gooijer and Hyndman, 2006, Ramnath et al., 2008):

1. Improvement of current statistical techniques.
2. Focused research in non-linear forecasting techniques.

3. Development of multivariate techniques that will focus on their practical application.

4. Further study on non-Gaussian methods.

5. Specification of the objectives of the forecasts and development of forecasting systems that address these objectives.

6. Development of quality assessment methods of the information that are used for prediction.

7. Development of model selection techniques that will incorporate effective use of both data and experts knowledge.

8. Further research in density forecasting (density forecasting is an alternative methodology that focuses on prediction of the probability density of future values rather than the average).

9. Improvement of combining forecasts and understanding of the reason why combining works better in some cases than in some others.

10. Improvement of forecast interval.

On the other hand, Fildes et al. (2008) identify four business application areas that have proved fertile to new ideas:

- Forecasting intermittent demand (Syntetos et al., 2009, Syntetos and Boylan, 2010).
- Sales response modeling and direct marketing (Chun, 2012).
- Credit scoring and credit risk management (Malik et al. 2010, Thomas et al. 2011).
- Dealing with the bullwhip effect in a supply chain (Barlas and Gunduz, 2011).

However, about the progress of forecasting research during the last years I endorse the opinion of Keith Ord in Daws et al. (1995), who states that:

> "*we cannot expect to model change (of forecasting research) perfectly, or even to be very good at anticipating it. The best we can hope for is to recognize situations where the potential for change exists. We may then guide policy decisions in the light of that potential, examining the consequences of such changes, if they were to take place.*"

According to Ord, it is not very likely to specify how the current forecasting methods will be improved in the future. Nevertheless, the opinion of academics and business experts is a good guide to estimate the trends of the research in this area for the next years.

2.8 CONCLUSION

The relevant literature of the field was presented in this chapter. As we can see, there is potential to explore LP applications for forecasting. Past studies show that LP is very applicable to theoretical statistics and there are many cases where LP works better when compared with the OLS (e.g. Kiountouzis, 1973, Feigin et al., 1996). Nonetheless, studies on LP applications on forecasting are very limited. All the applications are in the field of causal forecasting and there is no research on time series forecasting. In addition, the review indicates that LP application on combining forecasts could be explored further. After the literature review, the methodology of the study follows.

3 RESEARCH METHODOLOGY

This study belongs to the scientific field of OR. Thus, the methodology is based on the OR model building procedure. In the first section I analyse the development of the linear programs. The second section focuses on the selection of the series, which the LP approaches are tested on. The selection method, decomposition and the statistical analysis of the series are presented. The third section presents the accuracy testing methodology, the comparison methodology of the LP-based approaches against traditional techniques that are found in the literature and the evaluation of the whole process.

3.1 DEVELOPMENT

All the techniques are based on linear programming formulations. The initial basis for them is the proposed LP formulation for regression analysis, as it was presented by Wagner (1959). However, in this study the objective function and the constraints differ according to the objectives of each model. All the AR models are of Order 6 and 12 regardless of the approach. It is possible to conduct identification test to determine the optimal order, using the PACF, AIC, BIC or FPE, as I described in the previous chapter. However, I did not estimate the optimal order of the AR models because the research focuses on the comparison of methods for estimating the parameters of known models rather than finding the optimal model for a series.

3.2 DATA SELECTION AND ANALYSIS

Two data samples were used, one of 60 series and one of 25 series from the 3003 series of the M3 Competition[2]. The reason that I selected a sample of the total 3003 series of the M3 competition is the same reason for the limitation of the series of the NN3 competition. Crone et al. (2011) state that:

> "*To determine the degree of automation or manual tuning required, and to address prevailing concerns on the computational demands of predicting a large number of time series with CIs, we allowed participants to choose between two (disguised) datasets of different sizes. The contestants were asked to predict either a reduced dataset of 11 time series or the complete set of 111 series (which included the reduced set) as accurately as possible. As a fully automated methodology could be applied to large datasets just as*

[2] The M3 competition data are available at the Principles of Forecasting website http://www.forecastingprinciples.com.

> *easily as smaller sets, more submissions for the reduced dataset would indicate the limitations of the automation through need for manual of extremely computational intensive approaches, and indicate the need for further research into methodologies.*"

For the above reason, the series of the competition were limited, and most of the practitioners took the option to use only the small dataset of 11 series. Similarly with the CI forecasting, there are not any statistical/forecasting packages with the option to estimate the parameters of forecasting models using linear programming. Thus, due to the absence of automation, I had to automate a significant part of the process myself. However, manual tuning was still required. In addition, 458 linear programs were formulated for this study, 422 where applied on the 60 series sample and 36 on 35 series (the 25 series of the second sample and 10 series from the first). Hence, in total 26,580 LPs were solved and investigated. If I had used all the 3003 series, I would have to do 1,375,374 LPs, and spend much more time on automation, which was not possible due to the time limitation of a doctoral study.

The initial data sample consists of 60 randomly selected monthly time series and consists of:

- 12 Microeconomic series
- 12 Macroeconomic series
- 12 Series of monthly demand
- 12 Financial series
- 6 Series related to Demographic data
- 6 Series under the category "Other"

The descriptive statistics of the series are presented in Table 3.1. In the first line of the table shows the number of the series as stated in the M3 Competition section of the *Principles of Forecasting* website.

Table 3. 1 Descriptive statistics - Initial sample

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Series Number	N2719	N2524	N2475	N2209	N2718	N2736	N2737	N2572
Mean	7837.1	2836.87	4756.6	3163.96	9577	3226.24	5287.79	4566.62
Standard Error	40.48	92.52	44.03	36.52	49.87	70.23	111.98	125.19
Median	8059	2914.89	4900	3175	9808	3002	5053	4234.47
Mode	8177	-	5300	2950	-	2690	4084	3888.3
Standard Deviation	468.59	1110.2	528.36	438.27	577.29	813.03	1296.23	1449.16
Sample	219579	1232547	279169	192084	333269	661014	1680221	2100051

Variance								
Kurtosis	-0.41	-1.56	0.64	-0.31	-0.68	2.01	1.35	0.13
Skewness	-0.81	-0.25	-1.09	0.13	-0.7	1.29	1.04	0.84
Range	1802	3142.22	2550	2160	2159	4190	6756	6397.5
Minimum	6658	1098.11	3000	2250	8208.5	2032	3160	2025.3
Maximum	8460	4240.33	5550	4410	10367.5	6222	9916	8422.8
Sum	1050172	408509.6	684950	455610	1283318	432316	708564	611927.1
Count	134	144	144	144	134	134	134	134
Largest	8460	4240.33	5550	4410	10367.5	6222	9916	8422.8
Smallest	6658	1098.11	3000	2250	8208.5	2032	3160	2025.3

	9	10	11	12	13	14	15	16
Series Number	N2743	N2738	N2079	N2080	N2056	N2075	N2278	N2284
Mean	7064.46	7377.39	4684.65	3925.21	5697.49	4139.24	4754.07	5247.8
Standard Error	84.01	195.65	49.34	45.33	65.07	22.87	31.67	58.3
Median	6851.75	6997.5	4590	4015	5643	4173	4842.5	5490
Mode	6311.5	7925	4460	3870	-	4178	4795	5550
Standard Deviation	972.45	2264.81	592.05	543.99	780.81	274.44	366.64	674.83
Sample Variance	945667	5129385	350520	295922	609668	75317	134422	455395
Kurtosis	-0.61	-0.15	2.32	-0.76	-0.66	0.44	-0.03	-0.26
Skewness	0.56	0.77	1.26	-0.42	0.16	-0.61	-0.79	-0.03
Range	3522.5	9620	3280	2270	3454.3	1412	1575	3105
Minimum	5657.5	3920	3590	2780	4062.2	3286	3710	3700
Maximum	9180	13540	6870	5050	7516.5	4698	5285	6805
Sum	946637	988570	674590	565230	820438.1	596050	637045	703205
Count	134	134	144	144	144	144	134	134
Largest	9180	13540	6870	5050	7516.5	4698	5285	6805
Smallest	5657.5	3920	3590	2780	4062.2	3286	3710	3700

	17	18	19	20	21	22	23	24
Series Number	N2545	N2617	N2739	N2740	N2081	N2477	N2525	N2741
Mean	8579.62	5121.14	5065.91	7499.57	4465.49	4701.74	6483.97	5052.54
Standard Error	142.21	97.86	111.01	87.45	29.78	72.28	297.09	94.87
Median	8934.11	5281.4	4812.75	7289	4475	4725	5946.8	4820
Mode	-	5721.2	4761.5	7145.5	4110	4300	-	4660
Standard Deviation	1646.24	1132.85	1285.08	1012.3	357.41	867.39	3565.04	1098.16
Sample Variance	2710112	1283353	1651442	1024745	127744	752357	12709495	1205965
Kurtosis	-0.84	-1.22	1.16	-1.36	-0.7	0.89	-1.59	-0.68
Skewness	-0.63	-0.29	0.96	0.19	-0.05	-0.85	0.03	0.56
Range	5522.9	4191	6907	3232	1650	4150	10284.25	4520
Minimum	5197.29	3020.6	2787	6043.5	3560	2150	1533.3	3380
Maximum	10720.19	7211.6	9694	9275.5	5210	6300	11817.55	7900
Sum	1149669	686232.4	678832.5	1004943	643030	677050	933691.1	677040
Count	134	134	134	134	144	144	144	134
Largest	10720.19	7211.6	9694	9275.5	5210	6300	11817.55	7900
Smallest	5197.29	3020.6	2787	6043.5	3560	2150	1533.3	3380

	25	26	27	28	29	30	31	32
Series Number	N2347	N2393	N2523	N2527	N2742	N1420	N1421	N1422
Mean	5651.28	5275.59	3825	4940.66	4942.76	3778.26	4860.14	6085.51

Standard Error	63.9	42.98	122.64	156.92	93.75	159.59	195.92	324.03
Median	5815.425	5208.2	3759.445	4290.65	4800	3550	4600	5500
Mode	-	4942.4	-	-	4180	4000	3950	8900
Standard Deviation	739.69	497.54	1471.69	1883.03	1085.25	1325.64	1627.44	2691.56
Sample Variance	547144	247541	2165857	3545818	1177762	1757315	2648572	7244493
Kurtosis	-0.37	-1.54	-1.57	-1.45	-0.59	0.55	2.95	-0.61
Skewness	-0.16	0.23	-0.16	0.14	0.52	0.71	1.2	0.54
Range	3358	1611.4	4085.35	5825.05	4465	6150	9300	11100
Minimum	3951.1	4522.2	1623.3	2121.1	3115	1300	1900	2100
Maximum	7309.1	6133.6	5708.65	7946.15	7580	7450	11200	13200
Sum	757271.9	706928.6	550800.2	711455	662330	260700	335350	419900
Count	134	134	144	144	134	69	69	69
Largest	7309.1	6133.6	5708.65	7946.15	7580	7450	11200	13200
Smallest	3951.1	4522.2	1623.3	2121.1	3115	1300	1900	2100

	33	*34*	*35*	*36*	*37*	*38*	*39*	*40*
Series Number	N1423	N2095	N2245	N2402	N2600	N2606	N1424	N2745
Mean	5059.42	5216.68	5215.3	5726.29	4874.61	4961.34	3437.9	3462.39
Standard Error	367.88	43.34	51.68	57.1	283.63	157.13	108.46	225.38
Median	4700	5131.4	5297.5	5488.65	3464.75	5059	3150	2280
Mode	1700	4838.2	4420	5246.8	-	6152	3090	2080
Standard Deviation	3055.87	520.06	598.25	660.95	3403.52	1818.96	900.94	2609.01
Sample Variance	9338329	270462	357907	436858	11583959	3308614	811684	6806930
Kurtosis	0.74	0.25	-1.33	-1.49	0.63	-1.3	1.69	1.4
Skewness	0.89	0.52	0	0.22	1.25	-0.08	1.07	1.67
Range	14800	2952.4	1910	2138.3	14800.5	6592.4	4845	10420
Minimum	600	3989.6	4295	4692.5	1139	2241.6	1890	1340
Maximum	15400	6942	6205	6830.8	15939.5	8834	6735	11760
Sum	349100	751201.6	698850	767322.8	701944	664820.2	237215	463960
Count	69	144	134	134	144	134	69	134
Largest	15400	6942	6205	6830.8	15939.5	8834	6735	11760
Smallest	600	3989.6	4295	4692.5	1139	2241.6	1890	1340

	41	*42*	*43*	*44*	*45*	*46*	*47*	*48*
Series Number	N2478	N2084	N1426	N2744	N1428	N2285	N2549	N2551
Mean	4672.57	7964.44	3512.99	2700.07	3718.55	4898.96	4623.3	4596.36
Standard Error	69.4	37.27	79.65	133.87	126.26	47.92	76.6	183.2
Median	4725	8020	3468	1990	3580	5055	4868.14	3936.825
Mode	4350	8040	3090	2020	2480	4890	-	-
Standard Deviation	832.82	447.29	661.58	1549.62	1048.79	554.71	886.76	2120.7
Sample Variance	693595	200070	437690	2401315	1099957	307700	786351	4497381
Kurtosis	0.9	0.09	-0.33	0.88	-0.78	0.54	-0.24	-0.63
Skewness	-0.85	-0.36	0.41	1.48	0.42	-1.07	-0.81	0.76
Range	4050	2480	2918	5760	3900	2455	3514.12	7051.71
Minimum	2250	6640	2466	1250	2100	3280	2610.72	2025.93
Maximum	6300	9120	5384	7010	6000	5735	6124.84	9077.64
Sum	672850	1146880	242396	361810	256580	656460	619521.8	615912.8
Count	144	144	69	134	69	134	134	134
Largest	6300	9120	5384	7010	6000	5735	6124.84	9077.64

Smallest	2250	6640	2466	1250	2100	3280	2610.72	2025.93

	49	50	51	52	53	54	55	56
Series Number	N1429	N1430	N2481	N2493	N1959	N1961	N2304	N2305
Mean	4137.68	6981.88	4815.97	14204.58	4361.34	7357.24	6838.51	4998.88
Standard Error	173.61	302.94	66.03	532.69	73.53	77.54	103.97	84.5
Median	3820	7100	4875	10920	4180	7292.5	6870.925	5210.45
Mode	4900	7350	4100	10600	5300	8110	-	-
Standard Deviation	1442.15	2516.43	792.37	6392.28	851.17	897.57	1203.52	978.19
Sample Variance	2079783	6332424	627855	40861250	724494	805637	1448471	956865
Kurtosis	0.25	1.28	-0.48	-0.66	-0.68	-0.59	-1.12	-0.6
Skewness	0.8	0.06	-0.13	0.7	0.22	0.19	-0.08	-0.16
Range	6600	14300	3700	26740	3720	4070	4476.35	4175.1
Minimum	1800	200	2700	3410	2660	5245	4576.05	2933.7
Maximum	8400	14500	6400	30150	6380	9315	9052.4	7108.8
Sum	285500	481750	693500	2045460	584420	985870	916359.8	669850.3
Count	69	69	144	144	134	134	134	134
Largest	8400	14500	6400	30150	6380	9315	9052.4	7108.8
Smallest	1800	200	2700	3410	2660	5245	4576.05	2933.7

	57	58	59	60
Series Number	N2568	N2569	N2722	N1431
Mean	7611.03	4677.44	5294.87	7750.72
Standard Error	208.65	83.55	12.82	387.76
Median	7057.45	4461.9	5332.9	7900
Mode	6480.5	5367.75	5497.2	12100
Standard Deviation	2415.26	967.13	148.39	3221.01
Sample Variance	5833476	935348	22020	10374889
Kurtosis	0.13	-0.56	-1.46	-0.28
Skewness	0.84	0.34	-0.24	-0.02
Range	10662.5	5069.95	478.8	14800
Minimum	3375.5	2561.7	5062	600
Maximum	14038	7631.65	5540.8	15400
Sum	1019879	626777.2	709512	534800
Count	134	134	134	69
Largest	14038	7631.65	5540.8	15400
Smallest	3375.5	2561.7	5062	600

The first step of the analysis of the series is to calculate the ACF, given by the formula:

$$R(k) = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_{i+k} - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \qquad \textbf{(3. 1)}$$

with lag *k* (i.e. the correlation between data which are time-spaced by *k* time periods) with $\bar{x}$ the mean of all observations x_i for *i* [1, *n*]. The analysis shows that 18 series have a clear seasonal pattern because their autocorreleogram plot of the $R(k)$ values as a function of k shows a peak (maximum) for some value of k. Hence, the value of k determines the level of seasonality (Makridakis, 1998). In figures 3.1 and 3.2 we can see a graph of one of these 18 series and the autocorrelogram of the series. As we can see, the autocorrelation function shows a significant peak for lag 12; hence, we determine 12 periods in a season. The seasonality is clear in the graph of the series.

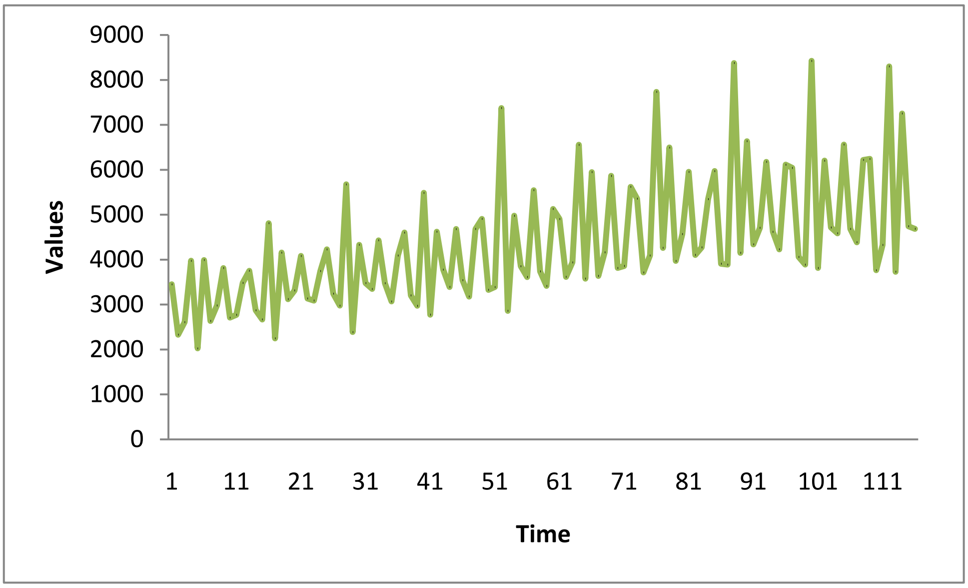

Figure 3. 1 Series with seasonality

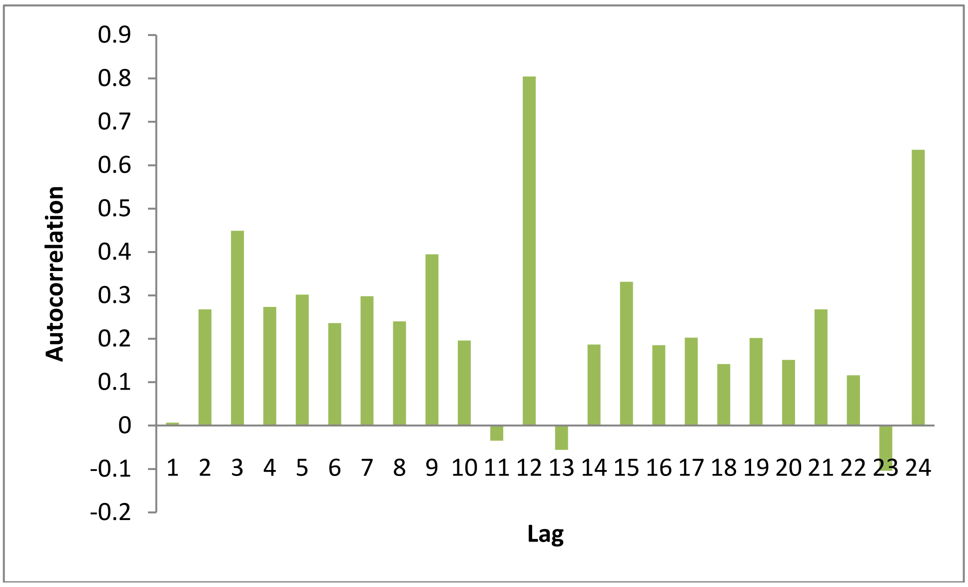

Figure 3. 2 Autocorrelation graph: Series with seasonality

Table 3.2 shows the results of the significance test of the seasonality for these 18 series. Colum *R(k)* shows the autocorrelation coefficient of lag k, where seasonality is observed, and the other two columns show the *t-statistic* and *P-value*. In order to accept the H_0 hypothesis (there is statistical significant seasonality of period *k*) and reject the H_1 hypothesis (there is not statistical significant seasonality) the *P-value* should be less than 0.01 for 99% confidence level.

Table 3. 2 Seasonality - Statistical significance test

Seasonality - Statistical significance test							
Series	*R(k)*	*t Stat*	*P-value*	*Series*	*R(k)*	*t Stat*	*P-value*
3	0.843	17.87	8.2E-37	22	0.841	17.72	1.78E-36
4	0.871	20.18	7.05E-42	24	0.793	14.27	1.26E-27
6	0.802	14.72	1.2E-28	29	0.750	12.41	2.95E-23
7	0.670	9.89	3.12E-17	34	0.648	9.70	4.64E-17
8	0.971	44.14	4.87E-76	41	0.799	15.17	1.48E-30
10	0.754	12.56	1.32E-23	51	0.926	27.90	9.84E-57
13	0.929	28.61	6.05E-58	53	0.803	14.76	9.64E-29
14	0.860	19.20	8.99E-40	57	0.971	44.14	4.87E-76
19	0.652	9.43	3.89E-16	58	0.816	15.45	2.56E-30

As we can see, for all of the 18 series *R(k)* is close to 1 and *P-value* is significant lower than 0.01; thus, the seasonal pattern of the series is statistically significant.

On the other hand, for 32 of the series the *R(k)* is monotoneously decreasing; hence this is a indication of strong trend. In Figures 3.3 and 3.4 we can see a graph of a series with significant trend and the autocorrelogram of the series. Here, the values of the autocorrelation are monotonously decreasing with increasing lag; hence we conclude no strong seasonality, but a significant trend.

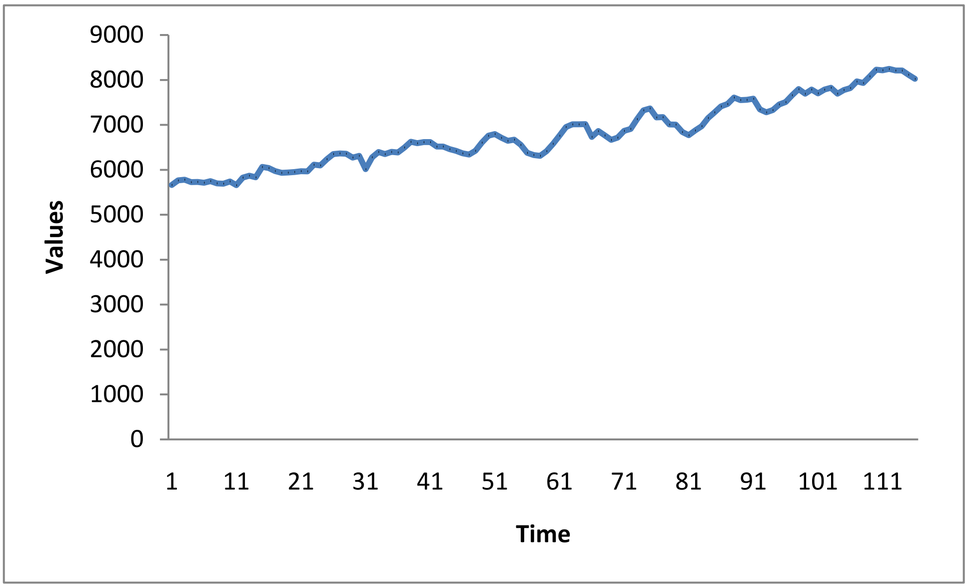

Figure 3. 3 Series with strong trend

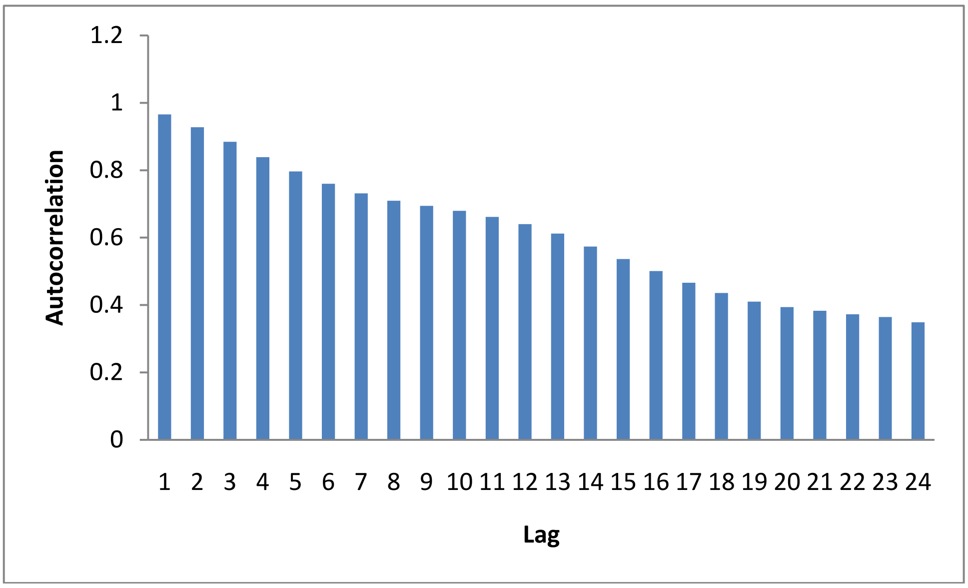

Figure 3. 4 Autocorrelation graph: Series with trend

Finally, for the last 10 series the autocorrelation graph is neither monotonously decreasing nor showing a strong seasonality. This indicates mainly randomness. (All these series have a high degree of variability as shown later in this chapter). Figures 3.5 and 3.6 show the graph and autocorrelogram for one of these series.

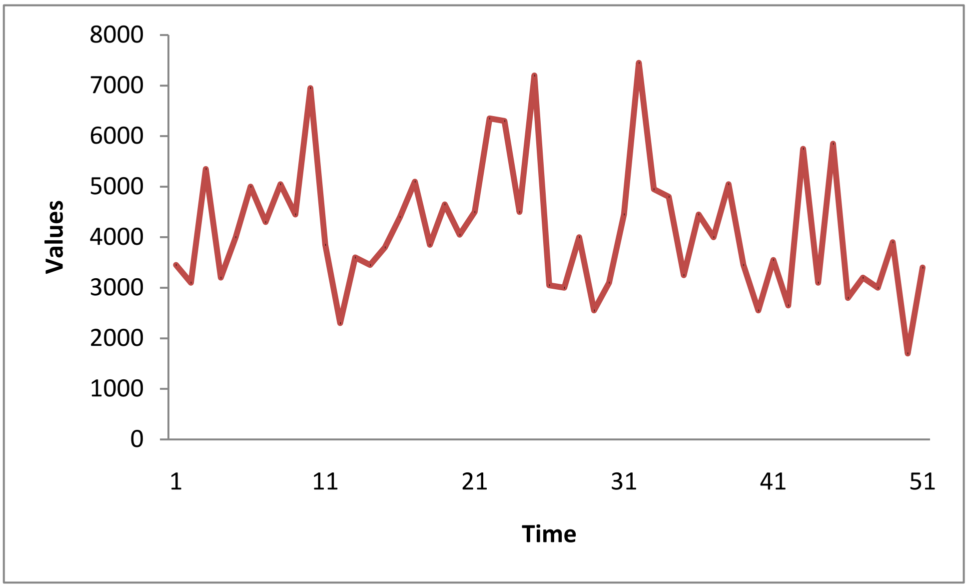

Figure 3. 5 Series with indication of high variability

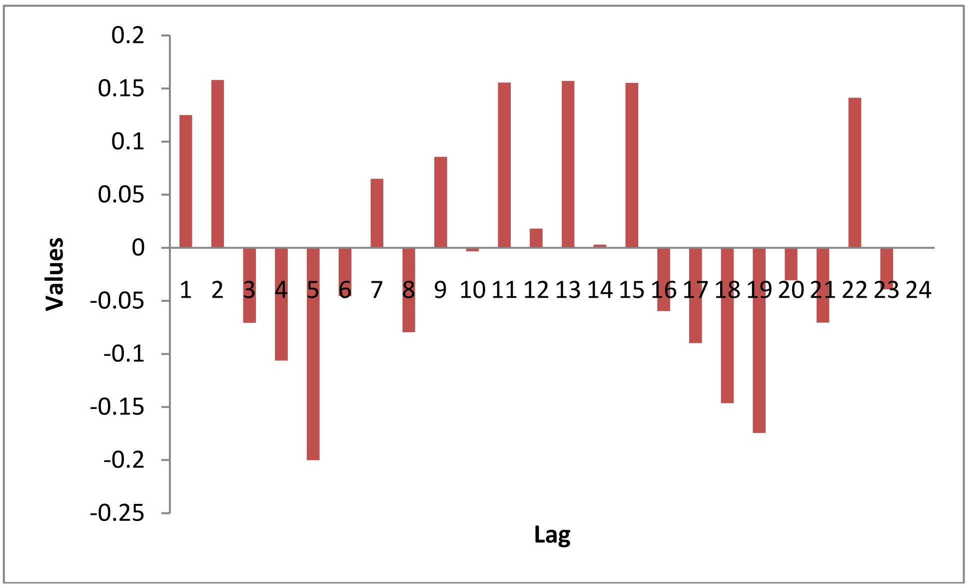

Figure 3. 6 Autocorrelation graph: Series with indication of high variability

The second step is the trend analysis. In order to estimate the trend (or cycle) pattern of the series I smooth the series using a double moving average of order 4 × 4 (Makridakis, 1998). Figure 3.7 shows the graph of one of the series and its smoothed trend.

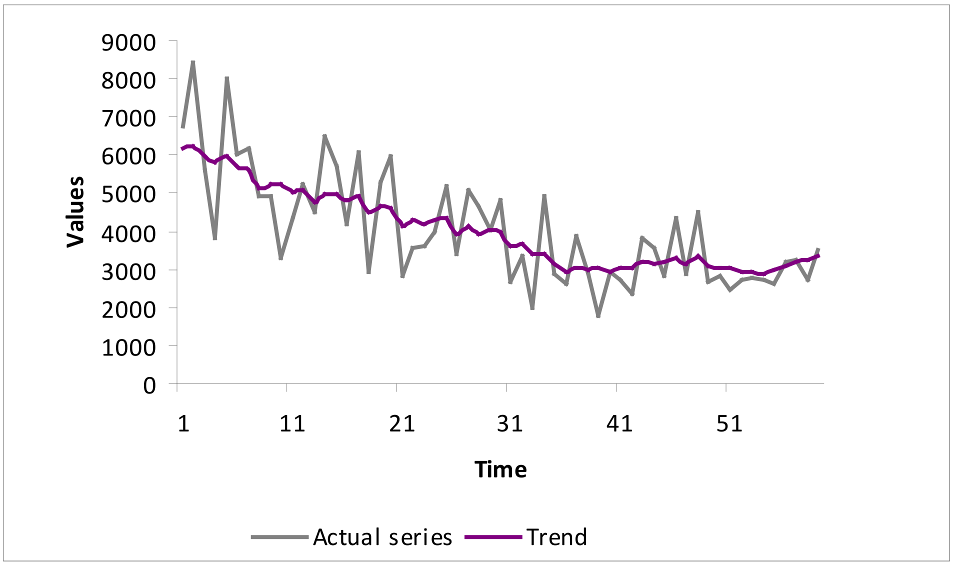

Figure 3. 7 Series with smoothed trend

Next, I estimate the correlation between the smoothed series and time *t*. Table 3.3 shows the correlation (*R*) between trend and time for all the series of the sample (seasonal series are in green). As we can see, most of the series show a significant correlation between the smoothed series and time, indicating a significant trend. Five series show low correlation and *P-value* more than 0.01, three of them are seasonal (series 6 with *R* = - 0.012 and *P* = 0.895, series 7 with *R* = 0.135 and *P* =

0.126, series 19 with *R* = 0.144 and *P* = 0.102) and two non seasonal (series 42 with *R* = 0.143 and *P* = 0.104, series 54 with *R* = 0.004 and *P* = 0.963).

Table 3. 3 Trend - Statistical significance test - Initial sample

Trend - Statistical significance test							
Series	*R*	*t Stat*	*P-value*	Series	*R*	*t Stat*	*P-value*
1	0.918	26.26	2.16E-53	31	-0.886	-15.14	1.14E-22
2	0.978	53.34	2.72E-89	32	-0.874	-14.27	2.12E-21
3	-0.58	-8.06	4.55E-13	33	-0.782	-9.96	1.48E-14
4	0.827	16.65	8.11E-34	34	0.709	11.36	4.12E-21
5	0.941	31.49	3.91E-62	35	0.992	86.36	2.4E-115
6	-0.012	-0.13	0.895	36	0.947	33.23	8.53E-65
7	0.135	1.54	0.126	37	0.824	16.47	2.01E-33
8	0.973	47.58	3.04E-83	38	0.881	21.07	1.92E-43
9	0.967	43.12	4.34E-78	39	-0.831	-11.86	1.05E-17
10	0.606	8.62	2.11E-14	40	0.633	9.26	6.31E-16
11	-0.5	-6.53	1.4E-09	41	0.76	13.23	1.05E-25
12	0.913	25.25	1.51E-51	42	0.143	1.64	0.104
13	0.918	26.10	4.23E-53	43	-0.527	-4.05	0.00014
14	0.779	14.07	9.49E-28	44	0.562	7.68	3.64E-12
15	0.762	13.30	6.87E-26	45	-0.772	-9.66	4.83E-14
16	0.958	37.95	1.68E-71	46	0.786	14.40	1.53E-28
17	0.943	32.14	3.8E-63	47	0.926	27.74	5.67E-56
18	0.925	27.58	1.03E-55	48	0.958	38.01	1.38E-71
19	0.144	1.65	0.102	49	-0.86	-13.35	4.94E-20
20	0.991	83.27	2.4E-113	50	0.523	3.12	0.00027
21	0.681	10.52	5.03E-19	51	0.954	35.90	1.09E-68
22	0.79	14.60	5.06E-29	52	0.738	12.36	1.36E-23
23	0.988	71.20	8E-105	53	-0.454	-5.76	5.91E-08
24	0.593	8.33	1.09E-13	54	0.004	0.05	0.963
25	0.955	36.61	1.14E-69	55	0.995	118.10	1.6E-132
26	0.964	40.77	3.45E-75	56	0.971	45.59	5.3E-81
27	0.979	54.68	1.28E-90	57	0.973	47.58	3.04E-83
28	0.951	34.95	2.48E-67	58	0.984	62.50	8.81E-98
29	0.485	6.27	5.11E-09	59	-0.961	-39.56	1.22E-73
30	-0.698	-7.74	1.01E-10	60	0.727	8.41	6.77E-12

The last step is the analysis of variability. The ACF shows that there is an indication that ten of the series have randomness only. First of all, the trend and seasonal components need to be removed. The element of randomness of a series can be estimated with the decomposition of the series for a period *i* [1, *n*]. For the decomposition of the series, I chose the unobserved decomposition model, because forecasting practice has shown that this model is very applicable on most time series regardless of their characteristics (Newbold and Bos, 1994):

$$Y_i = (T_i + E_i) \times S_i \quad \textbf{(3. 2)}$$

Hence:

$$E_i = \frac{Y_i}{S_i} - T_i \quad \textbf{(3. 3)}$$

Where E_i is the random pattern of period i, Y_i the actual observation, T_i the trend pattern that is the double moving average of period i, and S_i is seasonal pattern. For a seasonality of k periods there are k different seasonal indices that are repeated every k periods. The value of S_i is given as the average value of a k long period of trend adjusted ($Y_i - T_i$) observations (Makridakis, 1998). For series with without seasonality (k = 0) there are no seasonal indices. The series E_i is random noise with average 0.

In order to estimate the variability of the series the initial series has to be readjusted in a stationary series with stable mean and variance through time. This can be done by simply adding the mean of the initial series to the E_i series. Figures 3.8 and 3.9 present a series of the sample with the counterpart readjusted series (with the trend and seasonality removed) respectively.

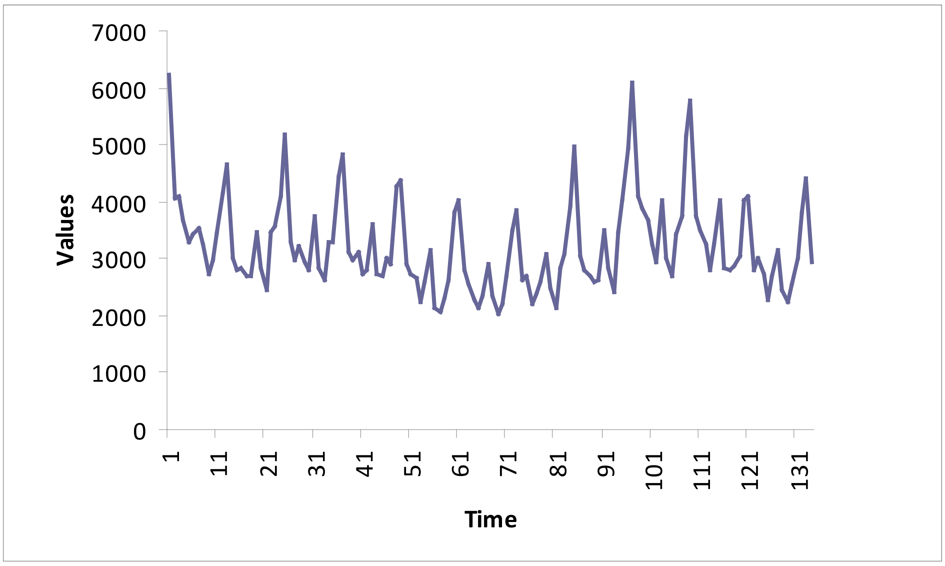

Figure 3. 8 Series before the trend and seasonal adjustment

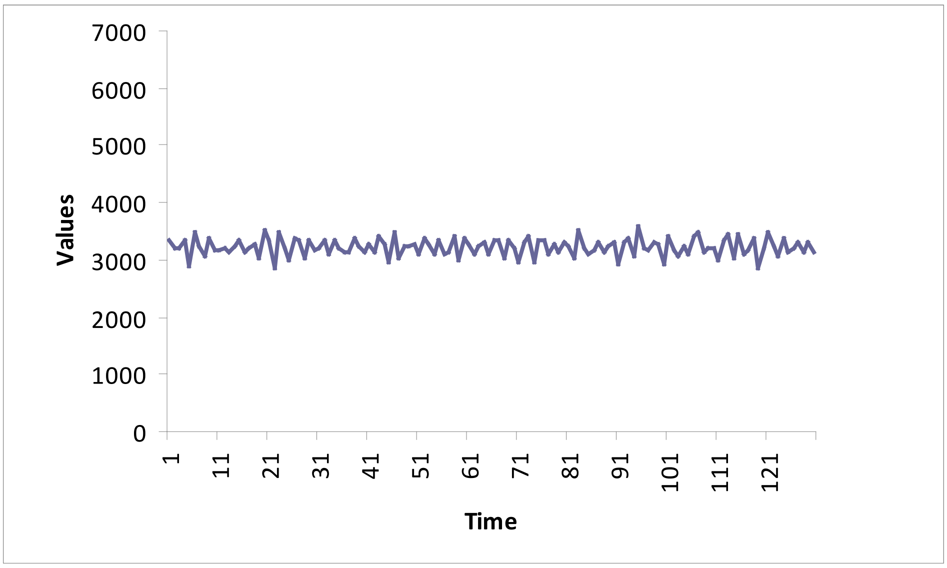

Figure 3. 9 Seasonal and trend adjusted series

The level of variability can be measured by the *Index of Dispersion* (IoD, variance over mean) of the readjusted series. The index of dispersion of all the series is presented in table 3.4.

Table 3. 4 Index of dispersion - Initial sample

Index of dispersion							
Series	IoD	Series	IoD	Series	IoD	Series	IoD
1	0.051	16	0.125	31	243.315	46	0.654
2	0.289	17	0.544	32	369.107	47	2.112
3	4.936	18	2.764	33	736.334	48	1.117
4	5.302	19	14.333	34	7.940	49	179.632
5	0.052	20	0.150	35	0.717	50	438.151
6	7.014	21	3.519	36	0.095	51	1.428
7	2.984	22	5.114	37	26.007	52	59.736
8	30.745	23	3.687	38	2.523	53	12.609
9	0.474	24	4.008	39	119.608	54	5.467
10	21.738	25	0.924	40	21.876	55	0.434
11	8.699	26	0.115	41	6.382	56	1.926
12	6.345	27	0.255	42	7.166	57	25.753
13	2.549	28	3.351	43	101.773	58	26.649
14	0.947	29	14.580	44	6.300	59	0.063
15	0.369	30	221.866	45	129.740	60	537.361

As we can see, the ten series (for which the autocorrelation graph indicated high variability), have a very high IoD (always more than 100) in comparison with the other series (where the highest IoD is 59.73). We can conclude that these are series of high variability. All the series with high variability are series with trend, because they show significant correlation between their trend pattern and time (Table 3.3).

According to the above analysis the series of the first sample can be classified in three main categories.

- 32 smooth series
- 18 series with strong seasonal pattern
- 10 series with high variability (hard series)

For Chapter 6, I selected an additional data sample of 25 series from the M3-Competition with high variability with selection criteria the autocorrelation graph (indication of no seasonality and no strong trend) and the IoD of the readjusted series (over 100). All these series have 69 data points. The sample that was used in Chapter 6 is the "hard series" group, which consists of the 10 series with high variability of the first data sample and the 25 series of the additional sample.

The descriptive statistics of the additional sample are presented in Table 3.5.

Table 3. 5 Descriptive statistics - Second sample

	1	2	3	4	5	6	7	8
Series Number	N1437	N2506	N1678	N1676	N1672	N1661	N1659	N1655
Mean	5549.28	15465.87	3597.83	4219.57	5090.58	2469.28	4901.45	3592.75
Standard Error	221.63	464.24	183.54	184.42	296.9	186.83	243.96	157.39
Median	5400	15330	3500	3950	4750	2180	4700	3550
Mode	5550	11695	3500	3000	4800	2800	3250	3400
Standard Deviation	1840.97	3856.3	1524.64	1531.94	2466.26	1551.96	2026.46	1307.4
Sample Variance	3389154	14871088	2324517	2346854	6082447	2408583	4106542	1709285
Kurtosis	0.39	-1.11	1.67	-0.12	1.24	3.96	-0.35	0.48
Skewness	0.66	-0.03	0.7	0.61	0.98	1.68	0.47	0.66
Range	8800	14930	7700	7050	11900	8240	9350	6250
Minimum	2350	7595	800	1400	1300	400	850	1300
Maximum	11150	22525	8500	8450	13200	8640	10200	7550
Sum	382900	1067145	248250	291150	351250	170380	338200	247900
Count	69	69	69	69	69	69	69	69
Largest	11150	22525	8500	8450	13200	8640	10200	7550
Smallest	2350	7595	800	1400	1300	400	850	1300

	9	10	11	12	13	14	15	16
Series Number	N1645	N1642	N1639	N1637	N1629	N1622	N1617	N1616
Mean	3843.48	7439.13	3192.75	4939.13	7214.49	3025.22	4984.78	4848.84
Standard Error	141.23	405.52	182.73	243.82	349.38	120.37	294.1	130.02
Median	3700	7800	2880	5100	6900	3000	4800	4680
Mode	5150	8700	2880	3600	6300	2820	5400	4440
Standard Deviation	1173.13	3368.51	1517.84	2025.3	2902.18	999.89	2442.97	1080.06
Sample Variance	1376244	11346829	2303847	4101829	8422655	999784	5968111	1166522
Kurtosis	0	-0.78	1.58	0.36	-0.42	0.25	-0.14	-0.15

Skewness	0.63	0.04	1.06	0.57	0.32	0.27	0.42	0.01
Range	5150	13500	7800	9600	12250	4920	11400	5040
Minimum	1900	900	360	1200	1850	1020	600	2100
Maximum	7050	14400	8160	10800	14100	5940	12000	7140
Sum	265200	513300	220300	340800	497800	208740	343950	334570
Count	69	69	69	69	69	69	69	69
Largest	7050	14400	8160	10800	14100	5940	12000	7140
Smallest	1900	900	360	1200	1850	1020	600	2100

	17	*18*	*19*	*20*	*21*	*22*	*23*	*24*
Series Number	N1611	N1604	N1569	N1551	N1540	N1537	N1519	N1508
Mean	4741.45	3073.33	3476.09	5489.86	4045.51	4065.51	4834.06	5406.52
Standard Error	215.68	178.63	104.18	110.95	99.12	123.08	111.14	114.57
Median	4680	2760	3450	5350	4120	4050	4900	5350
Mode	4920	2280	2850	5100	3760	3660	5050	5350
Standard Deviation	1791.57	1483.81	865.37	921.66	823.34	1022.34	923.21	951.7
Sample Variance	3209721	2201678	748868	849454	677881	1045181	852316	905729
Kurtosis	0.61	-0.42	-0.02	0.82	-0.29	-0.09	0.28	0.31
Skewness	0.47	0.71	0.18	0.64	-0.33	0.14	0.19	0.64
Range	8760	5500	4350	4650	3460	4620	4300	4450
Minimum	720	980	1650	3750	2000	1770	2850	3850
Maximum	9480	6480	6000	8400	5460	6390	7150	8300
Sum	327160	212060	239850	378800	279140	280520	333550	373050
Count	69	69	69	69	69	69	69	69
Largest	9480	6480	6000	8400	5460	6390	7150	8300
Smallest	720	980	1650	3750	2000	1770	2850	3850

	25
Series Number	N1480
Mean	4041.74
Standard Error	99.73
Median	4000
Mode	3380
Standard Deviation	828.41
Sample Variance	686256
Kurtosis	0.49
Skewness	0.57
Range	4040
Minimum	2360
Maximum	6400
Sum	278880
Count	69
Largest	6400
Smallest	2360

The correlation between time and the smoothed series (trend) can be found in Table 3.6.

Table 3. 6 Trend - Statistical significance test - Second sample

Trend – Statistical significance test							
Series	*R*	*t Stat*	*P-value*	Series	*R*	*t Stat*	*P-value*
1	-0.578	-5.62	4.672E-07	14	0.546	4.08	0.00013
2	0.702	7.82	7.455E-11	15	-0.495	-3.76	0.00045
3	-0.827	-11.70	1.943E-17	16	-0.569	-3.15	0.00246
4	-0.505	-4.64	1.815E-05	17	-0.557	-5.33	1.42E-06
5	-0.542	-5.12	3.082E-06	18	-0.843	-12.42	1.37E-18
6	-0.503	-2.82	0.0004155	19	0.641	6.63	8.77E-09
7	-0.639	-6.59	1.022E-08	20	-0.518	-0.95	0.347666
8	-0.594	-2.45	0.0172506	21	0.629	6.43	1.95E-08
9	-0.573	-5.55	6.097E-07	22	0.648	3.98	0.00018
10	-0.729	-8.46	5.595E-12	23	-0.676	-7.28	6.5E-10
11	0.616	6.21	4.637E-08	24	-0.688	-7.53	2.4E-10
12	0.546	3.37	7.154E-05	25	0.506	3.50	0.00062
13	-0.514	-3.93	0.00022				

As we can see all the additional series have a statistically significant trend pattern (all the *R* values are significant and all the *P-values* are smaller than 0.01).

The IoD can be found in Table 3.6 and shows that the new sample consist of series with high variability, since it is more than 100 for all the series.

Table 3. 7 Index of dispersion - Second sample

Index of dispersion					
Series	IoD	Series	IoD	Series	IoD
1	359.276	11	581.117	21	103.969
2	127.913	12	608.203	22	112.468
3	328.912	13	609.194	23	151.285
4	327.696	14	244.120	24	104.868
5	600.741	15	748.488	25	104.035
6	785.349	16	191.661		
7	296.150	17	488.783		
8	317.867	18	196.242		
9	197.637	19	132.905		
10	875.098	20	102.685		

3.3 TESTING, COMPARISON AND EVALUATION

3.3.1 Testing

In order to test the forecast I considered using statistical significance testing. However, Armstrong (2007) strongly supports that significance tests are not only unnecessary even when properly done, but also harm progress in forecasting. Armstrong states that significance tests harm the development of science in many ways. One reason is that there is a bias against publishing studies that fail to reject the null hypothesis, even if they might contain important findings. In addition, there is the fact that the null hypothesis is generally selected because its simplicity and not for its truth and importance. The author concludes that researchers should avoid tests of statistical significance when reporting their findings and journals should discourage them.

Schmit and Hunter (1997) state that reliance on significance test is indefensible and makes it difficult for the researchers to develop knowledge. They conclude that "*Statistical significance testing retards the growth of scientific knowledge; it never makes a scientific contribution.*" These opinions about significance testing are appreciated by many researchers (e.g. Goodwin, 2007).

Because of the conclusions of the above papers I have not used statistical significance testing. I test the performance of the forecasts with the M3-Competition "out of sample accuracy check" procedure as it is found in the M3-Competition. This procedure follows five steps:

1. Selection of a series.
2. The selected series is separated into a training data set and test data set.
3. The training set is used to estimate the parameters of the model.
4. After the estimation, a number of forecasts are produced for the test set period.
5. These forecasts are compared with the actual data of the test set to measure the performance of the model.

The training data set comprises the first n-18 data points, where n is the total number of the data points of each series; the test set is made up of the remaining 18 data points. The models are tested for *short-term* forecasting (one step ahead), *intermediate-term* forecasting (six steps ahead), and *long-term* forecasting (twelve steps ahead). The multiple periods ahead forecasts are single point forecasts.

A graph of a time series with both the training and test sets and the forecasts is shown in Figure 3.10.

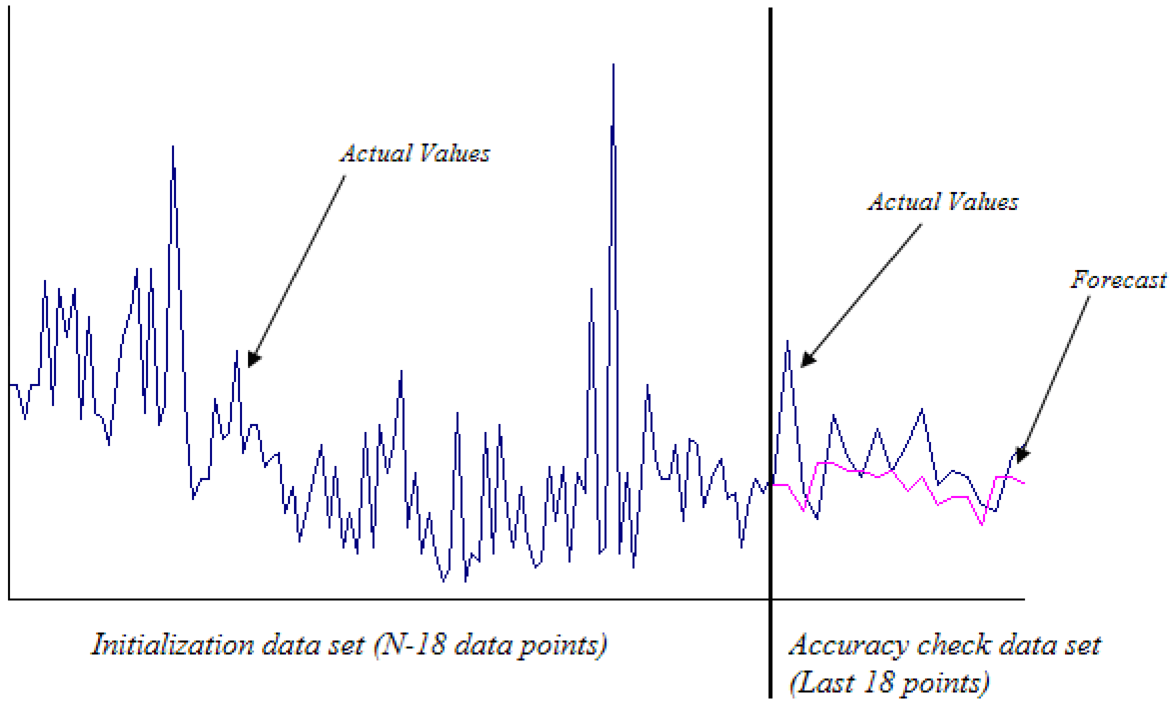

Figure 3. 10 The out of sample accuracy check

For the computational tests, the linear programming software *Lingo* is used.

3.3.2 Comparison and evaluation

All the LP-based forecasts are compared with traditional techniques that are found in the literature. All the techniques are applied on the same series and are tested in the same way (out of sample accuracy check). The models are compared in terms of accuracy, convenience to use and forecasting cost. For the comparison of the simple forecasts (Chapters 4, 5 and 6) the MAPE and sMAPE are used. I decided to use both, because there is an open conversation about which of the two types of accuracy measurements should be preferred. In addition I decided to use the MASE because according to Hyndman and Koehler (2006) are widely applicable and are always defined and finite. All the three accuracy measures can be found in Chapter 2 For the combined forecasts (Chapter 7) I have used three additional accuracy measures, the percentage difference between the combined forecast and the best individual, the percentage difference between the combined forecast and the worst individual and the percentage difference between the combined forecast and the average individual (they are presented in detail in Chapter 7). This is because while the initial three measure the forecasting accuracy independently, the latter measure the accuracy of the combinations in comparison with the individual techniques that are combined. For the cost minimisation (Chapter 8) I use the cost according to its underestimation and overestimation asymmetry and the MAPE. The reason that I use only one accuracy measure is that the analysis in chapter 7 is not focused on the accuracy of the forecasts, but on their performance according to the cost of the forecasting error. Hence, the use of one accuracy measure as an indicator is sufficient.

The approaches for the first, the second and the fifth research question are compared with the counterpart approaches based on the OLS method (minimising SSE). The approaches for the third research question are compared with the OLS and the following:

- Simple moving average (MA, Order 4, 11 and 12)
- Weighted moving average (WMA, Order 7, 8 and 12)
- Simple exponential smoothing (SES, 0.6, 0.7 and 0.8)
- Holt's exponential smoothing (Holt, 0.1-0.1, 0.2-0.1 and 0.1-0.2)

I want to ensure that LP-based forecasts are compared with the most accurate formulations of the other techniques. Thus, I do not simply optimise the parameters of the models, order of the moving average and the smoothing parameters, using the grid search of a statistical package. I explored different parameter values and I selected the three alternatives that were on average the most accurate on the test set of the thirty five series.

In the experiments for the fourth question eight individual techniques are used:

- Naive 1
- Simple moving average (Order 4)
- Simple exponential smoothing (0.8)
- Holt's exponential smoothing (0.2-0.1)
- Holt – Winters (0.2-0.1-0.8)
- Adaptive exponential smoothing
- Autoregressive (Order 6)

- Seasonal autoregressive (Order 6)

To estimate the coefficients of the simple and seasonal autoregressive models I used the method of the OLS and not the counterpart LP approach because I wanted to use a more mainstream approach, since the objective of the chapter is to test LP as a tool to compare forecasts. I neither select the most appropriate techniques for each of the series nor optimising their parameters because these forecasts would gave good results of similar accuracy; hence the accuracy of the combination would be similar and the comparison between the alternative combinations would be difficult. On the other hand, over a random selection of techniques, the accuracy of the results varies and conclusions of the comparison of the combinations are much clearer. Nonetheless, the objective of this part of the study is not to do the most accurate forecast for each series but the best combination of a number of forecasts.

In addition, the performance of the models is tested and compared with five traditional combinations methods (the can be found in Chapter 2):

- Simple Average
- Inverse Proportion of the MAD
- Inverse Proportion of the MAPE
- Inverse Proportion of the MSE
- Average Inverse Proportion of the MAD, MAPE and MSE
- Weighting based on the absolute error

The OLS approaches are developed in *STATA* while the other techniques were implemented in spreadsheets (Excel). For the individual forecasting techniques that are mentioned above I used the following formulation. Let Y_t be the actual value, F_t the forecast and e_t the forecasting error of period t ($1 \leq t \leq T$), then (all the formulations are for one step ahead forecasts):

Naïve 1:

$$F_t = Y_{t-1} \qquad \textbf{(3. 4)}$$

Simple Moving Average of Order *n* (1 ≤ *i* ≤ *n*):

$$F_t = \frac{Y_{t-1} + Y_{t-2} + \ldots + Y_{t-n}}{n} \quad \textbf{(3. 5)}$$

$$F_t = \frac{1}{n}\sum_{i=1}^{n} Y_{t-i} \quad \textbf{(3. 6)}$$

Weighted Moving Average of Order *n* (1 ≤ *i* ≤ *n*):

$$F_t = \frac{nY_{t-1} + (n-1)Y_{t-2} + \ldots + 2Y_{t-n+1} + Y_{t-n}}{n + (n-1) + \ldots + 2 + 1} \quad \textbf{(3. 7)}$$

$$F_t = \sum_{i=1}^{n} w_i Y_{t-i} \quad \textbf{(3. 8)}$$

for

$$w_i = \frac{2(n-i+1)}{n(n+1)} \quad \textbf{(3. 9)}$$

Simple Exponential Smoothing with smoothing factor *α* [0, 1]:

For *t* = 2: $$F_t = Y_{t-1} \quad \textbf{(3. 10)}$$

For *t* > 2: $$F_t = (1-\alpha)Y_{t-1} + \alpha F_{t-1} \quad \textbf{(3. 11)}$$

$$F_t = Y_{t-1} + \alpha(F_{t-1} - Y_{t-1}) \quad \textbf{(3. 12)}$$

$$F_t = Y_{t-1} + \alpha e_{t-1} \qquad \textbf{(3. 13)}$$

Holt's Exponential Smoothing with smoothing factors *α* [0, 1] and β [0, 1]:

$$F_{t+1} = S_t + G_t \qquad \textbf{(3. 14)}$$

For *t* = 2: base: $S_t = Y_{t-1}$ **(3. 15)**

For *t* > 2: base: $S_t = \alpha Y_t + (1-\alpha)(S_{t-1} + G_{t-1})$ **(3. 16)**

For *t* = 2: trend: $G_t = 0$ **(3. 17)**

For *t* > 2 trend: $G_t = \beta(S_t - S_{t-1}) + (1-\beta)G_{t-1}$ **(3. 18)**

Holt – Winters with smoothing factors *α* [0, 1], β [0, 1] and *γ* [0, 1] and *k* the seasonality level:

$$F_{t+1} = \left(S_t + G_t\right) \times W_t \qquad \textbf{(3. 19)}$$

For *t* = 12: base: $S_t = Y_{t-1}$ **(3. 20)**

For *t* > 12: base: $S_t = \alpha\left(Y_t / W_{t-k}\right) + (1-\alpha)(S_{t-1} + G_{t-1})$ **(3. 21)**

For *t* = 12: trend: $G_t = 0$ **(3. 22)**

For *t* > 12 trend: $G_t = \beta(S_t - S_{t-1}) + (1-\beta)G_{t-1}$ **(3. 23)**

For t = 12: seasonality: $W_t = \frac{12Y_{12}}{\sum_{t=1}^{12} Y_t}$ **(3. 24)**

For t > 12 seasonality: $W_t = \gamma(Y_t / S_t) + (1-\gamma)W_{t-k}$ **(3. 25)**

Adaptive Exponential Smoothing with smoothing factor α [0, 1]:

For t = 2: $F_t = Y_{t-1}$ **(3. 26)**

For t > 2: $F_{t+1} = F_t + \frac{ESME_t}{ESMAD_t} e_t$ **(3. 27)**

Exponential Smoothing of the Mean Error:

$$ESME_t = \alpha ESME_{t-1} + (1-\alpha)AD_{t-1} \quad \textbf{(3. 28)}$$

Exponential Smoothing of the Mean Absolute Deviation:

$$ESMAD_t = \alpha ESMAD_{t-1} + (1-\alpha)AD_{t-1} \quad \textbf{(3. 29)}$$

let AD_t be the absolute error for period t.

The last step is the evaluation of the forecasts according to its performance. This step is the feedback of the whole process; hence, if the forecast is shown to be inapplicable or very inaccurate, we go back to the first step and try to improve it. Finally, general conclusions about the performance of the techniques are made.

3.4 CONCLUSION

In this chapter, the methodology of the study was presented. Several LP-based approaches for estimating the parameters of time series forecasting models have been developed. These approaches are tested on two time series samples from the data set of the M3 Competition. The first sample consists of 60 series that are classified in three categories (smooth, seasonal and hard) and the second consists of 25 series that are all of them are classified as hard. The out of sample accuracy check is used for testing, and the accuracy is measured according to several error indices. The performance of the LP approaches are compared with the OLS and other techniques that are found in the literature (single and combined forecasting techniques) and they are evaluated according to the comparison.

4 OPTIMISING AUTOREGRESSIVE BASED FORECASTS

In this chapter I apply single objective LP to estimate the parameters of autoregressive and seasonal autoregressive forecasting models. In the initial formulation the objectives are: minimise the sum of absolute deviations, minimise the sum of absolute percentage errors and minimise the maximum absolute deviation. The first two objectives are common forecast accuracy measures; the third objective was tried as an attempt to avoid large errors. Then, in order to further improve results, and more specifically, in order to avoid or remove bias, I test the same linear programs with the additional constraint where the sum of the errors is equal to zero or the sum of the percentage errors is equal to zero. At this stage I also introduce two other, less common, objectives: minimising the absolute differences between deviations and minimising the absolute differences between percentage deviations. The rationale for testing these objectives is to avoid 'large swings' between positive and/or negative errors or, in other words, to obtain 'smoother' (and hopefully better) forecasts.

The linear programs are tested on the 60 series of the initial sample for order 6 and 12 models. Performance is evaluated through the out of sample MAPE, sMAPE and MASE over the 60 series and also by comparison with the OLS approach (minimising SSE). I report results for one, six and twelve step ahead forecasts.

4.1 SIMPLE OBJECTIVE MODELS: INITIAL FORMULATION

The first LP formulations minimise SAD, SAPE or MaxAD for simple autoregressive (AR) and autoregressive models with an additional, additive seasonal coefficients (ARS). The latter is applied to series with seasonal pattern, both autoregressive models without constant term b_0. The well-known equation for an autoregressive model of order m is:

$$Y_i = b_1 Y_{i-1} + b_2 Y_{i-2} + ... + b_m Y_{i-m} + e_i \tag{4. 1}$$

or

$$Y_i = \sum_{j=1}^{m} b_j Y_{i-j} + e_i \tag{4. 2}$$

Where Y_i is the predicted variable, Y_{i-j} are the explanatory variables, b_j is the coefficient of Y_{i-j}, e_i is the forecasting error, i ($1 \le i \le n$) is the index of the forecasting period and j ($1 \le j \le m$) is the order.

The above formulation produces one period ahead forecasts. For longer term forecasts the formulation is (s steps ahead):

$$Y_i = \sum_{j=1}^{m} b_j Y_{i-s-j} + e_i \tag{4. 3}$$

The formulation of the program for minimising the SAD for one period ahead is (assuming that Y_{i-j} is not defined for $j > i$):

$$MinSAD$$

$$Min \sum_{i=2}^{n} \left| \sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} - Y_i \right| \tag{4. 4}$$

$$Min \sum_{i=2}^{n} |e_i| \tag{4. 5}$$

If e_{1i} is the underestimation and e_{2i} is the overestimation error:

$$Min \sum_{i=2}^{n} e_{1i} + \sum_{i=2}^{n} e_{2i} \tag{4. 6}$$

Subject to the following constraints:

$$\sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} + e_{1i} - e_{2i} = Y_i \quad i = 2, \dots, n \tag{4. 7}$$

e_{1i}, e_{2i} non-negative and b_j unrestricted in sign.

An alternative formulation of the above program is

$$Min\sum_{i=2}^{n} e_i \quad \textbf{(4. 8)}$$

Where e_i is the absolute error of period *i.* Subject to the constraints:

$$\sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} + e_i \geq Y_i \quad i = 2, \ldots , n \quad \textbf{(4. 9)}$$

$$\sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} - e_i \leq Y_i \quad i = 2, \ldots , n \quad \textbf{(4. 10)}$$

e_i non-negative and b_j unrestricted in sign.

In the same way, the formulation for minimising SAPE is (assuming that Y_{i-j} is not defined for $j > i$):

$$MinSAPE$$

$$Min\sum_{i=2}^{n}\left|\left(\sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} - Y_i\right)\Big/ Y_i\right| \quad \textbf{(4. 11)}$$

$$Min\sum_{i=2}^{n}\left[\left(e_{1i} + e_{2i}\right)/Y_i\right] \quad \textbf{(4. 12)}$$

Subject to the following constraints:

(4. 7)

e_{1i}, e_{2i} non-negative and b_j unrestricted in sign.

And the alternative formulation is

$$Min\sum_{i=2}^{n} e_i \quad \textbf{(4. 13)}$$

Where e_i is the absolute percentage error of period *i.* Subject to the constraints:

$$\sum_{j=1}^{\min\{m,i-1\}} \frac{\left(b_j Y_{i-j}\right)}{Y_i} + e_i \geq 1 \quad i = 2, \ldots, n \tag{4. 14}$$

$$\sum_{j=1}^{\min\{m,i-1\}} \frac{\left(b_j Y_{i-j}\right)}{Y_i} - e_i \leq 1 \quad i = 2, \ldots, n \tag{4. 15}$$

e_i non-negative and b_j unrestricted in sign.

The formulation of the programme that minimises MaxAD is

$$MinMaxAD$$

$$Min \sum_{i=2}^{n} e \tag{4. 16}$$

Where e is the maximum absolute deviation. Subject to the constraints

$$\sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} + e \geq Y_i \quad i = 2, \ldots, n \tag{4. 17}$$

$$\sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} - e_i \leq Y_i \quad i = 2, \ldots, n \tag{4. 18}$$

e non-negative and b_j unrestricted in sign.

And the alternative formulation is

(4. 16)

Subject to:

(4. 7)

$$e - e_{1i} - e_{2i} \geq 0 \quad i = 2, \dots , n \tag{4. 19}$$

e, e_{1i}, e_{2i} non-negative and b_j unrestricted in sign.

Seasonal autoregressive models are simple autoregressive models for series with the addition of coefficients that represents the seasonality of the series. Additive seasonal coefficients are fixed amounts that are added or subtracted according to the seasonal pattern of the series (e.g. in a monthly series with annual seasonality, the seasonal coefficients repeat every 12 periods). Moreover, they should be zero in case there is no clear seasonal pattern. The equation for the seasonal autoregressive model is:

$$Y_i = b_1 Y_{i-1} + b_2 Y_{i-2} + \dots + b_k Y_{i-k} + S_i + e_i \tag{4. 20}$$

or

$$Y_i = \sum_{j=1}^{m} b_j Y_{i-j} + S_i + e_i \tag{4. 21}$$

with S_i is the seasonal coefficient of period i.

The seasonality level of the series is determined using the ACF, as explained in chapter 3. In analogy with the simple autoregressive model, LP can be applied to minimise the SAD, SAPE and MaxAD.

The linear program for minimising the SAD is (assuming that Y_{i-j} is not defined for $j > i$):

$$MinSAD$$

$$Min \sum_{i=2}^{n} \left| \sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} + S_i - Y_i \right| \tag{4. 22}$$

(4. 5), **(4. 6)**

If e_{1i} is the underestimation and e_{2i} is the overestimation errors, subject to:

$$\sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} + S_i + e_{1i} - e_{2i} = Y_i \quad i = 2, \dots, n \tag{4. 23}$$

$$S_i = S_{i+k} \quad i = 2, \dots, n \tag{4. 24}$$

e_{1i}, e_{2i} non-negative and b_j and S_i unrestricted in sign.

In the same way, the formulation for minimising MAPE is (assuming that Y_{i-j} is not defined for $j > i$):

$$MinSAPE$$

$$Min \sum_{i=2}^{n} \left| \left(\sum_{j=1}^{\min\{m,i-1\}} b_j Y_{i-j} + S_i - Y_i \right) \Big/ Y_i \right| \tag{4. 25}$$

(4. 11)

subject to:

(4. 23), **(4. 24)**

e_{1i}, e_{2i} non-negative and b_j and S_i unrestricted in sign.

And the formulation of the programme that minimises MaxAD is

$$MinMaxAD$$

(4. 16)

subject to:

(4. 23), **(4. 24)**, **(4. 19)**

e, e_{1i}, e_{2i} non-negative and b_j and S_i unrestricted in sign.

4.2 FIRST RESULTS

As it is mentioned in the research methodology, to facilitate the analysis, I divided the sample in three main groups:

- 32 series with low variability (smooth series, predictable)
- 18 series with strong seasonal pattern
- 10 series with high variability (hard series, difficult to predict)

Table 4.1 shows the results of the initial AR and ARS models for one period ahead forecasts for order 6 and 12. The table shows the average out of sample MAPE of all 60 series as well as the average MAPE for the smooth (white), hard (red) and seasonal (green) series. The LP results (minimising SAD, SAPE and MaxAD) are compared with these of the OLS (minimising MSE).

Table 4. 1 Initial results - MAPE

MAPE (%)	LP			OLS
1 step ahead	MinSAD	MinSAPE	MinMaxAD	
AR 6				
All	8.94	8.95	44.10	8.94
Smooth	3.22	3.02	24.61	3.19
Hard	25.11	25.89	116.25	25.01
Seasonal	10.13	10.08	36.56	10.24
AR 12				
All	8.47	8.69	26.57	8.29
Smooth	3.16	3.05	17.47	3.23
Hard	26.9	29.07	53.11	25.01
Seasonal	7.67	7.37	28.56	8
ARS 6				
All	7.57	7.67	36.48	7.66
Smooth	3.22	3.02	24.61	3.19
Hard	25.11	25.89	116.25	25.01
Seasonal	5.56	5.83	11.91	5.97
ARS 12				
All	7.75	8.11	21.56	7.54
Smooth	3.16	3.05	17.47	3.23
Hard	26.9	29.07	53.11	25.01
Seasonal	5.26	5.45	10.69	5.5

When comparing the LP approaches, MinSAD outperforms MinSAPE and MinMaxAD in general; however, MinSAPE performs slightly better on the smooth series and on the seasonal series. MinSAD performs better on the hard series for the AR/ARS 12 models, and slightly better on the hard series for the AR/ARS 6 models and on the seasonal series for the ARS models. In addition, the AR/ARS 12 models perform overall better than the AR/ARS 6; but the AR/ARS 6 models work better on the hard series.

Finally, the seasonal model improves the results on the seasonal series significantly. I note that I applied the seasonal model to the seasonal series only; hence only the results on the seasonal series are affected. The results on the non seasonal series are the same with the simple model, because no seasonality is observed in any lag; thus, no seasonal coefficient is added. MinMaxAD is the worst approach and performs significantly worse on all of the series.

Table 4. 2 Initial results - sMAPE

sMAPE (%)	**LP**			**OLS**
1 step ahead	**MinSAD**	**MinSAPE**	**MinMaxAD**	
AR 6				
All	8.57	9.28	24.83	8.66
Smooth	3.08	2.93	11.32	3.09
Hard	24.58	28.86	60.31	23.89
Seasonal	9.43	9.69	29.16	10.10
AR 12				
All	8.56	9.40	23.01	8.13
Smooth	3.00	2.93	10.74	3.08
Hard	28.69	34.42	56.57	24.35
Seasonal	7.25	7.01	26.20	8.11
ARS 6				
All	7.41	8.09	19.30	7.42
Smooth	3.08	2.93	11.32	3.09
Hard	24.58	28.86	60.31	23.89
Seasonal	5.57	5.72	10.70	5.97
ARS 12				
All	7.86	8.77	18.40	7.35
Smooth	3.00	2.93	10.74	3.08
Hard	28.69	34.42	56.57	24.35
Seasonal	4.94	4.90	10.83	5.48

Comparing the LP approaches with the OLS method, it is obvious that the second gives slightly better results, except of the MinSAD ARS 6 model. The OLS performs better on the hard series, while LP (only MinSAD and MinSAPE) performs better on the smooth and seasonal series for most of the models. Nevertheless, the differences between all the approaches are relatively small (except for the MinMaxAD which produces the highest MAPE overall).

Table 4.2 shows the average sMAPE on the test set of the 60 series (AR and ARS, Order 6 and 12). The performance of the methods according to sMAPE is similar to the comparison according MAPE. MinSAD outperforms MinSAPE overall but this is mainly due to better performance on the hard series. MinSAPE is the best performing approach on the smooth series. MinMaxAD is again the worst technique overall and for each group separately. In comparison with the OLS, LP performs better for order 6 models (MinSAD) while the OLS outperforms the LPs for order 12 models. LP performs better on the smooth and seasonal series and OLS performs better on the hard series.

Table 4. 3 Initial results - MASE

MASE	**LP**			**OLS**
1 step ahead	**MinSAD**	**MinSAPE**	**MinMaxAD**	
AR 6				
All	0.93	0.94	4.14	0.93
Smooth	0.96	0.95	4.09	0.97
Hard	0.91	0.99	5.36	0.89
Seasonal	0.88	0.89	3.54	0.90
AR 12				
All	0.90	0.91	4.64	0.91
Smooth	0.95	0.95	3.82	0.95
Hard	1.04	1.16	9.94	0.92
Seasonal	0.74	0.71	3.14	0.81
ARS 6				
All	0.82	0.84	3.41	0.83
Smooth	0.95	0.94	4.09	0.97
Hard	0.91	1.01	5.36	0.89
Seasonal	0.55	0.55	1.13	0.56
ARS 12				
All	0.83	0.85	4.03	0.82
Smooth	0.95	0.95	3.82	0.95
Hard	0.98	1.11	9.94	0.92
Seasonal	0.52	0.52	1.13	0.54

Table 4.3 shows the comparison of the results according MASE. When the average MASE is 1, the performance of the technique is similar to the Naive 1 model. If it is smaller than 1 the technique performs better than the Naive 1 method and if it is bigger than 1, it performs worse.

The comparison of the approaches according to MASE is along the same lines as with sMAPE. MinSAD outperforms MinSAPE overall, but mainly due to better performance on the hard series. MinSAPE performs very well on the smooth and seasonal series. MinMaxAD is again the worst technique overall and for each group separately. In comparison with the OLS, LP (MinSAD and MinSAPE) perform better on the smooth and seasonal series, whereas the OLS performs better on the hard series. All the techniques perform better than Naive 1 (MASE < 1) except MinMaxAD, and MinSAPE on the hard series.

4.3 INGNORING THE FIRST s+m-1 DATA POINTS IN THE OBJECTIVE FUNCTION

One of the first things that were observed through the initial experiments was that the models may give very large errors for the first data points in the training set (Figure 5.9). For the first *m* periods of the set – assuming a one step ahead forecast – there are not sufficient data to produce an autoregressive based forecast of order m. Hence, there is only a partial model for the first m periods and this may impact on the accuracy of the model. Therefore, I decided to run the experiments again ignoring the first *m* points of the training set in the objective function. Similarly, for an *s* – step ahead forecast with an order m model, we ignore the first *s*+*m*-1 data points in the objective function.

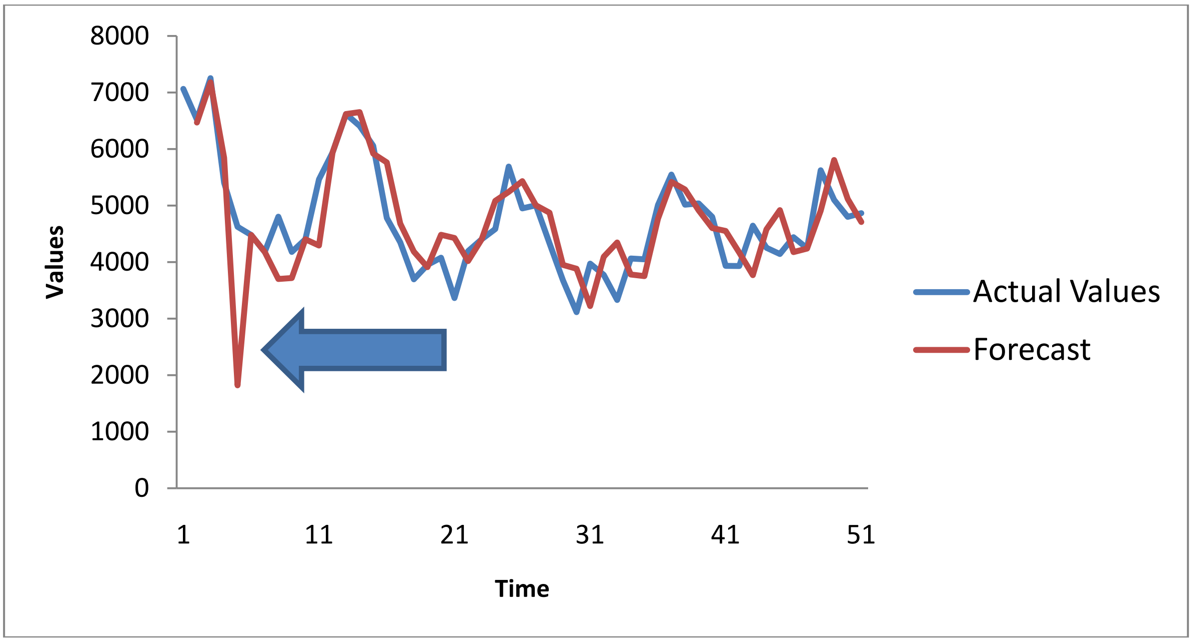

Figure 4. 1 Performance of an approach with the first s+m-1 point in the objective function

The new formulation of the linear program, minimising SAD, for the AR one step ahead model is (assuming that Y_{i-j} is not defined for $j > i$):

$$MinSAD$$

$$Min \sum_{i=m+1}^{n} \left| \sum_{j=1}^{m} b_j Y_{i-j} - Y_i \right| \quad \textbf{(4. 24)}$$

$$Min \sum_{i=m+1}^{n} |e_i| \quad \textbf{(4. 25)}$$

if e_{1i} is the underestimation and e_{2i} is the overestimation error:

$$Min \sum_{i=m+1}^{n} e_{1i} + \sum_{i=m+1}^{n} e_{2i} \quad \textbf{(4. 26)}$$

subject to:

$$\sum_{j=1}^{m} b_j Y_{i-j} + e_{1i} - e_{2i} = Y_i \quad i = m+1, \dots , n \quad \textbf{(4. 27)}$$

e_{1i}, e_{2i} non-negative and b_j unrestricted in sign.

In the same way, the formulation of the program for the ARS models are minimising the SAD is (assuming that Y_{i-j} is not defined for $j > i$):

$$MinSAD$$

$$Min \sum_{i=m+1}^{n} \left| \sum_{j=1}^{m} b_j Y_{i-j} + S_i - Y_i \right| \quad \textbf{(4. 28)}$$

(4. 25), (4. 26)

subject to:

$$\sum_{j=1}^{m} b_j Y_{i-j} + S_i + e_{1i} - e_{2i} = Y_i \quad i = m+1, \dots, n \tag{4. 29}$$

$$S_i = S_{i+k} \quad i = m+1, \dots, n \tag{4. 30}$$

e_{1i}, e_{2i} non-negative and b_j and S_i unrestricted in sign.

The linear programs that minimise SAPE and MaxAD are formulated in the same way.

4.4 RESULTS: IGNORING THE FIRST s+m-1 DATA POINTS

Table 4.4 shows the results of the same models, ignoring the first s+m-1 data points of the training set. The first s+m-1 data points are also ignored in the OLS model (minimising SSE). In brackets there is the difference between the new results and the previous results (Table 4.1). The results of the ARS on the hard and smooth series are the same with the AR; hence, I do not present these.

Table 4. 4 Results ignoring the first s+1-m data points - MAPE

MAPE (%)	**LP**						**OLS**	
1 step ahead	**MinSAD**		**MinSAPE**		**MinMaxAD**			
AR 6								
All	9.02	(0.08)	9.06	(0.11)	27.58	(-16.52)	8.69	(-0.25)
Smooth	3.19	(-0.03)	2.98	(-0.04)	21.43	(-3.18)	3.14	(-0.05)
Hard	24.96	(-0.15)	26.44	(0.55)	47.10	(-69.14)	24.47	(-0.54)
Seasonal	10.54	(0.41)	10.22	(0.14)	27.69	(-8.87)	9.79	(-0.45)
AR 12								
All	8.25	(-0.22)	8.57	(-0.12)	21.33	(-5.25)	8.01	(-0.28)
Smooth	3.17	(0.01)	3.06	(0.01)	18.77	(1.3)	3.22	(-0.01)
Hard	25.36	(-1.54)	27.88	(-1.19)	30.19	(-22.92)	23.69	(-1.32)
Seasonal	7.75	(0.08)	7.64	(0.27)	20.79	(-7.77)	7.80	(-0.2)
ARS 6								
All	7.54	(-0.03)	7.71	(0.04)	22.68	(-13.8)	7.39	(-0.27)
Seasonal	5.60	(0.04)	5.73	(-0.1)	10.68	(-1.23)	5.45	(-0.52)
ARS 12								
All	7.43	(-0.32)	7.78	(-0.33)	18.55	(-3)	7.22	(-0.32)
Seasonal	5.03	(-0.23)	5	(-0.45)	11.29	(0.6)	5.19	(-0.31)

By ignoring the first s+m-1 periods of the training set the results are generally improved. The OLS give better results, having on average 0.2% to 0.3% lower MAPE compared with table 4.1. LPs are improved in the AR/ARS 12, but they are slightly worse in the AR 6. The results of the LP approaches are generally improved on the hard series. Among the LP models MinSAD performs best on the hard series and MinSAPE performs best on the smooth series. MinMaxAD performs significantly worse compared with the other methods.

Table 4.5 presents the average out of sample sMAPE. According to sMAPE, the comparison of the methods is similar to the comparison according to MAPE (Table 4.4). Comparing LP approaches, MinMaxAD has the worst performance. MinSAD is overall the best LP approach and also on the hard series, while MinSAPE is the best on the smooth series.

Table 4. 5 Results ignoring the first s+1-m data points - sMAPE

sMAPE (%)	**LP**						**OLS**	
1 step ahead	**MinSAD**		**MinSAPE**		**MinMaxAD**			
AR 6								
All	8.64	(0.07)	9.40	(0.12)	19.41	(-5.43)	8.48	(-0.18)
Smooth	3.05	(-0.03)	2.89	(-0.04)	10.47	(-0.85)	3.03	(-0.05)
Hard	24.43	(-0.15)	29.47	(0.61)	42.66	(-17.65)	24.01	(0.12)
Seasonal	9.81	(0.38)	9.83	(0.14)	22.38	(-6.78)	9.54	(-0.57)
AR 12								
All	8.31	(-0.25)	9.25	(-0.15)	15.75	(-7.27)	7.81	(-0.32)
Smooth	3.01	(0.01)	2.94	(0.01)	9.72	(-1.01)	3.07	(-0.01)
Hard	27.05	(-1.64)	33.01	(-1.41)	32.89	(-23.67)	23.74	(-0.61)
Seasonal	7.33	(0.08)	7.26	(0.25)	16.93	(-9.27)	7.40	(-0.71)
ARS 6								
All	7.35	(-0.06)	8.17	(0.08)	15.61	(-3.69)	7.24	(-0.18)
Seasonal	5.51	(-0.05)	5.72	(2.79)	9.72	(-0.98)	5.39	(-0.57)
ARS 12								
All	7.59	(-0.27)	8.54	(-0.23)	13.71	(-4.69)	7.12	(-0.22)
Seasonal	4.94	(0.01)	4.90	(0)	10.15	(-0.68)	5.10	(-0.38)

Comparing OLS with LP, the differences are relatively low. LP performs better on smooth and seasonal series, and OLS performs better on the hard series.

Table 4.6 shows performance of the methods according to MASE. The comparison of the results is similar with the other two accuracy measures. The overall differences are low. LP is better than the OLS on the smooth and seasonal series and OLS is better on the hard series. MinSAD and MinSAPE perform similar on the seasonal series, MinSAD performs better on the hard series (MinSAPE performs worse than naive, MASE > 1) and MinSAPE performs better on the smooth. MinMaxAD is the worst and it is the only method that has an overall MASE larger than 1.

Table 4. 6 Results ignoring the 1st s+1-m data points - MASE

MASE	LP						OLS	
1 step ahead	MinSAD		MinSAPE		MinMaxAD			
AR 6								
All	0.93	(0)	0.94	(0)	3.15	(-0.99)	0.92	(-0.02)
Smooth	0.95	(-0.01)	0.94	(-0.01)	3.75	(-0.34)	0.95	(-0.02)
Hard	0.91	(-0.01)	1.01	(0.02)	2.40	(-2.96)	0.88	(-0.01)
Seasonal	0.92	(0.04)	0.91	(0.02)	2.50	(-1.04)	0.88	(-0.02)
AR 12								
All	0.90	(-0.01)	0.92	(0.01)	2.64	(-2)	0.88	(-0.03)
Smooth	0.95	(0)	0.95	(0)	3.39	(-0.44)	0.95	(0)
Hard	0.98	(-0.06)	1.11	(-0.05)	1.68	(-8.26)	0.88	(-0.05)
Seasonal	0.75	(0.01)	0.74	(0.03)	1.85	(-1.29)	0.75	(-0.06)
ARS 6								
All	0.82	(0)	0.84	(0)	2.71	(-0.7)	0.81	(-0.02)
Seasonal	0.54	(0)	0.55	(0)	1.05	(-0.08)	0.52	(-0.03)
ARS 12								
All	0.83	(0)	0.85	(0)	2.43	(-1.6)	0.81	(-0.02)
Seasonal	0.52	(0)	0.52	(0)	1.15	(0.02)	0.52	(-0.02)

4.5 SUM OF ERRORS EQUAL TO ZERO

One of the characteristics of the OLS regression is that the sum of the errors is always equal to zero. This removes overestimation or underestimation bias in the regression models. The sum of errors equal to zero is not guaranteed in initial LP formulations. This is also mentioned by Kiountouzis (1973) where linear programming based regression introduces a small bias even in symmetrical error distributions. Thus, I decide to run the experiments again adding this constraint to the linear program ($\Sigma e = 0$) in order to have a better comparison of the LP and OLS approaches. In the same way, I add a constraint where the sum of the k seasonal coefficients of the ARS model are equal to zero ($\Sigma S = 0$) in order to remove possible seasonal bias from the models. The new formulation of the program for both AR and ARS models are:

The linear program for the AR model minimising the SAD is (assuming that Y_{i-j} is not defined for $j > i$):

$$MinSAD$$

(4. 24), **(4. 25)**, **(4. 26)**

subject to:

(4. 27)

$$\sum_{i=m+1}^{n}\left(e_{1i} - e_{2i}\right) = 0 \qquad \textbf{(4. 31)}$$

e_{1i}, e_{2i} non-negative and b_j unrestricted in sign.

The linear program for the ARS model is:

$$MinSAD$$

(4. 28), **(4. 25)**, **(4. 26)**

subject to:

(4. 29), **(4. 30)**, **(4. 31)**

$$\sum_{i=m+1}^{m+k} S_i = 0 \qquad \textbf{(4. 32)}$$

e_{1i}, e_{2i} non-negative and b_j and S_i unrestricted in sign.

The linear programs that minimise SAPE and MaxAD are formulated in the same way.

I also test LPs with two other objective functions: minimising the *Absolute Differences Between the Deviations* minimising the *Absolute Difference Between Percentage Deviations*. These objectives were identified at a later stage in the study and the rationale for testing these is to avoid 'large swings' between positive and/or negative errors or, in other words, to obtain 'smoother' (and hopefully better) forecasts.

The formulation of the AR model for minimising the ADBD for one period ahead forecast is (assuming that Y_{i-j} is not defined for $j > i$):

$$MinADBD$$

$$Min \sum_{i=m+1}^{n-1} \sum_{l=i+1}^{n} \left| e_i - e_l \right| \quad \textbf{(4. 33)}$$

if e_{1il} and e_{2il} are the absolute positive and negative differences between the errors e_i and e_l, then this can be rewritten as:

$$Min \sum_{i=m+1}^{n-1} \sum_{l=i+1}^{n} \left(e_{1il} + e_{2il} \right) \quad \textbf{(4. 34)}$$

subject to:

$$e_i - e_l - e_{1il} + e_{2il} = 0 \quad i = m+1, \ldots, n-1 \text{ and } l = i+1, \ldots, n \quad \textbf{(4. 35)}$$

$$\sum_{j=1}^{m} b_j Y_{i-j} + e_i = Y_i \quad i = m+1, \ldots, n \quad \textbf{(4. 36)}$$

$$\sum_{i=m+1}^{n} e_i = 0 \quad \textbf{(4. 37)}$$

e_{1il}, e_{2il} non-negative and e_i and b_j unrestricted in sign.

In the same way, the formulation of the ARS model that minimises the ADBD for one period ahead forecast is (assuming that Y_{i-j} is not defined for $j > i$):

$$MinADBD$$

(4. 33), **(4. 34)**

subject to:

(4. 35)

$$\sum_{j=1}^{m} b_j Y_{i-j} + S_i + e_i = Y_i \quad i = m+1, \dots, n \tag{4. 38}$$

(4. 30), (4. 32), (4. 37)

e_{1il}, e_{2il} non-negative and e_i and b_j unrestricted in sign.

The LP formulation of the AR model that minimises the *Absolute Difference Between Percentage Errors* is as follows.

$$MinADBPE$$

$$Min \sum_{i=m+1}^{n-1} \sum_{l=i+1}^{n} \left| \frac{e_i}{Y_i} - \frac{e_l}{Y_l} \right| \tag{4. 39}$$

if e_{1il} and e_{2il} are the absolute positive and negative differences between the percentage errors e_i/Y_i and e_l/Y_l, this can be rewritten as:

(4. 34)

subject to:

$$\frac{e_i}{Y_i} - \frac{e_l}{Y_l} - e_{1il} + e_{2il} = 0 \quad i = m+1, \dots, n-1 \text{ and } l = i+1, \dots, n \tag{4. 40}$$

(4. 36), (4. 37)

e_{1il}, e_{2il} non-negative and e_i and b_j unrestricted in sign.

The formulation of the ARS MinADBPE model is similar.

4.6 RESULTS: SUM OF ERRORS EQUAL TO ZERO

Table 4.7 shows the average out of sample MAPE for the models with the additional two constraints: sum of the errors and sum of the seasonal coefficients equal to zero. The OLS results remain the same as in the table 4.4.

Adding the constraints improves the results and especially on the smooth and seasonal series. On the hard series the results are less conclusive. The best performing LP approaches on the smooth and seasonal series are MinADBD and MinADBPD. The best performing approach for the hard series is the OLS. The order 12 models perform generally better than the order 6 models (except for the hard series) and the models with seasonal coefficients perform much better than those without. MinMaxAD is again the worst technique; however, its performance improved considerably compared with the earlier results (Table 4.4).

Table 4. 7 Sum of errors equal to zero - MAPE - 1 step ahead

MAPE 1 STEP Σe = 0	LP										OLS
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
AR 6											
All	8.64	(-0.13)	8.70	(-0.34)	9.27	(-1.82)	8.76	(-0.03)	8.78	(-0.01)	8.69
Smooth	3.02	(-0.15)	3.05	(-0.15)	3.98	(0)	2.99	(-0.19)	2.99	(-0.21)	3.14
Hard	24.94	(0.14)	25.20	(-1.06)	25.14	(-6.05)	25.71	(0.94)	26.09	(1.12)	24.47
Seasonal	9.58	(-0.22)	9.57	(-0.28)	9.85	(-2.7)	9.59	(-0.29)	9.46	(-0.27)	9.79
AR 12											
All	8.36	(-0.01)	8.50	(-0.29)	8.89	(-0.47)	8.16	(0.08)	8.42	(-0.15)	8.01
Smooth	2.98	(-0.18)	3.09	(-0.07)	4.19	(-0.04)	2.91	(-0.27)	2.89	(-0.19)	3.22
Hard	26.83	(1.05)	27.42	(-1.09)	25.81	(0.66)	26.08	(1.7)	27.79	(0.06)	23.69
Seasonal	7.68	(-0.29)	7.60	(-0.25)	7.85	(-1.89)	7.55	(-0.18)	7.48	(-0.19)	7.80
ARS 6											
All	7.41	(-0.07)	7.47	(-0.27)	7.96	(-1.7)	7.52	(0.05)	7.59	(0.06)	7.39
Seasonal	5.46	(-0.05)	5.47	(-0.04)	5.51	(-2.29)	5.46	(-0.03)	5.49	(-0.05)	5.45
ARS 12											
All	7.52	(0.07)	7.67	(-0.22)	7.99	(-0.78)	7.36	(0.13)	7.61	(-0.1)	7.22
Seasonal	4.86	(-0.03)	4.85	(0)	4.85	(-2.89)	4.86	(-0.04)	4.78	(-0.05)	5.19

Table 4.8 shows the average out of sample sMAPE. According to this accuracy index, the best approaches on the smooth and seasonal series are the LPs with objective MinADBD and MinADBPD. MinMaxAD is again the worst technique.

As before, in comparison with the LP approaches, the OLS approach yields slightly better results, mainly due to its good performance on the hard series. The LP models (except MinMaxAD) typically perform better than the OLS on the smooth and seasonal series. For smooth and seasonal series the order 12 models are better than order 6, and the ARS models are better than the AR counterparts on the seasonal series.

Table 4. 8 Sum of errors equal to zero - sMAPE - 1 step ahead

sMAPE 1 STEP Σe = 0	LP										OLS
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
AR 6											
All	8.88	(0.41)	8.98	(0.23)	9.04	(-2.05)	8.69	(0.22)	8.72	(0.07)	8.48
Smooth	2.94	(-0.09)	2.94	(-0.1)	3.90	(-0.09)	2.93	(-0.12)	2.91	(-0.14)	3.03
Hard	26.38	(2.64)	26.99	(1.73)	24.24	(-6.94)	25.71	(1.87)	26.09	(1.12)	24.01
Seasonal	9.71	(0.05)	9.71	(-0.02)	9.73	(-2.84)	9.47	(-0.11)	9.40	(-0.12)	9.54
AR 12											
All	8.45	(0.24)	8.59	(-0.24)	8.60	(-0.91)	8.01	(0.17)	8.26	(-0.1)	7.81
Smooth	2.87	(-0.12)	2.97	(-0.03)	4.10	(-0.21)	2.83	(-0.18)	2.79	(-0.13)	3.07
Hard	28.37	(2.26)	29.05	(-0.94)	25.11	(-1.26)	26.08	(1.87)	27.79	(0.06)	23.74
Seasonal	7.31	(-0.23)	7.23	(-0.21)	7.43	(-1.98)	7.19	(-0.13)	7.12	(-0.14)	7.40
ARS 6											
All	7.58	(0.38)	7.68	(0.23)	7.74	(-1.86)	7.46	(0.24)	7.52	(0.1)	7.24
Seasonal	5.38	(-0.04)	5.38	(-0.03)	5.41	(-2.18)	5.39	(-0.02)	5.41	(-0.04)	5.39
ARS 12											
All	7.69	(0.31)	7.85	(-0.17)	7.80	(-1.16)	7.29	(0.21)	7.53	(-0.07)	7.12
Seasonal	4.78	(-0.01)	4.76	(0.01)	4.75	(-2.8)	4.77	(-0.03)	4.69	(-0.04)	5.10

Table 4. 9 Sum of errors equal to zero - MASE - 1 step ahead

MASE 1 STEP Σe = 0	LP										OLS
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
AR 6											
All	0.92	(0.01)	0.93	(0)	1.04	(-0.16)	0.91	(0)	0.91	(0)	0.92
Smooth	0.94	(0)	0.94	(0)	1.18	(-0.09)	0.93	(-0.01)	0.94	(-0.01)	0.95
Hard	0.94	(0.06)	0.95	(0.03)	0.88	(-0.23)	0.92	(0.04)	0.92	(0.02)	0.88
Seasonal	0.89	(-0.01)	0.88	(-0.01)	0.89	(-0.23)	0.88	(-0.01)	0.87	(-0.01)	0.88
AR 12											
All	0.90	(0)	0.91	(0)	1.07	(-0.06)	0.87	(0)	0.88	(-0.01)	0.88
Smooth	0.95	(0)	0.96	(0)	1.29	(0.01)	0.93	(-0.01)	0.93	(-0.01)	0.95
Hard	1.02	(0.06)	1.05	(0.03)	0.91	(-0.07)	0.94	(0.04)	0.97	(0.01)	0.88
Seasonal	0.75	(-0.02)	0.74	(-0.02)	0.76	(-0.18)	0.73	(-0.01)	0.73	(-0.01)	0.75
ARS 6											
All	0.81	(0.01)	0.82	(0)	0.93	(-0.17)	0.81	(0)	0.81	(0)	0.81
Seasonal	0.53	(0)	0.52	(0)	0.53	(-0.26)	0.52	(0)	0.52	(0)	0.52
ARS 12											
All	0.83	(0.01)	0.84	(0.01)	0.99	(-0.07)	0.80	(0)	0.80	(0)	0.81
Seasonal	0.51	(0)	0.51	(0)	0.50	(-0.22)	0.50	(0)	0.49	(0)	0.52

Table 4.9 shows the average MASE. The differences between all the methods are very small and most approaches perform much better than the Naive 1 forecast method. The exceptions are the MinMaxAD formulation on the smooth series and the AR12 models with MinSAD and MinSAPE on the hard series.

The general conclusion of the analysis of the initial results is that LP is very useful for the development and optimisation of (seasonal) autoregressive models in case the series behave well (smooth or seasonal); the weak point of LP approach is the performance on the series with high variability. The addition of the two constraints improved the performance of the LP models.

I decided to run the experiments for six (medium term) and twelve periods ahead (long term) forecasting. Table 4.10 shows the average MAPE of the LP approaches, as well as OLS, for six steps ahead forecasts. Here, both AR 6 and AR 12 models give very similar results in general. On the other hand, ARS 6 models perform better than the ARS 12 models. ARS models are still better than simple AR, except the OLS ARS 12 which performs slightly worse than the corresponding AR 12.

Table 4. 10 Sum of errors equal to zero - MAPE - 6 steps ahead

MAPE	LP					OLS
6 STEP	MinSAD	MinSAPE	MinMaxAD	MinADBD	MinADBPD	
Σe = 0						
AR 6						
All	11.99	12.03	14.33	11.55	11.68	11.41
Smooth	7.34	7.39	8.36	7.26	7.54	7.09
Hard	26.64	26.84	27.53	24.24	24.35	24.03
Seasonal	12.12	12.04	17.60	12.12	12.01	12.08
AR 12						
All	11.94	12.13	14.14	11.20	11.69	11.17
Smooth	7.87	7.91	7.76	7.83	7.93	7.76
Hard	29.08	30.20	30.40	25.12	27.81	25.35
Seasonal	9.64	9.60	16.46	9.44	9.42	9.34
ARS 6						
All	10.84	10.84	12.09	10.27	10.57	10.96
Seasonal	8.29	8.09	10.12	7.87	8.29	10.59
ARS 12						
All	11.49	11.69	12.15	10.71	11.61	11.27
Seasonal	8.14	8.13	9.84	7.83	9.15	9.67

OLS outperforms LP overall, except for the ARS 6 models. MinADBD is the best LP model. OLS performs better on the hard series; however, LP performs better on the seasonal series for the ARS models. Similarly as for the 1 step ahead methods, MinMaxAD is the worst method compared with the other LP approaches and the OLS.

Table 4.11 shows the average sMAPE for 6 steps ahead forecasts. The performance of order 6 models is typically better than order 12 models. The best performing LP approaches are again those with objective MinADBD and MinADBPD. However, the MinADBPD model of order 12 performs rather poorly on the hard series. Detailed investigation revealed that this was due to poor performance on one of the series, where the forecasts were negative (making the denominator of the sMAPE very small, and hence resulting in a large error).

Table 4. 11 Sum of errors equal to zero - sMAPE - 6 steps ahead

sMAPE 6 STEP Σe = 0	LP					OLS
	MinSAD	MinSAPE	MinMaxAD	MinADBD	MinADBPD	
AR 6						
All	11.40	11.42	14.14	10.97	11.03	10.88
Smooth	6.63	6.68	7.76	6.55	6.80	6.42
Hard	27.23	27.29	30.40	24.86	24.64	24.80
Seasonal	11.08	11.02	16.46	11.10	11.00	11.08
AR 12						
All	11.65	14.45	13.37	10.79	17.99	11.32
Smooth	6.87	6.92	8.87	6.86	6.94	6.84
Hard	31.75	48.50	32.26	27.02	69.99	30.32
Seasonal	8.99	8.94	10.86	8.78	8.77	8.73
ARS 6						
All	10.44	10.41	12.15	9.89	9.92	10.58
Seasonal	7.87	7.68	9.84	7.51	7.28	10.08
ARS 12						
All	11.27	14.08	13.20	10.38	19.02	11.45
Seasonal	7.70	7.68	10.29	7.39	12.17	9.17

Comparing LP with OLS, the former performs better for ARS and AR 6 and the latter performs better for AR 12. LP performs significantly better on the seasonal series of ARS and OLS performs slightly better on the hard series.

Table 4.12 shows the average MASE for 6 periods ahead forecasts. Here, the performance of the LP and OLS approaches is compared with Naive 6. The results are in line with the previous observations. The best performing LP methods are MinADBD and MinADBPD. The worst LP approach is MinMaxAD, which performs occasionally worse than the forecast. The OLS is the better technique for the hard series but the LP approaches (except MinMaxAD) do better on the seasonal series (ARS models).

Table 4. 12 Sum of errors equal to zero - MASE - 6 steps ahead

MASE 6 STEP $\Sigma e = 0$	LP					OLS
	MinSAD	MinSAPE	MinMaxAD	MinADBD	MinADBPD	
AR 6						
All	0.88	0.89	1.05	0.86	0.87	0.85
Smooth	0.97	0.98	1.10	0.95	0.98	0.93
Hard	0.87	0.88	0.89	0.80	0.79	0.79
Seasonal	0.72	0.72	1.06	0.73	0.72	0.72
AR 12						
All	0.87	0.87	1.03	0.83	0.86	0.83
Smooth	0.98	0.99	1.21	0.95	0.98	0.96
Hard	0.95	0.96	0.98	0.87	0.93	0.84
Seasonal	0.62	0.62	0.74	0.61	0.61	0.60
ARS 6						
All	0.83	0.83	0.93	0.80	0.81	0.84
Seasonal	0.54	0.52	0.66	0.52	0.51	0.69
ARS 12						
All	0.84	0.85	1.02	0.81	0.92	0.84
Seasonal	0.55	0.55	0.70	0.53	0.81	0.64

After 6 steps ahead, the 12 steps ahead analysis follows. Table 4.13 shows the application of the LP, as well as OLS, for 12 periods ahead forecasts (average MAPE). Both AR and ARS models give very similar results; thus, the seasonal coefficient does not improve the accuracy of the forecast for long term forecasting. Additionally, order 6 models do better than order 12 models. MinADBD gives the best results; MinSAD comes second and MinADBPD third, while MinMaxAD is the worst technique; however, the difference between the LP approaches (ignoring MinMaxAD) is smaller compared with the 6 and 1 step ahead forecasts. OLS performs in general a bit better than LP.

Tables 4.14 and 4.15 show the average sMAPE and MASE (comparison with Naive 12) for 12 steps ahead respectively. The conclusion of the comparison between LP and OLS approaches is similar as with the comparison according to the MAPE. Most approaches perform worse than the naive method (Table 4.15).

Table 4. 13 Sum of errors equal to zero - MAPE - 12 steps ahead

MAPE 12 STEP Σe = 0	LP					OLS
	MinSAD	MinSAPE	MinMaxAD	MinADBD	MinADBPD	
AR 6						
All	13.08	13.31	15.28	12.85	13.24	12.55
Smooth	10.56	10.72	11.82	10.53	10.81	10.04
Hard	25.72	26.73	32.03	24.76	26.31	25.07
Seasonal	10.52	10.47	12.13	10.37	10.29	10.06
AR 12						
All	14.22	15.05	21.47	14.19	14.78	13.85
Smooth	11.34	11.51	22.17	11.42	11.54	11.04
Hard	28.99	33.74	34.38	29.10	32.68	28.97
Seasonal	11.13	10.98	13.07	10.83	10.59	10.44
ARS 6						
All	13.16	13.37	15.46	12.87	13.24	12.58
Seasonal	10.81	10.66	12.73	10.43	10.29	10.14
ARS 12						
All	14.13	14.94	21.90	14.15	14.79	13.61
Seasonal	10.84	10.59	14.49	10.70	10.63	9.65

Table 4. 14 Sum of errors equal to zero - sMAPE - 12 steps ahead

sMAPE 12 STEP Σe = 0	LP					OLS
	MinSAD	MinSAPE	MinMaxAD	MinADBD	MinADBPD	
AR 6						
All	12.49	12.86	14.93	12.31	12.80	12.18
Smooth	9.52	9.54	10.58	9.62	9.77	9.25
Hard	27.03	29.28	35.78	25.95	28.55	26.85
Seasonal	9.69	9.63	11.07	9.53	9.45	9.26
AR 12						
All	13.92	14.18	17.94	13.96	13.08	13.48
Smooth	10.22	10.21	14.60	10.27	10.36	10.17
Hard	32.44	34.30	39.26	33.07	27.82	31.03
Seasonal	10.19	10.06	12.03	9.92	9.70	9.61
ARS 6						
All	12.57	12.92	15.12	12.33	12.79	12.42
Seasonal	9.96	9.83	11.71	9.59	9.40	10.05
ARS 12						
All	13.85	14.08	18.22	13.93	13.08	13.44
Seasonal	9.98	9.74	12.95	9.80	9.71	9.49

Table 4. 15 Sum of errors equal to zero - MASE - 12 steps ahead

MASE 12 STEP Σe = 0	LP					OLS
	MinSAD	MinSAPE	MinMaxAD	MinADBD	MinADBPD	
AR 6						
All	1.03	1.05	1.19	1.02	1.05	0.98
Smooth	1.07	1.10	1.20	1.06	1.12	1.00
Hard	0.91	0.92	1.06	0.87	0.90	0.90
Seasonal	1.04	1.04	1.25	1.02	1.02	1.00
AR 12						
All	1.10	1.13	1.41	1.08	1.13	1.07
Smooth	1.14	1.18	1.52	1.12	1.20	1.10
Hard	0.99	1.08	1.17	1.00	1.08	1.04
Seasonal	1.08	1.08	1.34	1.06	1.04	1.04
ARS 6						
All	1.06	1.07	1.22	1.03	1.06	0.98
Seasonal	1.11	1.09	1.34	1.06	1.05	1.00
ARS 12						
All	1.11	1.13	1.44	1.08	1.14	1.06
Seasonal	1.11	1.09	1.46	1.08	1.08	1.00

In summary, the conclusion is that the significance of the seasonal coefficients decreases as we move from short-term forecasting to longer-term. In the same way, additional variables in the autoregressive part become less meaningful (order 6 models perform better than order 12). Hence, it seems that simpler models tend to perform better when the forecasting horizon is longer. LP is shown to be a good tool for short-term forecasting (and especially for smooth and seasonal series) but the value of optimisation for longer term forecasts is less clear. The differences between LP and OLS are small.

4.7 SUM OF PERCENTAGE ERRORS EQUAL TO ZERO

Finally, I decided to run the experiments again changing the constraint that removes the underestimation and overestimation bias. I change the constraint sum of errors equal to zero ($\Sigma e = 0$) to sum of percentage errors equal to zero ($\Sigma e/Y = 0$). I apply this constraint to all five LP formulations and compare by the results with the OLS approach. The sum of percentage errors equal to zero constraint may not completely remove the bias; it is a milder condition than the sum of errors equal to zero constraint and can be seen as some kind of relaxation.

Tables 4.16, 4.17 and 4.18 show the results for one period ahead forecasts (average MAPE, sMAPE and MASE respectively); the OLS results remain the same. As we can see, the new constraint improves the results in many cases and specifically, on smooth and seasonal series for all the models, and on the hard series for MinSAPE (order 6 and 12)and MinMaxAD (order 6). On the other hand, it gives worse results on the hard series for the other LP approaches. MinMaxAD remains the worst method, yet in comparison with the other LP approaches, the performance of MinMaxAD improved a lot.

In comparison with the OLS, the relationship remains the same. LP performs better on the smooth and seasonal series, while OLS performs better on the hard series. OLS outperforms LP according to MAPE and sMAPE (except forthe AR/ARS 6 model with objective MinSAD). MinADBD and ADBPD outperform OLS according to the MASE accuracy measure.

Table 4. 16 Sum of percentage errors equal to zero - MAPE - 1 step ahead

MAPE	**LP**										**OLS**
1 STEP	**MinSAD**		**MinSAPE**		**MinMaxAD**		**MinADBD**		**MinADBPD**		
Σ%e = 0											
AR 6											
All	8.64	(-0.13)	8.70	(-0.34)	9.27	(-1.82)	8.76	(-0.03)	8.78	(-0.01)	8.69
Smooth	3.02	(-0.15)	3.05	(-0.15)	3.98	(0)	2.99	(-0.19)	2.99	(-0.21)	3.14
Hard	24.94	(0.14)	25.20	(-1.06)	25.14	(-6.05)	25.71	(0.94)	26.09	(1.12)	24.47
Seasonal	9.58	(-0.22)	9.57	(-0.28)	9.85	(-2.7)	9.59	(-0.29)	9.46	(-0.27)	9.79
AR 12											
All	8.36	(-0.01)	8.50	(-0.29)	8.89	(-0.47)	8.16	(0.08)	8.42	(-0.15)	8.01
Smooth	2.98	(-0.18)	3.09	(-0.07)	4.19	(-0.04)	2.91	(-0.27)	2.89	(-0.19)	3.22
Hard	26.83	(1.05)	27.42	(-1.09)	25.81	(0.66)	26.08	(1.7)	27.79	(0.06)	23.69
Seasonal	7.68	(-0.29)	7.60	(-0.25)	7.85	(-1.89)	7.55	(-0.18)	7.48	(-0.19)	7.80
ARS 6											
All	7.41	(-0.07)	7.47	(-0.27)	7.96	(-1.7)	7.52	(0.05)	7.59	(0.06)	7.39
Seasonal	5.46	(-0.05)	5.47	(-0.04)	5.51	(-2.29)	5.46	(-0.03)	5.49	(-0.05)	5.45
ARS 12											
All	7.52	(0.07)	7.67	(-0.22)	7.99	(-0.78)	7.36	(0.13)	7.61	(-0.1)	7.22
Seasonal	4.86	(-0.03)	4.85	(0)	4.85	(-2.89)	4.86	(-0.04)	4.78	(-0.05)	5.19

Table 4. 17 Sum of percentage errors equal to zero - sMAPE - 1 step ahead

sMAPE	LP										OLS
1 STEP	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
Σ%e = 0											
AR 6											
All	8.88	(0.41)	8.98	(0.23)	9.04	(-2.05)	8.69	(0.22)	8.72	(0.07)	8.48
Smooth	2.94	(-0.09)	2.94	(-0.1)	3.90	(-0.09)	2.93	(-0.12)	2.91	(-0.14)	3.03
Hard	26.38	(2.64)	26.99	(1.73)	24.24	(-6.94)	25.71	(1.87)	26.09	(1.12)	24.01
Seasonal	9.71	(0.05)	9.71	(-0.02)	9.73	(-2.84)	9.47	(-0.11)	9.40	(-0.12)	9.54
AR 12											
All	8.45	(0.24)	8.59	(-0.24)	8.60	(-0.91)	8.01	(0.17)	8.26	(-0.1)	7.81
Smooth	2.87	(-0.12)	2.97	(-0.03)	4.10	(-0.21)	2.83	(-0.18)	2.79	(-0.13)	3.07
Hard	28.37	(2.26)	29.05	(-0.94)	25.11	(-1.26)	26.08	(1.87)	27.79	(0.06)	23.74
Seasonal	7.31	(-0.23)	7.23	(-0.21)	7.43	(-1.98)	7.19	(-0.13)	7.12	(-0.14)	7.40
ARS 6											
All	7.58	(0.38)	7.68	(0.23)	7.74	(-1.86)	7.46	(0.24)	7.52	(0.1)	7.24
Seasonal	5.38	(-0.04)	5.38	(-0.03)	5.41	(-2.18)	5.39	(-0.02)	5.41	(-0.04)	5.39
ARS 12											
All	7.69	(0.31)	7.85	(-0.17)	7.80	(-1.16)	7.29	(0.21)	7.53	(-0.07)	7.12
Seasonal	4.78	(-0.01)	4.76	(0.01)	4.75	(-2.8)	4.77	(-0.03)	4.69	(-0.04)	5.10

Table 4. 18 Sum of percentage errors equal to zero - MASE - 1 step ahead

MASE	LP										OLS
1 STEP	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
Σ%e = 0											
AR 6											
All	0.92	(0.01)	0.93	(0)	1.04	(-0.16)	0.91	(0)	0.91	(0)	0.92
Smooth	0.94	(0)	0.94	(0)	1.18	(-0.09)	0.93	(-0.01)	0.94	(-0.01)	0.95
Hard	0.94	(0.06)	0.95	(0.03)	0.88	(-0.23)	0.92	(0.04)	0.92	(0.02)	0.88
Seasonal	0.89	(-0.01)	0.88	(-0.01)	0.89	(-0.23)	0.88	(-0.01)	0.87	(-0.01)	0.88
AR 12											
All	0.90	(0)	0.91	(0)	1.07	(-0.06)	0.87	(0)	0.88	(-0.01)	0.88
Smooth	0.95	(0)	0.96	(0)	1.29	(0.01)	0.93	(-0.01)	0.93	(-0.01)	0.95
Hard	1.02	(0.06)	1.05	(0.03)	0.91	(-0.07)	0.94	(0.04)	0.97	(0.01)	0.88
Seasonal	0.75	(-0.02)	0.74	(-0.02)	0.76	(-0.18)	0.73	(-0.01)	0.73	(-0.01)	0.75
ARS 6											
All	0.81	(0.01)	0.82	(0)	0.93	(-0.17)	0.81	(0)	0.81	(0)	0.81
Seasonal	0.53	(0)	0.52	(0)	0.53	(-0.26)	0.52	(0)	0.52	(0)	0.52
ARS 12											
All	0.83	(0.01)	0.84	(0.01)	0.99	(-0.07)	0.80	(0)	0.80	(0)	0.81
Seasonal	0.51	(0)	0.51	(0)	0.50	(-0.22)	0.50	(0)	0.49	(0)	0.52

I also ran the experiments for six and twelve steps ahead forecasts. Table 4.19, 4.20 and 4.21 show the results of the LP, as well as OLS, for six steps ahead forecasts. Adding the new constraint improves the LP results on the smooth and seasonal series. The results on the hard series are less conclusive. MinADBPD is significantly worse than the other approaches according to sMAPE due to a negative forecast for one of the hard series. As before, the OLS performs well on the hard series, while the LP models perform better on the seasonal and smooth series. The performance of the order 6 models tends to be better than the order 12 models.

Table 4. 19 Sum of percentage errorrs equal to zero - MAPE - 6 steps ahead

MAPE	LP										OLS
6 STEP	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
Σ%e = 0											
AR 6											
All	11.29	(-0.7)	11.24	(-0.79)	14.42	(0.09)	10.97	(-0.58)	11.10	(-0.58)	11.41
Smooth	6.53	(-0.81)	6.72	(-0.67)	7.89	(-0.47)	6.44	(-0.82)	6.77	(-0.77)	7.09
Hard	26.41	(-0.23)	25.79	(-1.05)	31.49	(3.96)	25.11	(0.87)	24.99	(0.64)	24.03
Seasonal	11.36	(-0.76)	11.19	(-0.85)	16.55	(-1.05)	11.18	(-0.94)	11.09	(-0.92)	12.08
AR 12											
All	11.06	(-0.88)	11.13	(-1)	13.57	(-0.57)	10.84	(-0.36)	11.09	(-0.6)	11.17
Smooth	6.96	(-0.91)	7.08	(-0.83)	9.85	(2.09)	6.84	(-0.99)	7.28	(-0.65)	7.76
Hard	27.58	(-1.5)	27.57	(-2.63)	29.21	(-1.19)	26.75	(1.63)	26.85	(-0.96)	25.35
Seasonal	9.16	(-0.48)	9.18	(-0.42)	11.48	(-4.98)	9.09	(-0.35)	9.11	(-0.31)	9.34
ARS 6											
All	10.28	(-0.56)	10.24	(-0.6)	12.47	(0.38)	9.92	(-0.35)	10.06	(-0.51)	10.96
Seasonal	7.99	(-0.3)	7.85	(-0.24)	10.05	(-0.07)	7.67	(-0.2)	7.61	(-0.68)	10.59
ARS 12											
All	10.72	(-0.77)	10.76	(-0.93)	13.34	(1.19)	10.38	(-0.33)	10.63	(-0.98)	11.27
Seasonal	8.03	(-0.11)	7.95	(-0.18)	10.71	(0.87)	7.58	(-0.25)	7.59	(-1.56)	9.67

Table 4. 20 Sum of percentage errors equal to zero - sMAPE - 6 steps ahead

sMAPE 6 STEP Σ%e = 0	LP										OLS
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
AR 6											
All	11.41	(0.01)	11.31	(-0.11)	12.67	(-1.47)	11.14	(0.17)	11.19	(0.16)	10.88
Smooth	6.18	(-0.45)	6.36	(-0.32)	7.60	(-0.16)	6.11	(-0.44)	6.44	(-0.36)	6.42
Hard	29.60	(2.37)	28.70	(1.41)	23.43	(-6.97)	28.49	(3.63)	27.94	(3.3)	24.80
Seasonal	10.61	(-0.47)	10.43	(-0.59)	15.71	(-0.75)	10.43	(-0.67)	10.35	(-0.65)	11.08
AR 12											
All	11.49	(-0.16)	11.90	(-2.55)	13.37	(0)	11.21	(0.42)	16.01	(-1.98)	11.32
Smooth	6.39	(-0.48)	6.62	(-0.3)	8.87	(0)	6.29	(-0.57)	8.09	(1.15)	6.84
Hard	33.01	(1.26)	34.72	(-13.78)	32.26	(0)	31.80	(4.78)	41.50	(-28.49)	30.32
Seasonal	8.60	(-0.39)	8.61	(-0.33)	10.86	(0)	8.51	(-0.27)	15.93	(7.16)	8.73
ARS 6											
All	10.52	(0.08)	10.43	(0.02)	10.89	(-1.26)	10.21	(0.32)	10.27	(0.35)	10.58
Seasonal	7.63	(-0.24)	7.50	(-0.18)	9.78	(-0.06)	7.34	(-0.17)	7.28	(0)	10.08
ARS 12											
All	11.20	(-0.07)	11.58	(-2.5)	13.20	(0)	10.81	(0.43)	13.39	(-5.63)	11.45
Seasonal	7.62	(-0.08)	7.55	(-0.13)	10.29	(0)	7.18	(-0.21)	7.21	(-4.96)	9.17

Table 4. 21 Sum of percentage errors equal to zero - MASE - 6 steps ahead

MASE 6 STEP Σ%e = 0	LP										OLS
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
AR 6											
All	0.87	(-0.01)	0.88	(-0.01)	1.05	(0)	0.85	(-0.01)	0.87	(0)	0.85
Smooth	0.96	(-0.01)	0.98	(0)	1.10	(0)	0.94	(-0.01)	0.98	(0)	0.93
Hard	0.91	(0.04)	0.89	(0.01)	0.97	(0.08)	0.87	(0.07)	0.86	(0.07)	0.79
Seasonal	0.69	(-0.03)	0.68	(-0.04)	1.01	(-0.05)	0.69	(-0.04)	0.68	(-0.04)	0.72
AR 12											
All	0.85	(-0.02)	0.87	(0)	1.03	(0)	0.83	(0)	1.31	(0.45)	0.83
Smooth	0.96	(-0.02)	1.00	(0.01)	1.21	(0)	0.93	(-0.02)	1.55	(0.57)	0.96
Hard	0.97	(0.02)	0.95	(-0.01)	0.98	(0)	0.94	(0.07)	1.08	(0.15)	0.84
Seasonal	0.60	(-0.02)	0.60	(-0.02)	0.74	(0)	0.59	(-0.02)	1.00	(0.39)	0.60
ARS 6											
All	0.82	(-0.01)	0.83	(0)	0.95	(0.02)	0.80	(0)	0.82	(0.01)	0.84
Seasonal	0.52	(-0.02)	0.51	(-0.01)	0.67	(0.01)	0.50	(-0.02)	0.50	(-0.01)	0.69
ARS 12											
All	0.84	(0)	0.85	(0)	1.02	(0)	0.81	(0)	1.16	(0.24)	0.84
Seasonal	0.54	(-0.01)	0.54	(-0.01)	0.70	(0)	0.51	(-0.02)	0.51	(-0.3)	0.64

Finally Tables 4.22, 4.23 and 4.24 show the results, for 12 steps ahead forecasting. Similarly, the results are improved on the smooth and seasonal series. The only results that are worse are the performance on the hard series. In comparison with the OLS, all the LP models are generally better according to MAPE, and the OLS is better according to sMAPE and MASE. LP performs better on the smooth series and OLS on the hard series.

Table 4. 22 Sum of percentage errors equal to zero - MAPE - 12 steps ahead

MAPE 12 STEP Σ%e = 0	LP										OLS
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
AR 6											
All	12.48	(-0.59)	12.42	(-0.9)	15.28	(0)	12.32	(-0.53)	12.40	(-0.83)	12.55
Smooth	9.26	(-1.29)	9.49	(-1.23)	11.82	(0)	9.17	(-1.36)	9.53	(-1.28)	10.04
Hard	27.36	(1.64)	26.24	(-0.49)	32.03	(0)	26.84	(2.09)	26.19	(-0.12)	25.07
Seasonal	9.94	(-0.59)	9.94	(-0.53)	12.13	(0)	9.86	(-0.51)	9.86	(-0.43)	10.06
AR 12											
All	13.66	(-0.56)	14.07	(-0.98)	21.47	(0)	13.31	(-0.88)	13.83	(-0.95)	13.85
Smooth	10.24	(-1.1)	10.58	(-0.93)	22.17	(0)	9.86	(-1.56)	10.32	(-1.22)	11.04
Hard	30.35	(1.36)	31.74	(-2)	34.38	(0)	29.76	(0.66)	31.64	(-1.05)	28.97
Seasonal	10.44	(-0.69)	10.47	(-0.5)	13.07	(0)	10.31	(-0.52)	10.18	(-0.41)	10.44
ARS 6											
All	12.61	(-0.55)	12.53	(-0.85)	15.33	(-0.13)	12.37	(-0.5)	12.45	(-0.79)	12.58
Seasonal	10.35	(-0.46)	10.30	(-0.36)	12.28	(-0.45)	10.04	(-0.4)	10.01	(-0.28)	10.14
ARS 12											
All	13.64	(-0.49)	14.03	(-0.91)	21.88	(-0.02)	13.29	(-0.86)	13.78	(-1.01)	13.61
Seasonal	10.38	(-0.46)	10.33	(-0.26)	14.42	(-0.07)	10.22	(-0.48)	10.02	(-0.62)	9.65

Table 4. 23 Sum of percentage errors equal to zero - sMAPE - 12 steps ahead

sMAPE 12 STEP Σ%e = 0	LP										OLS
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
AR 6											
All	13.03	(0.54)	12.83	(-0.02)	14.93	(0)	12.73	(0.42)	12.75	(-0.05)	12.18
Smooth	9.11	(-0.4)	9.19	(-0.34)	10.58	(0)	9.07	(-0.55)	9.26	(-0.51)	9.25
Hard	32.40	(5.37)	30.99	(1.71)	35.78	(0)	30.93	(4.97)	30.45	(1.91)	26.85
Seasonal	9.22	(-0.47)	9.21	(-0.42)	11.07	(0)	9.13	(-0.4)	9.11	(-0.33)	9.26
AR 12											
All	14.99	(1.07)	12.29	(-1.88)	17.94	(0)	14.21	(0.25)	14.21	(1.14)	13.48
Smooth	9.77	(-0.45)	10.00	(-0.2)	14.60	(0)	9.64	(-0.63)	9.64	(-0.72)	10.17
Hard	41.30	(8.86)	24.33	(-9.96)	39.26	(0)	37.29	(4.22)	37.29	(9.47)	31.03
Seasonal	9.64	(-0.55)	9.68	(-0.38)	12.03	(0)	9.51	(-0.4)	9.51	(-0.19)	9.61
ARS 6											
All	13.25	(0.68)	13.02	(0.1)	18.98	(3.86)	12.78	(0.44)	14.30	(1.52)	12.42
Seasonal	9.96	(0)	9.83	(0)	24.57	(12.86)	9.29	(-0.3)	9.26	(-0.14)	10.05
ARS 12											
All	15.09	(1.24)	12.31	(-1.77)	22.98	(4.77)	14.19	(0.26)	16.23	(3.15)	13.44
Seasonal	9.98	(0)	9.74	(0)	28.84	(15.89)	9.43	(-0.37)	9.26	(-0.45)	9.49

Table 4. 24 Sum of percentage errors equal to zero - MASE - 12 steps ahead

MASE 12 STEP Σ%e = 0	LP										OLS
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD		
AR 6											
All	1.05	(0.01)	1.07	(0.02)	1.19	(0)	1.02	(0.01)	1.06	(0.01)	0.98
Smooth	1.10	(0.03)	1.14	(0.04)	1.20	(0)	1.06	(0)	1.14	(0.02)	1.00
Hard	0.97	(0.06)	0.96	(0.04)	1.06	(0)	0.97	(0.09)	0.95	(0.05)	0.90
Seasonal	1.00	(-0.04)	1.00	(-0.04)	1.25	(0)	0.99	(-0.04)	0.99	(-0.03)	1.00
AR 12											
All	1.11	(0.01)	1.17	(0.04)	1.41	(0)	1.07	(-0.01)	1.07	(-0.06)	1.07
Smooth	1.17	(0.03)	1.27	(0.1)	1.52	(0)	1.11	(-0.01)	1.11	(-0.09)	1.10
Hard	1.06	(0.07)	1.08	(0)	1.17	(0)	1.05	(0.05)	1.05	(-0.03)	1.04
Seasonal	1.04	(-0.05)	1.04	(-0.03)	1.34	(0)	1.02	(-0.04)	1.02	(-0.02)	1.04
ARS 6											
All	1.08	(0.02)	1.10	(0.03)	1.20	(-0.02)	1.04	(0.01)	1.09	(0.03)	0.98
Seasonal	1.11	(0)	1.09	(0)	1.28	(-0.07)	1.03	(-0.03)	1.03	(-0.02)	1.00
ARS 12											
All	1.13	(0.03)	1.19	(0.05)	1.44	(0)	1.08	(-0.01)	1.09	(-0.05)	1.06
Seasonal	1.11	(0)	1.09	(0)	1.45	(-0.01)	1.04	(-0.03)	1.02	(-0.06)	1.00

As we can see, the new constraint improves the overall performance (MAPE) of the LP models and brings the level of accuracy to the same or better level as the OLS. The improvement is mainly observed in the smooth and seasonal series.

4.8 CONCLUSIONS

The objective of the chapter was to explore if linear programming can be used as a tool for optimising the parameters of autoregressive forecasting models. LP has been suggested as an alternative method to estimate the parameters of a linear regression in the past, but it was never tested on forecasting or compared with OLS. The analysis shows that LP can be used to optimise time series forecast and that it performs as good as the OLS. The additional constraints (sum of errors equal to zero and sum of percentage errors equal to zero) have helped to overcome the problem of biased forecasts. The formulation of a linear program is quite easy and the computational time for the calculation of the solution is less than a second. Hence, it should be easy for decision makers to implement and use these approaches. The findings of this chapter are summarised as follows:

- LP is a good alternative to the OLS for optimising autoregressive based forecasts.
- LP performs better than the OLS on series with low variability (smooth and seasonal), while it is worse on series with high variability (hard).
- The addition of constraints that remove bias improve the results in the test set (for seasonal and smooth series).
- The constraint that the sum of the percentage errors should be zero gives better result compared with the constraint where the sum of errors is zero on the smooth and seasonal series.
- ARS models are much better than AR models for short term forecasts in seasonal series. For long term forecasting it seems that the addition of a seasonal coefficient in the seasonal series does not improve the forecast.
- The best performing LP approaches are MinADBD and MinADBPD, but depending on the characteristics of the series, the decision maker may need to focus on different optimisation objectives, such as MinSAD, MinSAPE, MinADBD, MinADBPD (LP) and MinSSE (OLS).

However, MinMaxAD seems to be a bad choice as an objective because it always performs worse compared with the others.

- The differences between all the models are rather small.

In this chapter we show that simple linear programming can be used to estimate the parameters of autoregressive forecasting models. The objectives of the linear programs are to minimise one accuracy measure. The next step is to use linear goal programming, where the objective would be to minimise two accuracy measures.

5 GOAL PROGRAMMING FOR TIME SERIES FORECASTING

According to Makridakis et al. (1984) "*The performance of a technique may differ according to different accuracy measures*". Traditional tools aim to optimise one accuracy index (e.g. the least squares method minimises the SSE). LP, in contrast with other techniques, can be applied for multi-objective optimisation. I tested two types of goal programming formulations: Pre-emptive Goal Programming and Weighted (non pre-emptive) Goal Programming., The objective is to minimise both SAD and MaxAD (MinSum-MinMax). The first part of the chapter presents the pre-emptive goal programming formulation followed by the results (average MAPE, sMAPE and MASE on the test set) and comparison with OLS and the single objective approaches (MinSAD and MinMaxAD). The same programme is reformulated in a weighted goal programming form. The approaches are tested for one, six and twelve steps ahead forecasting and the results are compared with OLS and single objective LPs.

5.1 PRE-EMPTIVE GOAL PROGRAMMING

The first formulation is a MinMax-MinSum pre-emptive goal program (GP), which minimises MaxAD as a first objective and SAD as a second. The GP is applied to optimise the parameters of simple autoregressive **(4. 2)** and seasonal autoregressive **(4. 21)** models of both order 6 and 12. The GP formulations are similar to the final single objective LP formulations of the previous chapter, ignoring the first $s + m - 1$ data points and with the constraints sum of errors equal to zero or sum of percentage errors equal to zero.

For the simple autoregressive model (without base) for one period ahead forecast, the linear goal program for minimising MaxAD and SAD is (assuming that Y_{i-j} is not defined for $j > i$):

$$Min\ \ P_1\left(d_1^+\right)+P_2\left(d_2^+\right) \tag{5. 1}$$

subject to:

$$e + d_1^- - d_1^+ = 0 \quad \textbf{(5. 3)}$$

$$\sum_{i=m+1}^{n} \left(e_{1i} + e_{2i}\right) + d_2^- - d_2^+ = 0 \quad \textbf{(5. 4)}$$

where d_1^+, d_1^-, d_2^+ and d_2^- are the deviational variables, e_{1i} is the underestimation error, e_{2i} the overestimation error in period i and e the MaxAD.

(4. 27), **(4. 19)**

With **(4. 31)** for sum of errors equal to zero or

$$\sum_{i=m+1}^{n} \left(e_{1i} - e_{2i}\right) / Y_1 = 0 \quad \textbf{(5. 5)}$$

For sum or percentage errors equal to zero

$$e_{1i}, e_{2i}, e, d_1^+, d_1^-, d_2^+, d_2^- \geq 0$$

and

b_j unrestricted in sign.

In the same way, for the seasonal autoregressive model, the linear goal program for minimising MaxAD and SAD is (assuming that Y_{i-j} is not defined for $j > i$):

$$Min \;\; P_1\left(d_1^+\right) + P_2\left(d_2^+\right) \quad \textbf{(5. 6)}$$

subject to:

(5. 3), **(5. 4)**, **(4. 29)**, **(4. 19)**, **(4. 32)**

With **(4. 31)** for sum of errors equal to zero or **(5. 5)** for sum of percentage errors equal to zero.

Initial experiments showed that setting the optimal solution of the first goal as a hard constraint does not leave much room for optimising the second goal, and typically results in the same solution as obtained by single objective optimisation. To increase the flexibility of the GP the experiments are ran again whilst relaxing the optimal value of the first goal. A relaxation coefficient is used equal to 1.1, 1.2, 1.3 and 1.4. If *opt_value_1* denotes the optimal solution value of the first optimisation (Min d_1^+); then in the second optimisation we change the objective function to Min d_2^+ and add a constraint $d_1^+ \leq$ *relaxation coefficient* × *opt_value_1*.

5.2 RESULTS: PREEMPTIVE GOAL PROGRAMMING

Table 5.1 shows the average MAPE of the Pre-emptive Goal Programming approach on the test set, which minimises MaxAD as a first objective and SAD as a second objective for one period ahead forecasts for sum of errors equal to zero and for the four relaxation factors.

Table 5. 1 Pre-emptive goal programming - sum of errors equal to zero - MAPE

MAPE	LP				Single Objective LP		OLS
1 STEP	MinMax-MinSAD				MinSAD	MinMaxAD	
Relaxation factor	1.1	1.2	1.3	1.4			
Σe = 0							
AR 6							
All	9.13	8.91	8.91	8.98	8.64	9.27	8.69
Smooth	3.33	3.19	3.17	3.18	3.02	3.98	3.14
Hard	25.14	24.40	24.44	24.83	24.94	25.14	24.47
Seasonal	10.55	10.46	10.51	10.49	9.58	9.85	9.79
AR 12							
All	8.60	8.40	8.24	8.28	8.36	8.89	8.01
Smooth	3.65	3.61	3.51	3.43	2.98	4.19	3.22
Hard	24.47	23.73	23.51	24.11	26.83	25.81	23.69
Seasonal	8.59	8.39	8.16	8.12	7.68	7.85	7.80
ARS 6							
All	7.73	7.47	7.45	7.53	7.41	7.96	7.39
Seasonal	5.86	5.67	5.61	5.66	5.46	5.51	5.45
ARS 12							
All	8.02	7.69	7.61	7.50	7.52	7.99	7.22
Seasonal	6.65	6.04	5.74	5.52	4.86	4.85	5.19

Factor 1.2 gives the best results for the AR 6 model, 1.3 for the AR 12 and ARS 6 and 1.4 for the ARS 12; however, in the last case, the best results on the hard series are given by the factor 1.3.

If we compare this GP with the single objective LP models, we can see that the GP improves the results on the hard series for the AR/ARS 6 models with factors 1.2 and 1.3 and for the AR/ARS 12 models.

On the other hand, OLS performs generally better than this GP; however, the latter performs better on the hard series for the AR/ARS 6 models with factors 1.2 and 1.3 and for the AR/ARS 12 with factor 1.3.

Table 5.2 shows the average sMAPE for sum of errors equal to zero and one step ahead forecasts.

Table 5. 2 Pre-emptive goal programming - sum of errors equal to zero - sMAPE

sMAPE	LP				Single Objective LP		OLS
1 STEP	MinMax-MinSAD				MinSAD	MinMaxAD	
Relaxation factor	1.1	1.2	1.3	1.4			
Σe = 0							
AR 6							
All	8.85	8.48	8.41	8.48	8.88	9.04	8.48
Smooth	3.26	3.10	3.05	3.06	2.94	3.9	3.03
Hard	24.83	23.43	23.34	23.79	26.38	24.24	24.01
Seasonal	9.91	9.74	9.63	9.62	9.71	9.73	9.54
AR 12							
All	8.45	8.21	8.12	8.16	8.45	8.6	7.81
Smooth	3.53	3.49	3.38	3.33	2.87	4.1	3.07
Hard	24.64	23.94	23.94	24.45	28.37	25.11	23.74
Seasonal	8.19	7.86	7.75	7.70	7.31	7.43	7.40
ARS 6							
All	7.58	7.21	7.15	7.23	7.58	7.74	7.24
Seasonal	5.67	5.50	5.44	5.45	5.38	5.41	5.39
ARS 12							
All	7.78	7.42	7.37	7.36	7.69	7.8	7.12
Seasonal	5.97	5.24	5.25	5.05	4.78	4.75	5.10

In contrast with the comparison according to the MAPE, factor 1.3 gives the most accurate results overall and on the smooth and hard series. The relaxation factor 1.4 is the most accurate on the seasonal series.

In comparison with the single objective techniques, GP improves the results overall. Similarly as in the comparison according to the MAPE, the main improvement can be found on the hard series, while MinSAD performs better on smooth and seasonal series.

OLS performs better for order 12, whereas GP performs better for order 6.

Table 5.3 shows the average MASE for sum of errors equal to zero. Similarly as in the MAPE and sMAPE tables, the best GP approaches are these with relaxation factor 1.3 and 1.4. GP performs better than simple LP and OLS on the hard series, but worse on the smooth and seasonal series. However, the differences between all the approaches are small.

Table 5. 3 Pre-emptive goal programming - sum of errors equal to zero - MASE

MASE	LP				Single Objective LP		OLS
1 STEP	MinMax-MinSAD				MinSAD	MinMaxAD	
Relaxation factor	1.1	1.2	1.3	1.4			
Σe = 0							
AR 6							
All	0.97	0.93	0.92	0.92	0.92	1.04	0.92
Smooth	1.00	0.97	0.95	0.95	0.94	1.18	0.95
Hard	0.92	0.87	0.86	0.88	0.94	0.88	0.88
Seasonal	0.93	0.90	0.89	0.89	0.89	0.89	0.88
AR 12							
All	0.99	0.96	0.93	0.93	0.9	1.07	0.88
Smooth	1.11	1.07	1.03	1.03	0.95	1.29	0.95
Hard	0.92	0.87	0.86	0.88	1.02	0.91	0.88
Seasonal	0.82	0.80	0.78	0.78	0.75	0.76	0.75
ARS 6							
All	0.86	0.82	0.81	0.81	0.81	0.93	0.81
Seasonal	0.55	0.53	0.52	0.53	0.53	0.53	0.52
ARS 12							
All	0.92	0.88	0.86	0.85	0.83	0.99	0.81
Seasonal	0.59	0.54	0.54	0.52	0.51	0.5	0.52

The next three tables (5.4, 5.5 and 5.6) show the performance of the approaches with the constraint of sum of percentage errors equal to zero according to MAPE, sMAPE and MASE respectively. The results are compared with the single objective LPs with sum of percentage errors equal to zero and OLS. The results are also compared with those in Table 5.1 (GP with sum of errors equal to zero constraint). The number in brackets shows the difference in MAPE (negative for an improvement and positive for a worse result).

Similarly as for the single objective approaches (Chapter 4), adding the sum of percentage errors equal to zero constraint, improves the results on the smooth and seasonal series but worsens the results on the hard series (except for relaxation factor 1.4).GP with relaxation factor 1.4 is the best formulation overall and for each type of series separately. GP outperforms simple LP overall and OLS for order 6. The former performs better on smooth and seasonal series, while the latter performs better on the hard series.

Table 5. 4 Pre-emptive goal programming - sum of percentage errors equal to zero - MAPE

MAPE 1 STEP R. F. Σ%e = 0	LP								Single Objective LP		OLS
	MinMax-MinSAD								MinSAD	MinMaxAD	
	1.1		1.2		1.3		1.4				
AR 6											
All	9.32	(0.19)	8.85	(-0.06)	8.66	(-0.25)	8.57	(-0.41)	8.64	9.27	8.69
Smooth	3.19	(-0.14)	3.06	(-0.13)	3.00	(-0.17)	2.98	(-0.2)	3.02	3.98	3.14
Hard	27.85	(2.71)	25.88	(1.48)	24.99	(0.55)	24.71	(-0.12)	24.94	25.14	24.47
Seasonal	9.92	(-0.63)	9.69	(-0.77)	9.65	(-0.86)	9.56	(-0.93)	9.58	9.85	9.79
AR 12											
All	8.82	(0.22)	8.36	(-0.04)	8.08	(-0.16)	8.06	(-0.22)	8.36	8.89	8.01
Smooth	3.58	(-0.07)	3.42	(-0.19)	3.35	(-0.16)	3.26	(-0.17)	2.98	4.19	3.22
Hard	26.45	(1.98)	25.11	(1.38)	23.92	(0.41)	24.13	(0.02)	26.83	25.81	23.69
Seasonal	8.36	(-0.23)	7.83	(-0.56)	7.70	(-0.46)	7.67	(-0.45)	7.68	7.85	7.80
ARS 6											
All	8.09	(0.36)	7.62	(0.15)	7.41	(-0.04)	7.34	(-0.19)	7.41	7.96	7.39
Seasonal	5.82	(-0.04)	5.59	(-0.08)	5.49	(-0.12)	5.46	(-0.2)	5.46	5.51	5.45
ARS 12											
All	8.13	(0.11)	7.67	(-0.02)	7.38	(-0.23)	7.29	(-0.21)	7.52	7.99	7.22
Seasonal	6.04	(-0.61)	5.52	(-0.52)	5.35	(-0.39)	5.11	(-0.41)	4.86	4.85	5.19

The results are similar comparing the approaches according to sMAPE (Table 5.5). The best GP is with factor 1.4 with only exception AR 12, where the best is the approach with factor 1.3. GP outperforms simple LP overall (smooth, hard and seasonal) and performs better than the OLS on the smooth and hard series.

Table 5.6 shows the comparison of the techniques according to MASE. The differences with the sum of errors equal to zero constraint are relatively small. However, a small improvement is observed on the smooth and seasonal series, but it is worse on the hard series. On the other hand, it seems that GP performs worse overall compared with the simple LP and OLS.

Table 5. 5 Pre-emptive goal programming - sum of percentage errors equal to zero - sMAPE

sMAPE	LP								Single Objective LP		OLS
1 STEP	MinMax-MinSAD								MinSAD	MinMaxAD	
R. F.	1.1		1.2		1.3		1.4				
Σ%e = 0											
AR 6											
All	9.58	(0.73)	9.03	(0.55)	8.85	(0.45)	8.78	(0.3)	8.88	9.04	8.48
Smooth	3.16	(-0.1)	3.02	(-0.08)	2.96	(-0.09)	2.92	(-0.14)	2.94	3.90	3.03
Hard	29.66	(4.84)	27.09	(3.66)	26.23	(2.89)	26.01	(2.21)	26.38	24.24	24.01
Seasonal	9.85	(-0.07)	9.68	(-0.06)	9.68	(0.05)	9.62	(0)	9.71	9.73	9.54
AR 12											
All	9.20	(0.75)	8.68	(0.47)	8.29	(0.17)	8.33	(0.17)	8.45	8.60	7.81
Smooth	3.52	(-0.01)	3.37	(-0.13)	3.29	(-0.09)	3.19	(-0.13)	2.87	4.10	3.07
Hard	29.44	(4.8)	27.71	(3.78)	25.93	(1.99)	26.53	(2.09)	28.37	25.11	23.74
Seasonal	8.04	(-0.15)	7.54	(-0.32)	7.39	(-0.36)	7.35	(-0.35)	7.31	7.43	7.40
ARS 6											
All	8.34	(0.77)	7.77	(0.57)	7.57	(0.42)	7.51	(0.27)	7.58	7.74	7.24
Seasonal	5.71	(0.04)	5.49	(-0.01)	5.39	(-0.05)	5.37	(-0.08)	5.38	5.41	5.39
ARS 12											
All	8.56	(0.78)	8.04	(0.61)	7.65	(0.28)	7.63	(0.26)	7.69	7.80	7.12
Seasonal	5.93	(-0.03)	5.41	(0.17)	5.25	(0)	5.01	(-0.04)	4.78	4.75	5.10

Table 5. 6 Pre-emptive goal programming - sum of percentage errors equal to zero - MASE

MASE	LP								Single Objective LP		OLS
1 STEP	MinMax-MinSAD								MinSAD	MinMaxAD	
R. F.	1.1		1.2		1.3		1.4				
Σ%e = 0											
AR 6											
All	0.98	(0.01)	0.94	(0.01)	0.93	(0.01)	0.92	(0)	0.92	1.04	0.92
Smooth	1.00	(0)	0.96	(0)	0.95	(-0.01)	0.94	(-0.01)	0.94	1.18	0.95
Hard	1.01	(0.09)	0.94	(0.08)	0.92	(0.06)	0.92	(0.04)	0.94	0.88	0.88
Seasonal	0.93	(0)	0.90	(0)	0.89	(0)	0.88	(0)	0.89	0.89	0.88
AR 12											
All	1.00	(0.01)	0.95	(0)	0.93	(0)	0.92	(-0.01)	0.90	1.07	0.88
Smooth	1.11	(0)	1.06	(-0.01)	1.03	(0)	1.01	(-0.01)	0.95	1.29	0.95
Hard	0.99	(0.08)	0.95	(0.08)	0.92	(0.06)	0.92	(0.04)	1.02	0.91	0.88
Seasonal	0.80	(-0.02)	0.77	(-0.03)	0.75	(-0.03)	0.75	(-0.03)	0.75	0.76	0.75
ARS 6											
All	0.87	(0.01)	0.83	(0.01)	0.82	(0.01)	0.81	(0)	0.81	0.93	0.81
Seasonal	0.56	(0)	0.53	(0)	0.52	(0)	0.52	(0)	0.53	0.53	0.52
ARS 12											
All	0.93	(0.01)	0.89	(0.01)	0.86	(0.01)	0.85	(0)	0.83	0.99	0.81
Seasonal	0.59	(0)	0.54	(0.01)	0.54	(0)	0.52	(0)	0.51	0.50	0.52

Pre-emptive goal programming gives good results on the hard series (in Table 5.1 MinMaxAD&MAD give the best results on the hard series). It seems that according to the characteristic of the series, we may have to focus on different optimisation objectives.

5.3 WEIGHTED GOAL PROGRAMMING

An alternative way to relax the first goal of the linear goal program is by applying an approach similar to Lagangian relaxation. In pre-emptive GP the decision maker optimises the initial goal first and then the second goal, given the optimal value of the first goal as a constraint. Thus, the formulation of the above program that optimise the second goal is (assuming that Y_{i-j} is not defined for $j > i$):

$$Min \;\; d_2^+ \tag{5. 7}$$

subject to:

$$\sum_{i=m+1}^{n}\left(e_{1i} + e_{2i}\right) + d_2^- - d_2^+ = 0 \tag{5. 4}$$

$$e + d_1^- - d_1^+ = 0 \tag{5. 3}$$

$$d_1^+ = e' \tag{5. 8}$$

where d_1^+, d_1^-, d_2^+ and d_2^- are the deviational variables, e_{1i} is the underestimation error, e_{2i} the overestimation error in period i and e the MaxAD and e' is the minimum maximum absolute deviation obtained from the first optimisation.

Subject to:

(4. 29), **(4. 19)**

With **(4. 31)** for sum of errors equal to zero or **(5. 5)** for sum of percentage errors equal to zero.

Bringing constraint (5.8) into the objective function with weight w_1 yields:

$$Min\ \ d_2^+ + w_1\left(d_1^+ - e'\right) \qquad \textbf{(5. 9)}$$

The objective function is penalised if the constraint (5. 8) is violated. The weight is selected by the decision maker. In addition, e’ is a pre-calculated constant number. It can be removed from the objective function without affecting the results. Hence, the linear program is reformulated as:

$$Min\ \ w_1 d_1^+ + w_2 d_2^+ \qquad \textbf{(5. 10)}$$

Subject to:

(5. 3), **(5. 4)**, **(4. 29)**, **(4. 19)**

With **(4. 31)** for sum of errors equal to zero or **(5. 5)** for sum of percentage errors equal to zero.

Because the minimal maximum absolute deviation (e’) is removed from the formulation of the linear program, there is no need for calculating it as an initial objective. Thus, the pre-emptive goal programming is reformulated as a weighted goal programming model, where the objective is to minimise SAD and MaxAD with weights w_1 and w_2 = 1). I perform experiments for *w1* = 1, 2, 3 and 4. The formulation of the counterpart ARS models is similar.

5.4 WGP RESULTS: SUM OF ERRORS EQUAL TO ZERO

Table 5.7 shows the results of the weighted goal programming models for one period ahead forecasts with weights w_1 = 1, 2, 3 and 4. The last column shows the single objective results of the OLS that minimises the MSE. The weighted GP improves the performance of the simple LP overall. Specifically weighted GP performs better than the simple LP in every group of series except on the seasonal series for the AR 6 and AR/ARS 12. The biggest improvement is observed on the hard series, while the differences in the other groups are small.

If we compare the WGP models with the OLS, then OLS performs better for AR/ARS 6 and AR 12, while WGP works better for ARS 12. Specifically, OLS works better on the hard series. On the other hand, weighted GP works better on the smooth series and the seasonal series for the ARS models. Nevertheless, the differences are low.

Table 5. 7 Weighted goal programming - sum of errors equal to zero - MAPE - 1 step ahead

MAPE	WGP				Single Objective LP		OLS
1 STEP	MinSum-MinMaxAD				MinSAD	MinMaxAD	
w1	1	2	3	4			
Σe = 0							
AR 6							
All	8.76	8.80	8.76	8.87	8.64	9.27	8.69
Smooth	3.15	3.15	3.14	3.14	3.02	3.98	3.14
Hard	24.78	25.03	24.84	25.44	24.94	25.14	24.47
Seasonal	9.84	9.82	9.82	9.85	9.58	9.85	9.79
AR 12							
All	8.09	8.22	8.05	8.05	8.36	8.89	8.01
Smooth	3.15	3.17	3.17	3.15	2.98	4.19	3.22
Hard	24.25	24.96	23.93	23.94	26.83	25.81	23.69
Seasonal	7.89	7.91	7.91	7.94	7.68	7.85	7.80
ARS 6							
All	7.44	7.47	7.46	7.56	7.41	7.96	7.39
Seasonal	5.43	5.40	5.48	5.47	5.46	5.51	5.45
ARS 12							
All	7.18	7.33	7.17	7.17	7.52	7.99	7.22
Seasonal	4.86	4.93	4.96	5.00	4.86	4.85	5.19

The results are similar comparing the approaches according to sMAPE (Table 5.8). The approach with w1 = 3 performs in general very well. However, the differences between the approaches are small. WGP outperforms simple LP and OLS overall, except on the hard series with AR 12, where WGP is outperformed by OLS.

Finally, table 5.9 shows the comparison of the approaches according to MASE. The results of all the approaches are very similar and the differences rather small (apart from MinMaxAD which performs worse).

Table 5. 8 Weighted goal programming - sum of errors equal to zero - sMAPE - 1 step ahead

sMAPE	WGP				Single Objective LP		OLS
1 STEP	MinSum-MinMaxAD				MinSAD	MinMaxAD	
w1	1	2	3	4			
Σe = 0							
AR 6							
All	8.49	8.53	8.48	8.59	8.88	9.04	8.48
Smooth	3.02	3.03	3.02	3.02	2.94	3.9	3.03
Hard	23.80	24.07	23.86	24.43	26.38	24.24	24.01
Seasonal	9.69	9.66	9.65	9.68	9.71	9.73	9.54
AR 12							
All	7.90	8.08	7.87	7.88	8.45	8.6	7.81
Smooth	2.99	3.01	3.02	3.01	2.87	4.1	3.07
Hard	24.40	25.39	24.06	24.14	28.37	25.11	23.74
Seasonal	7.46	7.49	7.49	7.53	7.31	7.43	7.40
ARS 6							
All	7.18	7.22	7.20	7.30	7.58	7.74	7.24
Seasonal	5.34	5.31	5.38	5.37	5.38	5.41	5.39
ARS 12							
All	7.09	7.28	7.08	7.10	7.69	7.8	7.12
Seasonal	4.77	4.83	4.87	4.90	4.78	4.75	5.10

Table 5. 9 Weighted goal programming - sum of errors equal to zero - MASE - 1 step ahead

MASE	WGP				Single Objective LP		OLS
1 STEP	MinSum-MinMaxAD				MinSAD	MinMaxAD	
w1	1	2	3	4			
Σe = 0							
AR 6							
All	0.92	0.92	0.92	0.92	0.92	1.04	0.92
Smooth	0.94	0.94	0.95	0.95	0.94	1.18	0.95
Hard	0.88	0.89	0.88	0.90	0.94	0.88	0.88
Seasonal	0.89	0.89	0.89	0.89	0.89	0.89	0.88
AR 12							
All	0.89	0.90	0.89	0.90	0.9	1.07	0.88
Smooth	0.96	0.96	0.97	0.97	0.95	1.29	0.95
Hard	0.90	0.92	0.87	0.88	1.02	0.91	0.88
Seasonal	0.76	0.77	0.77	0.77	0.75	0.76	0.75
ARS 6							
All	0.80	0.81	0.81	0.81	0.81	0.93	0.81
Seasonal	0.52	0.52	0.52	0.52	0.53	0.53	0.52
ARS 12							
All	0.81	0.82	0.81	0.82	0.83	0.99	0.81
Seasonal	0.51	0.50	0.51	0.51	0.51	0.5	0.52

Table 5.10 shows the results of the WGP models for 6 periods ahead forecasting. Weight 4 performs better for all the models on most of the series. In addition, the comparison with the single objective LP is very similar to the one step ahead models. Specifically, WGP performs better in general, except on the seasonal series (where WGP outperforms the simple LP for the ARS 12 model only). In comparison with the OLS, the OLS performs better for AR6 and AR12 models, whereas the WGP does better on the ARS6 and ARS12 models. The WGP performs significantly better on the seasonal series with the ARS models.

Table 5. 10 Weighted goal programming - sum of errors equal to zero - MAPE - 6 steps ahead

MAPE	WGP				Single Objective LP		OLS
6 STEPS	MinSum-MinMaxAD				MinSAD	MinMaxAD	
w1	1	2	3	4			
Σe = 0							
AR 6							
All	11.86	11.82	11.76	11.70	11.99	14.33	11.41
Smooth	7.31	7.31	7.30	7.29	7.34	8.36	7.09
Hard	25.89	25.67	25.38	24.96	26.64	27.53	24.03
Seasonal	12.15	12.14	12.12	12.17	12.12	17.60	12.08
AR 12							
All	11.83	11.88	11.77	11.77	11.94	14.14	11.17
Smooth	7.87	7.87	7.90	7.90	7.87	7.76	7.76
Hard	28.43	28.74	28.08	28.08	29.08	30.40	25.35
Seasonal	9.63	9.64	9.60	9.60	9.64	16.46	9.34
ARS 6							
All	10.71	10.65	10.60	10.53	10.84	12.09	10.96
Seasonal	8.32	8.26	8.28	8.29	8.29	10.12	10.59
ARS 12							
All	11.34	11.36	11.27	11.28	11.49	12.15	11.27
Seasonal	8.02	7.92	7.95	7.96	8.14	9.84	9.67

The comparison according sMAPE (Table 5.11) and MASE (Table 5.12) shows very similar results. WGP does well on the seasonal series (ARS models); OLS performs better on the hard series.

Table 5. 11 Weighted goal programming - sum of errors equal to zero - sMAPE - 6 steps ahead

sMAPE	WGP				Single Objective LP		OLS
6 STEPS	MinSum-MinMaxAD				MinSAD	MinMaxAD	
w1	1	2	3	4			
Σe = 0							
AR 6							
All	11.29	11.26	11.19	11.13	11.40	14.14	10.88
Smooth	6.62	6.61	6.59	6.59	6.63	7.76	6.42
Hard	26.56	26.43	26.10	25.69	27.23	30.40	24.80
Seasonal	11.11	11.10	11.08	11.12	11.08	16.46	11.08
AR 12							
All	11.35	11.43	11.50	11.50	11.65	13.37	11.32
Smooth	7.02	7.02	6.90	6.89	6.87	8.87	6.84
Hard	29.36	29.81	30.83	30.83	31.75	32.26	30.32
Seasonal	9.04	9.05	8.95	8.95	8.99	10.86	8.73
ARS 6							
All	10.33	10.29	10.23	10.16	10.44	12.15	10.58
Seasonal	7.91	7.86	7.87	7.88	7.87	9.84	10.08
ARS 12							
All	10.92	10.97	11.07	11.07	11.27	13.20	11.45
Seasonal	7.61	7.52	7.51	7.52	7.70	10.29	9.17

Table 5. 12 Weighted goal programming - sum of errors equal to zero - MASE - 6 steps ahead

MASE	WGP				Single Objective LP		OLS
6 STEPS	MinSum-MinMaxAD				MinSAD	MinMaxAD	
w1	1	2	3	4			
Σe = 0							
AR 6							
All	0.88	0.87	0.87	0.87	0.88	1.05	0.85
Smooth	0.97	0.97	0.96	0.96	0.97	1.10	0.93
Hard	0.86	0.85	0.83	0.82	0.87	0.89	0.79
Seasonal	0.72	0.72	0.72	0.72	0.72	1.06	0.72
AR 12							
All	0.86	0.86	0.86	0.86	0.87	1.03	0.83
Smooth	0.97	0.97	0.97	0.97	0.98	1.21	0.96
Hard	0.96	0.95	0.94	0.94	0.95	0.98	0.84
Seasonal	0.62	0.62	0.62	0.62	0.62	0.74	0.60
ARS 6							
All	0.82	0.82	0.81	0.81	0.83	0.93	0.84
Seasonal	0.54	0.54	0.54	0.54	0.54	0.66	0.69
ARS 12							
All	0.84	0.84	0.83	0.83	0.84	1.02	0.84
Seasonal	0.54	0.53	0.53	0.53	0.55	0.70	0.64

Tables 5.13, 5.14, 5.15 show the results of the WGP models for 12 periods ahead forecasts (average MAPE, sMAPE and MASE respectively). The w1 = 1 approach typically performs better.

Table 5. 13 Weighted goal programming - sum of errors equal to zero - MAPE - 12 steps ahead

MAPE	WGP				Single Objective LP		OLS
12 STEPS	MinSum-MinMaxAD				MinSAD	MinMaxAD	
w1	1	2	3	4			
Σe = 0							
AR 6							
All	13.08	13.21	13.13	13.14	13.08	15.28	12.55
Smooth	10.55	10.60	10.57	10.49	10.56	11.82	10.04
Hard	25.83	26.45	26.00	26.30	25.72	32.03	25.07
Seasonal	10.51	10.51	10.51	10.53	10.52	12.13	10.06
AR 12							
All	14.10	14.30	14.30	14.44	14.22	21.47	13.85
Smooth	11.25	11.27	11.33	11.15	11.34	22.17	11.04
Hard	28.60	29.82	29.65	30.90	28.99	34.38	28.97
Seasonal	11.11	11.07	11.05	11.13	11.13	13.07	10.44
ARS 6							
All	13.15	13.29	13.15	13.15	13.16	15.46	12.58
Seasonal	10.73	10.78	10.59	10.57	10.81	12.73	10.14
ARS 12							
All	14.02	14.25	14.27	14.39	14.13	21.90	13.61
Seasonal	10.86	10.90	10.96	10.99	10.84	14.49	9.65

Comparing it with the simple LP, it performs slightly better for the AR 12 and ARS models and slightly worse with the AR 6 model; however, the differences are small. The performance of order 6 models is better than order 12 models for longer term forecasts. The best performing WGP the OLS is the best performing approach. The best performing WGP approach is this with weight w1 = 1, and its performance is a little bit better than the single objective LP models.

Table 5. 14 Weighted goal programming - sum of errors equal to zero - sMAPE - 12 steps ahead

sMAPE	WGP				Single Objective LP		OLS
12 STEPS	MinSum-MinMaxAD				MinSAD	MinMaxAD	
w1	1	2	3	4			
Σe = 0							
AR 6							
All	12.46	12.62	12.58	12.57	12.49	14.93	12.18
Smooth	9.43	9.45	9.48	9.40	9.52	10.58	9.25
Hard	27.16	28.04	27.77	27.90	27.03	35.78	26.85
Seasonal	9.67	9.67	9.67	9.69	9.69	11.07	9.26
AR 12							
All	13.76	13.97	13.95	14.26	13.92	17.94	13.48
Smooth	10.18	10.13	10.13	10.11	10.22	14.60	10.17
Hard	31.65	33.13	33.04	34.81	32.44	39.26	31.03
Seasonal	10.18	10.15	10.13	10.21	10.19	12.03	9.61
ARS 6							
All	12.52	12.69	12.60	12.58	12.57	15.12	12.42
Seasonal	9.88	9.93	9.74	9.72	9.96	11.71	10.05
ARS 12							
All	13.70	13.93	13.93	14.22	13.85	18.22	13.44
Seasonal	9.99	10.03	10.08	10.09	9.98	12.95	9.49

Table 5. 15 Weighted goal programming - sum of errors equal to zero - MASE - 12 steps ahead

MASE	WGP				Single Objective LP		OLS
12 STEPS	MinSum-MinMaxAD				MinSAD	MinMaxAD	
w1	1	2	3	4			
Σe = 0							
AR 6							
All	1.03	1.04	1.04	1.03	1.03	1.19	0.98
Smooth	1.06	1.07	1.07	1.06	1.07	1.20	1.00
Hard	0.91	0.92	0.92	0.92	0.91	1.06	0.90
Seasonal	1.04	1.04	1.04	1.04	1.04	1.25	1.00
AR 12							
All	1.10	1.10	1.10	1.10	1.10	1.41	1.07
Smooth	1.13	1.14	1.13	1.11	1.14	1.52	1.10
Hard	1.01	1.03	1.03	1.06	0.99	1.17	1.04
Seasonal	1.08	1.08	1.08	1.09	1.08	1.34	1.04
ARS 6							
All	1.05	1.05	1.05	1.04	1.06	1.22	0.98
Seasonal	1.09	1.10	1.08	1.07	1.11	1.34	1.00
ARS 12							
All	1.10	1.11	1.11	1.10	1.11	1.44	1.06
Seasonal	1.10	1.11	1.12	1.11	1.11	1.46	1.00

5.5 WGP RESULTS: SUM OF PERCENTAGE ERRORS EQUAL TO ZERO

Table 5.16 shows the results of the weighted goal programming approaches, compared with the simple LP and OLS for one period ahead forecasts according to the MAPE.

Table 5. 16 Weighted goal programming - sum of percentage errors equal to zero - 1 step ahead

MAPE	WGP								Single Objective LP		OLS
1 STEP	MinSum-MinMaxAD								MinSAD	MinMaxAD	
w1	1		2		3		4				
Σ%e = 0											
AR 6											
All	8.60	(-0.16)	8.68	(-0.12)	8.74	(-0.02)	8.79	(-0.08)	8.64	9.27	8.69
Smooth	3.01	(-0.14)	2.95	(-0.2)	2.93	(-0.2)	2.95	(-0.19)	3.02	3.98	3.14
Hard	24.73	(-0.05)	25.33	(0.3)	25.82	(0.99)	25.94	(0.5)	24.94	25.14	24.47
Seasonal	9.59	(-0.25)	9.61	(-0.21)	9.56	(-0.26)	9.64	(-0.21)	9.58	9.85	9.79
AR 12											
All	7.99	(-0.1)	8.01	(-0.21)	7.87	(-0.18)	7.91	(-0.14)	8.36	8.89	8.01
Smooth	2.95	(-0.2)	3.00	(-0.17)	3.03	(-0.14)	3.03	(-0.12)	2.98	4.19	3.22
Hard	24.71	(0.46)	24.66	(-0.3)	23.79	(-0.14)	23.95	(0.02)	26.83	25.81	23.69
Seasonal	7.66	(-0.23)	7.66	(-0.25)	7.64	(-0.27)	7.67	(-0.27)	7.68	7.85	7.80
ARS 6											
All	7.34	(-0.1)	7.43	(-0.05)	7.50	(0.04)	7.61	(0.06)	7.41	7.96	7.39
Seasonal	5.37	(-0.06)	5.44	(0.04)	5.44	(-0.04)	5.72	(0.25)	5.46	5.51	5.45
ARS 12											
All	7.14	(-0.03)	7.17	(-0.15)	7.06	(-0.1)	7.23	(0.06)	7.52	7.99	7.22
Seasonal	4.84	(-0.02)	4.88	(-0.05)	4.95	(-0.01)	5.41	(0.41)	4.86	4.85	5.19

In brackets I show the improvement in MAPE over the approaches with sum of errors equal to zero constraint. As before a negative value corresponds to a decrease in MAPE; a positive value indicates an increase.

The improvement over the approaches with sum of errors equal to zero constraint is obvious. Specifically, all the approaches give better results with the only exception the ARS 6 with w1 = 3 and 4. Similar as for the single objective LPs, the improvement is mainly observed on the smooth and seasonal series.

In comparison with the single objective LP with sum of percentage errors equal to zero, the WGP gives better results than the best single objective model. Similarly to the models with sum of errors equal to zero, the improvement is observed on the hard series.

In comparison with the OLS, the WGP gives better results, mainly due to better results on the smooth and seasonal series. It is difficult to draw conclusions about the impacts of the weights.

Table 5.17 shows the comparison of the WGP with the simple LP and OLS according to sMAPE. The best WGP approaches are with w1 = 1 for order 6 and w1 = 3 for order 12. The results are generally improved on the smooth and seasonal series, but they are worse on the hard series. WGP outperforms simple LP and performs better than the OLS on the smooth and seasonal series.

Table 5. 17 Weighted goal programming - sum of percentage errors equal to zero - sMAPE - 1 step ahead

sMAPE	WGP								Single Objective LP		OLS
1 STEP	MinSum-MinMaxAD								MinSAD	MinMaxAD	
w1	1		2		3		4				
Σ%e = 0											
AR 6											
All	8.84	(0.35)	8.91	(0.38)	8.95	(0.47)	9.01	(0.42)	8.88	9.04	8.48
Smooth	2.93	(-0.1)	2.89	(-0.14)	2.89	(-0.13)	2.92	(-0.1)	2.94	3.90	3.03
Hard	26.18	(2.37)	26.70	(2.62)	27.03	(3.17)	27.15	(2.72)	26.38	24.24	24.01
Seasonal	9.71	(0.02)	9.73	(0.06)	9.68	(0.03)	9.74	(0.06)	9.71	9.73	9.54
AR 12											
All	8.18	(0.28)	8.23	(0.14)	8.04	(0.17)	8.10	(0.21)	8.45	8.60	7.81
Smooth	2.85	(-0.14)	2.91	(-0.1)	2.95	(-0.07)	2.95	(-0.05)	2.87	4.10	3.07
Hard	26.83	(2.43)	26.92	(1.53)	25.72	(1.66)	25.98	(1.83)	28.37	25.11	23.74
Seasonal	7.29	(-0.18)	7.29	(-0.2)	7.28	(-0.21)	7.32	(-0.21)	7.31	7.43	7.40
ARS 6											
All	7.51	(0.33)	7.60	(0.38)	7.66	(0.45)	7.77	(0.47)	7.58	7.74	7.24
Seasonal	5.29	(-0.05)	5.35	(0.05)	5.36	(-0.02)	5.63	(0.25)	5.38	5.41	5.39
ARS 12											
All	7.42	(0.33)	7.48	(0.19)	7.32	(0.24)	7.50	(0.4)	7.69	7.80	7.12
Seasonal	4.76	(-0.01)	4.79	(-0.04)	4.87	(0)	5.31	(0.42)	4.78	4.75	5.10

Table 5.18 shows the comparison of the approaches according to MASE. Similar as for the sum of errors equal to zero constraint, the differences are relatively small. There is an improvement on the smooth and seasonal series. WGP improves the results of simple LP, with only exception MinMaxAD on the hard series and it performs as good as the OLS.

Table 5. 18 Weighted goal programming - sum of percentage errors equal to zero - MASE - 1 step ahead

MASE	WGP								Single Objective LP		OLS
1 STEP	MinSum-MinMaxAD								MinSAD	MinMaxAD	
w1	1		2		3		4				
Σ%e = 0											
AR 6											
All	0.92	(0)	0.92	(0)	0.92	(0.01)	0.93	(0)	0.92	1.04	0.92
Smooth	0.93	(-0.01)	0.93	(-0.01)	0.94	(-0.01)	0.94	(-0.01)	0.94	1.18	0.95
Hard	0.93	(0.05)	0.94	(0.05)	0.95	(0.06)	0.95	(0.05)	0.94	0.88	0.88
Seasonal	0.89	(-0.01)	0.89	(0)	0.88	(-0.01)	0.89	(0)	0.89	0.89	0.88
AR 12											
All	0.89	(0)	0.89	(0)	0.89	(0)	0.89	(0)	0.90	1.07	0.88
Smooth	0.95	(-0.01)	0.96	(-0.01)	0.96	(-0.01)	0.96	(-0.01)	0.95	1.29	0.95
Hard	0.95	(0.04)	0.96	(0.04)	0.92	(0.05)	0.93	(0.05)	1.02	0.91	0.88
Seasonal	0.75	(-0.02)	0.75	(-0.02)	0.75	(-0.02)	0.75	(-0.02)	0.75	0.76	0.75
ARS 6											
All	0.81	(0)	0.81	(0.01)	0.82	(0.01)	0.82	(0.01)	0.81	0.93	0.81
Seasonal	0.52	(-0.01)	0.52	(0.01)	0.52	(0)	0.55	(0.03)	0.53	0.53	0.52
ARS 12											
All	0.82	(0)	0.82	(0)	0.82	(0)	0.83	(0.02)	0.83	0.99	0.81
Seasonal	0.50	(0)	0.50	(0)	0.51	(0)	0.55	(0.04)	0.51	0.50	0.52

Table 5.19 shows the results (MAPE) for six steps ahead forecasts. In comparison with the LPs with sum of errors equal to zero constraint, the results are better. WGP outperforms simple LP overall, mainly because of the better performance on the smooth and hard series. WGP outperforms OLS overall and on the smooth and seasonal series. OLS performs better on the hard series.

The comparison of the approaches according to sMAPE (table 5.20) shows that WGP outperforms simple LP overall; however, the differences are very small. OLS outperforms WGP on the hard series; however, the WGP performs significantly better on the smooth and seasonal series.

Table 5. 19 Weighted goal programming - sum of percentage errors equal to zero - MAPE - 6 steps ahead

MAPE	LP								Single Objective LP		OLS
6 STEPS	MinSum-MinMaxAD								MinSAD	MinMaxAD	
w1	1		2		3		4				
Σ%e = 0											
AR 6											
All	11.17	(-0.69)	11.07	(-0.74)	11.02	(-0.73)	11.07	(-0.63)	11.29	14.42	11.41
Smooth	6.49	(-0.82)	6.47	(-0.83)	6.40	(-0.89)	6.42	(-0.87)	6.53	7.89	7.09
Hard	25.88	(-0.02)	25.35	(-0.32)	25.24	(-0.14)	25.42	(0.46)	26.41	31.49	24.03
Seasonal	11.33	(-0.83)	11.32	(-0.82)	11.34	(-0.78)	11.36	(-0.81)	11.36	16.55	12.08
AR 12											
All	10.97	(-0.85)	11.00	(-0.88)	11.09	(-0.68)	11.17	(-0.6)	11.06	13.57	11.17
Smooth	7.01	(-0.87)	7.00	(-0.87)	7.04	(-0.86)	7.04	(-0.86)	6.96	9.85	7.76
Hard	26.94	(-1.49)	27.03	(-1.71)	27.60	(-0.48)	28.04	(-0.04)	27.58	29.21	25.35
Seasonal	9.15	(-0.48)	9.21	(-0.44)	9.13	(-0.47)	9.16	(-0.44)	9.16	11.48	9.34
ARS 6											
All	10.18	(-0.53)	10.11	(-0.54)	10.07	(-0.53)	10.10	(-0.43)	10.28	12.47	10.96
Seasonal	8.01	(-0.31)	8.11	(-0.15)	8.16	(-0.12)	8.15	(-0.14)	7.99	10.05	10.59
ARS 12											
All	10.62	(-0.72)	10.59	(-0.77)	10.70	(-0.57)	10.77	(-0.51)	10.72	13.34	11.27
Seasonal	7.99	(-0.03)	7.84	(-0.08)	7.83	(-0.11)	7.82	(-0.14)	8.03	10.71	9.67

Table 5. 20 Weighted goal programming - sum of percentage errors equal to zero - sMAPE - 6 steps ahead

sMAPE	LP								Single Objective LP		OLS
6 STEPS	MinSum-MinMaxAD								MinSAD	MinMaxAD	
w1	1		2		3		4				
Σ%e = 0											
AR 6											
All	11.30	(0.01)	11.21	(-0.04)	11.16	(-0.03)	11.20	(0.06)	11.41	12.67	10.88
Smooth	6.15	(-0.46)	6.14	(-0.47)	6.07	(-0.53)	6.07	(-0.51)	6.18	7.60	6.42
Hard	29.06	(2.5)	28.59	(2.16)	28.49	(2.39)	28.65	(2.96)	29.60	23.43	24.80
Seasonal	10.58	(-0.54)	10.58	(-0.52)	10.59	(-0.49)	10.61	(-0.52)	10.61	15.71	11.08
AR 12											
All	11.43	(0.08)	11.43	(0)	11.55	(0.05)	11.68	(0.18)	11.49	13.37	11.32
Smooth	6.43	(-0.59)	6.42	(-0.6)	6.46	(-0.44)	6.45	(-0.44)	6.39	8.87	6.84
Hard	32.53	(3.18)	32.45	(2.65)	33.19	(2.36)	33.95	(3.12)	33.01	32.26	30.32
Seasonal	8.59	(-0.44)	8.65	(-0.4)	8.58	(-0.37)	8.60	(-0.35)	8.60	10.86	8.73
ARS 6											
All	10.42	(0.09)	10.36	(0.08)	10.32	(0.09)	10.35	(0.19)	10.52	10.89	10.58
Seasonal	7.65	(-0.26)	7.74	(-0.12)	7.79	(-0.08)	7.79	(-0.09)	7.63	9.78	10.08
ARS 12											
All	11.13	(0.2)	11.07	(0.1)	11.21	(0.14)	11.32	(0.26)	11.20	13.20	11.45
Seasonal	7.59	(-0.03)	7.44	(-0.09)	7.43	(-0.08)	7.41	(-0.1)	7.62	10.29	9.17

The comparison according to the MASE (table 5.21) shows that the differences are very small. OLS performs better on the hard series, and WGP performs significantly better on the seasonal series for ARS.

Table 5. 21 Weighted goal programming - sum of percentage errors equal to zero - MASE - 6 steps ahead

MASE	LP								Single Objective LP		OLS
6 STEPS	MinSum-MinMaxAD								MinSAD	MinMaxAD	
w1	1		2		3		4				
Σ%e = 0											
AR 6											
All	0.86	(-0.01)	0.86	(-0.01)	0.86	(-0.01)	0.86	(-0.01)	0.87	1.05	0.85
Smooth	0.95	(-0.01)	0.95	(-0.02)	0.94	(-0.02)	0.94	(-0.02)	0.96	1.10	0.93
Hard	0.89	(0.04)	0.88	(0.03)	0.87	(0.04)	0.87	(0.06)	0.91	0.97	0.79
Seasonal	0.69	(-0.03)	0.69	(-0.03)	0.69	(-0.03)	0.69	(-0.03)	0.69	1.01	0.72
AR 12											
All	0.85	(-0.02)	0.84	(-0.02)	0.85	(-0.01)	0.85	(-0.01)	0.85	1.03	0.83
Smooth	0.96	(-0.01)	0.95	(-0.02)	0.95	(-0.02)	0.95	(-0.02)	0.96	1.21	0.96
Hard	0.95	(-0.01)	0.95	(-0.01)	0.97	(0.04)	0.99	(0.05)	0.97	0.98	0.84
Seasonal	0.60	(-0.03)	0.60	(-0.02)	0.59	(-0.03)	0.60	(-0.02)	0.60	0.74	0.60
ARS 6											
All	0.81	(-0.01)	0.81	(-0.01)	0.81	(-0.01)	0.81	(0)	0.82	0.95	0.84
Seasonal	0.52	(-0.02)	0.53	(-0.01)	0.53	(-0.01)	0.53	(-0.01)	0.52	0.67	0.69
ARS 12											
All	0.83	(-0.01)	0.82	(-0.01)	0.83	(-0.01)	0.83	(-0.01)	0.84	1.02	0.84
Seasonal	0.54	(0)	0.53	(0)	0.53	(-0.01)	0.53	(-0.01)	0.54	0.70	0.64

Finally Tables 5.22, 5.23 and 5.24 show the results of the WGP for 12 periods ahead forecasts. The improvement in the 12 step ahead models is similar. The results are improved on the smooth and seasonal series, but they are worse on the hard series. The single objective LPs, tend to perform a bit better on the hard series. OLS performs better on the hard and seasonal series but worse on the smooth. As before, for longer term forecasts the order 6 models do better than the order 12 models.

Table 5. 22 Weighted goal programming - sum of percentage errors equal to zero - MAPE - 12 steps ahead

MAPE	LP								Single Objective LP		OLS
12 STEPS	MinSum-MinMaxAD								MinSAD	MinMaxAD	
w1	1		2		3		4				
$\Sigma\%e = 0$											
AR 6											
All	12.47	(-0.61)	12.46	(-0.75)	12.52	(-0.61)	12.54	(-0.6)	12.48	15.28	12.55
Smooth	9.28	(-1.27)	9.31	(-1.28)	9.29	(-1.28)	9.28	(-1.21)	9.26	11.82	10.04
Hard	27.26	(1.43)	27.08	(0.64)	27.43	(1.42)	27.55	(1.25)	27.36	32.03	25.07
Seasonal	9.93	(-0.58)	9.94	(-0.57)	9.98	(-0.54)	10.00	(-0.53)	9.94	12.13	10.06
AR 12											
All	13.79	(-0.31)	13.72	(-0.58)	13.87	(-0.43)	13.91	(-0.53)	13.66	21.47	13.85
Smooth	10.30	(-0.95)	10.20	(-1.07)	10.15	(-1.18)	10.17	(-0.98)	10.24	22.17	11.04
Hard	31.00	(2.4)	30.83	(1)	31.83	(2.18)	31.86	(0.96)	30.35	34.38	28.97
Seasonal	10.45	(-0.67)	10.47	(-0.6)	10.51	(-0.54)	10.58	(-0.55)	10.44	13.07	10.44
ARS 6											
All	12.59	(-0.55)	12.60	(-0.69)	12.64	(-0.51)	12.66	(-0.49)	12.61	15.33	12.58
Seasonal	10.34	(-0.39)	10.41	(-0.38)	10.38	(-0.21)	10.40	(-0.17)	10.35	12.28	10.14
ARS 12											
All	13.74	(-0.28)	13.69	(-0.56)	13.87	(-0.4)	13.91	(-0.48)	13.64	21.88	13.61
Seasonal	10.28	(-0.58)	10.37	(-0.53)	10.50	(-0.46)	10.58	(-0.41)	10.38	14.42	9.65

Table 5. 23Weighted goal programming - sum of percentage errors equal to zero - sMAPE - 12 steps ahead

sMAPE	LP								Single Objective LP		OLS
12 STEPS	MinSum-MinMaxAD								MinSAD	MinMaxAD	
w1	1		2		3		4				
$\Sigma\%e = 0$											
AR 6											
All	12.96	(0.51)	12.95	(0.33)	13.02	(0.44)	13.08	(0.51)	13.03	14.93	12.18
Smooth	9.07	(-0.35)	9.12	(-0.33)	9.10	(-0.37)	9.09	(-0.31)	9.11	10.58	9.25
Hard	32.15	(4.99)	31.92	(3.88)	32.37	(4.6)	32.70	(4.8)	32.40	35.78	26.85
Seasonal	9.21	(-0.46)	9.21	(-0.46)	9.25	(-0.43)	9.27	(-0.42)	9.22	11.07	9.26
AR 12											
All	14.68	(0.93)	14.30	(0.33)	14.56	(0.61)	14.62	(0.37)	14.99	17.94	13.48
Smooth	9.81	(-0.36)	9.71	(-0.42)	9.68	(-0.45)	9.69	(-0.42)	9.77	14.60	10.17
Hard	39.33	(7.68)	37.29	(4.16)	38.89	(5.86)	39.11	(4.3)	41.30	39.26	31.03
Seasonal	9.64	(-0.54)	9.67	(-0.47)	9.71	(-0.42)	9.78	(-0.43)	9.64	12.03	9.61
ARS 6											
All	13.07	(0.56)	13.08	(0.39)	13.14	(0.54)	13.19	(0.61)	13.25	18.98	12.42
Seasonal	9.59	(-0.29)	9.66	(-0.27)	9.62	(-0.11)	9.66	(-0.06)	9.96	24.57	10.05
ARS 12											
All	14.65	(0.95)	14.28	(0.34)	14.57	(0.63)	14.62	(0.4)	15.09	22.98	13.44
Seasonal	9.53	(-0.46)	9.61	(-0.42)	9.73	(-0.35)	9.78	(-0.31)	9.98	28.84	9.49

Table 5. 24 Weighted goal programming - sum of percentage errors equal to zero - MASE - 12 steps

MASE	LP								Single Objective LP		OLS
12 STEPS	MinSum-MinMaxAD								MinSAD	MinMaxAD	
w1	1		2		3		4				
Σ%e = 0											
AR 6											
All	1.04	(0.01)	1.05	(0.01)	1.04	(0)	1.04	(0.01)	1.05	1.19	0.98
Smooth	1.09	(0.03)	1.09	(0.02)	1.09	(0.01)	1.08	(0.02)	1.10	1.20	1.00
Hard	0.97	(0.06)	0.97	(0.05)	0.97	(0.06)	0.97	(0.06)	0.97	1.06	0.90
Seasonal	1.00	(-0.04)	1.00	(-0.04)	1.00	(-0.04)	1.01	(-0.04)	1.00	1.25	1.00
AR 12											
All	1.12	(0.02)	1.10	(0)	1.11	(0.01)	1.11	(0.01)	1.11	1.41	1.07
Smooth	1.17	(0.04)	1.15	(0.01)	1.14	(0.01)	1.14	(0.03)	1.17	1.52	1.10
Hard	1.08	(0.08)	1.08	(0.05)	1.11	(0.08)	1.11	(0.05)	1.06	1.17	1.04
Seasonal	1.04	(-0.04)	1.04	(-0.04)	1.05	(-0.03)	1.06	(-0.04)	1.04	1.34	1.04
ARS 6											
All	1.06	(0.02)	1.07	(0.01)	1.06	(0.01)	1.06	(0.02)	1.08	1.20	0.98
Seasonal	1.07	(-0.03)	1.07	(-0.03)	1.07	(-0.01)	1.07	(-0.01)	1.11	1.28	1.00
ARS 12											
All	1.12	(0.02)	1.11	(0)	1.12	(0.01)	1.11	(0.01)	1.13	1.44	1.06
Seasonal	1.06	(-0.04)	1.07	(-0.04)	1.08	(-0.04)	1.06	(-0.06)	1.11	1.45	1.00

5.6 CONCLUSIONS

The objective of the chapter is to expand the simple LP formulations of chapter 4 into GP formulations. The initial formulation is a pre-emptive GP that minimises MaxAD as a first objective and SAD as a second. The program is tested for the AR and ARS approaches, order 6 and 12, one step ahead forecasts. The MinMaxAD-MinSAD pre-emptive GP approach improves the results of the single objective LP approaches and performs better or similar than OLS on the hard series. The initial pre-emptive GP approach is reformulated as a weighted GP and it is tested for one, six and twelve steps ahead forecasts. The results show that WGP improves simple LP overall and occasionally outperforms the OLS. The advantages of the WGP are that it is a very simple formulation, consisting of one linear program. In contrast, pre-emptive GP required solving two LPs. The different approaches give the decision maker the option to select the most appropriate approach according to the characteristics of the problem. The conclusions of this chapter are summarised as follows:

- Goal programming can improve the results of the single objective linear programming. The pre-emptive goal programming model improved the results on the hard series. The weighted goal programming approaches (with sum of percentage errors equal to zero constraint) performed very well on the seasonal and smooth series (and do typically better than the OLS for short and medium term forecasts). For the hard series, the OLS gives the best results in general. For longer term forecasts, order 6 models do better than order 12 models, and the benefit from goal programming is less clear.

- Similar to the single objective LP models, the constraint that the sum of the percentage errors should be zero gives better result compared with the constraint where the sum of errors is zero (they perform better on the smooth and seasonal series, but the perform worse on the hard series).

- Single objective MinMaxAD is the worst model overall, but setting MinMaxAD as an additional GP goal improves the results of the single objective LP models.

- It is difficult to draw conclusions about the impact of the relaxation factors and the weights on the performance of the GP models.

- According to the characteristics of the series and the forecasting horizon (one step ahead and multiple steps ahead) we may need to use different methods (single objective LP, pre-emptive GP, WGP and OLS) and different optimisation objectives.

A general conclusion of chapters 4 and 5 is that LP based approaches (simple and GP) perform better on series with low variability, while the OLS performs better on series with high variability. The next step is to examine if the flexibility of LP can improve the performance of the LP based approaches on series with high variability.

6 LINEAR PROGRAMMING FOR FORECASTING SERIES WITH HIGH VARIABILITY

In this chapter the objective of the study is to improve the performance of the LP approaches of the previous two chapters (single objective LPs and WGP, for one step ahead forecasting) on series with high variability. For this reason, extra constraints are added in the initial linear programming formulation. The results are compared with the OLS and six other simple forecasting techniques, which are shown to perform well on series with high variability.

6.1 APPROACHES

The weak point of the LP based approaches developed so far was their performance on the hard series. AR in general seems to result in very big errors on series with high levels of randomness. The main reason is that the autocorrelation between observations in these series is low (Chapter 3); hence, the explanatory value of the model based on the training data set is low.

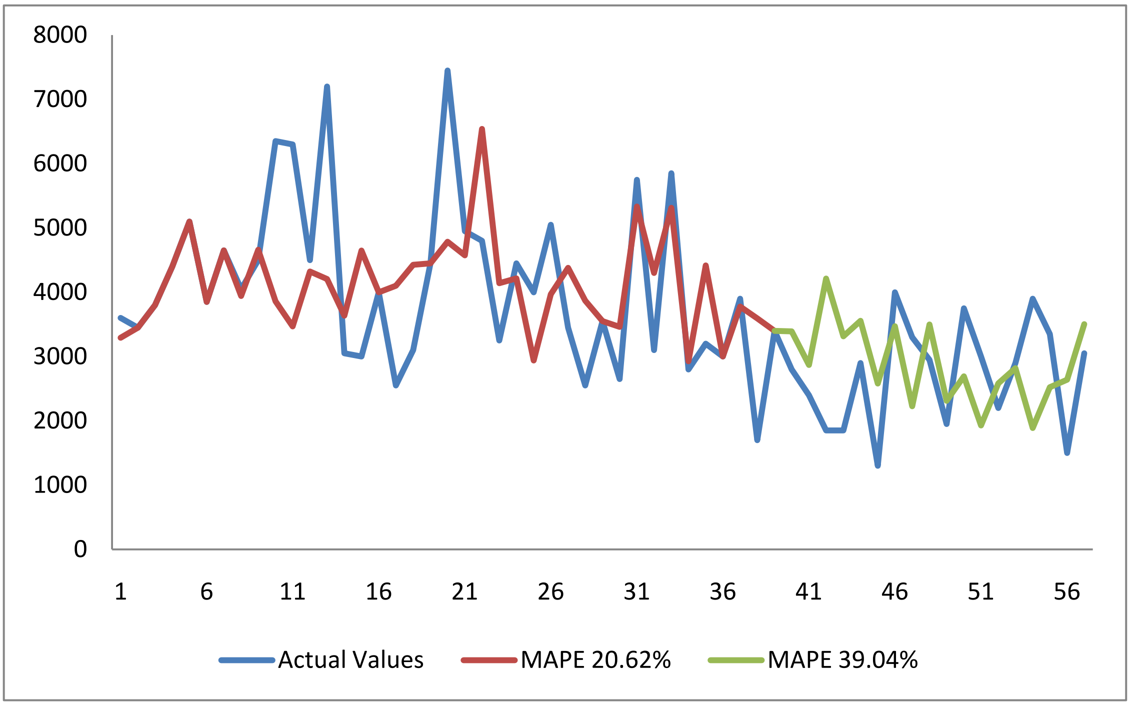

Figure 6. 1 Performance of an approach on a hard series on the training set and test set

Figure 6.1 shows a characteristic example of the performance of a MinSAPE on a series with high variability. The MAPE based on the training set is 20.62%, while the MAPE on the test set data is almost double (39.04%).

The M-Competition showed that in cases with high randomness the decision maker should prefer simpler techniques, like moving averages and exponential smoothing. In this chapter I show that the flexibility of the LP approaches can be exploited to do also well on series with high randomness. I used the techniques that are presented in the previous chapters with additional constraints. These new approaches may sacrifice some of the accuracy of the training set, in favour of a better accuracy on the test set. The results are compared with those obtained by simple techniques including exponential smoothing and moving averages .

I tested the simple AR for one period ahead forecast, order 6 and 12, with the sum of errors equal to zero constraint or the sum of the percentage errors equal to zero constraint. I add one or two additional constraints. The first is that all the weights must be positive: $b_j \geq 0$. The second is to put a hierarchy on the weights: $b_j \geq b_{j-1}$.

6.2 TESTS

The new approaches are tested on two data sets. The first consists of the ten hard series in the initial dataset that I used in the previous chapters. The second consists of twenty five additional monthly series with high variability from the dataset of the M3 Competition. The selection and statistical analysis of the series can be found in chapter 3. I tested both the initial and the new approaches with the additional constraints.

The results are compared with the results by OLS (minimising MSE), and four other forecasting techniques: Simple Moving Average (MA), Weighted Moving Average (WMA), Simple Exponential Smoothing (SES) and Holt's Exponential Smoothing (Holt).[3] I want to compare the new LP approaches with the most accurate technique. Thus, for the selection of the order of the moving average and the smoothing parameters, I explored different parameter values and I selected the three alternatives that were on average the most accurate on the test set of the thirty five series.

[3] The formulae for these approaches can be found in Chapter 3.

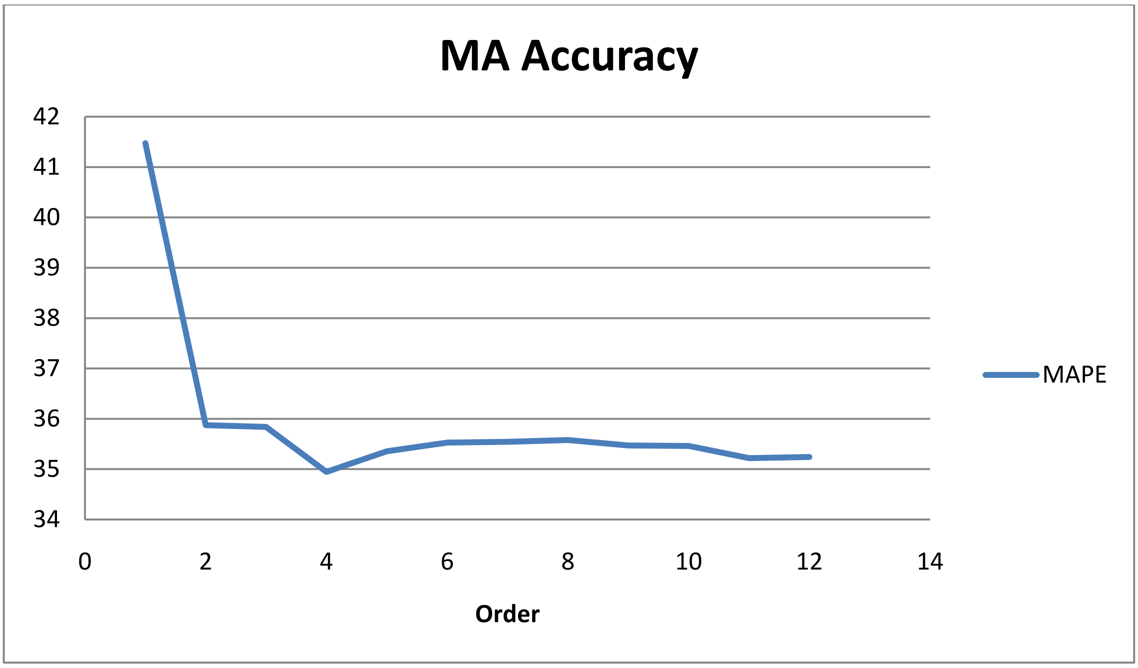

Figure 6. 2 Performance of MA of order 1 - 12

The MA accuracy is shown in Figure 6.2. All the alternatives are calculated for order 1 to 12. According to the average percentage error, the most accurate were orders 4, 11 and 12. However, the average MAPE is not very sensitive to the change in the order, for an order equal to or bigger than 4.

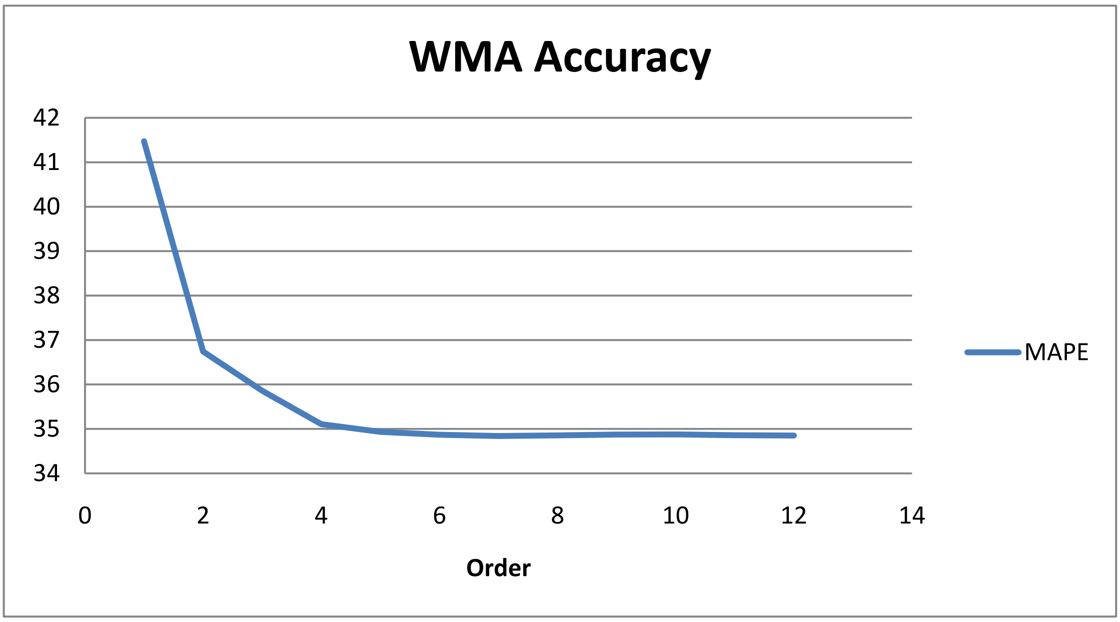

Figure 6. 3 Performance of WMA of order 1 - 12

Similarly, the WMA accuracy graph shows that the most accurate are for order 7, 8 and 12. However, the differences are very low (for an order equal to or bigger than 4).

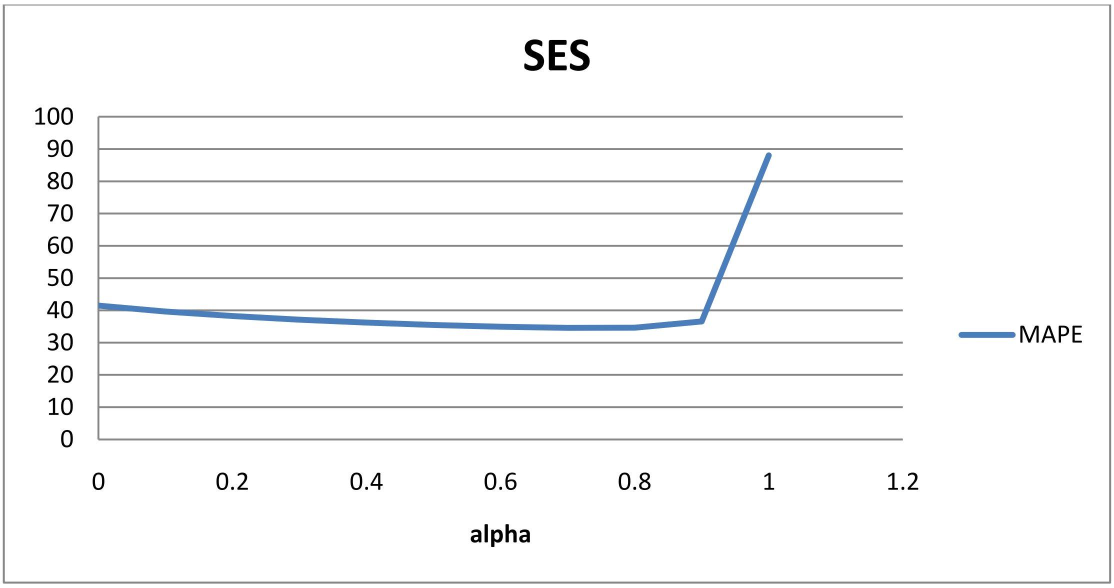

Figure 6. 4 Performance of SES with factor 0 - 1

For the smoothing parameter of the SES we evaluate all the alternatives between 0 and 1 with step size 0.1. According to the graph (Figure 6.4), the smoothing parameters 0.8, 0.7 and 0.6 are the three most accurate. Compared with MA and WMA, the performance of the SES is more sensitive to changes in α.

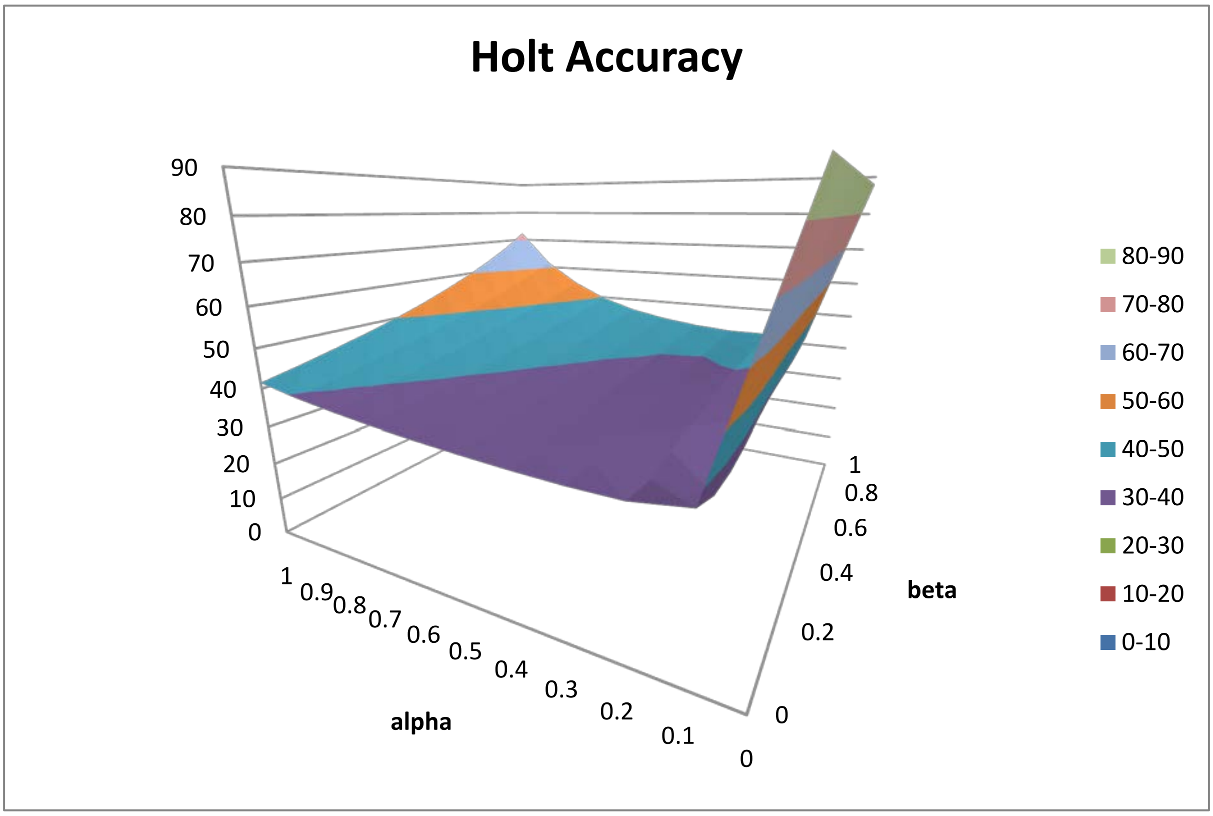

Figure 6. 5 Performance of Holt's with factors 0 - 1, 0 - 1

Similarly, all the alternative combinations of the two smoothing parameters are evaluated (alpha: base, beta: trend) for Holt's method (values between 0 and 1 with step size 0.1). The surface graph in Figure 6.5 presents the average MAPE over all the series. The three best smoothing factor combinations are 0.1-0.1, 0.2-0.1 and 0.1-0.2.

Table 6.1 shows the results (MAPE) of the OLS, the moving average and the exponential smoothing techniques. The first line of the results shows the average MAPE on the test set in all the series, the second the average MAPE in the old set of series (hard series of the initial set) and the third the average MAPE on the new series.

OLS is the least accurate technique both for AR 6 and AR 12. Exponential smoothing is slightly better than moving average. Holt is the most accurate technique but only slightly better than SES. Specifically, Holt (0.1-0.2) is the best model, Holt (0.1-0.1) is the second and Holt (0.2-0.1) is the third. SES is the second most accurate (0.7, 0.8, 0.6 in order of accuracy). WMA is the most accurate of the moving average techniques and the difference between the order levels is insignificant. MA is least accurate of the four with MA 4, MA 11 and MA 12 in order of accuracy.

Table 6. 1 Traditional techniques - MAPE

	Moving Average					
	MA			WMA		
Order	**4**	**11**	**12**	**7**	**8**	**12**
All	34.94	35.22	35.24	34.84	34.85	34.85
Old	24.06	24.16	24.26	23.98	24.13	24.06
New	39.30	39.64	39.63	39.18	39.14	39.17

	Exponential Smoothing					
	SES			Holt		
S. F.	**0.8**	**0.7**	**0.6**	**0.1 - 0.1**	**0.2 - 0.1**	**0.1 - 0.2**
All	34.64	34.59	34.93	33.43	33.67	33.17
Old	23.75	23.81	24.03	22.18	22.92	23.12
New	39.00	38.90	39.28	37.93	37.97	37.19

	OLS	
Order	**6**	**12**
All	35.06	36.84
Old	24.47	23.69
New	39.29	42.09

Table 6.2 shows the results of the traditional techniques according to the sMAPE. Here, OLS remains the least accurate technique. The other four techniques produce very similar results and the differences are very small. The most accurate is SES 0.8, second and third are MA 12 and 11, then WMA 12 and Holt 0.1 – 0.1. On the other hand, the worst techniques are MA 4 and SES 0.6. The differences between the moving average and smoothing are too small to make clear conclusions; however, all of them perform significantly better than the OLS.

Table 6. 2 Traditional techniques - sMAPE

	Moving Average					
	MA			WMA		
Order	4	11	12	7	8	12
All	28.34	27.15	27.13	27.94	27.80	27.29
Old	23.04	22.72	22.74	22.96	23.09	22.83
New	30.46	28.92	28.88	29.93	29.68	29.07

	Exponential Smoothing					
	SES			Holt		
S. F	0.8	0.7	0.6	0.1 - 0.1	0.2 - 0.1	0.1 - 0.2
All	27.09	27.53	28.08	27.77	27.83	28.03
Old	22.37	22.64	22.86	24.54	23.50	24.85
New	28.98	29.49	30.17	29.06	29.56	29.29

	OLS	
Order	6	12
All	29.07	29.71
Old	24.01	23.74
New	31.10	32.09

Table 6.3 shows the comparison of the traditional techniques according to the MASE. Similarly with sMAPE, the differences between the moving average and exponential smoothing techniques are rather small. All of them mainly perform between 0.82 and 0.84 with only exception MA 4 that performs 0.86; hence, and it is hard to draw firm conclusions. However, all of them perform significantly better than the OLS, which performs 0.88 and 0.89 for order 6 and 12 respectively.

Table 6. 3 Traditional techniques - MASE

	Moving Average					
	MA			WMA		
Order	4	11	12	7	8	12
All	0.86	0.83	0.82	0.84	0.84	0.83
Old	0.86	0.83	0.83	0.84	0.84	0.83
New	0.86	0.83	0.82	0.84	0.84	0.82

	Exponential Smoothing					
	SES			Holt		
S. F	0.8	0.7	0.6	0.1 - 0.1	0.2 - 0.1	0.1 - 0.2
All	0.82	0.82	0.83	0.84	0.83	0.84
Old	0.82	0.83	0.84	0.87	0.85	0.88
New	0.81	0.82	0.83	0.82	0.83	0.83

	OLS	
Order	6	12
All	0.88	0.89
Old	0.88	0.88
New	0.88	0.90

6.3 RESULTS: SUM OF ERRORS EQUAL TO ZERO

The results of the single objective LP with the sum of errors equal to zero constraint are presented in Table 6.4. The first group of rows is for the original approaches, the second is with the additional constraint of positive weights and the third for the approaches with two additional constraints (positive weights and progressively decreasing, i.e. the oldest observations have lowest weights).

First of all, it is obvious that the first additional constraint improves the initial approaches and gives results of similar accuracy as with the OLS. The second additional constraint improves the results further and gives results slightly better than the OLS and the simple MA. The performance of all the approaches is similar.

As we can see, similar to the performance of the WGP in the previous chapter, the weights 3 and 4 give the best results in general. Like in the single objective formulations, the additional constraints improve the results of the initial approaches. The results of the WGP are better than the counterpart single objective LPs. The non negative weight LPs give similar results as the MA, slightly better than the OLS and slightly worse than the exponential smoothing methods. Moreover, the order 12 of the progressive approaches slightly outperforms SES and is only slightly worse than Holt.

Table 6. 4 Single objective - sum of errors equal to zero - MAPE

MAPE	LP									
e = 0	MinMAD		MinMAPE		MinMaxAD		MinADBD		MinADBPD	
Order	6	12	6	12	6	12	6	12	6	12
All	37.31	40.24	39.48	46.60	41.57	42.48	36.32	38.16	38.12	43.09
Old	24.81	25.78	26.26	28.52	28.94	24.95	24.77	24.38	24.97	27.73
New	42.31	46.03	44.77	53.84	46.12	48.79	40.94	43.67	43.38	49.23
NON NEG										
All	36.48	35.42	36.83	36.98	36.73	36.11	36.02	35.10	36.83	36.68
Old	25.42	22.62	25.71	23.74	24.92	21.98	24.84	22.34	25.32	23.42
New	40.90	40.54	41.27	42.27	41.45	41.77	40.49	40.21	41.44	41.99
PROGRES										
All	35.18	34.27	35.37	34.37	34.99	34.71	35.11	34.34	35.39	34.54
Old	23.84	22.48	23.96	22.95	23.86	23.61	23.58	22.91	23.76	23.28
New	39.72	38.98	39.93	38.94	39.44	39.15	39.73	38.91	40.04	39.05

Table 6.5 shows the results of the WGP with the sum of errors equal to zero constraint.

Table 6. 5 Weighted goal programming - sum of errors equal to zero - MAPE

MAPE	MinMAD&MaxAD							
e = 0	1		2		3		4	
	6	12	6	12	6	12	6	12
All	37.15	40.01	37.04	39.36	36.64	39.20	37.01	39.06
Old	24.78	24.39	25.03	24.04	24.84	23.93	25.44	23.94
New	42.10	46.26	41.85	45.49	41.36	45.31	41.64	45.10
NON NEG								
All	36.04	35.36	35.77	35.26	35.70	35.06	35.45	34.57
Old	24.91	22.52	24.99	22.32	24.79	21.93	24.78	21.91
New	40.49	40.49	40.08	40.44	40.07	40.31	39.72	39.63
PROGRES								
All	35.19	34.02	35.03	34.07	34.93	33.96	34.97	33.95
Old	23.81	22.55	23.80	22.72	23.67	22.70	23.71	22.66
New	39.75	38.61	39.53	38.61	39.43	38.47	39.47	38.46

Table 6.6 shows the performance of the simple LP with sum of errors equal to zero according to the sMAPE. Similar to the comparison according to the MAPE, the first additional constraint improves the initial approaches and outperforms OLS and Holt. It performs slightly worse than WMA. The second additional constraint improves the results further and the LP with progressive weights outperforms the best (SES 0.8) of the traditional techniques.

Table 6. 6 Single objective - sum of errors equal to zero - sMAPE

sMAPE	LP									
e = 0	MinMAD		MinMAPE		MinMaxAD		MinADBD		MinADBPD	
Order	6	12	6	12	6	12	6	12	6	12
All	29.96	31.63	31.44	44.91	35.41	34.87	29.25	30.12	30.77	36.25
Old	23.74	26.11	25.26	29.98	29.15	26.31	23.84	24.22	24.97	27.73
New	32.45	33.84	33.90	50.88	37.66	37.95	31.41	32.48	33.09	39.66
NON NEG										
All	29.15	27.37	28.64	30.96	29.82	28.91	29.15	27.37	29.60	28.60
Old	24.26	22.36	24.24	25.11	24.71	22.45	24.26	22.36	24.88	23.48
New	31.11	29.37	30.40	33.29	31.86	31.50	31.11	29.37	31.50	30.64
PROGRES										
All	28.52	27.02	28.69	27.28	28.77	27.81	28.45	27.17	28.66	27.46
Old	23.22	22.26	23.33	22.50	23.26	23.15	23.05	22.63	23.22	22.86
New	30.64	28.92	30.84	29.19	30.98	29.67	30.62	28.98	30.84	29.30

Table 6.7 presents the performance of the WGP according to the sMAPE. The results are very similar. WGP seems not to improve the performance of the simple LP approaches, comparing them according to the sMAPE and the results of the two approaches are very similar. However, WGP still performs better than the OLS and the other four traditional techniques.

Table 6. 7 Weighted goal programming - sum of errors equal to zero - sMAPE

sMAPE	MinMAD&MaxAD							
e = 0	1		2		3		4	
	6	12	6	12	6	12	6	12
All	30.21	31.72	29.71	31.00	29.37	31.12	29.69	32.06
Old	24.83	24.64	23.43	23.94	23.34	23.94	23.79	24.45
New	32.37	34.55	32.22	33.83	31.78	33.99	32.05	35.10
NON NEG								
All	29.16	27.78	28.99	27.81	28.93	27.59	28.74	27.46
Old	24.29	22.55	24.38	22.48	24.16	22.19	24.16	22.17
New	31.11	29.88	30.83	29.94	30.83	29.75	30.57	29.58
PROGRES								
All	28.75	27.02	28.62	27.08	28.55	27.09	28.61	27.10
Old	23.24	22.23	23.23	22.43	23.15	22.41	23.17	22.34
New	30.95	28.93	30.78	28.94	30.70	28.96	30.79	29.01

Tables 6.8 and 6.9 show the performance of the approaches according to MASE. Similar to the comparison according to MAPE and sMAPE, the additional constraint on the simple LP approaches improves the results and outperforms the OLS. The second additional constraint improves the results further and performs as good as the best of the traditional techniques (0.82). The counterpart WGP approaches improve the results further and performs better (0.8) than the most accurate traditional technique.

Table 6. 8 Single objective - sum of errors equal to zero - MASE

MASE	LP									
e = 0	MinMAD		MinMAPE		MinMaxAD		MinADBD		MinADBPD	
Order	6	12	6	12	6	12	6	12	6	12
All	0.91	0.96	0.95	1.07	1.06	1.05	0.89	0.92	0.93	1.00
Old	0.88	0.95	0.92	1.02	1.09	1.01	0.88	0.90	0.91	0.96
New	0.93	0.97	0.96	1.09	1.04	1.07	0.90	0.93	0.94	1.01
NON NEG										
All	0.88	0.84	0.85	0.91	0.92	0.88	0.88	0.84	0.90	0.87
Old	0.88	0.84	0.88	0.91	0.92	0.86	0.88	0.84	0.89	0.86
New	0.88	0.85	0.84	0.92	0.92	0.89	0.88	0.85	0.90	0.88
PROGRES										
All	0.86	0.82	0.86	0.83	0.86	0.83	0.86	0.83	0.86	0.83
Old	0.85	0.81	0.86	0.82	0.86	0.85	0.84	0.83	0.85	0.84
New	0.86	0.82	0.87	0.83	0.86	0.82	0.86	0.83	0.87	0.83

Table 6. 9 Weighted goal programming - sum of errors equal to zero - MASE

MASE	MinMAD&MaxAD							
e = 0	1		2		3		4	
	6	12	6	12	6	12	6	12
All	0.91	0.94	0.89	0.92	0.88	0.92	0.89	0.92
Old	0.92	0.92	0.87	0.87	0.86	0.86	0.88	0.88
New	0.91	0.95	0.90	0.94	0.89	0.94	0.89	0.94
NON NEG								
All	0.88	0.86	0.88	0.86	0.88	0.85	0.87	0.85
Old	0.88	0.84	0.89	0.84	0.88	0.83	0.88	0.83
New	0.88	0.87	0.87	0.86	0.88	0.86	0.87	0.85
PROGRES								
All	0.85	0.80	0.84	0.80	0.84	0.80	0.84	0.80
Old	0.85	0.82	0.85	0.83	0.85	0.83	0.85	0.83
New	0.84	0.79	0.84	0.79	0.84	0.79	0.84	0.79

6.4 RESULTS: SUM OF PERCENTAGE ERRORS EQUAL TO ZERO

Table 6.10 presents the results of the single objective approaches with sum of percentage errors equal to zero constraint.

Table 6. 10 Single objective - sum of percentage errors equal to zero - MAPE

MAPE %e = 0	LP									
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD	
SIMPLE										
Order 6										
All	34.33	(-2.98)	33.83	(-5.65)	39.89	(-1.69)	33.17	(-3.14)	34.29	(-3.84)
Old	24.94	(0.14)	25.20	(-1.06)	25.14	(-3.8)	25.71	(0.94)	26.09	(1.12)
New	38.08	(-4.23)	37.28	(-7.49)	45.78	(-0.33)	36.16	(-4.78)	37.56	(-5.82)
Order 12										
All	38.91	(-1.34)	41.87	(-4.73)	43.34	(0.87)	36.20	(-1.95)	36.69	(-6.4)
Old	26.83	(1.05)	27.42	(-1.09)	25.81	(0.85)	26.08	(1.7)	27.79	(0.06)
New	43.74	(-2.29)	47.65	(-6.19)	50.36	(1.58)	40.25	(-3.42)	40.25	(-8.98)
NON NEG										
Order 6										
All	33.07	(-3.41)	33.55	(-3.28)	32.71	(-4.02)	32.08	(-3.93)	32.60	(-4.23)
Old	24.22	(-1.2)	24.22	(-1.5)	25.43	(0.51)	23.51	(-1.32)	23.69	(-1.62)
New	36.61	(-4.29)	37.28	(-4)	35.62	(-5.83)	35.51	(-4.98)	36.16	(-5.27)
Order 12										
All	32.66	(-2.76)	32.00	(-4.98)	34.71	(-1.41)	31.80	(-3.3)	32.09	(-4.59)
Old	23.29	(0.67)	22.92	(-0.82)	23.61	(1.62)	22.13	(-0.21)	22.28	(-1.14)
New	36.40	(-4.13)	35.63	(-6.64)	39.15	(-2.62)	35.67	(-4.54)	36.01	(-5.97)
PROGRES										
Order 6										
All	31.19	(-3.99)	31.35	(-4.02)	31.26	(-3.72)	31.01	(-4.11)	31.31	(-4.07)
Old	22.74	(-1.1)	22.73	(-1.23)	23.61	(-0.25)	22.59	(-0.99)	22.72	(-1.04)
New	34.57	(-5.15)	34.80	(-5.13)	34.33	(-5.11)	34.37	(-5.35)	34.75	(-5.29)
Order 12										
All	30.54	(-3.73)	30.48	(-3.89)	30.95	(-3.75)	30.38	(-3.96)	30.61	(-3.93)
Old	22.43	(-0.05)	22.17	(-0.78)	23.12	(-0.48)	22.17	(-0.74)	22.33	(-0.95)
New	33.79	(-5.2)	33.81	(-5.13)	34.09	(-5.06)	33.67	(-5.24)	33.92	(-5.12)

As in the previous chapter, the approaches with the sum of the percentage errors equal to zero constraint, give betters results (MAPE) compared with the sum of errors equally to zero constraint. Here, the non negative weight approaches outperform all the other techniques. All the non negative approaches weight give very similar results; however, MinADBD is slightly better and MinMaxAD slightly worse. The constraint of progressive weights improves the results further. Here all the approaches give almost the same results, the average MAPE close to 31% for order 6 and close to 30% for order 12. The average improvement over the OLS is significant (reduction of MAPE by 4% - 6%), and so is the average improvement over Holt (reduction of MAPE by 3%).

Table 6.11 presents the results of the WGP with sum of percentage errors equal to zero constraint. The results are improved compared with the WGP with sum of errors equal to zero constraint. However, the accuracy of both the non negative and progressive approaches is comparable with the accuracy of the single objective LP in the previous table.

Table 6. 11 Weighted goal programming - sum of percentage errors equal to zero - MAPE

MAPE	MinSAD&MaxAD							
%e = 0	1		2		3		4	
SIMPLE								
Order 6								
All	34.14	(-3.01)	34.02	(-3.02)	33.99	(-2.65)	33.76	(-3.25)
Old	24.73	(-0.05)	25.33	(0.3)	25.82	(0.99)	25.94	(0.5)
New	37.90	(-4.2)	37.50	(-4.34)	37.26	(-4.1)	36.89	(-4.75)
Order 12								
All	38.79	(-1.23)	38.60	(-0.76)	38.40	(-0.79)	36.81	(-2.24)
Old	24.71	(0.31)	24.66	(0.62)	23.79	(-0.14)	23.95	(0.02)
New	44.42	(-1.84)	44.17	(-1.32)	44.25	(-1.06)	41.96	(-3.15)
NON NEG								
Order 6								
All	33.14	(-2.9)	32.96	(-2.81)	32.22	(-3.49)	31.99	(-3.46)
Old	24.18	(-0.74)	23.95	(-1.05)	23.93	(-0.86)	24.13	(-0.65)
New	36.73	(-3.77)	36.57	(-3.52)	35.53	(-4.54)	35.13	(-4.59)
Order 12								
All	32.59	(-2.77)	32.58	(-2.68)	31.82	(-3.25)	31.73	(-2.83)
Old	22.94	(0.42)	22.74	(0.42)	22.38	(0.44)	22.44	(0.53)
New	36.45	(-4.04)	36.52	(-3.92)	35.59	(-4.72)	35.45	(-4.18)
PROGRES								
Order 6								
All	31.11	(-4.09)	31.00	(-4.03)	31.14	(-3.78)	31.06	(-3.91)
Old	22.86	(-0.95)	22.74	(-1.06)	22.77	(-0.91)	22.71	(-1)
New	34.41	(-5.34)	34.31	(-5.22)	34.50	(-4.93)	34.40	(-5.08)
Order 12								
All	30.51	(-3.51)	30.35	(-3.71)	30.34	(-3.63)	30.43	(-3.51)
Old	22.42	(-0.13)	22.32	(-0.4)	22.46	(-0.24)	22.46	(-0.2)
New	33.75	(-4.86)	33.57	(-5.04)	33.49	(-4.98)	33.62	(-4.84)

Tables 6.12 and 6.13 show the average sMAPE of the approaches on the test set. In contrast with the comparison according to the MAPE, the performance of these approaches worsens comparing the sMAPE. The result are similar to the approaches presented in the previous chapters, since the constraint where the sum of errors should be equal to zero improves the results overall, but the improvement is mainly observed on the smooth and seasonal series, where they worsen on the hard series. Nevertheless, both simple LP and WGP outperform the OLS, but not the other four traditional techniques.

Table 6. 12 Single objective - sum of percentage errors equal to zero - sMAPE

sMAPE %e = 0	LP									
	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD	
SIMPLE										
Order 6										
All	31.60	(1.63)	32.13	(0.69)	38.65	(3.24)	30.16	(0.91)	30.96	(0.19)
Old	26.38	(2.64)	26.99	(1.73)	24.24	(-4.91)	25.71	(1.87)	26.09	(1.12)
New	33.69	(1.23)	34.18	(0.28)	44.41	(6.75)	31.94	(0.53)	32.90	(-0.19)
Order 12										
All	47.26	(15.63)	41.84	(-3.07)	83.43	(48.56)	32.04	(1.92)	36.72	(0.46)
Old	28.37	(2.26)	29.05	(-0.94)	25.11	(-1.2)	26.08	(1.87)	27.79	(0.06)
New	54.82	(20.98)	46.96	(-3.92)	106.76	(68.81)	34.42	(1.94)	40.29	(0.63)
NON NEG										
Order 6										
All	30.54	(1.39)	30.88	(2.24)	30.83	(1.01)	29.74	(0.59)	30.03	(0.43)
Old	26.05	(1.79)	26.09	(1.85)	27.62	(2.9)	25.38	(1.13)	25.67	(0.79)
New	32.33	(1.22)	32.80	(2.4)	32.11	(0.25)	31.49	(0.38)	31.78	(0.28)
Order 12										
All	29.00	(1.63)	28.69	(-2.27)	30.87	(1.95)	28.52	(1.16)	28.60	0.00
Old	25.16	(2.79)	24.97	(-0.14)	25.44	(2.99)	24.08	(1.71)	24.37	(0.9)
New	30.54	(1.17)	30.18	(-3.12)	33.04	(1.54)	30.30	(0.93)	30.28	(-0.36)
PROGRES										
Order 6										
All	28.91	(0.39)	29.34	(0.65)	29.69	(0.92)	29.00	(0.54)	29.24	(0.57)
Old	23.60	(0.39)	24.59	(1.26)	25.33	(2.08)	24.42	(1.37)	24.56	(1.34)
New	31.03	(0.38)	31.24	(0.4)	31.44	(0.46)	30.83	(0.21)	31.11	(0.27)
Order 12										
All	28.22	(1.21)	28.16	(0.89)	28.82	(1.02)	28.02	(0.85)	28.33	(0.87)
Old	24.18	(1.92)	23.97	(1.47)	24.81	(1.66)	23.96	(1.33)	24.13	(1.27)
New	29.84	(0.92)	29.84	(0.65)	30.43	(0.76)	29.65	(0.66)	30.00	(0.71)

Table 6. 13 Weighted goal programming - sum of percentage errors equal to zero sMAPE

sMAPE	MinSAD&MaxAD							
%e = 0	1		2		3		4	
SIMPLE								
Order 6								
All	32.56	(2.35)	31.68	(1.98)	31.50	(2.13)	31.41	(1.72)
Old	29.66	(4.84)	27.09	(3.66)	26.23	(2.89)	26.01	(2.21)
New	33.72	(1.35)	33.52	(1.3)	33.61	(1.83)	33.57	(1.52)
Order 12								
All	36.49	(4.77)	35.78	(4.78)	35.60	(4.48)	34.75	(2.69)
Old	29.44	(4.8)	27.71	(3.78)	25.93	(1.99)	26.53	(2.09)
New	39.31	(4.76)	39.01	(5.18)	39.47	(5.48)	38.04	(2.94)
NON NEG								
Order 6								
All	30.83	(1.66)	30.60	(1.61)	30.17	(1.24)	30.10	(1.36)
Old	26.00	(1.71)	25.63	(1.24)	25.59	(1.43)	25.79	(1.63)
New	32.76	(1.65)	32.59	(1.76)	32.00	(1.17)	31.82	(1.25)
Order 12								
All	29.27	(1.49)	29.21	(1.4)	28.82	(1.23)	29.00	(1.54)
Old	24.84	(2.29)	24.63	(2.15)	24.25	(2.07)	24.34	(2.16)
New	31.04	(1.16)	31.05	(1.11)	30.65	(0.9)	30.86	(1.29)
PROGRES								
Order 6								
All	30.83	(2.08)	30.60	(1.98)	30.17	(1.62)	30.10	(1.48)
Old	26.00	(2.75)	25.63	(2.39)	25.59	(2.44)	25.79	(2.62)
New	32.76	(1.81)	32.59	(1.81)	32.00	(1.3)	31.82	(1.03)
Order 12								
All	29.27	(2.25)	29.21	(2.13)	28.82	(1.73)	29.00	(1.9)
Old	24.84	(2.61)	24.63	(2.2)	24.25	(1.84)	24.34	(2)
New	31.04	(2.11)	31.05	(2.11)	30.65	(1.68)	30.86	(1.86)

Finally, Tables 6.14 and 6.15 present the performance of the simple LP and WGP according the MASE. Similar to the comparison according to MASE, it seems that the results are slightly worse. However, the differences between the approaches are very small.

Table 6. 14 Single objective - sum of percentage errors equal to zero - MASE

MASE	LP									
%e = 0	MinSAD		MinSAPE		MinMaxAD		MinADBD		MinADBPD	
SIMPLE										
Order 6										
All	0.93	(0.01)	0.93	(-0.02)	1.02	(-0.03)	0.89	(0)	0.91	(-0.02)
Old	0.94	(0.06)	0.95	(0.03)	0.88	(-0.22)	0.92	(0.04)	0.92	(0.02)
New	0.92	(0)	0.92	(-0.04)	1.08	(0.04)	0.88	(-0.01)	0.90	(-0.04)
Order 12										
All	0.99	(0.03)	1.04	(-0.02)	1.07	(0.02)	0.92	(0)	0.95	(-0.05)
Old	1.02	(0.06)	1.05	(0.03)	0.91	(-0.09)	0.94	(0.04)	0.97	(0.01)
New	0.98	(0.01)	1.04	(-0.05)	1.13	(0.06)	0.91	(-0.02)	0.94	(-0.07)
NON NEG										
Order 6										
All	0.90	(0.01)	0.90	(0.05)	0.92	(0)	0.88	(0)	0.88	(-0.01)
Old	0.93	(0.05)	0.93	(0.05)	1.00	(0.08)	0.91	(0.03)	0.91	(0.01)
New	0.89	(0)	0.89	(0.05)	0.89	(-0.03)	0.87	(-0.02)	0.87	(-0.02)
Order 12										
All	0.87	(0.02)	0.86	(-0.05)	0.90	(0.02)	0.85	(0.01)	0.85	(-0.03)
Old	0.93	(0.09)	0.92	(0.01)	0.95	(0.09)	0.89	(0.05)	0.89	(0.03)
New	0.84	(0)	0.84	(-0.08)	0.88	(-0.01)	0.83	(-0.01)	0.83	(-0.05)
PROGRES										
Order 6										
All	0.85	(0)	0.86	(0)	0.86	(0)	0.85	(0)	0.86	(0)
Old	0.87	(0.02)	0.88	(0.03)	0.90	(0.05)	0.88	(0.03)	0.88	(0.03)
New	0.84	(-0.01)	0.85	(-0.02)	0.84	(-0.02)	0.84	(-0.02)	0.85	(-0.02)
Order 12										
All	0.83	(0.01)	0.83	(0)	0.84	(0.01)	0.83	(0)	0.84	(0)
Old	0.88	(0.07)	0.87	(0.05)	0.90	(0.05)	0.87	(0.05)	0.88	(0.04)
New	0.82	(-0.01)	0.81	(-0.02)	0.81	(-0.01)	0.81	(-0.02)	0.82	(-0.02)

Table 6. 15 Weighted goal programming - sum of percentage errors equal to zero - MASE

MASE	MinSAD&MaxAD							
%e = 0	1		2		3		4	
SIMPLE								
Order 6								
All	0.93	(0)	0.91	(0)	0.91	(0)	0.91	(0)
Old	1.01	(0.01)	0.94	(0.04)	0.92	(0.04)	0.92	(0.03)
New	0.91	(0.1)	0.90	(0.04)	0.90	(0.03)	0.90	(0.03)
Order 12								
All	0.97	(0)	0.96	(0)	0.96	(0)	0.94	(0)
Old	0.99	(0.06)	0.95	(0.09)	0.92	(0.09)	0.92	(0.06)
New	0.97	(0.04)	0.97	(0.01)	0.97	(-0.02)	0.95	(-0.02)
NON NEG								
Order 6								
All	0.89	(0)	0.89	(0)	0.88	(0)	0.88	(0)
Old	0.93	(0.01)	0.92	(0)	0.92	(0)	0.93	(0)
New	0.88	(0.04)	0.88	(0.05)	0.86	(0.04)	0.86	(0.06)
Order 12								
All	0.86	(0)	0.86	(0)	0.85	(0)	0.85	(0)
Old	0.92	(0.02)	0.91	(0.01)	0.90	(0.02)	0.91	(0.02)
New	0.84	(0.05)	0.83	(0.05)	0.83	(0.04)	0.83	(0.05)
PROGRES								
Order 6								
All	0.85	(0)	0.85	(0)	0.85	(0)	0.85	(0)
Old	0.89	(0)	0.88	(-0.01)	0.88	(0)	0.88	(0)
New	0.84	(0.05)	0.83	(0.05)	0.84	(0.05)	0.83	(0.04)
Order 12								
All	0.83	(0)	0.83	(0)	0.82	(0)	0.82	(0)
Old	0.88	(0.01)	0.88	(0)	0.89	(0)	0.88	(0)
New	0.81	(0.09)	0.80	(0.09)	0.80	(0.09)	0.80	(0.09)

6.5 CONCLUSION

I exploited the flexibility of linear programming to improve the results of the initial approaches for series with special characteristics, specifically high level of randomness. In linear programming it is very easy to add additional constraints and this is one of the main virtues of this technique, compared with OLS or other methods. The analysis shows that the additional constraints improve the performance of the approaches and outperform OLS and the other traditional techniques that are compared. The conclusions of this chapter are summarized as follows:

1. The non negative weight approaches outperform the initial LP formulations; however, the progressive weight approaches outperform the non negative weight approaches.
2. WGP perform slightly better than the simple LP; however, the differences are small.
3. The sum of percentage error equal to zero constraint gives a lower MAPE than the sum of errors equal to zero constraint, but the latter giver lower sMAPE and MASE.
4. The approaches with sum of errors equal to zero outperform all the traditional techniques according to sMAPE and MASE.
5. The approaches with sum of percentage errors equal to zero constraint outperform all the other techniques according to MAPE.

LP has been proved a very good tool for optimising autoregressive based time series forecasts. The next step is to test the applicability of LP as an approach for combining forecasts.

7 LINEAR PROGRAMMING FOR COMBINED FORECASTING

In this chapter, linear programming is applied to combine forecasts. I use linear programming to minimise one error index (SAD, SAPE and MaxAD) and the average of the three. In addition, I apply weighted goal programming to minimise two error indices (MaxAD and SAD, similar to chapter 5). The approaches combine eight simple individual techniques and the results are compared with six traditional combination techniques that are found in the literature.

7.1 SINGLE OBJECTIVE AND AVERAGE LINEAR PROGRAMMING

The first LP formulations minimise SAD, SAPE, MaxAD for combined forecasting models. Let F_c be the combined forecast, F_i the forecast by method i ($1 \leq i \leq k$), w_i the weight of forecast i, Y_t the actual observation at time t, e_{1t} is the error of underestimation and e_{2t} the error of overestimation and t ($1 \leq t \leq T$) is the time. The mathematical expression is:

$$F_{ct} = \sum_{i=1}^{k} w_i F_{it} \quad \textbf{(7. 1)}$$

Where $\sum_{i=1}^{k} w_i = 1$ and $0 \leq w_i \leq 1$

Then, the linear program for minimising the *SAD* is:

Objective function:

$$Min \sum_{t=1}^{T} \left| F_{ct} - Y_t \right| \quad \textbf{(7. 2)}$$

$$Min \sum_{t=1}^{T} \left| \sum_{i=1}^{k} w_i F_{it} - Y_t \right| \quad \textbf{(7. 3)}$$

$$Min \sum_{t=1}^{T} (e_{1t} + e_{2t}) \quad \textbf{(7. 4)}$$

subject to:

$$\sum_{i=1}^{k} w_i F_{it} + e_{1t} - e_{2t} = Y_t \quad \forall t \qquad \textbf{(7. 5)}$$

$$\sum_{i=1}^{k} w_i = 1 \qquad \textbf{(7. 6)}$$

$$0 \leq w_i \leq 1 \qquad \textbf{(7. 7)}$$

e_1, e_2 non-negative and w_i unrestricted in sign.

In the same way, the linear program for minimising the MAPE is:

Objective function:

$$Min \sum_{t=1}^{T} \left| (F_{ct} - Y_t)/Y_t \right| \qquad \textbf{(7. 8)}$$

$$Min \sum_{t=1}^{T} \left| \left(\sum_{i=1}^{k} w_i F_{it} - Y_t \right) \Big/ Y_t \right| \qquad \textbf{(7. 9)}$$

$$Min \sum_{t=1}^{T} \left[(e_{1t} + e_{2t})/Y_t \right] \qquad \textbf{(7. 10)}$$

subject to:

(7. 5), **(7. 6)**, **(7. 7)**

Finally the LP model that minimises the MaxAD is:

$$Min \;\; e \qquad \textbf{(7. 11)}$$

Where *e* is the *MaxAD,*

subject to:

$$\sum_{i=1}^{k} w_i F_{it} + e_t = Y_t \quad \forall t \tag{7. 12}$$

Where e_t is the forecasting error for period t

$$e - e_t \geq 0 \quad \forall t \tag{7. 13}$$

$$-e - e_t \leq 0 \quad \forall t \tag{7. 14}$$

(7. 6), **(7. 7)**

If F_{c1}, F_{c2} and F_{c3} are the forecasts of the MinSAD, MinSAPE and MinMaxAD models respectively, then the average forecast F_{ca} of the three model can be calculated:

$$F_{ca} = \frac{F_{c1} + F_{c2} + F_{c3}}{3} \tag{7. 15}$$

Finally, we ran the experiments again by adding the *Σe* = 0 and *Σ%e* = 0 constraints to the linear program for all the three models in order to remove underestimation or overestimation bias from our forecast; however the models became over-constrained and it was not possible to obtain an optimal solution for many of the series in the data sets.

7.2 TESTS

As presented in the methodology chapter, I used 8 different forecasting techniques for the experiment:

- Naïve 1
- Moving Average (Order 4)
- Simple Exponential Smoothing (0.2)
- Holt linear (0.2, 0.1)

- Holt Winters (0.2, 0.1, 0.8)
- Adaptive Exponential Smoothing (0.9)
- Autoregressive (Order 6)
- Seasonal Autoregressive (Order 6)

I applied them on the initial data set that was used in chapter 5. In addition, the performance of the models is tested in comparison with five traditional combined forecasting methods (The combination methods are described in detail in Chapter 2):

- Simple Average
- Inverse Proportion of the SAD
- Inverse Proportion of the MAPE
- Inverse Proportion of the MSE
- Average Inverse Proportion of the MAD, MAPE and MSE
- Weighting based upon the absolute error

To estimate the coefficients of the simple and seasonal autoregressive models we used the method of the ordinary least squares (calculated in *STATA*), and we implemented the other techniques in *Excel*. I run experiments by combining all forecasts of the eight individual forecasting techniques. In addition I have rerun the experiments using only the first seven. The reason is that the ARS model was a dominant technique for the series with strong seasonal pattern. The results of the second analysis can be found in APENDIX A. The dataset that I used for the experiments consists of the 60 randomly selected monthly time series that were used in the previous models.

7.3 FIRST RESULTS

The first table shows the results of all the eight individual techniques. I report the average MAPE over all the series and the average MAPE on the series in each subgroup. The AR and ARS models give the best results in general. Specifically, they are better on the smooth and seasonal series. ARS is significant by better than all other techniques. On the other hand, HOLT is the best technique on the hard series; however, SES and MA give good results too.

Table 7. 1 Performance of individual forecasts - MAPE

	MAPE							
TECHNIQUES	**NAIVE**	**MA**	**SES**	**HOLT**	**H-W**	**AES**	**AR**	**ARS**
All	10.19	10.02	10.26	10.3	11.22	10.17	8.69	7.39
Smooth	3.17	3.94	4.82	4.97	4.53	3.17	3.14	3.14
Hard	28.04	24.06	23.75	22.92	28.7	27.89	24.47	24.47
Seasonal	12.77	13.02	12.44	12.74	13.42	12.77	9.79	5.45

The comparison of the performance of the individual techniques according to sMAPE (Table 7.2) is similar. However, the best technique on the hard series is SES, second comes the MA and third the Holt.

Table 7. 2 Performance of individual forecasts - sMAPE

	sMAPE							
TECHNIQUES	**NAIVE**	**MA**	**SES**	**HOLT**	**H-W**	**AES**	**AR**	**ARS**
All	9.81	9.70	9.82	10.12	10.77	9.79	8.54	7.24
Smooth	3.11	3.86	4.72	4.75	4.36	3.11	3.03	3.03
Hard	26.62	23.04	22.37	23.50	27.40	26.48	24.01	24.01
Seasonal	12.37	12.68	11.90	12.22	12.94	12.37	9.75	5.41

Finally Table 7.3 shows the comparison of the techniques according to MASE.

Table 7. 3 Performance of individual forecasts - MASE

	MASE							
TECHNIQUES	**NAIVE**	**MA**	**SES**	**HOLT**	**H-W**	**AES**	**AR**	**ARS**
All	1.00	1.21	1.42	1.27	1.22	1.00	0.92	0.81
Smooth	1.00	1.36	1.81	1.51	1.34	1.00	0.95	0.95
Hard	1.00	0.86	0.82	0.85	1.05	0.99	0.88	0.88
Seasonal	1.00	1.13	1.06	1.08	1.10	1.00	0.88	0.52

Similarly to the comparison according to MAPE and sMAPE, AR and ARS are the best approaches overall and on the smooth and seasonal series. On the hard series, the best is SES followed by Holt and MA. Obviously, Naive has MASE equal to 1, because it is compared with itself. The above comparison does not aim to identify the most accurate technique. The order of MA, AR and ARS and the parameters of the smoothing techniques have not been optimised. It is an observation of the performance of the techniques that will be useful in the analysis of performance of the combined forecast.

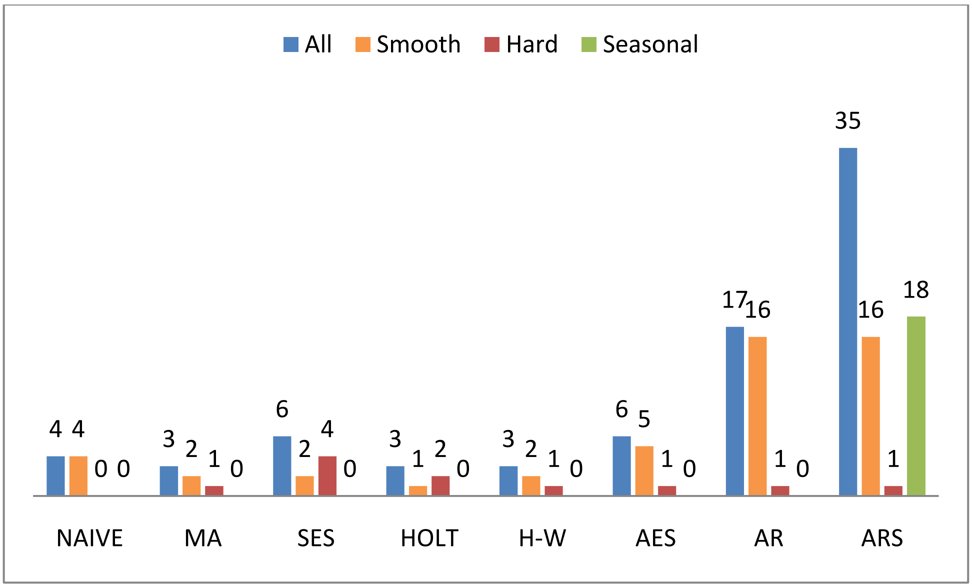

Figure 7. 1 Number of times each individual technique gives the best results

Figure 7.1 shows the number of times each individual forecasting method gives the best results. As we can see, AR and ARS approaches give the best results, much more often than the other techniques. ARS is the only technique that gives the best results on the seasonal series.

Figure 7.2 shows the number of times each individual forecasting method gives the worst result.

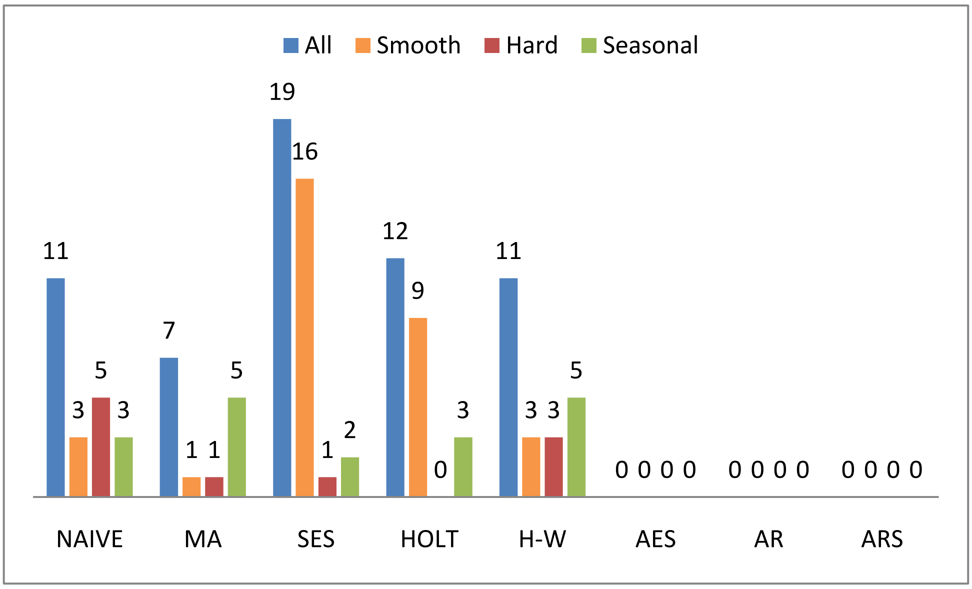

Figure 7. 2 Number of times each individual technique gives the worst results

As we can see, the AES, AR and ARS approaches do not perform worst on any of the series. SES is the technique that performs worst most times.

In the next 2 tables (7.4 and 7.5) we can see the results of the of the LP models for combined forecasting compared with the traditional combined forecasting methods. The first table shows the average MAPE of all the series in the test set as well as for each subgroup separately.

Table 7. 4 Traditional and single objective LP combinations - MAPE

PERFORMANCE OF COMBINED FORECASTS										
MAPE	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	8.99	8.31	8.30	7.63	8.07	7.53	7.28	7.30	7.51	7.20
Smooth	3.52	3.34	3.35	3.21	3.29	3.16	3.14	3.06	3.35	3.12
Hard	23.96	23.70	23.65	23.52	23.62	23.86	23.92	24.28	23.12	23.42
Seasonal	10.40	8.60	8.59	6.67	7.92	6.25	5.40	5.41	6.24	5.43

To begin with, the weighting based on the absolute deviation gives the best results compared with the other traditional techniques. This method is the best overall as well as for the smooth and seasonal series. On the other hand, the inverse proportion to MSE gives the best results compared with the other inverse proportion formulations. The inverse proportion to MAD and MAPE and the average inverse proportion method give very similar results. The Simple Average method gives the worst results in general.

We can see that the LP models give better results than the traditional combination techniques. The MinSAD model gives the best results of the three simple LP models as well as the best results on the seasonal series. The MinSAPE model gives slightly worse results than the MinSAD; however, it gives the best results on the smooth series. Both MinSAD and MinSAPE models perform slightly worse than the inverse proportion models on the hard series. The MinMaxAD model performs worse than the other two LP models; however, it is the technique that performs better on the hard series. The average LP model is the best combined forecasting method. It performs better than any other method overall and it gives very good results (slightly worse than the best technique for each group) for each of the subgroups separately. Finally, in Table 7.1 we can see that the best individual technique is the ARS with average MAPE 7.39. If we compare this with the combined forecasting models, we can see that only the LP models (with exception the MinMaxAD) outperform the best individual technique overall, as well as its performance in each subgroup. Thus, it seems that combined forecasting based on LP does not only minimise the probability of a bad forecast, but also improves the accuracy of a good forecast.

Table 7.5 shows the average sMAPE of all the series in the test set as well as for each subgroup separately.

Table 7. 5 Traditional and single objective LP combinations - sMAPE

sMAPE	PERFORMANCE OF COMBINED FORECASTS									
	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	8.70	8.07	8.07	7.44	7.85	7.31	7.12	7.12	7.42	7.04
Smooth	3.42	3.24	3.25	3.12	3.20	3.07	3.05	2.97	3.24	3.02
Hard	23.20	23.00	22.96	22.85	22.94	22.99	23.34	23.59	23.24	22.90
Seasonal	10.05	8.37	8.36	6.55	7.73	6.13	5.34	5.36	6.05	5.35

The results are very similar as in the comparison according to the MAPE. The only difference is that the traditional techniques outperform simple LP on the hard series, except the average LP, which gives the best results.

Table 7. 6 Traditional and single objective LP combinations - MASE

MASE	PERFORMANCE OF COMBINED FORECASTS									
	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	0.98	0.91	0.91	0.85	0.88	0.84	0.81	0.81	0.89	0.82
Smooth	1.10	1.02	1.03	0.99	1.01	0.99	0.95	0.96	1.07	0.97
Hard	0.86	0.85	0.85	0.84	0.85	0.85	0.86	0.87	0.84	0.85
Seasonal	0.85	0.73	0.73	0.60	0.68	0.57	0.52	0.52	0.58	0.52

Table 7. 6 shows the average MASE of all the series in the test set as well as for each subgroup separately. LP outperforms the traditional techniques overall and on each group separately. Specifically MinSAD, MinSAPE and the average LP outperform all the traditional techniques on the smooth and seasonal and MinMaxAD performs as good as the best traditional technique on the hard series (the Inverse Proportion according to the MSE).

An additional way to measure the performance of a combination method is to test how it performs in comparison with the best individual technique. In order to do this we can calculate the percentage difference between the MAD of the combination and the MAD of the best technique. That is:

$$average\ \%difference\ between\ the\ best\ MAD = \frac{1}{M}\sum_{m=1}^{M}\frac{MAD_{FC} - \min MAD_{im}}{\min MAD_{im}} \times 100 \qquad \textbf{(7. 16)}$$

m (= 1, … , *M*) is the series in our sample, MAD_{FC} the MAD of the combined forecast on the test set and $minMAD_{im}$ the MAD of the best individual technique *i* on the specific series *m*. This index shows the amount (in percent) by which is the MAD of the combination method improves the MAD of the best individual forecasting method. If the percentage difference is negative, then the index shows that the MAD of the combined forecast is actually smaller than this of the best individual method.

Table 7.7 shows the percentage difference between the MAD of a combined forecasting technique and the MAD of the best individual forecast for each series. As we can see, the average percentage difference between the MAD of the LP combination models is significantly smaller than this of the other combined forecasting techniques, except for the MinMaxAD model. According to this index the best LP combination model is MinSAD. In addition, both MinSAD and MinSAPE outperform the average LP combination approach. The good performance is mainly observed on the smooth series and especially on the seasonal series, where the average difference between the MAD is negative. On the other hand, the performance on the hard series is similar for all the combined forecasting techniques (where the average inverse proportion is the best and the MinSAPE is the worst).

Table 7. 7 Traditional and single objective LP combinations - Difference between the best

PERFORMANCE OF COMBINED FORECASTS

%BEST SERIES	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	38.86	24.88	24.91	12.60	20.46	11.15	4.96	5.18	15.74	5.84
Smooth	23.46	14.98	15.13	10.29	13.24	11.46	6.53	6.25	20.24	8.31
Hard	11.77	10.32	10.15	9.33	9.93	10.12	11.00	12.80	10.01	10.08
Seasonal	81.29	50.55	50.47	18.50	39.13	11.17	-1.18	-0.97	10.92	-0.93

One of the main reasons why we use combined forecasting is to minimise the risk of selecting a forecasting method that is inaccurate or inappropriate for the application that is selected for. Thus, the combinations help us to hedge this risk between several different techniques. Hence, the next table shows the percentage difference between the MAD of the worst individual forecasting technique and the MAD of the combined forecasting technique for each series. That is:

$$average\ \%difference\ between\ the\ worst\ MAD = \frac{1}{M}\sum_{m=1}^{M}\frac{\max MAD_{im} - MAD_{FC}}{\max MAD_{im}} \times 100 \qquad \textbf{(7. 17)}$$

m (= 1, … , *M*) is the series in our sample, MAD_{FC} the MAD of the combined forecast on the test set and $maxMAD_{im}$ the MAD of the worst individual technique *i* on the specific series *m*. This index indicates the amount (in percent) by which the MAD of the combined forecasting technique improves on the performance of the worst individual technique.

As we can see in Table 7.8, the results are similar to those for the previous index. MinSAD, MinSAPE and the average LP have greater average percentage differences compared to the other methods, and MinMaxAD has a smaller percentage difference compared with the weights based on absolute error and inverse proportion to the MSE method. LP models in general perform better on the smooth series and significantly better on the seasonal series. They perform slightly worse on the hard series. Nevertheless, the differences are not as significant as in the percentage difference in relation the best individual technique.

Table 7. 8 traditional and single objective LP combinations - Difference between the worse PERFORMANCE OF COMBINED FORECASTS

%WORST SERIES	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	32.43	37.15	37.18	41.24	38.64	41.92	43.58	43.42	38.64	43.27
Smooth	37.40	40.32	40.28	41.78	40.89	41.69	42.70	42.94	36.11	42.00
Hard	22.10	22.75	22.89	23.32	22.99	22.64	21.87	20.43	22.92	22.57
Seasonal	29.34	39.52	39.59	50.23	43.35	53.04	57.19	57.04	51.88	57.03

Finally, if we assume that the decision maker has no knowledge about which is the best individual forecasting technique on a specific series and the technique is selected randomly, the expected MAD of the forecast is the average MAD of all the individual techniques. An additional way to test the performance of a combined forecast is to compare it with the expected (average) MAD of all the individual techniques. Hence, the next table shows the percentage difference between the average MAD and the MAD of the combined forecasting technique for each series. That is:

$$average\ \%difference\ between\ the\ average\ MAD = \frac{1}{M}\sum_{m=1}^{M}\frac{MMAD_m - MAD_{FC}}{MMAD_m}\times 100 \quad \textbf{(7. 18)}$$

m (= 1, … , M) is the series in our sample, MAD_{FC} the MAD of the combined forecast on the test set and $MMAD_m$ the average MAD of the all the individual techniques on the specific series m. This index indicates the amount (in percent) by which the MAD of the combined forecasting technique improves on the performance of the "average" technique.

The results (Table 7.9) are similar to the percentage differences between the MAD of the worst technique and this of the combined forecasts. MinSAD, MinSAPE and the average LP have greater average percentage differences compared to the other methods, and MinMaxAD has a smaller percentage difference compared with the weights based on absolute error and inverse proportion to the MSE method. LP models in general perform better on the smooth series and significantly better on the seasonal series. They perform slightly worse on the hard series. Nevertheless, the differences are not as significant as in the percentage difference in relation the best individual technique.

Table 7. 9 Traditional and single objective LP combinations - Difference between the average PERFORMANCE OF COMBINED FORECASTS

%AVERAGE	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	10.51	17.23	17.23	22.81	19.26	19.26	26.22	26.04	19.29	25.69
Smooth	11.16	16.11	16.03	18.68	17.08	17.08	20.59	20.84	10.94	19.46
Hard	6.15	7.25	7.40	8.05	7.57	7.57	6.64	5.06	7.50	7.38
Seasonal	11.76	24.76	24.83	38.36	29.62	29.62	47.11	46.92	40.69	46.94

A general conclusion about the application of LP to combined forecasting is that LP based approaches, with the only exception of MinMaxAD, outperform the traditional combined forecasting techniques on the one hand, and that they also improve on the performance of the individual techniques on the other hand. MinMaxAD is the best combined forecasting method for series with high variability. The next step is to examine whether single objective LP models can be improved with a multi-objective Weighted Goal Programming approach.

7.4 WEIGHTED GOAL PROGRAMMING

Similar to the single LP the WGP formulation that minimises both SAD and MaxAD is the following:

$$Min \ \ a_1 d_1^+ + a_2 d_2^+ \tag{7. 19}$$

subject to:

$$\sum_{t=1}^{T} (e_{1t} + e_{2t}) + d_1^- - d_1^+ = 0 \tag{7. 20}$$

$$e + d_2^- - d_2^+ = 0 \tag{7. 21}$$

where e_{1t} is the underestimation error, e_{2t} the overestimation error and e the MaxAD.

$$e - e_{1t} - e_{2t} \geq 0 \quad \forall t \qquad \textbf{(7. 22)}$$

(7. 6), (7. 7)

$$e_1, e_2, e, d_1^+, d_1^-, d_2^+, d_2^- \geq 0$$

Similar to the WGP for optimising the parameters of autoregressive models, I run experiments for $a_1 = 1$ *and* a_2 = 1, 2, 3, 4, 5 and additionally a_2 = 6.

7.5 RESULTS

In Table 7.10 I show the performance of the WGP combined forecasting models. I explore the combinations of all eight individual forecasting methods and report the overall average MAPE and the average MAPE on different types of time series.

Table 7. 10 Weighted goal programming combinations - MAPE

	PERFORMANCE OF COMBINED FORECASTS					
MAPE	**WEIGHTED GOAL PROGRAMMING**					
a2	**1**	**2**	**3**	**4**	**5**	**6**
All	**7.27**	**7.30**	**7.28**	**7.23**	**7.25**	**7.25**
Smooth	**3.13**	**3.14**	**3.13**	**3.13**	**3.14**	**3.15**
Hard	**23.85**	**24.03**	**23.87**	**23.55**	**23.56**	**23.48**
Seasonal	**5.41**	**5.42**	**5.44**	**5.45**	**5.49**	**5.52**

As we can see, the WGP model results are very similar to the single objective LP (Table 7.4). We can observe a small improvement for weights 4, 5 and 6 with the best results weight 4. The only approach that outperforms WGP is the average LP combination technique, nevertheless, WGP is a bit simpler because we need to solve only one linear program compared to the latter where we need to solve three (MinSAD, MinSAPE and MinMaxAD) and calculate the average of them. However, the differences are very small.

Table 7.11 shows the performance of the WGP according to the sMAPE. The results are similar. The best approaches are with weight 4, 6 and 5. Weight 6 is the best on the hard series, weights 1, 2, 3 and 4 are the best on the smooth series and weights 1 and 2 on the seasonal series. WGP outperforms simple LP overall, but the latter is better on the smooth and seasonal series. However, the differences are small.

Table 7. 11 Weighted goal programming combinations - sMAPE

	PERFORMANCE OF COMBINED FORECASTS					
sMAPE	**WEIGHTED GOAL PROGRAMMING**					
a2	**1**	**2**	**3**	**4**	**5**	**6**
All	**7.08**	**7.11**	**7.09**	**7.03**	**7.05**	**7.03**
Smooth	**3.03**	**3.03**	**3.03**	**3.03**	**3.04**	**3.05**
Hard	**23.16**	**23.31**	**23.13**	**22.76**	**22.79**	**22.59**
Seasonal	**5.36**	**5.36**	**5.38**	**5.39**	**5.44**	**5.46**

The performance according to MASE is found in Table 7.12. All the alternative weights perform the same overall and on the smooth series. On the seasonal series the best are weights 1, 2,3 and 4 and on the hard series weights 4, 5 and 6. WGP outperforms simple LP overall and performs as good as MinMaxAD on the hard series and as good as MinSAD and MinSAPE on the seasonal series. On the other hand, it performs slightly worse than MinSAD on the smooth series.

Table 7. 12 Weighted goal programming combinations - MASE

	PERFORMANCE OF COMBINED FORECASTS					
MASE	**WEIGHTED GOAL PROGRAMMING**					
a2	**1**	**2**	**3**	**4**	**5**	**6**
All	**0.81**	**0.81**	**0.81**	**0.81**	**0.81**	**0.81**
Smooth	**0.96**	**0.96**	**0.96**	**0.96**	**0.96**	**0.96**
Hard	**0.85**	**0.86**	**0.86**	**0.84**	**0.84**	**0.84**
Seasonal	**0.52**	**0.52**	**0.52**	**0.52**	**0.53**	**0.53**

In Table 7.13 we can see the percentage difference between the MAD of the combined forecasting technique and this of the best individual technique.

Table 7. 13 Weighted goal programming combinations - Difference between the best

	PERFORMANCE OF COMBINED FORECASTS					
%BEST	**WEIGHTED GOAL PROGRAMMING**					
a2	**1**	**2**	**3**	**4**	**5**	**6**
All	**4.91**	**5.31**	**5.42**	**5.20**	**5.54**	**5.64**
Smooth	**6.40**	**6.72**	**6.90**	**6.95**	**7.14**	**7.34**
Hard	**10.48**	**11.54**	**10.89**	**9.24**	**9.54**	**8.95**
Seasonal	**-0.84**	**-0.65**	**-0.24**	**-0.15**	**0.46**	**0.78**

The results are very similar to the single objective LP models (Table 7.7). Weight 1 gives the smallest average percentage difference between MAD; weights 6 and 4 give the smallest average percentage difference on the hard series.

Table 7.14 shows the percentage difference between the MAD of the worst technique on each series and this of the WGP combinations.

Table 7. 14 Weighted goal programming combinations - Difference between the worst

All Techniques	PERFROMANCE OF COMBINED FORECASTS					
%WORST	WEIGHTED GOAL PROGRAMMING					
a2	1	2	3	4	5	6
All	43.60	43.33	43.29	43.47	43.30	43.26
Smooth	42.72	42.50	42.39	42.35	42.20	42.05
Hard	22.23	21.49	22.01	23.29	23.08	23.49
Seasonal	57.01	56.92	56.72	56.67	56.48	56.40

By comparison with Table 7.8, the difference between the WGP approach and the single objective MinSAD, MinSAPE and average LP are insignificant.

Table 7. 15 Weighted goal programming combinations - Difference between the average

All Techniques	PERFORMANCE OF COMBINED FORECASTS					
%AVERAGE	WEIGHTED GOAL PROGRAMMING					
a2	1	2	3	4	5	6
All	26.27	25.94	25.89	26.09	25.88	25.84
Smooth	20.67	20.40	20.27	20.22	20.06	19.88
Hard	7.06	6.17	6.73	8.14	7.88	8.36
Seasonal	46.89	46.78	46.53	46.48	46.24	46.14

Table 7.15 shows the percentage difference between the average MAD of all the individual techniques on each of the series and the MAD of the WGP combinations. The results are similar with Table 7.13. There WGP approaches perform as good as the single objective LP and the differences are small.

A general conclusion about the WGP approaches is that they slightly improve the overall results of the three LP approaches as well as for each subgroup of the series separately. WGP performs slightly worse than the average LP model according to the MAPE, but it is better according to the other accuracy indices. In addition, the first approach requires solving one program only. WGP seem to be stable and have the same good performance according to all the performance measurement indices. Nevertheless, the differences between all the LP approaches are small.

7.6 CONCLUSION

The aim of the chapter was to examine the applicability of LP as a tool to combine forecasts. For this reason four simple linear programs and one weighted goal program was formulated (the latter was tested with alternative weights). The performance of these approaches was compared with traditional combination approaches that are found in the literature. The analysis shows that linear programming is a very good tool for combined forecasting. The simple LP methods outperform the traditional combined forecasting techniques that are found in the literature. On the other hand, the WGP methods outperform the former. Moreover, it seems that the distribution of the more and less accurate individual forecasts does affect the accuracy of traditional combination techniques, but not this of the LP based combinations. This is because while traditional techniques tend to distribute the weights over all the individual forecasts, LP approaches tend to select only the most accurate approaches.

The conclusions of this chapter are summarised as follows:

1. LP models outperform the traditional combination techniques in general.
2. LP is the only method for combination that outperforms all the individual techniques.
3. The inverse proportion to MSE and the weights based on the absolute error are the most accurate traditional combination techniques.
4. WGP performs slightly better than single objective LP.
5. All LP models, except MinMaxAD, are better than the traditional techniques on the smooth and seasonal series, but they give similar results on the hard series.
6. MinMaxAD gives the best results on the hard series.

LP is a very good method for combing forecast. The last step of the study is to test LP as a tool to minimise forecasting cost.

8 MINIMISING FORECASTING COST

The aim of the chapter is to create forecasts for situations with asymmetries in the cost of the underestimation error and the cost of the overestimation error. Such situations have been presented in detail by Kahn (2003) in his study about how to measure the impact of forecasting cost on an enterprise. In this study the forecasting cost is separated into *over-forecasting* cost and *under-forecasting* course. The over-forecasting cost consists of excess in inventory, inventory holding cost, transhipment cost, obsolescence and reduced margin. The under-forecasting cost consists of order expediting cost, higher product cost, lost sales cost, lost companion product sales and reduced customer satisfaction (Kahn, 2003). In his study, Kahn presents the example of an unnamed company where the potential excess inventory cost due to 1% over-forecasts is $86,613, and the potential lost profit due to 1% under-forecast is $85,473. On the other hand, the cost asymmetries in forecasting the demand of a blood bank or a hospital inventory are much bigger, because uncover demand for blood or drugs may cost the loss of human lives (Pereira, 2004, Drackley et al., 2011).

In addition to inventory management, asymmetries between underestimation and overestimation cost is very important in fiscal policy forecasting. Specifically, underestimation cost tends to be much more serious than overestimation, especial for conservative, stability-oriented governments (Bretschneider et al., 1989, Keereman, 1999, Jonung and Larch 2006). For this reason, Auerbach (1999) argues that the sum of forecasting errors should be very close to zero (no symmetric bias) only when the cost of forecasting errors is symmetric.

In order to deal with cost asymmetry situations I have develop two alternative approaches. The first are simple LP formulations with the cost relationship in the objective function. The second are simple LP formulations that minimise SAD, SAPE and ADBD with an additional cost constraint. We run the experiments for autoregressive models of order 6 and 12. The aim of this chapter is to examine the sensitivity of the cost of the forecasting error as the difference between the overestimation and underestimation cost grows larger. Thus, in the experiments I explore cases where the underestimation cost is bigger than the overestimation cost by factor 1.5, 2, 4, 7 and 10.

8.1 COST RELATIONSHIP IN THE OBJECTIVE FUNCTION

The first LP formulations minimise the total cost and relative cost for simple autoregressive models (AR) and autoregressive models with an additional, additive seasonal coefficients (ARS).

Let Y_i be the predicted variable, Y_{i-j} the explanatory variables, b_j the coefficient of Y_{i-j}, e_i the forecasting error, i ($1 \le i \le n$) the index of the forecasting period, j ($1 \le j \le m$) the order, e_{1i} the underestimation error and e_{2i} the overestimation error and S_i the seasonal coefficient. The formulation of the program for minimising the total cost for one period ahead forecasts is (assuming that Y_{i-j} is not defined for $j \ge i$):

$$Min \sum_{i=m+1}^{n} a_1 e_{1i} + \sum_{i=m+1}^{n} a_2 e_{2i} \qquad \textbf{(8. 1)}$$

Where a_1 and a_2 are the coefficients of the underestimation and overestimation error respectively.

subject to:

(4.27) for the simple AR

and

(4.29), **(4.30)** and **(4.32)** for the ARS

e_{1i}, e_{2i} non-negative and b_i and S_i unrestricted in sign.

The formulation of the program for minimising the relative cost for one period ahead forecasts is (assuming that Y_{i-j} is not defined for $j \ge i$):

$$Min \sum_{i=m+1}^{n} \left[\left(a_1 e_{1i} + a_2 e_{2i} \right) / Y_i \right] \qquad \textbf{(8. 2)}$$

subject to:

(4.27) for the simple AR

and

(4.29), **(4.30)** and **(4.32)** for the ARS

e_{1i}, e_{2i} non-negative and b_i and S_i unrestricted in sign.

I run experiments for a_1 = 1.5, 2, 4, 7 and 10 assuming that a_2 is always 1.

8.2 COST RELATIONSHIP IN THE CONSTRAINTS

These LP formulations minimise SAD, SAPE and ADBD for simple autoregressive model (AR) and autoregressive models with an additive seasonal coefficient (ARS). The cost relationship is expressed in the constraints.

Let Y_i be the predicted variable, Y_{i-j} the explanatory variables, b_j the coefficient of Y_{i-j}, e_i the forecasting error, i ($1 \leq i \leq n$) the index of the forecasting period, j ($1 \leq j \leq m$) is the order, e_{1i} the underestimation error and e_{2i} the overestimation error, a_1 and a_2 the coefficients of the underestimation and overestimation cost, and S_i the seasonal coefficient. The formulation of the program for minimising SAD for one period ahead forecasts is (assuming that Y_{i-j} is not defined for $j \geq i$):

$$MinSAD$$

(4. 24), **(4. 25)**, **(4. 26)**

subject to:

(4. 27)

$$\sum_{i=m+1}^{n} a_1 e_{1i} + \sum_{i=m+1}^{n} a_2 e_{2i} = 0 \qquad \textbf{(8. 3)}$$

e_{1i}, e_{2i} non-negative and b_j unrestricted in sign.

The formulation for the ARS model is:

$$MinSAD$$

(4. 28), **(4. 25)**, **(4. 26)**

subject to:

(4. 29), **(4. 30)**, **(4. 32)**, **(8. 3)**

e_{1i}, e_{2i} non-negative and b_j and S_i unrestricted in sign.

The formulation for minimising SAPE is similar.

Finally, the formulation of the AR minimising the ADBD for one period ahead forecasts is (assuming that Y_{i-j} is not defined for $j \geq i$):

$$MinADBD$$

(4. 33), **(4. 34)**

subject to:

(4. 35), **(4. 36)**

$$e_{ui} + e_{oi} = e_i \quad i = m+1, \ldots, n-1 \qquad \textbf{(8. 4)}$$

Where e_{ui} and e_{oi} are the underestimation and overestimation errors for period i

(8. 3)

e_{1il}, e_{2il}, e_{ui}, e_{oi} non-negative and b_j unrestricted in sign.

In the same way, the formulation of the ARS that minimises the ADBD for one period ahead forecasts is (assuming that Y_{i-j} is not defined for $j \geq i$):

$$MinADBD$$

(4. 33), **(4. 34)**

subject to:

(4. 35), **(4. 38)**, **(4. 30)**, **(4. 32)**, **(8. 3)**, **(8. 4)**

e_{1il}, e_{2il}, e_{ui}, e_{oi} non-negative and b_j unrestricted in sign.

8.3 RESULTS

I used the same data set (60 series) that was used in the previous chapters. I compare the results of the above models in terms of accuracy (MAPE) and cost. The cost is the sum of the underestimation and overestimation errors multiplied by the cost coefficients a_1 and a_2. The results are compared with these of the OLS approach. Cost coefficients cannot be adapted using the OLS method; thus, the results remain the same in terms of accuracy, but the cost changes according to the change in the underestimation cost coefficient.

Table 8.1 present the result of all the models (AR 6, AR 12, ARS 6, ARS 12) for underestimation cost coefficient equal to 1.5.

Table 8. 1 Under/overestimation cost ratio 1.5

1.5	OLS		LP Objective Function				LP Constraints					
			MinSAD		MinSAPE		MinSAD		MinSAPE		MinADBD	
AR 6	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST
All	8.69	630777.03	9.51	627583.84	9.06	650564.32	9.24	620726.13	8.81	681513.93	9.26	621496.36
Smooth	3.14	185639.33	3.45	188007.82	3.24	179856.27	3.40	187257.80	3.10	191442.26	3.39	187292.85
Hard	24.47	231799.41	26.99	224994.56	25.58	250433.02	26.01	221371.25	25.84	258953.93	26.01	221915.53
Seasonal	9.79	213338.29	10.57	214581.46	10.24	220275.03	10.31	212097.08	9.50	231117.74	10.39	212287.98
AR 12												
All	8.01	632503.71	9.11	571621.39	8.86	585215.21	8.83	565442.66	9.05	572009.33	9.23	582056.64
Smooth	3.22	187366.00	3.38	184030.47	3.21	179339.51	3.39	185867.13	3.30	183569.01	3.45	187732.96
Hard	23.69	231799.41	28.62	233372.19	28.17	253939.72	27.27	227064.19	28.97	235244.94	25.09	210360.59
Seasonal	7.80	213338.29	8.43	154218.73	8.16	151935.99	8.26	152511.34	8.21	153195.38	10.69	183963.10
ARS 6												
All	7.39	540101.24	8.08	532343.91	7.76	551706.25	7.82	525100.07	7.64	567942.90	7.85	530367.42
Seasonal	5.45	122662.50	5.80	119341.53	5.89	121416.96	5.56	116471.02	5.61	117546.71	5.69	121159.04
ARS 12												
All	7.22	530239.74	8.17	521380.66	8.00	537381.74	7.88	513803.40	8.11	520285.07	7.55	498749.42
Seasonal	5.19	111074.33	5.32	103978.00	5.32	104102.51	5.07	100872.08	5.06	101471.11	5.09	100655.88

As we can see, the differences in both cost and accuracy are relative small. Typically, the ARS 12 outperform the ARS 6 both in accuracy and cost, whereas, ARS 12 performs better than the AR 12 model and the ARS 6 performs better than the AR 6.

OLS is the best in accuracy. However, in most cases, the cost obtained through the LP model is lower than the cost obtained through the OLS approach. The MinSAPE with the cost relation in the objective function performs well on the smooth series; while the MinADBD with the cost in the constraints performs well on the hard series.

The overall best performing models, in terms of cost, are the LP approaches with MinSAD objective and cost constraint (for AR 6, AR 12 and ARS 6) and the LP approach with MinADBD with cost constraint (for ARS 12). We can conclude that cost and accuracy differences between all the models for underestimation cost weight 1.5 are small; this was expected because the difference in overestimation and underestimation cost weights is small (0.5).

Table 8. 2 Under/overestimation cost ratio 2

2	OLS		LP Objective Function				LP Constraints					
			MinSAD		MinSAPE		MinSAD		MinSAPE		MinADBD	
AR 6	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST
All	8.69	758333.09	10.59	728623.44	9.53	738842.82	9.78	723059.35	9.84	721617.92	9.73	716334.72
Smooth	3.14	218265.43	3.73	217087.49	3.47	207025.56	3.58	213587.25	3.52	212743.87	3.57	213190.22
Hard	24.47	285836.87	29.67	261049.21	26.32	282822.14	27.52	261250.55	28.17	262320.96	27.26	258932.82
Seasonal	9.79	254230.79	12.18	250486.74	10.96	248995.13	10.97	248221.55	10.89	246553.09	10.93	244211.68
AR 12												
All	8.01	661821.13	9.65	657875.73	9.12	669074.48	9.13	649048.85	9.34	657832.19	9.43	649402.56
Smooth	3.22	218692.40	3.76	216493.77	3.44	206476.63	3.56	211126.57	3.42	207128.57	3.61	212521.64
Hard	23.69	275121.16	29.72	268901.78	28.17	293978.27	27.98	270135.45	29.71	280660.96	25.48	242557.83
Seasonal	7.80	168007.57	8.98	172480.18	8.63	168619.59	8.54	167786.83	8.54	170042.66	10.86	194323.08
ARS 6												
All	7.39	649743.85	8.81	617640.37	8.11	628252.26	8.25	607990.00	8.34	610858.74	8.21	609811.43
Seasonal	5.45	145641.55	6.26	139503.68	6.23	138404.56	5.85	133152.21	5.89	135793.91	5.87	137688.39
ARS 12												
All	7.22	624019.59	8.71	605370.90	8.23	618339.63	8.17	597515.16	8.39	605251.32	7.74	567245.37
Seasonal	5.19	130206.02	5.85	119975.35	5.65	117884.74	5.35	116253.14	5.36	117461.79	5.25	112165.90

Table 8.2 present the results for underestimation cost coefficient equal to 2. As expected, the accuracy of the LP approaches decreases; while, the cost differences of between OLS and LP approaches is more significant. Specifically, the forecasting cost with the LP approaches is for all but one case smaller than with the counterpart OLS.

Table 8. 3 Under/overestimation cost ratio 4

4	OLS		LP Objective Function				LP Constraints					
			MinSAD		MinSAPE		MinSAD		MinSAPE		MinADBD	
AR 6	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST
All	8.69	1268557.32	13.21	996174.33	11.67	986388.94	11.23	996727.94	11.32	1001286.44	11.26	981725.78
Smooth	3.14	348769.83	4.76	305377.45	4.49	291542.02	4.16	294001.88	4.06	289020.94	4.24	292225.08
Hard	24.47	501986.73	38.13	387071.50	31.13	392084.54	32.07	378819.73	32.97	387689.57	31.35	362413.22
Seasonal	9.79	417800.76	14.39	303725.38	13.62	302762.38	12.23	323906.32	12.21	324575.93	12.58	327087.48
AR 12												
All	8.01	1073051.41	11.59	903676.39	10.73	942650.62	9.97	899740.91	10.14	924038.68	10.37	855617.25
Smooth	3.22	343998.01	4.76	306872.26	4.37	294329.35	4.10	296335.79	3.93	282612.15	4.23	288552.03
Hard	23.69	489947.23	34.31	374851.66	31.30	435159.30	29.42	392837.63	30.89	427179.46	27.56	343452.50
Seasonal	7.80	239106.17	11.11	221952.48	10.61	213161.98	9.60	210567.49	9.65	214247.08	11.74	223612.72
ARS 6												
All	7.39	1088314.33	11.27	884641.56	9.88	874152.83	9.55	864940.20	9.62	866289.88	9.41	837291.57
Seasonal	5.45	237557.77	7.91	192192.61	7.65	190526.27	6.61	192118.58	6.52	189579.37	6.42	182653.27
ARS 12												
All	7.22	1040678.05	10.50	847943.26	9.61	883961.77	8.88	845677.77	11.14	1070042.30	8.60	776667.77
Seasonal	5.19	206732.82	7.47	166219.34	6.87	154473.12	5.96	156504.35	12.97	360250.69	5.83	144663.24

The best overall performing (in terms of cost) LP approaches are the LP models with objectives MinSAD or MinADBD and cost constraint. As before, the MinADBD performs well on the hard series, and the MinSAPE model with cost in the objective function performs well on the smooth series.

In Table 8.3 we can see the results for underestimation cost coefficient equal to 4. The accuracy of the LP approach decreases further, while the cost difference between the LPs and the OLS grows larger. MinADBD gives the best results overall. It is the best approach on the hard series and seasonal series and gives good results on the smooth series. MinSAPE with cost in the constraints gives the lower cost on the smooth series; however, it gives the worst results overall.

Table 8.4 show the situations where the underestimation cost is seven times bigger than the overestimation cost. The cost difference between OLS and LP is very big. In addition, the cost differences between the LP models are rather small. MinADBD typically outperforms the other models in cost and accuracy.

Table 8. 4 Under/overestimation cost ratio 7

7	OLS		LP Objective Function				LP Constraints					
			MinSAD		MinSAPE		MinSAD		MinSAPE		MinADBD	
AR 6	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST
All	8.69	2033893.66	16.01	1235475.48	14.03	1230026.72	12.67	1279857.02	12.69	1279484.79	12.85	1239668.18
Smooth	3.14	544526.42	5.89	389445.72	5.49	364682.09	4.79	380745.85	4.65	376620.38	4.96	378084.13
Hard	24.47	826211.51	46.58	507572.50	38.37	513604.29	36.60	519960.04	37.04	519353.38	35.74	472732.79
Seasonal	9.79	663155.73	17.04	338457.25	15.68	351740.34	13.39	379151.13	13.44	383511.03	14.17	388851.25
AR 12												
All	8.01	1689896.82	13.47	1138631.15	12.77	1213079.02	11.00	1182363.34	11.58	1247161.66	11.51	1079780.56
Smooth	3.22	531956.42	5.76	409217.43	5.53	395259.52	5.37	407190.61	5.23	389100.98	4.92	368244.73
Hard	23.69	812186.32	38.99	471452.90	36.26	555923.99	29.68	520384.64	33.49	604418.35	30.26	459974.50
Seasonal	7.80	345754.07	12.99	257960.82	12.60	261895.51	10.63	254788.09	10.68	253642.34	12.80	251561.32
ARS 6												
All	7.39	1746170.04	13.42	1121177.56	11.87	1111083.44	10.83	1153831.27	10.83	1150203.84	10.70	1075320.26
Seasonal	5.45	375432.10	8.38	224159.34	8.49	232797.06	7.24	253125.38	7.24	254230.09	7.01	224503.33
ARS 12												
All	7.22	1665665.75	12.12	1093851.19	11.50	1157120.67	9.77	1129843.81	10.33	1199090.18	9.77	1052722.56
Seasonal	5.19	321523.00	8.49	213180.86	8.36	205937.15	6.53	202268.56	6.51	205570.85	7.01	224503.33

Finally, when the underestimation cost is ten times bigger than the overestimation, the cost differences between OLS and LP are very large. In addition, the cost differences between all the LP models are small. The models with the cost function in the constraints give significantly smaller MAPE.

Figures 8.1, 8.2, 8.3 and 8.4 show the relationship between the cost of the forecasting error and the cost coefficient relationship on all the series (Figure 8.1) and for each series subgroup (Figures 8.2, 8.3 and 8.4). The figures show the performance of the ARS model, however, the performance of the other three models is similar.

Table 8. 5 Under/overestimation cost ratio 10

10	OLS		LP Objective Function				LP Constraints					
			MinSAD		MinSAPE		MinSAD		MinSAPE		MinADBD	
AR 6	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST	MAPE	COST
All	8.69	2799230.00	17.78	1404878.71	16.09	1408891.93	13.81	1524002.96	13.89	1532374.21	14.03	1434313.62
Smooth	3.14	740283.02	6.51	465802.40	6.33	427863.99	5.20	455662.48	5.13	452174.14	5.50	446975.07
Hard	24.47	1150436.29	52.44	574613.60	44.29	620558.13	40.47	644572.61	41.05	652598.73	39.05	563616.25
Seasonal	9.79	908510.69	18.54	364462.71	17.77	360469.81	14.31	423767.87	14.37	427601.35	15.31	423722.30
AR 12												
All	8.01	2306742.23	14.94	1374863.76	13.87	1378491.82	11.58	1413213.11	12.21	1483649.54	10.38	1424162.47
Smooth	3.22	719914.83	6.69	497435.46	6.34	473995.12	5.81	491870.52	5.72	463540.58	4.25	463512.14
Hard	23.69	1134425.42	42.99	599196.12	38.83	611942.41	30.55	638144.05	34.46	734216.11	27.56	671839.02
Seasonal	7.80	452401.97	14.04	278232.18	13.38	292554.28	11.29	283198.54	11.40	285892.85	11.74	288811.31
ARS 6												
All	7.39	2404025.74	15.26	1275442.95	13.60	1306102.28	11.83	1408096.97	11.83	1406873.29	11.68	1262918.95
Seasonal	5.45	513306.43	10.15	235026.95	9.47	257680.16	7.69	307861.88	7.49	302100.43	7.46	252327.63
ARS 12												
All	7.22	2290653.45	13.58	1337325.29	12.66	1314417.17	10.28	1365791.14	10.84	1426254.61	8.61	1368696.70
Seasonal	5.19	436313.19	9.49	240693.70	9.34	228479.63	6.98	235776.57	6.84	228497.91	5.83	233345.55

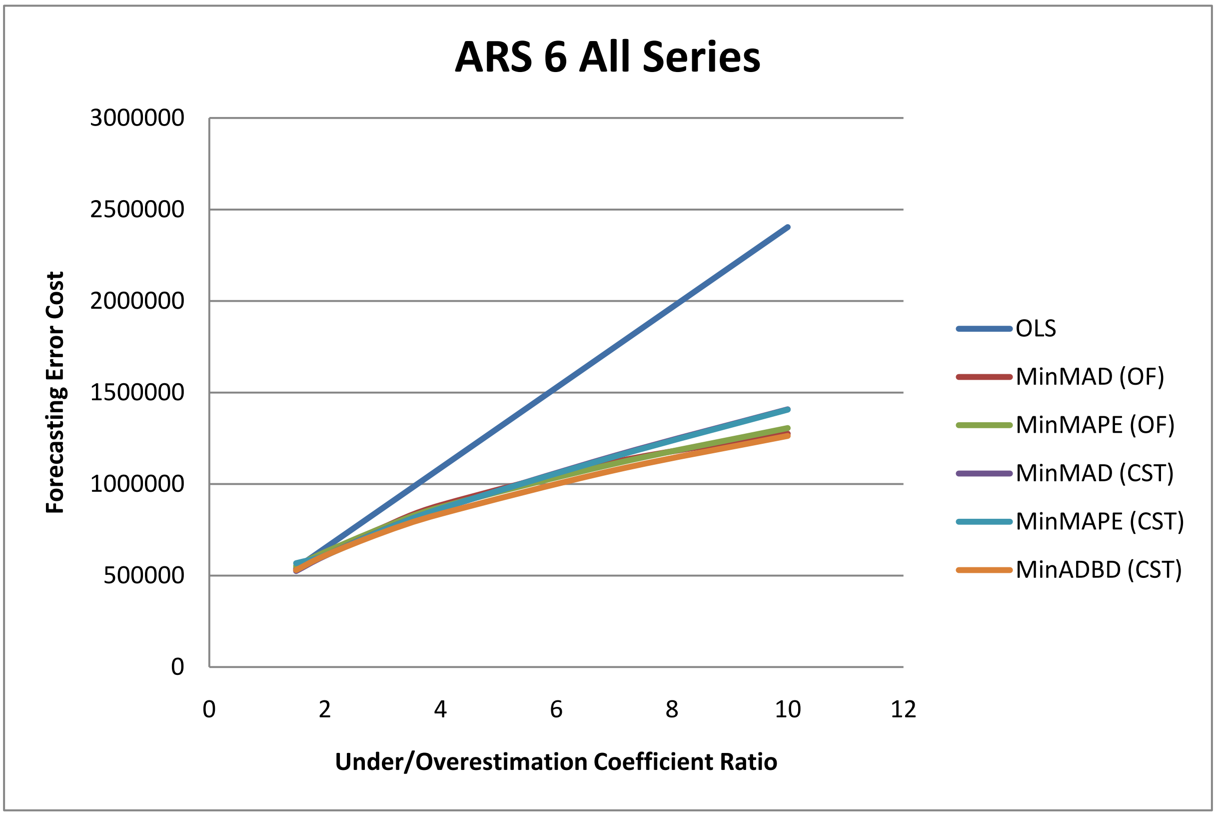

Figure 8. 1 Cost performance (all series)

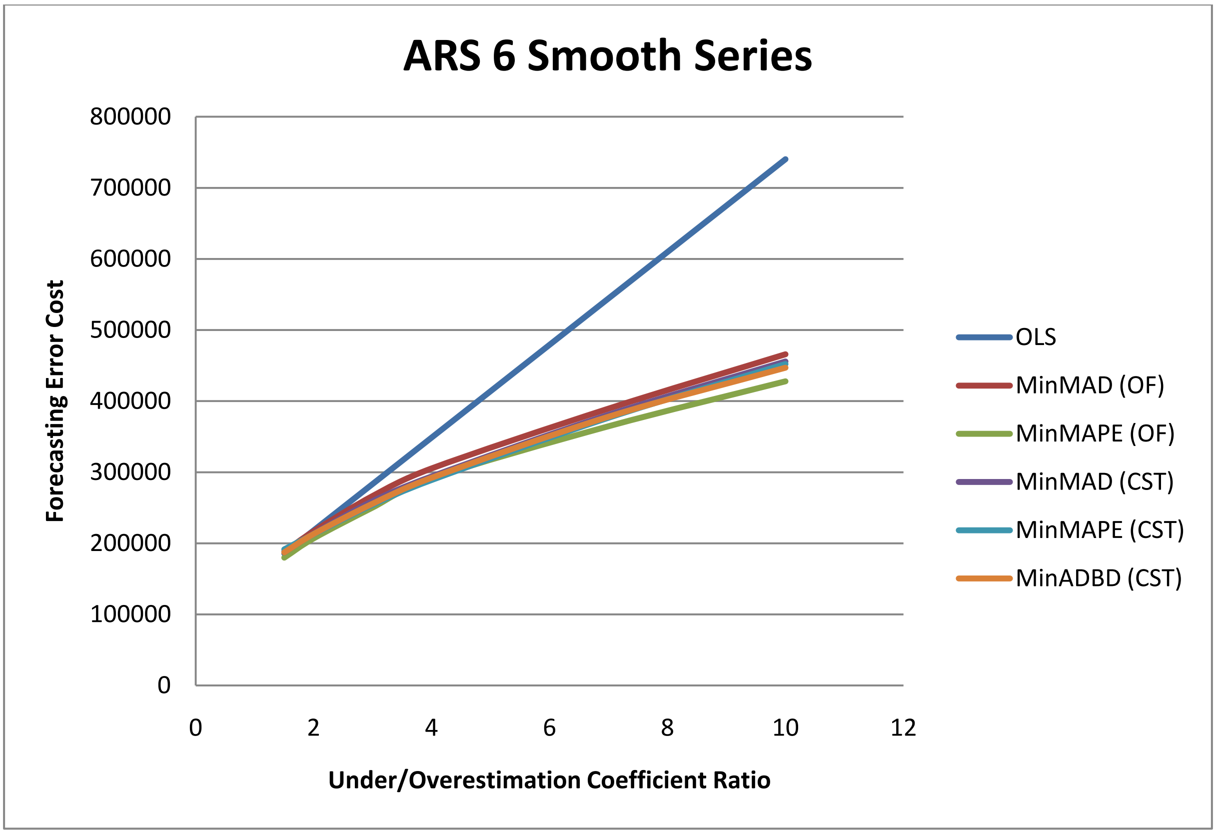

Figure 8. 2 Cost performance (smooth series)

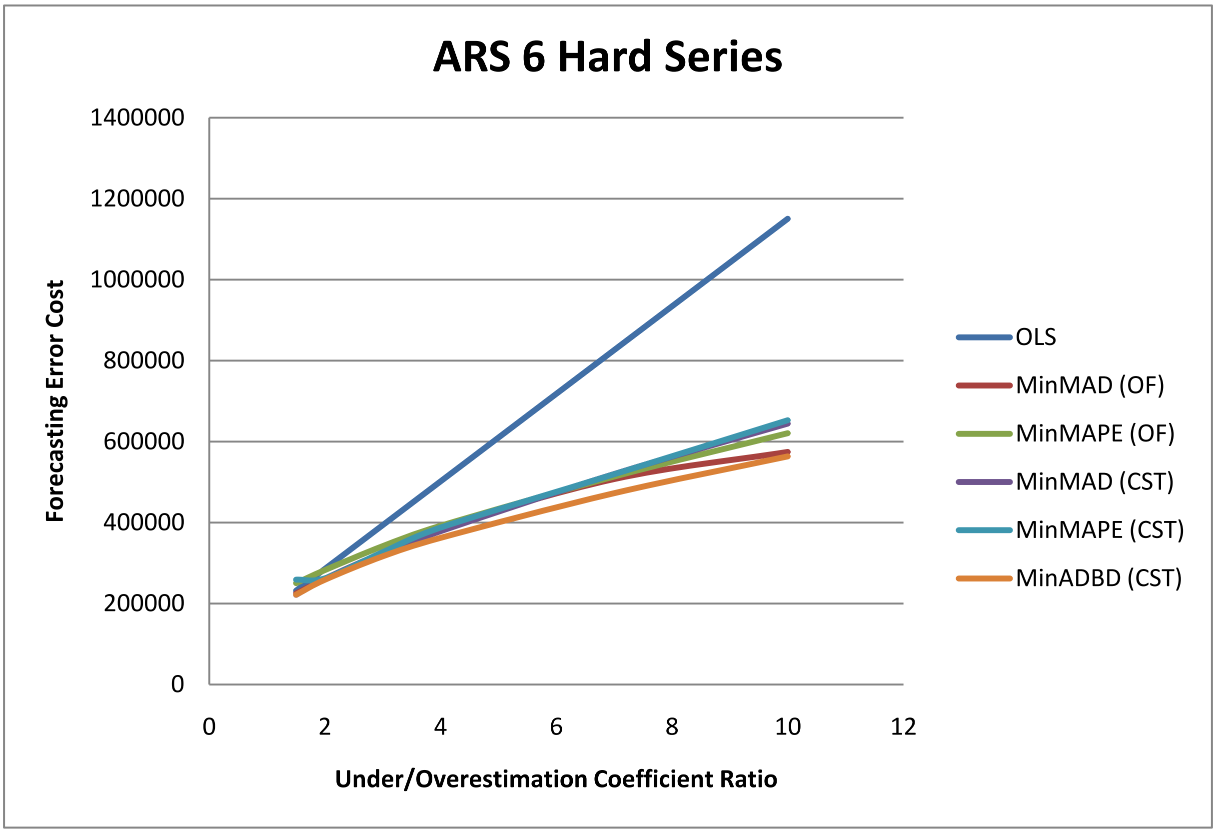

Figure 8. 3 Cost performance (hard series)

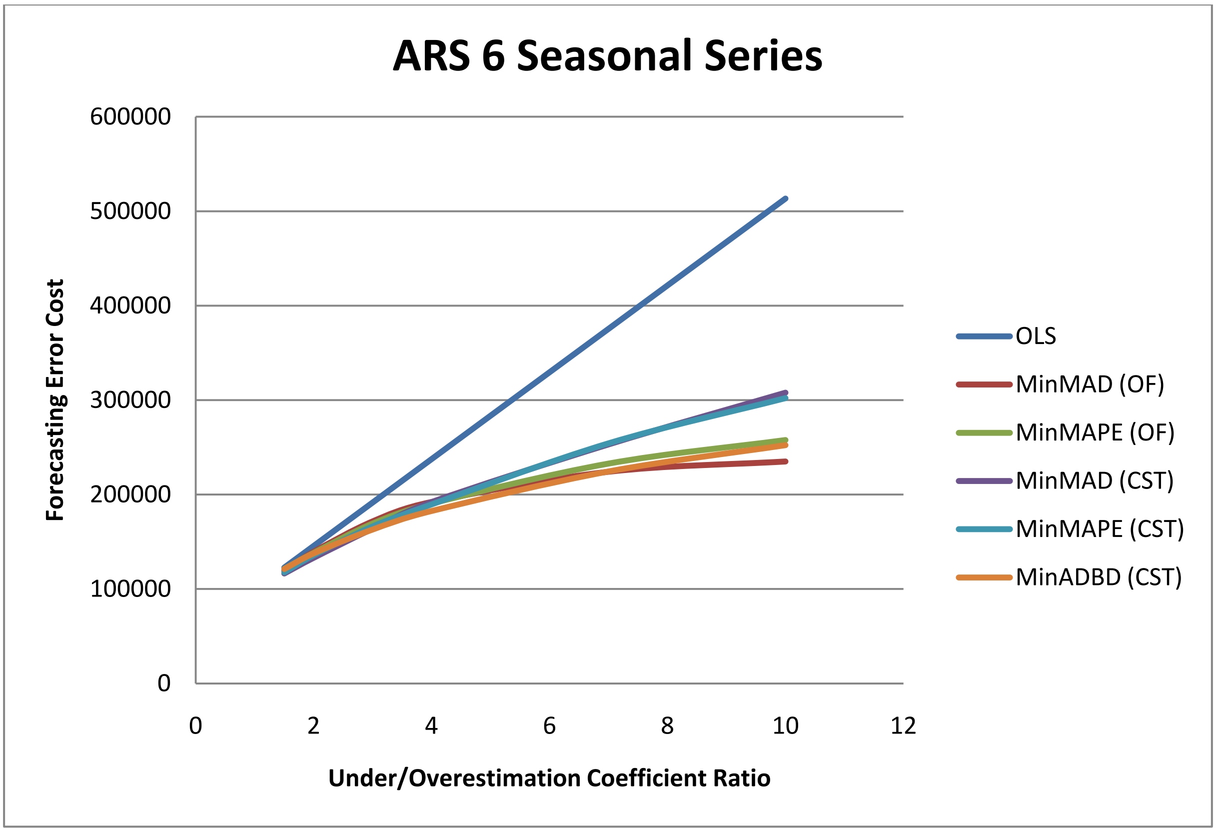

Figure 8. 4 Cost performance (seasonal series)

As we can see, the cost grows larger as the under/overestimation cost coefficient ratio grows bigger. This relationship is linear for the OLS and damped for the LP models, and the bigger the cost ratio is, the greater the difference between the cost performance of the OLS and the LP models. MinADBD seems to be the best model overall and on the hard series. MinSAPE with cost in the objective function performs better on the smooth series. On the seasonal series, MinADBD performs better for smaller cost coefficient ratio, while MinSAD with cost in the objective function performs better for bigger cost coefficient ratio.

8.4 CONCLUSIONS

The aim of the chapter was to develop linear programs that estimate the parameters of autoregressive forecasting models that incorporate asymmetries in the error of forecasting cost. There are two alternative formulations. The first minimise has the cost asymmetry relationship in the objective function (cost minimisation), whereas in the second it is expressed as a constraint. The LP approaches are tested for situations where the underestimation cost is slightly or significantly bigger than the overestimation cost. The performance of the LP approaches is compared with the OLS in terms of cost and accuracy.

In summary, we can say that LP is a good tool for making forecasts with cost asymmetries. The differences between the OLS and LP approaches are small, when the cost asymmetry is low; however, the differences are bigger when the cost asymmetry is more important. In addition, it seems that models with the cost relationship in the constraints perform better than those with the cost relationship in the objective function, both in cost and accuracy. The MinADBD tends to give better results compared with the other LP formulations.

9 CONCLUSION

9.1 SUMMARY

In this research I explored the usage of linear programming as a tool to optimise time series forecasting models. The performance of LP is validated by examining five potential applications.

First of all, I compared the accuracy of single objective linear programming for optimising the parameters of autoregressive models. The comparison shows that LP is a very good alternative to the OLS. Specifically, LP performs better than the OLS on series with low variability and seasonal series, while OLS performs better on series with high variability. However, the accuracy of the LP approaches can be improved by adding constraints where the sum of the errors should be zero, or the sum of the percentage errors should be zero, in order to remove the bias. The second approach seems to improve the results more, especially on smooth and seasonal series; however, the first seemed to work better on the hard series. The general conclusion of this comparison is that according to the characteristics of the series, the decision maker may have to focus on different optimisation objectives. Nevertheless, the differences between all the models are small.

Secondly, I examined the performance of goal programming based approaches, comparing them with the single objective LP and the OLS. The experiments show that goal programming improves the performance of the single objective approaches. Specifically, weighted goal programming formulations seems to improve the performance of simple LP overall, while, pre-emptive goal programming can improve the results on series with specific characteristics. The additional constraints (sum errors or sum percentage errors equal to zero) improve the results in the same way as in the simple LP models. In general, the decision maker should select the most appropriate method (simple LP, GP or OLS) according to the characteristics of the series and the forecasting horizon.

One of the main conclusions so far is that LP performed worse than the OLS on hard series. This was expected because OLS minimises the sum of squared errors; hence it gives more weight to bigger errors, while LP requires linear errors measures. For this reason, the next step was to explore whether or not the flexibility of LP can improve the performance on series with high variability. I tried to improve the single objective and goal programming approaches with one or two additional constraints. Firstly, by setting all the coefficients of the LPs as positive and secondly by adding a hierarchy relation. I compare the accuracy of the new models with the OLS, as well as with four other techniques that traditionally perform well on these series. The experiments show that the additional constraints significantly improve the performance of the initial models and outperform all the other techniques. Specifically, the best LP-based model performs better (reduction in MAPE by 3%) than the best model of all the other traditional techniques.

After testing the performance of LP as a tool for developing individual forecasts, I examined LP for combined forecasting. I developed single objective models (minimising SAP, SAPE and MaxAD) and WGP models (minimising both SAP and MaxAD) and I compared them with five traditional combination techniques. The results show that the LP approaches outperform all the traditional combination methods as well as the best individual technique. In more detail, the conclusion is that WGP performs better than the simple LP models (except the average LP approach); the approach that minimises SAPE performs better on the smooth series, the approach that minimises SAP is the best on the seasonal series and the one that minimises MaxAD is the best on the hard series.

Finally, LP is tested as a tool to minimise forecasting cost, instead of forecasting error. I applied the initial autoregressive forecasts, adding the cost relationship in the objective function or as a constraint and I compared the results with the OLS. The experiments show that the LP approaches perform significantly better, in terms of cost, than the latter. Moreover, it seems that the approaches with the cost relationship in the constraints outperform these with the cost relationship in the objective function, both in cost and accuracy.

The general conclusion is that LP is a very useful tool that can be used to develop accurate time series forecasts. Moreover, LP approaches are easy and in many cases perform better than the traditional approaches that are found in forecasting literature and practise. In addition, this study gave more general conclusions on the field of forecasting. Firstly, multi-objective approaches seem to outperform the counterpart single objective models. In addition, it seems that the performance of a quantitative forecast can be improved on the test set if sum of percentage errors is set the equal to zero on the training set. Finally, according to the specific characteristics of the forecasting problem, we may have to focus on models with different objectives.

9.2 IMPLICATIONS FOR FORECASTING THEORY

As it is mention in the first chapter, this research is rather a theoretical study than a real world application. Hence, the topic is interesting from a theoretical and mathematical perspective. It is the first time that linear programming was used to estimate the parameters of an autoregressive model and the analysis showed that it can be used for this purpose and that it is a good alternative to the methods that were used so far.

LP can help decision makers when they want to estimate the parameters of simple autoregressive models. The approaches presented in Chapters 4 and 5 should be preferred for series with low variability and these presented in Chapter 6 for series with high variability. In addition, LP was shown to be a very good tool for adding seasonal coefficients for forecasting series with strong seasonality. The analysis shows that in these situations the LP approaches perform better than the OLS or the other techniques that are found in the literature.

There is also a significant contribution to the field of combined forecasting. The LP based approaches for combined forecasts have a simple formulation and require short computational time (less than a second using any LP optimisation software). Thus, they can be easily implemented by decision makers. In addition, LP outperformed all the traditional combination techniques that are found in the literature.

There are three general implications for the field of forecasting (which are neither limited to LP models nor to estimating the parameters of autoregressive models and combining forecasts). First is the addition of logical constraints to the forecasts. The analysis shows that adding constraints, which sacrifice the performance on the training set, may improve the performance of the test set. Specifically, adding the constraint where the sum of percentage errors is equal to zero, improves the accuracy on series with low variability; on the other hand, setting non-negative and adding a hierarchy on the coefficients of the model improves the accuracy on series with high variability. Second is the application of multi-objective forecasting. The analysis shows that a MinSum-MinMax approach for forecasting tends to outperform a single objective approach. Finally, there is the implementation of biased forecasts, when the decision maker prefers underestimation to overestimation or vice versa. Linear programming was shown to be a powerful technique for dealing with asymmetries in the forecast error costs. It would be interesting to further verify these findings/issues in future research.

9.3 LIMITATIONS

The main limitation of LP for time series forecasting is the need of linearity. LP approaches cannot be used to estimate the parameters of models that are not linear. LP can be used only for linear regression based forecasts, either autoregressive, or causal. These LP approaches cannot be extended to estimate the parameters of ARMA or ARIMA forecasting models. Similarly, they cannot be extended to estimate the parameters of seasonal autoregressive models with multiplicative instead of additive seasonal coefficients. In addition, LP cannot be used to estimate the parameters of other forecasting techniques that do not have a linear mathematical formulation, such as the smoothing parameters of exponential smoothing models (SES, Holt, Holt-Winters etc.).

In addition, the minimisation objectives should be linear. LP cannot be applied for quadratic or non-linear accuracy measures, such as the Sum of Squared Errors, the Root of the Sum Squared Errors and the Root of the Sum of Squared Percentage Errors. Similarly, the approaches that deal with situation with asymmetry in the cost of the forecasting error cannot be applied to cases where the cost is quadratic or non-linear.

9.4 RECOMMENDATIONS FOR FURTHER RESEARCH

During this study several areas that deserve further investigation are identified. The potential topics for further research can be classified into two main categories. These are the extensions/improvements of the current models and the application of the existing approaches to different areas of forecasting.

About the first part, there are several potential ways that could improve the LP-based approaches. To begin with, LP formulations, due to their flexibility, could be extended by incorporating the knowledge of experts in additional constraints. In this way, a qualitative forecasting element can be added to the initial autoregressive approaches, which could further improve the forecasts. On the other hand, goal programming models that minimise only the SAP and the MaxAD have been developed. Many different combinations of objectives can be examined, using accuracy indices such as the ADBD and the MaxADBD.

Another promising area is the LP approach for combination. It was shown that in most situations adding the constraint where the sum of errors or the sum of percentage errors is equal to zero improves the accuracy of the models significantly. However, as it was mentioned in Chapter 7, the LP models for combined forecasting with these extra constraints were over-constrained and had in most cases no feasible solution. Hence, the possibility to relax these constraints can be examined.

A very interesting area is the development of non-linear programming approaches. It was shown that linear programming performs well for time series forecasting. The next step could be to try to exceed some of the LP limitations by using non-linear modelling. Non-linear programming will give the opportunity to extend the autoregressive models to ARMA or ARIMA formulations and compare their accuracy with the Maximum Likelihood Estimation and several other methods that are found in the literature. In addition, non-linear formulations will give the chance to develop multi-objective approaches where the weights of the goals are not selected by the decision maker, but they are determined by the programs. This could improve the accuracy of the WGP approaches further. Moreover, with non-linear programming cost minimisation models with quadratic cost functions can be explored.

About the second part, linear programming is used only for estimating the parameters of autoregressive models and for combining forecasts. The same LP approach could be used for causal modelling as well as for time series decomposition. In addition, an LP-based method could be useful for model selection.

Another interesting area for further research is the applications of multi-objective modelling for forecasting. This study shows that multi-objective approaches tend to be more accurate than single objective models. It seems very promising to explore the field of multi-objective optimisation in more depth, using more sophisticated techniques, such as artificial neural networks and genetic algorithms.

10 APPENDIX

As mentioned in Chapter 7, it is observed that the ARS method always gave the best results on the seasonal series and that it was significantly better than the other techniques. In order to remove the impact of the ARS model, I ran the experiments again ignoring the ARS method.

Figure A.1 shows the number of times each individual forecasting method gives the best results, but ignoring the ARS model.

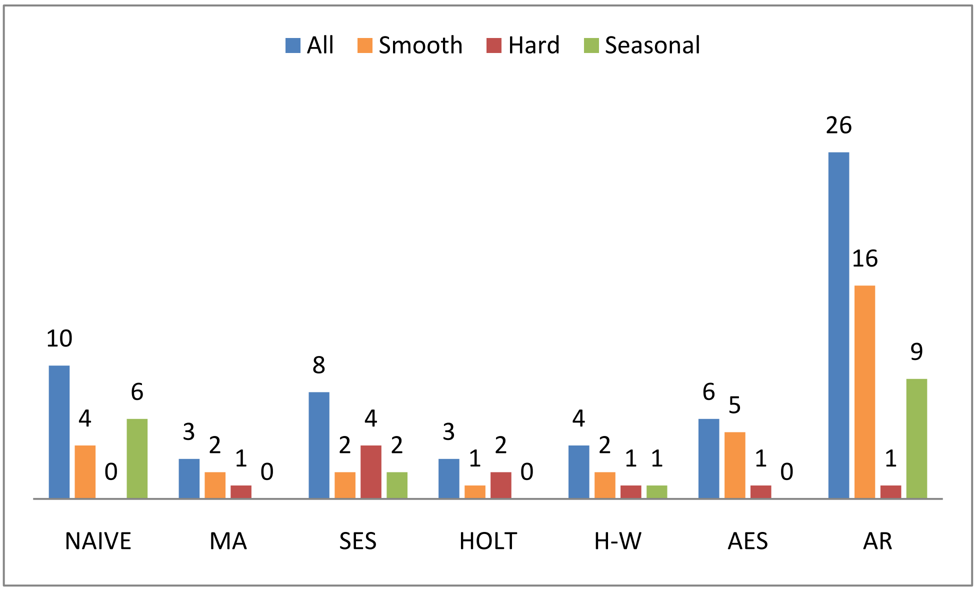

Figure 10. 1 Number of times each individual technique gives the best results (no ARS)

As we can see, AR gives the best results most of the times and on the smooth and seasonal series. On the other hand, SES gives most of the times the best results on the hard series. The number the individual forecast give the worst results can be found in Figure 7.2. AES and AR approaches do not perform worst on any of the series. SES is the technique that performs worst most times.

10.1 SIMPLE LP

Table 10.1 shows the average MAPE of the single objective LP combined forecasting models compared with the other five traditional methods.

As we can see, the performance of all the combined forecasting techniques and the differences between the LP and the other techniques remain similar. The Average LP remains the technique with the lowest MAPE. Significant differences are that all the LP models (and MinMaxAD) give better results than the other methods, as well as better than the best individual (that is the simple AR). In addition, the average inverse proportion to MSE method is the most accurate compared with the other traditional techniques.

Table 10. 1 Traditional and single objective LP combinations - MAPE

no ARS	PERFORMANCE OF COMBINED FORECASTS									
MAPE	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	9.40	9.15	9.14	8.95	9.08	9.04	8.55	8.55	8.65	8.47
Smooth	3.62	3.44	3.45	3.28	3.39	3.46	3.14	3.06	3.35	3.12
Hard	24.12	23.79	23.74	23.57	23.70	24.02	23.95	24.28	23.12	23.43
Seasonal	11.48	11.17	11.16	10.90	11.07	10.63	9.60	9.58	10.02	9.66

Table 10.2 shows the performance of the combination techniques according to the sMAPE.

Table 10. 2 Traditional and single objective LP combinations - sMAPE

no ARS	PERFORMANCE OF COMBINED FORECASTS									
sMAPE	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	9.08	8.86	8.86	8.67	8.80	8.75	8.35	8.38	8.41	8.27
Smooth	3.52	3.35	3.36	3.19	3.30	3.36	3.04	2.97	3.27	3.03
Hard	23.30	23.04	23.00	22.84	22.96	23.16	23.36	23.76	22.36	22.83
Seasonal	11.07	10.80	10.79	10.54	10.71	10.32	9.45	9.45	9.81	9.49

As we can see, the results are similar as in the comparison according to MAPE. LP formulations outperform the traditional techniques. Average LP gives the best results overall, MinSAPE performs best on the smooth series, MinMaxAD on the hard and MinSAD and MinSAPE on the seasonal.

Table 10. 3 Traditional and single objective LP combinations - MASE

no ARS	PERFORMANCE OF COMBINED FORECASTS									
MASE	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	1.03	0.99	0.99	0.96	0.98	0.98	0.91	0.91	0.99	0.92
Smooth	1.14	1.06	1.07	1.02	1.05	1.07	0.96	0.96	1.07	0.97
Hard	0.86	0.85	0.85	0.84	0.85	0.86	0.86	0.87	0.84	0.85
Seasonal	0.94	0.92	0.92	0.91	0.92	0.90	0.86	0.86	0.92	0.87

Table 10.3 shows the performance of the techniques according to MASE. LP outperforms the traditional techniques, except the MinMaxAD which is outperformed by inverse proportion to MSE, the average inverse proportion and the weighted according to the absolute deviations. MinMaxAD

gives the best results on the hard series. MinSAD and MinSAPE give the best results overall, and on the smooth and seasonal series.

In Table 10.4 we can see the percentage difference between the MAD of the combined forecasting method and the MAD of the best individual.

Table 10. 4 traditional and single objective LP combinations - Difference between the best

no ARS	PERFORMANCE OF COMBINED FORECASTS									
%BEST	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	21.94	16.20	16.24	12.20	14.84	15.08	5.60	5.68	14.64	6.65
Smooth	28.53	19.86	20.02	14.06	17.92	19.86	6.53	6.25	20.24	8.31
Hard	12.53	10.66	10.50	9.39	10.16	11.64	11.13	12.80	10.01	10.13
Seasonal	15.44	12.77	12.71	10.47	11.95	8.49	0.87	0.70	7.25	1.77

The results of the table are slightly different compared with those in Table A.1 MinSAD, MinSAPE and average LP give the lowest average percentage difference and perform significantly better on the smooth and seasonal series; however, the differences on the hard series remain the same. A general observation is that while the total average percentage difference in the *LP* models is slightly bigger, it is significantly smaller compared with the other methods of combination. More specifically, this difference is observed in the subgroup of the seasonal series.

Hence, it seems that when there is a dominant individual forecasting technique (ARS on the seasonal series) the traditional models of combination result in significantly bigger errors compared with the best technique; whereas, when the best forecast is distributed over different individual forecasting methods, the percentage deviation over the best technique is smaller. When there is a dominant individual forecasting technique, the LP models tend to have smaller (or negative) percentage differences over the best individual technique compared with the case when the best forecast is distributed over different individual techniques.

Table 10.5 show the percentage differences between the MAD of the worst technique and this of the combined forecasts. Similar as in the previous table, the results of the last table are slightly different compared with those of Table A.4: MinSAD, MinSAPE and average LP perform better; however, the difference between the LP models and the other techniques are slightly smaller. This difference is again observed on the seasonal series. Thus, it seems that the absence of a dominant individual technique tends to increase the performance of the LP combinations methods a bit. The average percentage difference between the worst technique and the combined forecasts remains bigger for the LP models compared to the traditional techniques.

Table 10. 5 Traditional and single objective LP combinations - Difference between the worst

no ARS	PERFORMANCE OF COMBINED FORECASTS									
%WORST	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	29.19	31.37	31.38	32.86	31.89	31.90	35.16	35.11	30.45	34.72
Smooth	35.48	38.50	38.47	40.40	39.15	38.64	42.70	42.94	36.11	42.00
Hard	21.75	22.63	22.76	23.37	22.93	22.07	21.78	20.43	22.92	22.55
Seasonal	22.14	23.54	23.56	24.72	23.96	25.38	29.17	29.36	24.59	28.54

Table 10.6 shows the percentage differences between the average MAD of the individual forecasting techniques on a series and this of the combined forecasts.

Table 10. 6 Traditional and single objective LP combinations - Difference between the average

no ARS	PERFORMANCE OF COMBINED FORECASTS									
%AVERAGE	SIMPLE AVERAGE	INVERSE PROPORTION				WEIGHTS BASED ON AD	SIMPLE LP			
SERIES		MAD	MAPE	MSE	AVERAGE		MinSAD	MinSAPE	MinMaxAD	AVERAGE
All	9.39	12.75	12.73	15.00	11.68	13.41	18.52	18.47	12.12	13.52
Smooth	10.71	15.59	15.51	18.67	14.01	15.72	22.62	22.87	13.29	16.62
Hard	6.20	7.60	7.74	8.60	7.23	6.84	7.04	5.57	8.10	7.99
Seasonal	8.82	10.56	10.59	12.04	10.01	12.94	17.62	17.82	12.28	11.09

The results are slightly different to the percentage differences between the MAD of the worst technique and this of the combined forecasts. MinSAD and MinSAPE perform better than the other combination techniques; however the average LP method is outperformed by the inverse proportion to MSE and the MinMaxAD is outperformed by all the traditional combination techniques except the simple average method. The MinSAD and MinSAPE LP models perform significantly better than the traditional techniques.

A general conclusion about the application of LP to combined forecasting is that LP based models, except MinMaxAD, outperform the traditional combined forecasting techniques, and that they also improve on the performance of the individual techniques. MinMaxAD is the best combined forecasting method for series with high variability. The next step is to examine whether single objective LP models can be improved with a multi-objective Weighted Goal Programming approach.

10.2 WEIGHTED GOAL PROGRAMMING

Similar as for the single objective LP models, I ran the experiments ignoring the ARS technique. The following table shows the average MAPE of the WGP models. As we can see, all the WGP approaches, except this with weight 2, outperform the MinSAD, MinSAPE and MinMaxAD. In addition, the model with weight 6 performs better than the average LP approach.

Table 10. 7 Weighted goal programming - MAPE

no ARS	PERFORMANCE OF COMBINED FORECASTS					
MAPE	WEIGHTED GOAL PROGRAMMING					
a2	1	2	3	4	5	6
All	8.53	8.56	8.54	8.48	8.48	8.45
Smooth	3.13	3.14	3.13	3.13	3.14	3.18
Hard	23.85	24.03	23.87	23.55	23.56	23.48
Seasonal	9.61	9.62	9.63	9.60	9.62	9.48

Table 10.8 shows the performance of WGP according to sMAPE. Here, WGP outperforms simple LP overall. However, MinMaxAD performs better on the hard series and MinSAPE on the smooth series, WGP with weight 6 performs better on the seasonal series and it is the best overall.

Table 10. 8 Weighted goal programming - sMAPE

no ARS	PERFORMANCE OF COMBINED FORECASTS					
sMAPE	WEIGHTED GOAL PROGRAMMING					
a2	1	2	3	4	5	6
All	8.31	8.34	8.31	8.24	8.25	8.17
Smooth	3.03	3.03	3.03	3.03	3.04	3.10
Hard	23.16	23.31	23.13	22.76	22.79	22.59
Seasonal	9.45	9.46	9.46	9.43	9.45	9.17

Table 10.9 shows the performance of WGP according to MASE. WGP slightly outperforms simple LP. It performs as good as simple LP on the smooth and hard series, but significantly better on the seasonal series (weight 6).

Table 10. 9 Weighted goal programming - MASE

no ARS	PERFORMANCE OF COMBINED FORECASTS					
MASE	WEIGHTED GOAL PROGRAMMING					
a2	1	2	3	4	5	6
All	0.91	0.91	0.91	0.91	0.91	0.90
Smooth	0.96	0.96	0.96	0.96	0.96	1.01
Hard	0.85	0.86	0.86	0.84	0.84	0.84
Seasonal	0.86	0.86	0.86	0.86	0.86	0.76

The next table presents the percentage difference between the MAD of the combined forecasting models and the MAD of the best individual technique.

Table 10. 10 Weighted goal programming - Difference between the best

no ARS	PERFRMANCE OF COMBINED FORECASTS					
%BEST	WEIGHTED GOAL PROGRAMMING					
a2	1	2	3	4	5	6
All	5.45	5.85	5.86	5.55	5.76	7.45
Smooth	6.40	6.72	6.90	6.95	7.14	14.13
Hard	10.48	11.54	10.89	9.24	9.54	8.95
Seasonal	0.97	1.15	1.21	1.01	1.21	-5.27

The performance of the WGP approach is similar to the performance of the best single objective approach. The model with weight 1 gives the smallest average percentage difference; the models with weight 4 and 6 give the smallest average percentage difference on the hard series.

Table 10.11 shows the percentage difference between the MAD of the worst individual forecasting method and the MAD of the combined forecasts.

Table 10. 11 Weighted goal programming - Difference between the worst

no ARS	PERFORMANCE OF COMBINED FORECASTS					
%WORST	WEIGHTED GOAL PROGRAMMING					
a2	1	2	3	4	5	6
All	35.23	34.94	34.96	35.18	35.02	36.80
Smooth	42.72	42.50	42.39	42.35	42.20	41.05
Hard	22.23	21.49	22.01	23.29	23.08	23.49
Seasonal	29.11	28.98	28.94	29.03	28.89	36.64

Similarly as when making combinations of eight techniques, the differences between the performance of the WGP approach and the MinSAD, MinSAPE and average LP models are small.

Table 10. 12 Weighted goal programming - Difference between the average

no ARS	PERFORMANCE OF COMBINED FORECASTS					
%AVERAGE	WEIGHTED GOAL PROGRAMMING					
SERIES	1	2	3	4	5	6
All	18.63	18.29	18.30	18.54	18.37	19.42
Smooth	22.69	22.43	22.29	22.25	22.09	19.20
Hard	7.58	6.70	7.26	8.67	8.42	8.90
Seasonal	17.54	17.38	17.33	17.43	17.27	25.67

Finally, Table 10.12 shows the percentage difference between the average MAD of all the individual techniques on each of the series and the MAD of the WGP combination models ignoring the ARS model. The WGP models slightly improve the results of single objective LP models. Nevertheless, the differences remain small.

As in Chapter 7, WGP approaches slightly improve the results of the three LP approaches and the average LP, as well as for each subgroup of the series separately. WGP seems to be stable and have the same good performance according to all the performance measurement indices. Nevertheless, the differences between all the LP models are small.

10.3 CONCLUSION

The conclusion of the Appendix is similar to this of Chapter 7. LP approaches for combination outperform the traditional techniques, even when there is not a dominant individual technique. MinSAD and MinSAPE perform better on the smooth and seasonal series, MinMaxAD performs better on the hard series and WGP performs better overall. The only difference in comparison with the results of Chapter 7 is that when there is not a dominant technique on a specific type of series (such as ARS on the seasonal series), LP formulations do not improve the results of the best technique, since the percentage difference between the combination and the best individual technique is not negative.

REFERENCES

Aaker D. A. and R. Jacobson (1987). The sophistication of 'naïve' modelling. *International Journal of Forecasting*, 3(3-4), pp. 449 – 451.

Ahuja R. K. and J. B. Orlin (2001). Inverse optimization. *Operations Research*, 49(5), pp. 771 – 783.

Akaike H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE*, 19(6), pp. 716 – 723.

Akaike H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21(1), pp. 243 – 247.

Amin G. R. and A. Emrouznejad (2007). Inverse forecasting: A new approach for predictive modelling. *Computers and Industrial Engineering*, 53(3), pp. 491 – 498.

Anderson D. R., D. J. Sweeney, T. A. Williams (1998). *Quantitative Methods for Business*, seventh edition, South-Western College Publishing, Cincinnati.

Andrawis R. R., A. F. Atiya and H. El-Shishiny (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting*, 27(3), pp. 870 – 886.

Armstrong J. S. (2001). *Principles of Forecasting*, Kluwer Academic Publishers, Massachusetts.

Armstrong J. S. (2007). Significance tests harm progress in forecasting, *International Journal of Forecasting*, 23(2), pp. 321 – 327.

Arthanari T. S. and Y. Dodge (1981). *Mathematical Programming in Statistics*, Wiley, New York.

Auerbach A. J. (1999). On the Performance and Use of Government Revenue Forecasts. *National Tax Journal,* 52, pp.765 – 782.

Banuls V. A. and M. Turoff (2011). Scenario Construction via Delphi and Cross Impact Analysis. *Technological Forecasting and Social Change*, 78(9), pp. 1487 – 1720.

Barlas Y. and B. Gunduz (2011). Demand forecasting and sharing strategies to reduce fluctuations and the bullwhip effect in supply chains. *Journal of the Operational Research Society*, 62(3), pp. 458 – 473.

Bates J. M. and C. W. J. Granger (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), pp. 451 – 468.

Belton V. And P. Goodwin (1996). Remarks on the application of the analytic hierarchy process to judmental forecasting. *International Journal of Forecasting*, 12(1), 155 – 161.

Bermudez J. D., J. V. Segura and E. Vercher (2006). Improving demand forecasting accuracy using nonlinear programming software. *Journal of the Operational Research Society*, 57(1), pp. 94 – 100.

Billio M. and R. Casarin (2010). Identifying business cycle turning points with sequential Monte Carlo methods: an online real-time application to the Euro area. *Journal of Forecasting*, 29(1-2), pp. 145 – 167.

Blattberg R. C. And S. J. Hoch (1990). Database models and managerial intuition – 50-percent model C 50-percent manager. *Management Science*, 36(8), 887 – 899.

Bollerslev T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), pp. 307 – 327.

Bolton G. E. (2002). Game theory's role in role-playing. *International of Forecasting*, 18(3), pp. 353 – 358.

Box G. E., and G. M. Jenkins (1970). *Time Series Analysis Forecasting and Control*, Holden Day, San Francisco.

Bretschneider S. I., W. L. Gorr, G. Grizzle and E. Klay (1989). Political and organizational influences on the accuracy of forecasting state government revenues. *International Journal of Forecasting*, 5(3), pp. 307 – 319

van Bruggen G. H., M. Spann, G. L. Lilien and B. Skiera (2010). Prediction Markets as institutional forecasting support systems. *Decision Support Systems*, 49(4), pp. 404 – 416.

Bunn D. W. (1996). Non-traditional methods of forecasting. *European Journal of Operational Research*, 92(3), pp. 528 – 536.

Chen C. W. S., R Gerlach, E. M. H. Lin and W. C. W. Lee (2011). Bayesian forecasting for financial risk management, pre and post global crisis. *Journal of Forecasting*, *In press.*

Chase B. R., F. R. Jacobs and J. N. Aquilano (2006). *Operations Management for Competitive Advantage*, McGraw-Hill/Irwin, New York.

Cheng L. and M. A. Duran (2004). Logistics for world-wide crude oil transportation using discrete event simulation and optimal control. *Computers and Chemical Engineering*, 28(6-7), pp. 897 – 911.

Chun Y. H. (2012). Monte Carlo analysis of estimation methods for the prediction of customer response patterns in direct marketing. *European Journal of Operational Research,* 217(3), pp. 673 – 678.

Clemen R. T. and R. L. Winkler (1986). Combining economic forecasts. *Journal of Business and Economic Statistics*, 4(1), pp. 39 – 46.

Collopy F. and J. A. Armstrong (1992). Rule-Based forecasting: Development and validation of an expert system approach to combining time series extrapolations. *Management Science*, 38(10), pp. 1394 – 1414.

Croce V. and K. W. Wober (2011). Judgemental forecasting support systems in tourism. *Tourism Economics*, 17(4), pp. 709 – 724.

Crone S. F., M. Hibon and K. Nikolopoulos (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, 27(3), pp. 635 – 660.

Dantzig G. B. (2002). Linear programming. *Operations Research*, 50(1), pp. 42 – 47.

Dhahri I. and H. Chabchoub (2007). Nonlinear goal programming models quantifying the bullwhip effect in supply chain based on ARIMA parameters. *European Journal of Operational Research*, 177(3), pp. 1800 – 1810.

De Gooijer G. J. and R. J. Hyndman (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3), pp. 443 – 473.

Delen D., D. Cogdell and N. Kasap (2011). A comparative analysis of data mining methods in predicting NCAA bowl outcomes. *International Journal of Forecasting, In press.*

Dixit A. and S. Skeath (1999). *Games of Strategy*, New York.

Dortmans P. J. and E. Eiffe (2004). An examination of future scenarios using historical analogy. *Futures*, 36(10), pp. 1049 – 1062.

Drackley A., B. Newbold, A. Paez and N. Heddle (2011). Forecasting Ontario's blood supply and demand. *Transfusion*, *In press.*

Dugdale M. (1971). Long term trend projections as an aid to management. *Journal of the Royal Statistical Society. Series D (The Statistitian)*, 20(3), pp. 45 – 54.

Engle R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, 50, pp. 987 – 1008.

Engle R. F., D. Lilien and R. Robins (1987). Estimation of time varying risk premia in term structure: the ARCH-M model. *Econometrica*, 55, pp. 391 – 407.

Eroglu C. And K. L. Croxton (2010). Bias in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, 26(1), pp. 116 – 133.

Fan H. and T. Söderström (1997). Parameter estimation of continuous time AR progress using integrated sampling. *Proceedings of the 36th conference on Decision and Control*, San Diego, California.

Feigin P. D. M. F. Kratz and S. L. Reisnick (1996). Parameter estimation for moving averages with positive innovations. *The Annals of Applied Probability*, 6(4), pp. 1157 – 1190.

Feigin P. D. and S. L. Reisnick (1992). Estimation of autoregressive processes with positive innovations. *Stochastic Models*, 8(3), pp. 685 – 717.

Feigin P. D. and S. L. Reisnick (1994). Limit distribution for linear programming time series estimators. *Stochastic Processes and their Applications*, 51(1), pp. 135 – 165.

Feigin P. D. and S. L. Reisnick (1995). Testing for interdependence in heavy tailed and positive innovation time series. *Stochastic Models*, 11(4), pp. 587 – 612.

Feigin P. D. and S. L. Reisnick (1997). Linear programming and bootstrapping for heavy tailed phenomena. *Advances in Applied Probability*, 29(3), pp. 759 – 805.

Feigin P. D. and S. L. Reisnick (1999). Pitfalls of fitting autoregressive models for heavy-tailed time series. *Extremes*, 1(4), pp. 391 – 422.

Fildes R. and P. Goodwin (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37(6), pp. 570 – 576.

Fildes R. and P. Goodwin (1999). Judgemental forecasts of time series affected by special events: does providing statistical forecast improve accuracy? *Journal of Behavioural Decision Making*, 12, pp. 37 – 53.

Fildes R., P. Goodwin and M. Lawrence (2006). The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, 42(1), pp. 351 – 361.

Fildes R., K. Nikolopoulos, S. F. Crone and A. A. Syntetos (2008). Forecasting and operational research: a review. *Journal of the Operational Research Society*, 59(9), pp. 1150 – 1172.

Fildes R., P. Goodwin, M. Lawrence and K. Nikolopoulos (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), pp. 3 – 23.

Freed N. and F. Glover (1981). A linear programming approach to the discriminant problem. *Decision Sciences*, 12(1), pp. 68 – 74.

Freed N. and F. Glover (1981). Simple but powerful goal programming models for discriminant analysis. *European Journal of Operational Research*, 7(1), pp. 44 – 60.

Glen J. J. (2006). A comparison of standard and two stage mathematical programming discriminant analysis methods. *European Journal of Operational Research*, 171(2), pp. 496 – 515.

Glen J. J. (2003). An interactive mixed integer programming method for classification accuracy maximizing discriminant analysis. *Computers and Operations Research*, 30(2), pp. 181 – 198.

Glover F. (1990). Improved linear programming models for discriminant analysis. *Decision Sciences*, 21(4), pp. 771 – 785.

Goodwind P. (2000). Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, 16(3), pp. 85 – 99.

Goodwind P. (2002). Forecasting games: Can game theory win? *International Journal of Forecasting*, 18(3), pp. 369 – 374.

Goodwind P. (2007). Should we be using significance tests in forecasting research? *International Journal of Forecasting*, 23(2), pp. 333 – 334.

Goodwind P. and R. Lawton (1999). On the asymmetry of the symmetric MAPE. *International Journal of Forecasting*, 15(4), pp. 405 – 408.

Galbraith C. S., A. F. DeNoble, S. B. Ehrlich and J. Mesmer – Magnus, (2010). Review panel consensus and post-decision commercial performance: a study of early stage technologies. *The Journal of Technology Transfer*, 35(2), pp. 253 – 281.

Grambsch P. and W. A. Stahel (1990). Forecasting demand for special telephone services: A case study. *International Journal of Forecasting*, 6(1), pp. 53 – 64.

Green K. C. (2002). Forecasting decisions in conflict situations: a comparison of game theory, role-playing and unaided judgement. *International Journal of Forecasting*, 18(3), pp. 321 – 344.

Green K. C. (2005). Game theory, simulated interaction, and unaided judgement for forecasting decisions in conflicts: Further evidence. *International Journal of Forecasting*, 21(3), pp. 463 – 472.

Green P. E. and V. Srinivasan (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54(4), pp. 3 – 19.

Gupta S. and P. C. Wilton (1987). Combination of forecasts: An extension. *Management Science*, 33(3), pp. 356 – 372.

Halme M. and M. Kallio (2011). Estimation methods for choice-based conjoint analysis of consumer preferences. *European Journal of Operational Research*, 214(1), pp. 160 – 167.

Hand D., H. Mannila and P. Smyth (2001). *Principles of Data Mining*, The MIT Press, Cambridge – London, UK.

Hibon M. and T. Evgeniou (2005). To combine or not to combine: selecting among the forecasts and their combinations. *International Journal of Forecasting*, 21(1), pp. 15 – 24.

Higuchi T. and M. D. Troutt (1994). Dynamic simulation of the supply chain for a short life cycle product – Lessons from the Tamagotchi case. *Computer and Operations Research*, 31(7), pp. 1097 – 1114.

Holden K., A. D. Peel and L. J. Thompson (1990). *Economic Forecasting: An Introduction*, Cambridge University Press, New York.

Hoogerheide L. and H. K. van Dijk (2011). Baysian forecasting of Value at Risk and Expected Shortfall using adaptive importance sampling. *International Journal of Forecasting*, 26(2), pp. 231 – 247.

Hyndman R. J. and B. A. Koehler (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), pp. 679 – 688.

Hyndman R. J., A. B. Koehler, J. K. Ord and R. D.Snyder (2008). *Forecasting with exponential smoothing: the state space approach*. New York, NY: Springer.

Johnston F. R., J. E. Boyland, M. Meadows and E. Shale (1999). Some properties of a simple moving average when applied to forecasting a time series. *The Journal of the Operational Research Society*, 50(12), pp. 1267 – 1271.

Jonung L. and M. Larch (2006). Fiscal policy in the EU: are official output forecasts biased? *Economic Policy*, July, pp. 491 – 534.

Jursa R and K. Rohring (2008). Short-term wind power forecasting using evolutionary algorithms for automated specification of artificial intelligence models. *International Journal of Forecasting*, 24(4), pp. 694 – 709.

Kanama D., (2010). A Japanese experience of a mission-oriented multi-methodology technology foresight process: an empirical trial of a new technology foresight process by integration of Delphi method and scenario writing. *International Journal of Technology Intelligence and Planning*, 6(3), pp. 253 – 267.

Kang H. (2003). How to measure the cost of a Forecast error on an enterprise? *Journal of Business Forecasting Methods & Systems*, 22(1), pp. 21 – 25.

Kang H. (1986). Unstable weights in the combination of forecasts. *Management Science*, 32(6), pp. 683 – 695.

Kantorovich L. V. (1960). Mathematical Methods of Organizing and Planning Production. *Management Science*, 6(4), pp. 366 – 422.

Keereman F. (1999). The track record of the Commission Forecasts. Economic Papers, pp. 137.

Kiountouzis E. A. (1973). Linear programming techniques in regression analysis. *Journal of Royal Statistical Society. Series C (Applied Statistics)*, 22(1), pp. 69 – 73.

Kirkpatrick C. D. and J. R. Dahlquist (2011). *Technical Analysis: The complete Resource for Financial Market Technicians*, FT Press.

Kolassa S. (2011). Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting*, 27(2), pp. 238 – 251.

Konishi S. and Kitagawa G. (2008). *Information criteria and statistical modeling*. New York, NY: Springer.

Lam K. F., H. W. Mui and H. K. Yuen (2001). A note of minimising absolute percentage error in combined forecasts. *Computers and Operations Research*, 28(11), pp. 1141 – 1147.

Lawrence S. M., P. Goodwin, M. O'Connor, and D. Onkal (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3), pp. 493 – 518.

Lawrence K. D., Dinesh R. Pai and S. M. Lawrence (2010). Forecasting using fuzzy objective linear programming. Into: K. D. Lawrence, R. K. Klimberg (ed.), *Advances in Business and Management Forecasting, Volume 7*, Emerald Group Publishing Limited, pp. 149 – 156.

Leung M. T., H. Daouk and A. Chen (2001). Using investment portfolio return to combine forecasts: A multiobjective approach. *European Journal of Operational Research*, 134(1), pp. 84 – 102.

Lilja K. K., K. Laakso and J. Palomki (2011). Using the Delphi method. *Technology Management in the Energy Smart World*, 28(11), pp. 1 – 10.

Liu S., J. Zhang, M. Keil and T. Chen (2010). Comparing senior executive and project manager perceptions of IT project risk: a Chinese Delphi study. *Information Systems Journal*, 20(4), pp. 319 – 355.

Lo A. W., H Mamaysky and J. Wang (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The Journal of Finance*, 55(4), pp. 1705 – 1765.

Luna I. and R. Ballini (2011). Top-down strategies on adaptive fuzzy rule-based systems for daily time series forecasting. *International Journal of Forecasting*, 27(3), pp. 708 – 724.

Malik M. and L. C. Thomas (2010). Modelling credit risk of portfolios of consumer loans. *Journal of the Operational Research Society*, 61(3), pp. 411 – 420.

Makridakis S. (1989). Why combining works? *International Journal of Forecasting*, 5(4), pp. 601 – 603.

Makridakis S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Leandowski, J. Newton, E. Parzen and R. Winkler (1982). The accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition. *Journal of Forecasting*, 1(2), pp. 111 – 153.

Makridakis S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Leandowski, J. Newton, E. Parzen and R. Winkler (1984). *The Forecasting Accuracy of Major Time Series Methods,* John Wiley & Sons, New York.

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

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

Makridakis S., C. S, Wheelwright, and J. R. Hyndman (1998). *Forecasting: Methods and Applications*, 3rd edition, Willey, New Jersey.

de Menezes L. M., D. W. Bumm and J. W. Taylor (2000). Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, 120(1), pp. 190 – 204.

Mohammadi K., H. R. Eslami and R Kahawita (2006). Parameter estimation of an ARMA model for river flow forecasting using goal programming. *Journal of Hydrology*, 331(1-2), pp. 293 – 299.

Mosheiov G. and A. Raveh (1997). On trend estimation of time-series: a simple linear programming approach. *Journal of the Operational Research Society*, 48(1), pp. 90 – 96.

Nahmias S. (2005). *Production and Operations analysis*, McGraw-Hill/Irwin, New York.

Newbold P. and T. Bos (1994). *Introductory Business and Economic forecasting*, 2nd edition, South Western Publishing, Cincinnati.

Nicholson W. and C. Snyder (2007). *Theory and Application of Intermediate Microeconomics*, 10th edition, Thomson South-Western.

Ozyildirim A, B. Schaitkin and V. Zarnowitz, (2010). Business cycles in the euro area defined with coincident economic indicators and predicted leading economic indicators. *Journal of Forecasting*, 29(1-2), pp. 6 – 28.

Parzen E. (1982). ARARMA models for time series analysis and forecasting, *Journal of forecasting*, 1(1), pp. 67 – 82.

Pereira A. (2004). Performance of time-series methods in forecasting the demand for red blood cell transfusion. *Transfusion*, 44(5), 739 – 746.

Perry M. B. (2010). The Weighted Moving Average Technique. *Wiley Encyclopaedia of Operations Research and Management Science*, Wiley.

Peterson R. T. (1993). Improving Sales Force composite: Forecasting by using scripts. *Journal of Business Forecasting*, 12(3), pp. 10 – 14.

Peterson R. T. (1989). Sales Force composite forecasting – An exploratory analysis. *Journal of Business Forecasting*, 8(1), pp. 23 – 27.

Pflaumer P. (1988). Confidence intervals for population projections based on Monte Carlo methods. *International Journal of Forecasting*, 4(1), pp. 135 – 142.

Pidd M. (1998). *Computer Simulation in Management Science*, Wiley.

Pierce D. A. (1980). A survey of recent developments in seasonal adjustment. *The American Statistician*, 34(3), pp. 125 – 134.

Ramnath S., S Rock and P. Shane (2008). The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting*, 24(1), pp. 34 – 75.

Reeves G. R. and K. D. Lawrence (1982). Combining multiple forecasts given multiple objectives. *Journal of Forecasting*, 1(3), pp. 271 – 279.

Reeves G. R. and K. D. Lawrence (1991). Combining forecasts given different types of objectives. *European Journal of Operational Research*, 51(1), pp. 65 – 72.

Reeves G. R. and K. D. Lawrence, S. M. Lawrence and J. B. Guerard (1988). Combining earnings forecasts using multiple objective linear programming. *Computers and Operations Research*, 15(6), pp. 551 – 559.

Reisman A. and F. Kirschnick (1995). Research Strategies Used by OR/MS Workers as Shown by an Analysis of Papers in Flagship Journals, *Operations Research*, 43(5), pp. 731 – 740.

Reisman A. and F. Kirschnick (1994). The Devolution of OR/MS: Implications from a Statistical Content Analysis of Papers in Flagship Journals, *Operations Research*, 42(4), pp. 577 – 588

Romero Morales D. and J. Wang (2010). Forecasting cancellation rates for services booking revenue management using data mining. *European Journal of Operational Research*, 202(2), pp. 554 – 562.

Russell T. D. and E. E. Adam Jr. (1987). An empirical evaluation of alternative forecasting combinations. *Management Science*, 33(10), pp. 1267 – 1276.

Sanders N. and K.B. Manrodt (1994). Forecasting practices in US corporations: Survey results. *Interfaces*, 24(2), 92 – 100.

Sanders N. R. and B. K. Manrodt (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice*. Omega: The International Journal of Management Science*, 31(6), pp. 511 – 522.

Schmidt F. L. and J. E. Hunter (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. Into: L. L Harlow, S. A. Mulaik and J. H. Steiger (ed.), *What if there were no significance tests?* Lawrence Erlbaun, London, pp. 37 – 64.

Schwarz G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, 6(2), pp. 461 – 464.

Segura J. V. and E. Vercher (2001). A spreadsheet modeling approach to the Holt-Winter Optimal forecasting. *European Journal of Operational Research*, 131(2), pp. 375 – 388.

Shah P. and A. Guez (2008). Mortality forecasting using neural networks and an application to cause-specific data for insurance purposes. *Journal of Forecasting*, 28(6), pp. 535 – 548.

Simpson P. K. (1992) (edited by Edgar Sánchez-Sinencio), Foundations of Neural Networks, into: *Artificial neural networks: paradigms, applications and hardware implementations*. Institute of Electrical and Electronics Engineers Press.

Smith J. Q and G. Freeman (2010). Distributional Kalman filters for Baysian forecasting and closed from recurrences. *Journal of Forecasting*, 30(1), pp. 210 – 224.

Soliman S. A., S. Persuad, K. El-Negar and M. E. El-Hawary (1997). Application of least absolute value parameter estimation based on linear programming to short-term load forecasting. *Electrical Power and Energy Systems*, 19(3), pp. 209 – 216.

Stevenson W. J. (2005). *Operations Management (8th edition)*, McGraw – Hill, Boston.

Swanson R. N. and H. White (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, 13(4), pp. 439 – 461.

Syntetos A. A., J. E. Boylan, K. Nikolopoulos J. E. Boylan, R. Fildes and P. Goodwin (2009). The effects of integrating management judgment in intermittent demand forecasts. *International Journal of Production Economics*, 118(1), pp. 72 – 81.

Syntetos A. A. and J. E. Boylan (2010). On the variance of intermittent demand estimates. *International Journal of Production Economics*, 128(2), pp. 546 – 555.

Tarantola A. (1987). *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, Amsterdam.

Thomas L. C., A. Matuszyk and A. Moore (2011). Comparing debt characteristics and LGD models for different collection policies. *International Journal of Forecasting*, in press.

Trapp J. N. (1986). A methodology for estimating integrated forecasting/decision model parameters using linear programming. *Southern Journal of Agricultural Economics*, 18(2), pp. 247 – 256.

Trigg D. W. and A. G. Leach (1967). Exponential smoothing with an adaptive response rate. *Operational Research*, 18(1), pp. 53 – 59.

Venkatesan R and V. Kumar (2002). A genetic algorithms approach to growth phase forecasting of wireless subscribers. *International Journal of Forecasting*, 18(4), pp. 625 – 646.

Wagner H. M. (1959). Linear programming techniques for regression analysis. *Journal of the American Statistical Association*, 54, pp. 206 – 212.

Webby R., M. O'Connor and B. Edmundson (2005). Forecasting support systems for the incorporation of event information: An empirical investigation. *International Journal of Forecasting*, 21(3), pp. 411 – 423.

Wong W. K., M. Xia and W. C. Chu (2010). Adaptive neural network model for time-series forecasting. *European Journal of Operational Research*, 207(2), pp. 807 – 816.

Wright G. and G. Rowe (2011). Group-based judgmental forecasting: An integration of extant knowledge and the development of priorities for a new research agenda. *International Journal of Forecasting*, 27(1), pp. 1 – 13.

Wu D. D., X. Kefam, L. Hua, Z. Shi and D. L. Oslon (2010). Modelling technological innovation risks of an entrepreneurial team using system dynamics: An agent-based perspective. *Technological Forecasting and Social Change*, 77(6), pp. 857 – 869.

Yelland P. M. (2010). Bayesian forecasting of parts demand. *International Journal of Forecasting*, 26(2), pp. 374 – 396.

Yokum J. T. and S. Armstrong (1995). Beyond accuracy: Comparison of criteria used to select forecasting methods. *International Journal of Forecasting*, vol. 11(4), pp. 591 – 597.

Zhang J., and Z. Liu (1996). Calculating some inverse linear programming problems. *Journal of Computational Applied Mathematics*, 72(2), pp. 261 – 273.

Zhou Z., Y. Shi and X. Hao (1999). An MC^2 linear programming approach to combined forecasting. *Mathematical and Computer Modeling*, 29(8), pp. 97 – 103.