Call Centre Forecasting

A comprehensive analysis of missing data, extreme values, holiday influences and different forecasting methods

Master Thesis Econometrics and Operations Research
Master Operations Research and Management Science

Mathijs Jansen

Supervisors:
Dr. O. Boldea - Tilburg University
Dr. F.C. Drost – Tilburg University
Ir. T. de Nooij MTD - Anago

Tilburg University
1 April 2010

ABSTRACT

In this thesis a model will be developed to forecast the number of calls in a call centre. Given the importance of these forecasts to get a high service rate with minimal costs it is vital that the forecasting method yields good results. By first analysing and correcting missing data the quality of the data will be improved. A method to detect and correct anomalies also improves the quality of the data. These missing values and anomalies are generally caused by system failures and can therefore be corrected without losing vital information. The hypothesis that holidays have a significant influence on the call volume is also tested in this thesis and it is shown that no significant influence is present. With this data we then continue with forecasting the call volume for the weeks to come. This forecasting is done with a new adaptation to the standard Holt-Winters model. For this adaptation we will introduce a new method to generate initial values. The model now incorporates the current level, trend and two seasonality factors. Most call centre projects have a seasonality effect in every week, i.e. on Mondays the call volume is higher and on Fridays the volume is smaller. In addition to that there can also be a monthly pattern, e.g. when invoices are being sent, which will be estimated by the second seasonality factor. For the optimisation of the parameters used in the extended Holt-Winters model a new method for optimisation is introduced and used. This method of optimisation uses Latin Hypercube Designs to find initial values for the parameters and then improves these values. The results of this entire forecasting procedure are compared to more computational intensive ARIMA models to see whether the performance of our new models is better or worse than the performance of the ARIMA models.

TABLE OF CONTENTS

1 INTRODUCTION

Call centres form an important part of service based companies. Estimates from 2005 indicate that over $300 billion is spent on call centres worldwide (Gilson & Khandelwal, 2005). It is also estimated that there are 4 million call centre agents in the US, 800 thousand in the UK, 500 thousand in Canada and 500 thousand in India (Holman, Batt, & Holtgrewe, 2007). Since call centres are such an important part of service, a lot of effort is put into the efficient management of the call centre. What you try to achieve is getting a high rate of service against minimum costs. Typically call centres use a service rate (e.g. 95%) and they set the goal that at least this percentage of calls is answered within a certain amount of time, or before the caller hangs up his telephone.

To achieve this goal it is needed to make a good planning of customer service representatives (CSR's) or call centre agents. These agents answer the calls that come in to the system. Usually there are several different types of agent. It can be that a service number first guides you through a multiple choice menu to get you to the right agent with the right knowledge. It is also possible that the agent you speak to first directs you to a higher placed agent with more knowledge or influence. To help all people that call the service number, there need to be enough agents at every station and at every level. Determining how many agents are needed depends on several factors. First you need to determine how many calls you are expecting at every moment in time. This is done by forecasting the number of calls based on historical data. Next you need to know how long it takes for an agent to answer the call and help the customer (both on the phone, and after the call). This determines how many callers an agent can help per time period. You also need to determine how impatient callers are, i.e. how quickly will they hang up the phone. All these factors make it hard to estimate the amount of work there is at a call centre, but getting a good estimate yields high service rates with low costs. Given the amount of money that is spent on call centres it is really beneficial to make advanced estimates of all the processes concerned in the call centre.

In this thesis we will look at the problem of forecasting the number of incoming calls in a call centre. Though the model developed here can also be used in other applications, we will limit our research to call centre forecasting.

Forecasting is used in many applications in business. Making good forecasts is important in many different industries. It can e.g. be used to forecast demand for a certain product, such that the production level can be determined. Also for repair centres for consumer electronics there should be a forecast for the expected number of faulty units, such that spare parts can be ordered and personnel can be hired. In the call centre world, forecasting is basically the same, i.e. it uses the same tools to derive the forecasts. From historical data we derive a forecast for future periods. Call centre forecasting is not only limited to forecasting the number of calls for the next period. First the data should be examined. A well know saying in any statistical analysis is: "garbage in, garbage out", so having a good dataset is essential to the rest of the process.

The internship connected to this thesis is accommodated by Anago Software B.V. Anago is a small software company, with their own software package. This software package consists of several 'building blocks' of which a custom application can be made for their customers. For one of their customers, Unamic/HCN, they made a call centre application. Unamic/HCN is one of the largest call centre firms in the Netherlands with several offices in the Netherlands and offices in Belgium, Suriname and Turkey. The software Anago installed keeps track of several processes within the company. It monitors all the incoming calls, makes forecasts based on the historical data and based on that forecast makes a planning for the CSR's. Staffing officers can indicate the availability of call centre agents and the software makes the planning on all constraints. Making a good forecast is essential for making a good planning, and therefore the forecasting module should be accurate. Currently forecasting is done on the historical data using three different methods (chosen by the call centre planner). These methods are: Trend forecasting, exponential smoothing and Holt-Winters forecasting. The first two methods are basic methods, which cannot deal with seasonal variation, and therefore might lead to unwanted results. The latter method can generate good results when data has a seasonal pattern. However, the seasonal pattern cannot be determined automatically and therefore most planners don't know how to use this seasonal parameter. Forecasting is done on week level. After a forecast is done, a week pattern is applied to the forecast. This week pattern is a pre determined distribution over the days of the week. After applying the week pattern a day pattern is applied as well, which completes the forecasting process.

The data we use in this thesis is data from Sanoma. Sanoma is a Finnish company, which publishes magazines all over Europe. In the Netherlands Sanoma is the biggest publisher of weekly magazines. They publish over eighty magazines, which include: Donald Duck, Libelle, Panorama, Story and Autoweek, some of the biggest weekly magazines in the Netherlands. For Sanoma the call centre is regulated by Unamic/HCN. Call centre data is handled by the Sanoma workgroup. This workgroup is divided into several sub projects, see 0. In the appendix you can find the availability of data per project and a short description of the project. We will mainly focus on the currently running projects. The data is available per quarter of an hour from 9.00 hours until 20.00 hours (the opening times of the call centre). In the data, it could be that there is data missing for a certain time period (usually because of system failures), or the data could be missing for the entire day (usually because of holidays). It is the goal of this thesis to make good forecasts based on the available data.

1.1 Problem Description

Currently the forecasting module is really basic. All data is available on quarter of an hour level, but forecasts are only made on week level. If we then have the forecast on week level, we determine the number of calls per quarter of an hour by applying subsequently a week filter and a day filter. The week filter divides the total number of calls over the five days, and the day filter divides the calls over the quarters. This method, though simple, has a big disadvantage. Some projects have a monthly pattern. This can be the case with a project for invoice questions. The invoices for the magazines are sent on a specific day each month and following those invoices the amount of calls will rise for that specific project. However when we use weekly forecasting, this effect cannot be modelled by our forecasting model. This is because, e.g. the first of every month is not every four weeks, but it could also be after five weeks. We therefore need to change the forecasting to a daily pattern, or even quarter of an hour forecasts.

Changing the forecasting time period also brings up another problem. Typically the call volume on Mondays is higher than in the rest of the week. On Fridays the call volume is again lower than on the other three days. To deal with this seasonality we need to incorporate the effects of the different days into the model.

Furthermore, no data analysis is done on the data before it is used in the forecast. Occasionally the system has a failure and then the number of calls isn't recorded. Calls keep coming into the system, but in our data the number of calls for that specific period (or periods) is zero. If we then would use this value in our forecast it would influence the results. Therefore we will investigate the data before it is used for forecasting. This investigation consists of two parts. First we will look at the problem of missing data. And after that we will look at the problem of outliers in the data.

To summarise the problem description is as follows:

How can we forecast the incoming call volume of a call centre?

With points of focus:
-Quality check on the data
-Influence of holiday on the remainder of the week
-Automatic detection of seasonal patterns
-Incorporate seasonal effects into the forecast
-Determine the quality of the forecast compared to other methods

With these points of focus we will give a comprehensive framework for forecasting call centre volume. We also introduce a new method to estimate the parameters of the Holt-Winters double seasonal model.

1.2 Overview

In this thesis we will start with a literature overview in chapter 2. Here we will elaborate on the research papers that have been written on the call centre world in general and on call centre forecasting in particular. We will also compare this literature to the investigation that is done in this thesis.

Chapter 3 deals with the problem of missing data. Because there can be data missing in our dataset, we need to investigate whether this data is really missing. In the dataset a missing data point can mean two things. First it can be that the number of calls for that time period is actually zero, or it can mean that there was a system failure and therefore the system didn’t record the number of calls. The research indicated that the missing data can be imputed by a random draw from the Poisson distribution with the average value of the ‘neighbourhood’ as the input value. This neighbourhood is determined to consist of at most twelve observations in the same day, and at most sixteen observations of the same time period on other days. In this chapter we will determine the size of this neighbourhood and show that the distribution of calls is Poisson. Subsequently we will determine whether a zero value can be considered a missing value. In that case a random draw from the Poisson distribution will be imputed into the dataset.

Chapter 4 deals with the problem of extreme value detection. Because of a problem in the system, extremely high or extremely low values can occur. In this chapter we will determine if these extreme values can be considered faults in the system and should therefore be corrected. This correction is done in the same way as the missing data correction. The anomalies will be corrected by a random draw from the Poisson distribution.

Chapter 5 examines the influence of holidays. If e.g. on Monday it is Easter and therefore the call centre is closed, how does this influence the call volume for the remainder of the week. It can be that people call more because they have one day less in that week to call the number. It can also be that they call less because they are on holiday, or are busy with other (more important) things. However, these two effects cancel out each other and therefore there is no indication of holiday influence in the data.

Chapter 6 is the chapter where we will develop the forecasting model. First a background is given about the forecasts used in practice and found in literature. Then we will develop our extension to the Holt-Winters model and our method to determine the initial values of this model. Next we will introduce a new method to optimise the parameters to the Holt-Winters model. This method uses Latin Hypercube designs (LHD's) to calculate the outcomes given different parameter settings. With the LHD's we determine initial parameter settings which yield good forecast results. These initial parameter settings will then be optimised by adapting the parameters slightly. This process of optimising the parameters stops when no forecast can be obtained by changing parameter values.

Chapter 7 is where we will investigate and determine the call volume per time period. The forecasts are made on day level to reduce the complexity of the forecast. However, we need forecasts on time period level. Therefore a distribution of the call volume over the day is determined such that we can find the forecasts on quarter of an hour level.

The final chapter 8 shows the results and the performance of the developed model. A comparison is made between the extended Holt-Winters model and the ARIMA models.

2 LITERATURE REVIEW

Since call centre processes are interesting to model, a lot of literature has been written about call centres. For an elaborate literature review see (Mandelbaum, 2004). The most important goal for call centres is to have enough call centre agents (or Customer Service Representatives, CSR's) available for the incoming call volume. For this goal, a lot of work has to be done. First it is important that the historical data is reliable by missing data analysis and outlier detection. The concept of missing data and some methods to deal with missing data are described in (Schafer & Graham, 2002) and (Jamshidian, 2004). Also the book (Little & Rubin, 2002) gives insights into the missing data problem in statistics. After handling with missing data we should further improve the quality of the data by detecting extreme values, see e.g. (Hardy & Bryman, 2004). Both these methods of improving the quality of the data depend on the assumption that call arrivals follow a Poisson distribution. This is generally the case when arrivals are independent of each other. Several articles have been written about the Poisson distribution of call arrivals. Brown and Zhao (Brown & Zhao, 2001) discuss several methods to test whether the arrival rate follows a Poisson distribution, as well as introduce a new test. In (Soyer & Tarimcilar, 2008) a modulated Poisson process model is presented to describe and analyse arrival data. This model takes time effects and covariate effects into account and makes it possible to assess the effectiveness of different advertising strategies. Shen and Huang (Shen & Huang, 2008) developed a forecasting method to model the underlying inhomogeneous Poisson processes. First the arrival data is reduced using Singular Value Decomposition (as introduced in (Shen & Huang, 2005)). Next a factor model is derived to forecast the arrival rates. In (Antipov & Meade, 2002) several forecasting methods are applied to a financial call centre. They developed a model with a dynamic level, multiplicative calendar effects and a multiplicative advertising response. Taylor (Taylor, 2008) compared several methods to forecast intraday arrivals. Intraday series most notable feature is the presence of both an intraweek and an intraday seasonal cycle. The methods considered include (amongst others) seasonal autoregressive integrated moving average (ARIMA) modelling; periodic autoregressive modelling and an extension of Holt-Winters exponential smoothing for the case of two seasonal cycles. This thesis will focus on the data cleaning and forecasting of calls in a call centre. The double seasonality model introduced by Taylor is improved with a method to easily determine initial values. Furthermore a new method is introduced to optimise the parameter settings of the Holt-Winters model. We also introduce a method to determine the length of the seasons in a dataset.

After this step of forecasting the arrival rate, a good forecast should be made of the workload. The workload of a call centre depends on the service time of calls and possibly the time needed to solve the costumer's problems. Forecasting of this workload is done in (Aldor-Noiman, Feigin, & Mandelbaum, To Be Published). Another interesting paper on the topic is (Brown, et al., 2002), which handles several types of queuing problems concerning call centres, i.e. service time, queuing time, waiting time for service or abandoning and prediction of the load. With forecasting the workload, also the problem of queuing customers arises. Designing your call centre in such a way that customers do not have to wait long at each station and hang up the phone is important for your service level. The paper (Koole & Mandelbaum, 2002) contains a survey of queuing theory used in telephone call centres.

The final part of call centre research deals with planning. For a call centre the volume of calls can differ a lot between days, but also between the periods in a day. Making sure that there are enough agents at each period calls for good planning. Call centres typically employ a lot of students or other people that are flexible in their own planning. Furthermore there are constraints to the problem, because people have restrictions in work hours, knowledge, they need breaks and cannot work more hours than what they are contracted for. This optimisation problem is featured in a lot of literature, see e.g. (Atlason, Epelman, & Henderson, 2004)

In this thesis the different fields of research will be combined. First we will clean up the data with missing data analysis and outlier detection. The methods used in those chapters are not new compared to the literature. The suggested forecasting method is an extension to the know Holt-Winters method. This method has been mentioned in literature before, but it was not elaborated. Therefore we define the model for the multiplicative Holt-Winters as well as the additive Holt-Winters version. Furthermore a new method of finding initial values has been developed.

The Holt-Winters forecasting method typically has the problem of over-parameterisation. Finding appropriate parameters is hard and time consuming. Therefore a method is developed to speed up the determining of parameter values. The concept of Latin Hypercube Designs is used to find initial estimates and then a local search optimisation is used. This method makes finding parameter values really quick and therefore the entire forecasting method can be quickly conducted for various datasets. Furthermore as input to the Holt-Winters forecasting method we need to determine the seasonality of the data. A method is introduced to determine the length of the seasonal patterns.

3 MISSING DATA

The first aspect of the problem is to handle with missing data. In the call centre application, missing data can occur for various reasons. If there is a holiday, some projects might be closed, while other projects (e.g. an incident number for an electricity company) remain open. On the other hand, data can also be missing because of a failure in the system. This can be a complete system failure, i.e. the entire call centre does not receive any calls and therefore the number of calls in this period is zero. These first two causes of missing data can not formally be considered as missing data, since the observation for that time period should really be zero. In this thesis we will estimate the effect of holidays and complete system failures in the next chapter. For forecasting purposes we will exclude this data from the dataset and correct for this influence later on. Furthermore, a complete system failure (i.e. no incoming calls at all) had not occurred in the Sanoma project (and is generally extremely rare) according to Unamic/HCN

The final cause of missing data is a data recording system. In this case the call centre is working as scheduled, but the number of incoming calls is not recorded. This final case is the most pure form of missing data, since the only reason of missing data is the system failure. To deal with this missing data, we will first establish a general theory about missing data and the causes of missing data. Then we will introduce several methods of handling missing data and we will pick the best option for our situation. Theory for this chapter is based on (Schafer & Graham, 2002), (Jamshidian, 2004) and (Little & Rubin, 2002)

3.1 The Missing Data Problem

The paper by (Efron, 1994) defines missing data as a class of problems made difficult by the absence of some part of a familiar data structure. Missing data can occur in all data experiments. The first step is to determine if the data is really missing, i.e. the underlying value exists. In the call centre data, this can be illustrated by the different system failures. When a complete system failure occurs, the call centre cannot handle any calls and therefore the data for that time period is missing. In that case the underlying value doesn’t exist either. For a recording system failure, the underlying data does exist, because the only problem was recording the number of incoming calls.

In survey data the cause of missing data might not be that straightforward. It could happen that the researcher forgot to ask a certain question, or the respondent decided not to answer a typical question. Respondents can also decide not to answer a question because they are too sensitive (e.g. age, income, drug use) or because they don’t know. In longitudinal studies the problem of missing data is even more complicated. If respondents die during the study, you would consider the values of subsequent observations missing if the death is just coincidence. But for health surveys, the occurrence of a death might be of importance to the survey.

The problem with missing data is that statistical analysis and subsequent operations like forecasting are influenced by missing data. If the missing data remains in the dataset with value zero, the forecasts are less reliable. Leaving out the missing values can yield significant loss of data and therefore information. Furthermore it can influence the structure of the dataset.

Another method of handling with missing data is imputation. With this technique, the missing values are estimated and filled in into the dataset. In paragraph 3.4 we will discuss several methods of handling with missing data.

3.2 Missing Data Patterns

Data can be missing in several patterns. In this paragraph we will discuss the different patterns of missing data and give examples for the missing data. It is useful to determine the pattern because some methods of handling with missing data are intended for particular patterns. Other methods are suitable for any pattern, but are more time consuming.

Let $Y=(y_{ij})$ denote an $(n \ x \ K)$ rectangular data set without missing values, with the ith row $y_i = (y_{i1},...,y_{iK})$ where y_{ij} is the value of variable Y_j for subject i. With missing data, define the *missing data indicator matrix* $M = (m_{ij})$, such that $m_{ij} = 1$ if y_{ij} is missing and $m_{ij} = 0$ if y_{ij} is present. The matrix defines the pattern of missing data. Figures 1-5 show some examples of missing data patterns. With several examples we will illustrate the different patterns.

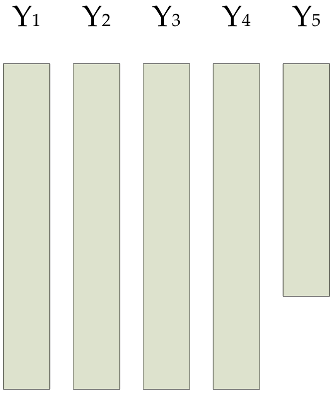


Figure 1: Single Univariate Missing Data

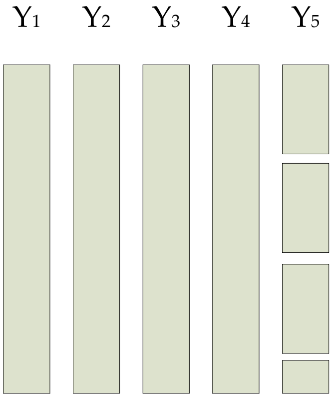


Figure 2: Call Centre Missing Data

The first pattern (Figure 1) is the pattern of univariate missing data. This occurs when observations for only one variable are missing. This is usually the case in designed experiments. All other variables are set by the researchers and only the fifth variable is observed to see the influence of the other variables. E.g. if there is a survey where you first ask whether someone has a job and later on you ask the income, the first question determines whether the second question will be asked. Interest is in the dependent variable Y_K on the set of factors $Y_1,......,Y_{K-1}$. In our application, we know the variables: *year, week number, day of the week, project* and *time period (in this case quarter of an hour time periods)* and try to estimate the *number of calls.* But the call centre dataset is not as straightforward as we might think. Because our data is a time series, the order of the observations matters as well. So the real pattern is better described by Figure 2.

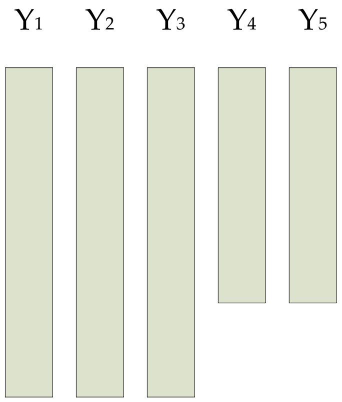


Figure 3: Multiple Univariate Missing Data

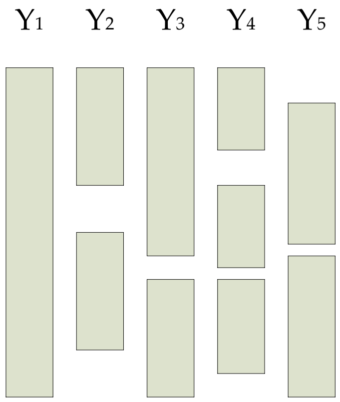


Figure 4: General Missing Data

Univariate missing data can also occur for multiple cases as illustrated in Figure 3. This could for example happen in a survey, where respondents don't have to answer parts of the survey. E.g. when there is the question: "Have you ever used drug?" and after a positive response some follow up questions are asked. If the respondent replied negatively, these follow up questions are never asked. Survey practitioners call missing values on particular items in the questionnaire *item nonresponse*. These missing values typically have a haphazard pattern, such as that in Figure 2 or Figure 4. These patterns can arise when people refuse to answer a question, or don't know the answer, or when a question is a follow up of a previous question. All these factors can lead to 'gaps' in our dataset.

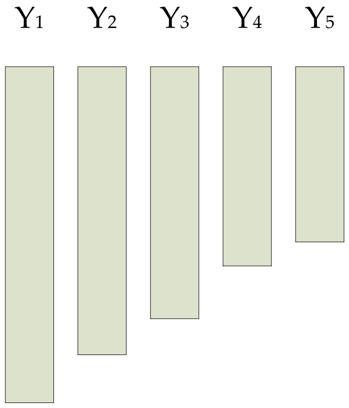


Figure 5: Monotone Missing Data

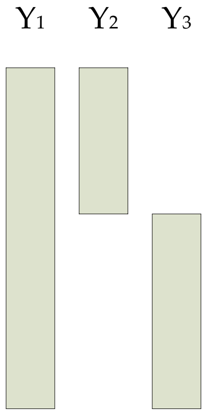


Figure 6: File Matching Missing Data

In longitudinal studies data is collected on a set of cases repeatedly over time. Respondents are asked to fill in a similar survey every time period. Researchers then examine this panel data to see if e.g. habits and views of the respondents have changed. During this study it could happen that respondents drop out of the survey, e.g. because they move, or, in a medical survey, they are cured. This phenomenon is called *attrition*. The pattern of attrition is an example of *monotone* missing data, where variables can be sorted so that for $Y_{j+1},\ldots,Y_K$ the value Y_j is missing, for all $j = j + 1, \ldots, K$(see Figure 5).

The final pattern we will discuss here is the File Matching problem. This problem occurs when two (or more) variables are never observed at the same time. This can happen when there is a large amount of data missing. It can also occur that a typical question (e.g. have you ever used drugs?) results in two different sets of questions (how often do you use drugs? Why haven't you tried it? Etc.). While working with this data, you should be aware of the fact that the relationship between these variables is not in the data. Making estimates based on relations to other variables may yield misleading results. Figure 6 illustrates such a problem. Problems like this could also arise because of the experimental design. If e.g. a medical experiment is conducted with two types of medicine, some variables might not be measured for all participants. Temperature might be relevant for the first group, because an increase in temperature is one of the side effects. For the other group it could be irrelevant and tests for other side effects will be conducted.

3.3 Missing Data Mechanisms

After discussing the patterns of missing data, we will look at the mechanisms that lead to missing data. It is especially of importance if the value of the underlying (missing) variable is related to the fact that the value is missing. By determining the missing data pattern and mechanism, we can choose the appropriate method of handling with missing data. Different missing data mechanisms require different methods. The different mechanisms are formalized in the theory of Rubin (Rubin, 1976). In this chapter we will describe the use of this theorem in modern statistical literature on missing data.

Define the complete data $Y = (y_{ij})$ and the missing data indicator matrix $M = (m_{ij})$ as defined in the previous paragraph. The missing data mechanism is characterized by the conditional distribution of M given Y, say $f(M \mid Y, \varphi)$, where φ denotes the unknown parameters. The easiest case is when missingness does not depend on any other data. This is defined as:

$$f(M|Y,\varphi) = f(M|\varphi) \text{ for all } Y, \varphi \qquad (3.1)$$

Data satisfying this equation are called M*issing Completely at Random* (MCAR) or *Completely Missing at Random* (CMAR). This doesn't mean that the pattern of missingness is random, but only that the fact that a value is missing doesn't depend on the data values. This is e.g. the case if the interviewer forgot to ask a question, or a thermometer fails to record the temperature because it is broken.

Let Y_{obs} denote the observed values of Y, and Y_{mis} the missing entries of the data set. Please note that for Y_{obs} we only know that the observation exists, the underlying value is not known (if it were known Y_{obs} would not be a random variable anymore). We define the next missing data mechanism:

$$f(M|Y,\varphi) = f(M|Y_{obs},\varphi) \text{ for all } Y_{mis}, \varphi \qquad (3.2)$$

This second definition is less restrictive than MCAR. Now the missingness only depends on the values of the observed variables and not on the values of the missing variables. This mechanism is called *Missing at Random* (MAR). In a survey with conditional questions (early questions that rule out some future questions) MAR often applies. A third missing data mechanism is *Missing Not at Random* (MNAR) or *Not Missing at Random* (NMAR). When a dataset is neither MCAR nor MAR, we consider the dataset NMAR. In that case, the missingness depends on the underlying value of the missing variable as well. This can e.g. happen when a participant refuses to answer a sensitive question about drug use. If the participant has never used drugs, he/she will answer the question, but frequent users might refuse to answer.

To illustrate these mechanisms of missing data we will use the following example. Consider the call centre with incomings calls for various projects. Missingness in the data can occur for various reasons. If e.g. the recording system failed to record the number of incoming calls at a certain period, while the call centre functions as normal, we call this MCAR. The cause of missingness is assumed to be an external factor and doesn't depend on the other data. Missingness can also originate from other causes. Because of unavailability of call centre agents due to high demand on other projects, a certain project might temporarily be shut down. There will still be incoming calls, but they will not enter the system and are therefore not recorded. In this case, the cause of missingness is the high number of incoming calls for other projects and is thus MAR. The mechanism of missingness MNAR can be illustrated as follows. If there is an extremely high demand for the observed project, the recording device might fail and therefore we have missing data. In this case the high number of calls causes the missing data and therefore the mechanism is MNAR.

In our application, missing values are MCAR. The reason data can be missing is because of a system failure, but these failures are not dependent on underlying data. This assumption is based on the experience of call centre agents. There haven't been other reasons for missing data in the past. So it is assumed to be impossible that there are so many calls at a certain moment that the system will shut down. Furthermore the hardware bounds of the system (i.e. maximum number of calls that can be handled by the hardware) are far larger than the typical peak volumes. Failures have (mostly unknown) software causes, which are external factors to the system.

3.4 Methods of Handling with Missing Data

For dealing with missing data several methods are available, ranging from less efficient to more efficient methods. In the following subchapters I will first describe *ad hoc* methods and then introduce Likelihood-based methods.

3.4.1 Case deletion

Case deletion methods are generally easy to use and require almost no effort. In this subchapter I will discuss the different methods and their advantages and disadvantages.

One of the oldest methods is *list wise deletion (LD)* or *complete-case (CC) analysis*. This method is still used in many statistical programmes by default. For further analysis only observations without missing values in any variable are observed, so if one of the observations is missing, the entire case will not be taken into account. If the data is MCAR, the complete cases are a random sample of the entire dataset. Estimates based on this subset are still consistent estimates of the entire set. In the case of sparse data, you will have to delete a lot of cases and therefore lose a lot of information. Because this subset is just a small part of the entire dataset the estimates might not be consistent. *Available-case (AC) analysis* in contrast to CC uses the largest possible set of available cases to estimate parameters. For example, we can use every observed value of X_j to estimate the expectation of X_j, and all observed combinations of (X_j , X_k) to estimate the covariance of X_j and X_k. Because AC uses more data for the estimation of the parameters, you would expect that, even in sparse datasets, the estimators are (close to) consistent. (Kim & Curry, 1977) Showed that for MCAR data with modest correlation between the variables, AC outperforms CC in parameter estimation. Other studies however show that, when correlation is large, CC yields more efficient estimates. See for example (Azen & Van Guilder, 1981).

Both *case deletion* methods (AC and CC) are generally valid under MCAR. For MAR or NMAR missing data, the parameter estimates may be inconsistent. Consider e.g. a study where education level and income are measured, but for some highly educated people the interviewer didn't dare to ask, or they refused to tell, how much they earned. For these observations the income data is missing, but because there is possibly a positive correlation with education level, we lose information. If we now estimate the average income based on the variables we have, it will yield a lower income than the real population income. So parameters can be biased if case deletion is used. The main reason why case deletion is still used is simplicity. When the number of missing values is low, this method will yield acceptable results, however when there is a large amount of data missing, the biased estimation will become a problem.

3.4.2 Imputation Methods

A second method for handling with missing values is imputation. This means that for the missing values an estimation will be made of the value of the missing data point. This estimation will then be used in the data. In this subchapter the different imputation methods will be described with their advantages and disadvantages.

Unconditional Mean Imputation (UMI) is one of the simplest imputation methods. The missing value is replaced by the mean of the known values for the specific variable (for categorical variables you could impute the mode). This method yields biased estimates of parameters. For time series this method however does not work, since there could be e.g. a trend in the data. If we estimate the variance, UMI underestimates the true variance. Since all missing values are imputed with a mean value, the variance of the dataset is reduced. Furthermore it can be that imputing the mean is not the best option if the cases in the survey differ a lot. If you for example want to impute the income of a certain person, you can impute the mean of the entire population, or you can impute the mean of the same group of people. This group can e.g. consist of people with the same age or gender. This greatly increases the performance of UMI if the population consists of several sub groups.

Conditional Mean Imputation (CMI) methods are developed to cope with the disadvantages of UMI methods. One of the most popular is Buck's method (Buck, 1960). First the mean and the covariance matrix are calculated using CC. The missing values are then predicted by regression models based on these parameters. This method usually outperforms the UMI method, since it is able to deal with e.g. a trend in the data. The estimated mean of the imputed values is a consistent estimate of the mean for MCAR data (Buck, 1960). In MAR data the CMI method underestimates the variance and covariance when the data is multivariate normal, but this underestimation is often less than when UMI is used.

Both methods mentioned in this subchapter underestimate the variance of the dataset. This is because we impute a best guess for that data point, instead of incorporating variability in the imputed variable. If the amount of missing values in the dataset is low, the effect of UMI or CMI on the variance is not significant. However, if the amount of missing data is high, the underestimation of the variance is more severe. *Hot Deck imputation* is one of the methods developed to deal with this problem. In this method the missing value is imputed by a random draw from the observed variables. So we take a random draw from all non missing observations of variable X to replace the missing values of variable X. The problem of underestimating variance is therefore partially covered, but still problems exist because correlation between variables is distorted. Again you can restrict the random draw set to observations from the same group (i.e. same age or gender) to make the imputations more reliable. Another method is: *imputing from a conditional distribution*, which can help deal with distorted covariances. We impute the missing variable by a random draw from the conditional distribution of Y given X. In a standard linear regression model, we add a residual error to $\hat{Y}$. This residual is a draw from a normal distribution with mean zero and variance equal to the residual variance of the regression model on de complete cases. More formally:

$$\hat{y}_{iK} = \tilde{\beta}_0 + \sum_{j=1}^{K-1} \tilde{\beta}_j y_{ij} + z_{iK} \tag{3.3}$$

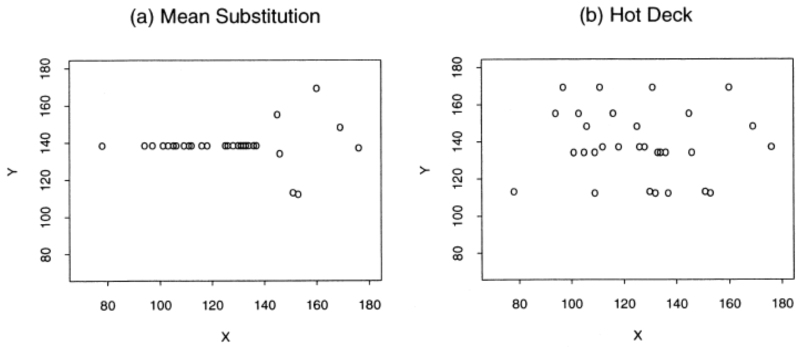
Where $(\tilde{\beta}_0,....,\tilde{\beta}_{K-1})$ are the parameter estimates of the regression model, with Y_K unknown. Z_{iK} is the normal variable with mean zero and variance equal to the residual variance mentioned above. This method can also be generalized for other distributions. Suppose we have data $Y = (Y_{obs},Y_{mis})$ from the distribution $P(Y_{obs},Y_{mis};\theta)$. Here note again that Y_{obs} is still a random variable. Imputing from the conditional distribution means simulation from:

$$P(Y_{mis}|Y_{obs};\theta) = \frac{P(Y_{obs,}Y_{mis};\theta)}{P(Y_{obs};\theta)} \tag{3.4}$$

Where the denominator is given by:

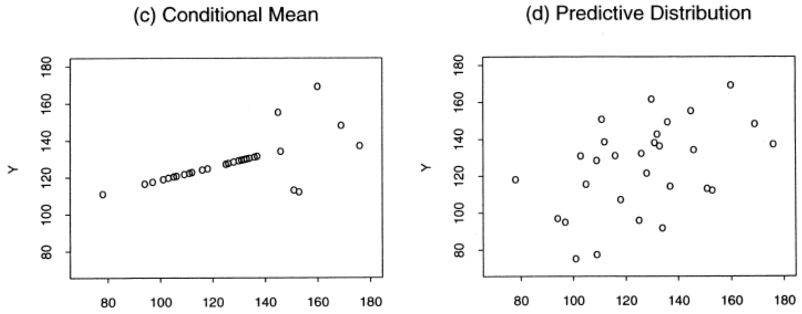
$$P(Y_{obs};\theta) = \int P(Y;\theta)dY_{mis} \tag{3.5}$$

The parameters (θ) for this distribution are unknown in practice and therefore have to be estimate from Y_{obs}. Imputation using this method yields good results, even if the missing data mechanism is MAR. However, the pattern of the data is really important here. If there is a univariate missing data pattern, the conditional distribution can easily be formulated, but for a general missing data pattern, it can be quite complicated. Drawing from such a complicated distribution can require just as much effort as more advanced (and better performing) methods discussed later. Figure 7 was taken from (Schafer & Graham, 2002) to illustrate the different imputation methods. These figures are based on a measurement of blood pressure on two occasions. The used missing data pattern was MAR and the amount of missing data was large. You can see that respondents with a high blood pressure in the first survey (X), also returned for the second measurement (Y). Using mean substitution or conditional mean substitution yields a dataset with far less variance than the original data. Hot Deck has better performance when it comes to variance, but the positive correlation between the variables cannot be seen in the data. Using a conditional distribution yields the best results for this dataset, considering the alternative methods.



(a) Mean Substitution

(b) Hot Deck



(c) Conditional Mean

(d) Predictive Distribution

Figure 7: The four imputation methods

3.4.3 Likelihood based models

The next class of methods for handling with missing data is based on the likelihood function. The likelihood function is a function that gives a value to the likelihood that the parameters have a certain value given the data. The likelihood function can be defined as follows:

$$L(\theta|Y) = f(Y|\theta) \text{ for any } \theta \in \Omega_\theta \tag{3.6}$$

With Ω_θ the possible values of θ. By definition $L(\theta|Y) = 0$ for any $\theta \notin \Omega_\theta$. In most problems we work with the *loglikelihood function* $l(\theta|Y)$, which is the natural logarithm (ln) of the likelihood function. The *maximum likelihood* (ML) estimate $\hat{\theta}$, is the value for which the loglikelihood is maximised (so it is most likely that the parameters have that true value). In large samples, this parameter estimation tends to be approximately consistent and efficient given that the distribution is correct. The expressions for ML estimates however, cannot be written down in most cases. Computing the ML estimates often requires an iterative process. One of the most widely used methods is the EM algorithm introduced by (Dempster, Laird, & Rubin, 1977). Each iteration of the EM algorithm consists of an E-step (estimation) and an M-step (maximisation). At each iteration step the loglikelihood $l(\theta|Y_{obs})$ increases (under general conditions). Also if the loglikelihood is bounded, the sequence $l(\theta^{(t)}|Y_{obs})$ converges to a fixed point $l(\theta|Y_{obs})$. A disadvantage is that the rate of convergence can be rather slow if there is a large fraction of missing data.

In most literature call centre arrival data is generally assumed to follow a Poisson distribution. The Poisson process expresses the probability of a number of arrivals occurring a fixed period of time. The arrival rate is known and subsequent arrivals are independent of each other. The probability density function of the Poisson process is given by:

$$f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{3.7}$$

Here λ is the arrival rate and $f(k,\lambda)$ will give the probability of k arrivals during one period. The inter arrival times between two consecutive arrivals follow an exponential distribution (by definition). The cumulative density function can be defined as follows.

$$f(x,\lambda) = \frac{\Gamma(\lfloor x+1 \rfloor, \lambda)}{\lfloor x \rfloor !} \; for\ x \geq 0\ or\ e^{-\lambda} \sum_{i=0}^{x} \frac{\lambda^i}{i!} \tag{3.8}$$

Where $\lfloor k \rfloor$ is the floor function and $\Gamma(x,y)$ is the upper Incomplete Gamma function, which is defined by:

$$\Gamma(x,y)=\int_{y}^{\infty} t^{x-1}e^{t}dt \tag{3.9}$$

This Incomplete Gamma function is almost the same as the normal Gamma function, except that the integral runs from y to ∞ instead of from *0* to ∞. Another interesting feature of the Poisson process is that the mean and variance are equal, namely λ. In call centre data, we also see that the variance and mean are equal for short intervals.

Next we need a Maximum Likelihood estimation of λ. Given a sample of n observations x_i, the log likelihood function is defined as follows:

$$\begin{aligned} L(\lambda) = log\prod_{i=1}^{n} f(x_i|\lambda) &= \sum_{i=1}^{n}\log\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right) \\ &= -n\lambda+\left(\sum_{i=1}^{n}x_i\right)\log(\lambda)-\sum_{i=1}^{n}\log(x_i!) \end{aligned} \tag{3.10}$$

After rewriting the log likelihood function we obtained equation (3.10). To solve this MLE we take the derivative of L with respect to λ and equate it to zero. This yields:

$$\frac{d}{d\lambda}L(\lambda)=0 \Leftrightarrow -n+\left(\sum_{i=1}^{n}x_i\right)\frac{1}{\lambda}=0 \tag{3.11}$$

Solving for λ yields the estimated mean. Obviously the second derivative is negative, since we assume that the arrival rate λ and the arrivals x_i are positive. The maximum likelihood estimator of λ then becomes:

$$\hat{\lambda}_{MLE}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{3.12}$$

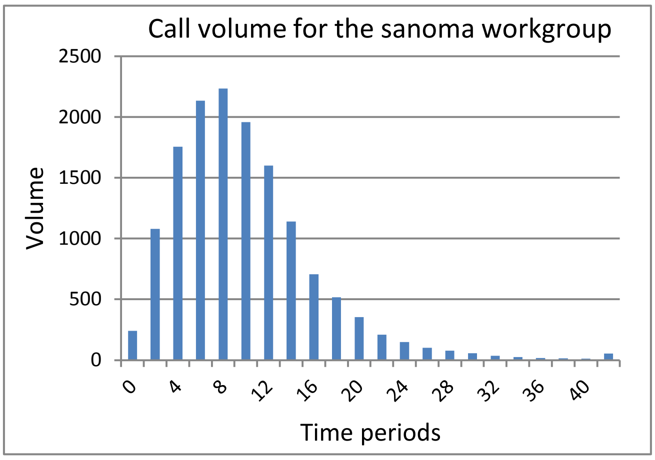
Which is the average of the observed arrivals. Every observation has expectation λ and the mean of this process is λ as well. Because the mean of the Poisson process is known to be λ this estimator is unbiased.

With this unbiased ML estimator, we can easily impute the missing values in our data.

3.5 Missing Data Analysis in call centre data.

In our data, we also have missing data. This can be for the entire day in case of a holiday (see Chapter 5) or system failure, but it can also be for certain time periods. In the data we have, there is no difference between a zero observation and missing data. It is thus important to find a way to indicate if a value is missing, or just zero. Especially in the evening, the call volume is a lot lower and can actually be zero. Most projects have a call volume between 200 and 400 a day, which leads to call volumes ranging from 0 to 20 (or even more) per time period. Because the call centre is open from 9.00 to 20.00 for the Sanoma projects, there are 44 (15 minute) time periods.

In our dataset (we use project 192 for the analysis in this chapter, see 0 and project 188 and 242 for verification), there is not much missing data. About 10% of the periods have a 0-value. The missing data analysis shown below only indicated 1% of total data as really missing and the other 9% as 0-values.



To detect the difference between a missing value and a 0-observation, we will first estimate the probability of a certain occurrence. To calculate the probability, we first have to make some assumptions on the call arrivals. In most literature, call arrivals are modelled as a Poisson arrival process. To see if this is also the case in our data, we plotted a histogram. You can see the call volume for the entire Sanoma workgroup (so all sub projects together) for the different time periods. This histogram follows the typical distribution of a Poisson arrival process. To prove this, we need the Likelihood Ratio Statistic test. We want to test H_0 vs H_1, where H_0 states that:

$$H_0 \colon X_i \sim Poiss(\lambda_i),\ \ \lambda_1 = \cdots = \lambda_n$$

And H_1states that:

$$H_1 \colon X_i \sim Poiss(\lambda_i), \qquad \sum \left(\lambda_i - \bar{\lambda}\right)^2 > 0$$

So we test whether the arrival rate λ is equal for a set of X_i. To determine the set of X_i we first have to make some assumptions. Because the call centre data changes rapidly across periods (in the evening call volume is significantly lower) and across days (call volume can differ per month or week) we have to define the set of X_i. For each data point in our matrix, we determine its neighbourhood. For that neighbourhood we test if H_0 holds. We assume the neighbourhood of the observations is at most 6 time periods before and 6 time periods after the observation, and at most 8 days before and 8 days after the observation. The way we determined these numbers will not be discussed right now, but will be explained later in this chapter. Once we determined the neighbourhood, we can test whether de observations are from a Poisson distribution. The Likelihood Ratio Test (Brown & Zhao, 2001) is defined as follows:

$$T_{LR} = 2\sum_{i=1}^{n} X_i \ln\left(\frac{X_i}{\bar{X}}\right) \tag{3.13}$$

Under the null hypothesis this statistic is asymptotically distributed as a Chi-Squared variable with $n - 1$ degrees of freedom. Hence, this test reject H_0 when $T_{LR} > \chi^2_{\mathrm{n-1;1-\alpha}}$

For our project (192) we construct the different neighbourhoods. So for every value in the project we construct the corresponding neighbourhood. For these neighbourhoods we calculated the Likelihood ratio and compared it to the Chi-Squared distribution with $\alpha = 0.05$. The result was that out of 5588 neighbourhoods, only 282 neighbourhoods reject H_0. This is only 5% of the number of neighbourhoods, which falls within the accuracy set in advance. We can therefore assume the arrival rate follows a Poisson distribution on the neighbourhood of the observation.

Now we can say that the arrival process follows a Poisson distribution on the neighbourhood of the observation, we can start by finding missing values. Missing values can arise when the recording module fails to record the number of incoming calls correctly. This means that for a certain time interval, the number of calls is recorded as a 0 value. It might also be possible that the incoming calls are no longer recorded within an interval (e.g. halfway they don't record anymore), but that cannot be seen from the data. If that is the case, we might find those anomalies in chapter 4. For now we assume that missing data only consists of 0-observations. We also observed that missing data usually spans more than one period. System failures to record the data generally takes longer than fifteen minutes to solve. We therefore make a subset of the data, where we observe all 0-sequences (note that a 0-sequence can also consist of just one observation). For this sequence we need to know the neighbouring observations, and the number of 0-values. With this subset of sequences, we can determine if a sequence of 0-values is just coincidence or really a missing value. First we determine the arrival rate of the Poisson process for that missing data point. This is done by calculating the average of the neighbouring observations. Neighbouring observations are:

-up to 6 time periods before the 0-sequence (if possible)

-up to 6 time periods after the 0-sequence (if possible)

-up to 8 observations of the same time period the days before (if possible)

-up to 8 observations of the same time period the days after (if possible)

These numbers are obtained in the following way. For each entry in a sample dataset of project 192, we made an estimation based on the number of neighbouring observations. E.g. with a scope of 3 on the same day and 5 in the same period we calculated the average over at most 16 observations. This was calculated for 1 to 10 time periods on the same day and 1 to 10 observations of the same time period. So we then have 100 matrices which estimate the initial dataset. With all these matrices, we should determine which choice for neighbouring observations yield the best result. Therefore we will introduce several measures to determine the difference between the real observation and the estimation. The first we will introduce is the Mean Squared Error, which is defined as:

$$MSE(\hat{x}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - x_i)^2 \tag{3.14}$$

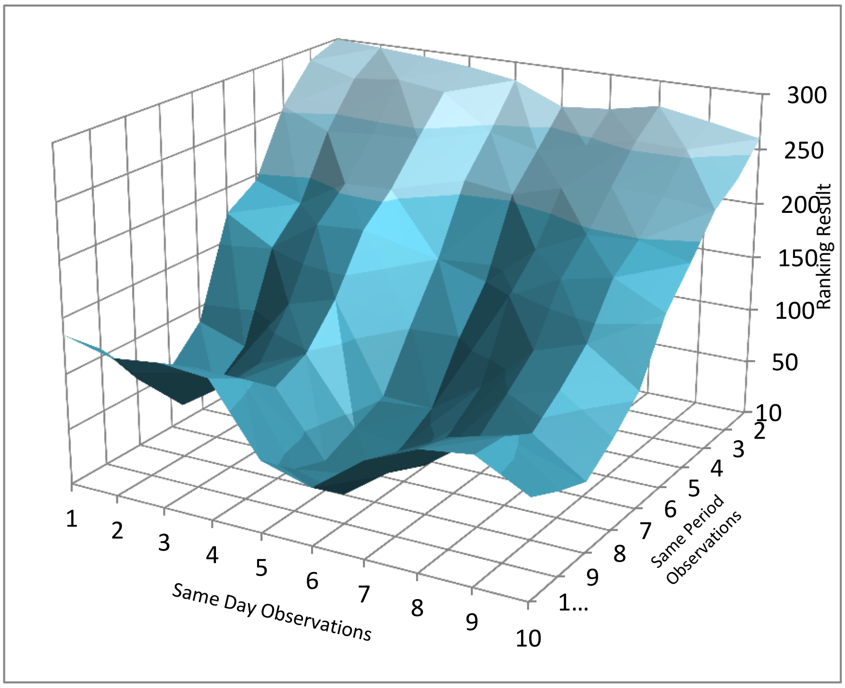
Where $\hat{x}_i$ is the estimation of x_i and n is the total number of observations in the dataset (in our sample database this was 5588). Because the differences in observations are squared, large deviations from the real value are punished more severely. Please note that this definition only holds when all observations have the same probability of occurrence. If that is not the case an adaptation should be made to weigh the observations and their errors. Another measure is the Mean Absolute Error, which is defined as:

$$MAE(\hat{x}) = \frac{1}{n}\sum_{i=1}^{n}|\,\hat{x}_i - x_i| \tag{3.15}$$

In contrary to the MSE, this measure first takes the absolute difference before taking the mean. This causes large differences between the estimation and the real value to be measured less strict. The third measure is defined as follows:

$$MPE(\hat{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{x}_i - x_i|}{x_i} \tag{3.16}$$

This is the Mean Percentage Error. Here we not only look at the difference between estimation and real value, but we take the relative difference into account. An error in forecasting of 5 calls is not so severe if the call volume is 100, but it is severe if the call volume is just 10. Especially when the call volume is changing a lot, this measure gives a better view on how large the errors are. The outcome is also more relevant. When you look at the outcome of the MSE or MAE, the number will not directly tell you something about the effect of errors. They can be used for comparing different models, but the number does not yield any information. An MPE of 0.15 means that the mean error is approximately 15%.



With these three measures we have to determine which choice for the neighbouring observations is most suitable for our problem. Since it is not directly clear which measure is most suitable for our data, we therefore want to incorporate all measures in the decision process. First we rank all options based on their performance. E.g. if a certain choice has the lowest value for the MSE, then it gets ranking 1. This is done for all three measures and all choices for the neighbouring observations. Finally, we add all these rankings, to obtain a number which indicates the performance. In Appendix B you can find the table with the total ranking results. On the right you can see a 3D plot of the resulting rankings. Darker colours mean lower ranking scores. The lowest part of the graph is on the coordinates (8,6), which corresponds to looking at 8 observations of the same time period, but a different day, and 6 observations of the same day, but a different time period. This parameter setting 'scores' 24 points, which consists of 3rd place in MSE results, 2nd in MAE results and 19th in MPE results. Because this setting has a bad score for the MPE measure we looked at the other settings which do score well for the MPE measure. However, on the settings where MPE scores well, both the other measures score badly. Furthermore the difference in MPE between the best setting and the introduced (8,6) setting is small (~1,5%).

As shown in the previous paragraph, the average of the observations from the same distribution is a MLE for the arrival rate of the Poisson process. It can be argued that these neighbouring observations are not from the same distribution, since the arrival rate of the process changes over time. However within the neighbourhoods the change in arrival rate is small and we therefore can assume that they follow the same Poisson distribution. For this size of the neighbourhood we also showed that the data follows a Poisson distribution.

With this estimated arrival rate, we estimate the likelihood of the 0-sequences. With the cumulative density function of the Poisson process, the probability of a 0 observation will be calculated. If a sequence consists of more than two 0-values, the probability of two 0-values occurring after each other will be calculated. With this probability we determine if a sequence is indeed a missing value. If the probability is below 0.0001, we mark the sequence as a missing value.

After we have determined which values are missing and which values are just 0-observations, we can impute the missing values. With the arrival rate determined before, we will impute the missing values with a random draw from the Poisson process. Because we use a random draw, the variance in the dataset will not get lost. Now we have a complete dataset and therefore we can continue our analysis.

4 ANOMALY DETECTION

In this chapter we will investigate extreme values found in the data. An extreme value is an observation which is strange or unusual compared to the rest of the data. In statistics anomalies are all observations that do not conform to the established normal behaviour of the process. This can e.g. be a different pattern or an extreme value. In this chapter we will use the term anomaly detection to indicate the search for extreme values. Extreme values can be an extremely high or low call volume for some time period. Before we continue with the anomaly detection, we will first have to deal with the missing data of chapter 3.

Anomalies can have several causes. Because of system failure, it could happen that for a certain time period the number of calls are no longer recorded. Since the missing data detection and correction only works for time periods where the number of calls was actually 0, we were not able to detect these anomalies. High values can also arise in several software errors, e.g. when the volume recording fails to identify the sub project and incorrectly assigns calls to a certain project. It can also be the case that for some strange occurrence (e.g. 9-11) the number of calls is significantly higher or lower for some time period.

With our anomaly detection we follow a similar approach as in the missing data chapter. For each observation in the dataset, we will look at the neighbouring observations (6 time periods before and after the observation on the same day, and 8 days before and after the observation of the same time period). For this neighbourhood we have proven that the data follows a Poisson distribution. Then we estimate the arrival rate for each point in the dataset by the average number of observations of their neighbourhood (which again is the MLE of the arrival rate of a Poisson process). The next thing we should do is compare the true number of arrivals with the expected number of arrivals. This is done by calculating the probability of the true number of calls, under the assumptions that arrivals are Poisson distributed with arrival rate λ equal to the average number of observations in the neighbourhood. We indicate an observation to be an anomaly if the probability of that occurrence is smaller than 0.0001. The final thing we do is impute the anomalies with a random draw from the Poisson distribution. This deletes the anomaly from the data and increases the predictive power of the forecasting model. In our analysis, the number of anomalies was usually less than 0.1% and therefore the impact of changing these values is small as well. If there would have been a large amount of anomalies, this is a feature of the data and cannot be omitted that easily. With this corrected dataset, we can continue our analysis.

In the data we only found anomalies of extremely high call volume. There were no anomalies of low values, so a system failure during a period has not caused extremely low values. A likely explanation is that the impact of system failure is small compared to the call volume, and therefore doesn't yield significantly different values. It can also be that when a system error occurs during a time period it will forget the number of calls up until then and generate a missing value (similarly than when the process is working again it begins recording the volume for every complete period)

Anomalies can also be detected on a day level. It is more likely that a certain event or system failure has a larger impact than only for some time periods. Especially when there is some strange occurrence, like 9/11, it can change the call volume for the entire day. Therefore, we also look at the difference between days. To determine the average calls for a certain period, we first have to estimate the size of the neighbourhood. First we will look at the observations directly before and after the particular arrival rate we want to forecast. We will not make a distinction between days, so to estimate e.g. the arrival rate for a Monday; we will use data from the other days as well. Next, we will again estimate the neighbourhood, but now we will only look at the same day. So forecasts for Monday will only be done with data from other Mondays.

Analysis for this is done on project number 192 (see Appendix A) and for verification project 188 and 242 is used as well. We start off with estimating arrival rates for each day, by looking at the neighbourhood without considering the day of the week. To determine the best size of the neighbourhood, we calculate the arrival rates for several sizes of the neighbourhood. As before, we look at 1 up to 10 days before and after the observation. So if we want to estimate the value for, say 5 June 2007, with the size of the neighbourhood set to 4, we look at 4 (working)days before 5 June and 4 days after 5 June. The average of these observations is then the estimate for the daily arrival rate. To determine the best value for the size of the neighbourhood, we will again calculate the MSE (3.14), the MAE (3.15) and the MPE (3.16). The results can be seen below:

Number of observations before and after										
	1	2	3	4	**5**	6	7	8	9	10
MSE	943.6	915.3	993.5	998.1	**900.3**	946.2	996.0	1040.6	1058.5	1041.6
MAE	22.39	22.42	23.26	22.81	**21.22**	21.67	22.44	23.03	23.10	22.72
MPE	0.182	0.186	0.191	0.186	**0.171**	0.174	0.180	0.185	0.185	0.182

As can be seen, the best size for this neighbourhood is 5. However, when we examine the data more carefully, we can improve the way the neighbourhood is defined in order to make the arrival rate estimation more accurate. If we calculate the MSE, MAE and MPE separately for each day, we see that on Mondays the errors are much larger (see Appendix D). These errors are the main contributors to the overall errors. Call volume on Monday is much larger than in the rest of the week, so estimating the Monday call volume by other days might not be a good idea (it is no surprise that a neighbourhood of size 5 yields the best result, since that will incorporate two Mondays into the average). The second method we will examine will estimate call volumes by looking at neighbouring observations of the same day. So only Monday data will be used to estimate the Monday arrival rate. The results can be found in Appendix D and overall results are in the table below:

Number of observations before and after										
	1	**2**	3	4	5	6	7	8	9	10
MSE	1115	**1086**	1146	1188	1259	1314	1380	1399	1431	1484
MAE	22.94	**22.19**	22.82	23.29	24.20	25.09	25.89	26.24	26.78	27.29
MPE	0.188	**0.181**	0.183	0.183	0.191	0.199	0.205	0.208	0.212	0.217

As you can see the overall value of the different error measures has increased. So this method of estimating the arrival rate of data within the dataset yields worse results. However, we can see that for Monday data the error measures yield better results (see Appendix D). Because the call volume is higher on Monday's the best estimation is done by looking at the observations of other Monday's. For the other days this doesn't hold. Since the call arrival pattern changes every period, i.e. within day changes and changes over weeks/months, we need to make a trade-off. On the one hand, we can use only data of the same day. This will give us good results if that day is significantly different from the others. The downside for this approach is that the length of the time period you are using for the estimation is longer, and since the rate changes a lot, this might give you worse results. On the other hand, we can use data from other days, but close to the estimated observation to make the estimation. This makes the time period over which the estimation is made smaller and therefore is less sensible to changes in rate over time. However, this method can only be done when the days are comparable. For our application, we have seen that looking at the same day is beneficial for Monday data and not for the other days. Therefore, we introduce a method to where we estimate arrival rates for Mondays by using the data of other Mondays and for the other week days, we will use data from all other days (except Mondays). The results are again given in Appendix D and the overall results are given below:

Number of observations before and after										
	1	**2**	3	4	5	6	7	8	9	10
MSE	862.9	**800.6**	849.6	868.7	925.5	950.5	981.1	995.9	1034	1081
MAE	20.32	**19.08**	19.53	19.5	20.26	20.71	21.25	21.49	21.92	22.35
MPE	0.167	**0.155**	0.157	**0.155**	0.16	0.164	0.168	0.17	0.173	0.176

As you can see, this decreases the values of all error estimators, and is therefore the best method of the methods we have seen. We see that the size of the neighbourhood is still a discussion, since the MSE and MAE indicate that the best size of the neighbourhood should be two observations before and two after, but the MPE indicates that the size should be four (though not very different from two). This can be better investigated by looking at the error measures for each day separately (See Appendix D).

Since Monday's arrival rate is estimated by other Mondays, this process of estimation is sensitive to changes in rate over time. In the data we can see that therefore the neighbourhood should be of size two. So we will only look at the two weeks before and the two weeks after the observation to make an estimate. Since we will only look at four observations, the estimate might be bad, but given that this yields the best results in our data, we will continue to use this. For the other weekdays this is different. Since the rate doesn't change that fast within a week, having more observations to base your estimate on seems beneficial. In the appendix you can see that having a neighbourhood of size three yields the best result. The results do vary depending on day and measure, but the use of a second dataset shows that in general using three as the neighbourhood size yields the best results. To estimate the arrival rate, we therefore use three observations before the rate we want to estimate (except Mondays) and three observations after the rate.

Now that we have determined the neighbourhoods of the data, we can continue with the anomaly analysis of the data. This is done in a similar way as before when we where detecting anomalies in the data on quarter level. First we determine which values can be considered anomalies in comparison to the Poisson distribution and the estimated arrival rate. If the probability of a value is really low (i.e. 0.0001), than we replace the value with a random draw from the Poisson distribution. Note that the chosen value for the probability is very low. This is because we don't want to replace a lot of values. The call volume is really volatile and we do not want to smooth out all irregularities. Therefore an anomaly will only be replaced if it is highly unlikely to have occurred. Furthermore, after a replacement has been made, the neighbourhoods will be recalculated. This is because an extremely high value has had its influence on the neighbourhoods of other observations, which can cause that other observations are considered outliers as well. If we change an extremely high value and recalculate the neighbourhoods of other observations, there is no longer any influence of that outlier. In the datasets we tested the occurrence of outliers varied a lot. In the beginning of a new project a lot of outliers where detected. This is probably due to the fact that the automated telephone choice menus are redesigned and flows of callers therefore change. In the first five weeks about 10% of the daily call volume was considered an outlier. Once the project had run for some weeks the amount of outliers dropped to about 1%. With this corrected dataset we can continue our analysis to forecast the call volume

If we found an anomaly on day level and have generated a random draw from the Poisson distribution to replace the volume for that day, this volume also needs to be distributed of the time periods. For a Monday we use the call volume of the Monday the weeks before and the weeks after to estimate the average. We then need to distribute the call volume for that day over the distinct time periods. This is done by calculating the pattern of the call volume for Mondays, see chapter 7 for the method to determine this pattern. With this pattern, we can then distribute the call volume over the time periods and can continue our forecasting procedure.

5 HOLIDAY INFLUENCE

After we filtered out the missing data and outliers, we will look at the influence of holidays in our data. The call centre we investigate is closed for some days a year, e.g. during Easter and Christmas. In this chapter we try to investigate the influence of these holidays on the data. If for example Monday is Easter and therefore the call centre is closed. We want to examine whether people, who would have normally called on that Monday, will call back and when. Perhaps they call on Tuesday or Wednesday, or they might call next Monday, because Mondays are most convenient for them.

We will conduct several analyses on the influence of holidays. These analyses will be done by a simple linear regression model to test the influence of holidays on the remainder of the week. Results for all testing procedures can be found in Appendix C. The first test we will conduct is testing whether all holidays have an influence on the call volume for the other weekdays. The following regression model was used:

$$Calls_i = \beta_0 + \beta_1 * Holiday_i + \epsilon_i \tag{5.1}$$

With *Calls* the vector of call volume for each day and *Holiday* a dummy variable which indicates whether there is a holiday in that certain week. Furthermore i indicates the position in the vectors. In this first model, all holidays are included. For this dataset (project 188) that means we include: Ascension 07, Pentecost 07, Christmas 07, New year 08, Easter 08, Queens day 08, Ascension 08, Pentecost 08, Christmas 08, New year 09, Easter 09, Queens day 09 and Ascension 09. In Appendix C you can see the results of this regression. Here the influence of *Holiday* yields a significant result. However, we are not completely convinced about the influence of holidays, since our model is still very basic. This can also be seen by the low value of R^2. Therefore we elaborate our model to the following:

$$Calls_i = \beta_0 + \beta_1 * Calls_{i-1} + \beta_2 * Holiday_i + \epsilon_i \tag{5.2}$$

In this new model we added a lag variable to include autocorrelation into the model. Results of this regression can again be found in Appendix C. Here we see that *Holiday* does not have a significant impact. The effect of the lag variable covers most of the variability. Also the R^2 has risen quite a lot.

Since there is also a difference between weekdays (Monday e.g. has a higher volume than a Friday) we elaborate our model to cope with these influences as well. This yields the following model:

$$\begin{aligned} Calls_i = \beta_0 + \beta_1 * Calls_{\mathrm{i}-1} + \beta_2 * Holiday_{\mathrm{i}} + \beta_3 * Monday + \beta_4 \\ * Tuesday + \beta_5 * Wednesday + \beta_6 * Thursday \\ + \epsilon_i \end{aligned} \tag{5.3}$$

This model includes dummy variables for the days. The dummy variables (*Monday, Tuesday, Wednesday and Thursday*) are 1 when it is that day and 0 otherwise. We do not include a dummy for Friday, since that would cause multicollinearity. Multicollinearity can arise in regression models when two or more variables are highly correlated. Again the result of this regression can be found in the appendix.

Here we can again see that *Holiday* does not have a significant influence on the number of calls in a certain period. For other projects in the dataset this was also tested and no significant results were found. Furthermore we also tested whether call from one Monday (e.g. Easter Monday) are transferred to the next Monday. Here again we could not find any significant results. An explanation to this could be that the effects of a holiday cancel out. On one hand, people could call less because they are away during the holidays. On the other hand, they could call more because they have a day off. In the remainder of this thesis we will therefore not take the influence of holidays into account.

We do however need to correct for the holidays. Because there is no call volume on the holidays, these days can be seen as missing values for each time period during that day. In the remainder of the thesis we will use a model which needs data in a specific format. Every working day (Monday till Friday) should be represented, even if it is a holiday. That way we can have a season of length five representing the influence of weekdays on call volume. To make the data complete again, we therefore impute values for the holidays the call centre was closed. This is done in the same way as was done with the anomaly detection. For Easter Monday we use the call volume of the Monday the two weeks before and the two weeks after to estimate the average. For the other weekdays, we estimate the average call volume by the three working days before and after that specific day (excluding Mondays). We then need to distribute the call volume for that day over the distinct time periods. This is done by calculating the pattern of the call volume for that day, see chapter 7 for the method to determine this pattern. With this pattern, we can distribute the call volume over the time periods and continue our forecasting procedure.

6 FORECASTING

After all the data cleaning processes we will continue with forecasting future call volumes. Forecasting is done in almost every industry. Whether it is a forecast for the sales in a shop or factory, the failure rate of a machine in a production process or the number of calls in a call centre, forecasting is useful in many ways. In this chapter we will investigate the forecasting method most suited for call centre forecasting. This method should be able to work with seasonal data and increasing or decreasing call volumes. First we will introduce several forecasting methods that are used in practice.

6.1 Forecasting methods: Moving averages and Smoothing methods

In this sub chapter several forecasting methods will be discussed, including their advantages and disadvantages. We start with the most basic forecasting methods. For more reference about these forecasting methods see (Hanke & Wichern, 2009) or (Brockwell & Davis, 1996)

6.1.1 Simple Average

Calculating a forecast based on Simple Average is done with the following equation:

$$\hat{Y}_{t+1} = \frac{1}{\mathrm{t}}\sum_{\mathrm{i}=1}^{\mathrm{t}} \mathrm{Y_i} \tag{6.1}$$

This equation is rather straightforward. The forecast for the next observation $(t + 1)$ is the average over all previous observations. This makes it an easy forecasting method, but doesn’t incorporate seasonal effects or trend effects. If the volume is increasing this method will underestimate the real call volume, since the forecast is merely an average over past (lower) volumes.

6.1.2 Moving Average

A similar forecasting method doesn’t look at all previous observations, but only looks at the last k observations. This yields the following equation:

$$\hat{Y}_{t+1} = \frac{\mathrm{Y_t} + \mathrm{Y_{t-1}} + \cdots + \mathrm{Y_{t-k+1}}}{\mathrm{k}} \tag{6.2}$$

This forecasting method is more sensitive to an increase (or decrease) in volumes. It will perform better than SA in case of an increasing volume; however it will still underestimate the real volume. If the call volume is strictly increasing the Moving Average forecast will always underestimate the next volume, since it is the average of past observations, which are all lower.

6.1.3 Double Moving Average

To deal with the problem of forecasting errors when the data follows an increasing trend (second derivative > 0), the method of Double Moving Averages can be used. As the name implies, two sets of moving averages are computed. First we compute the two moving average by:

$$\mathrm{M_t} = \hat{Y}_{t+1} = \frac{\mathrm{Y_t + Y_{t-1} + \cdots + Y_{t-k+1}}}{\mathrm{k}}$$

$$\mathrm{M'_t} = \frac{\mathrm{M_t + M_{t-1} + \cdots + M_{t-k+1}}}{\mathrm{k}}$$

The first moving average estimates the level the same way the normal Moving Average does. The second formula then calculates the average forecasts for the previous periods. With these two series of moving averages, we can calculate the forecast:

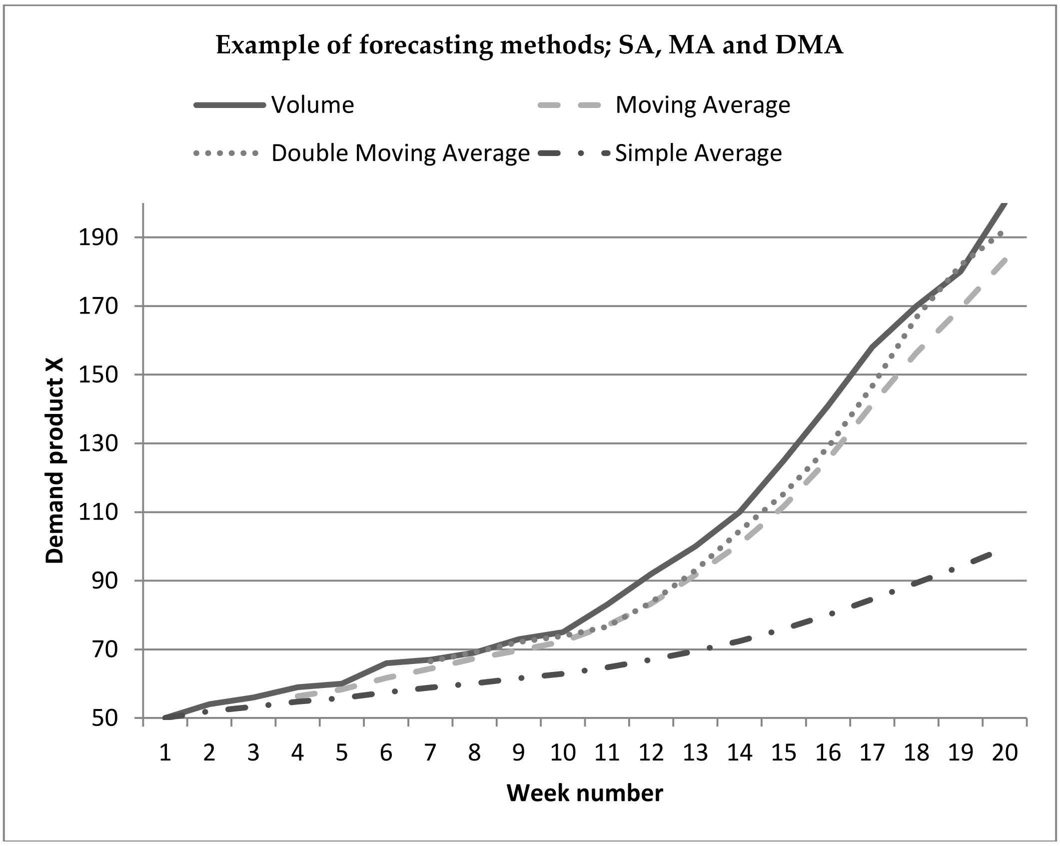
$$\hat{Y}_{t+p} = \mathrm{a_t + b_t p} \tag{6.3}$$

Where

$$a_t = M_t + (M_t - M'_t) = 2M_t - M'_t$$
$$b_t = \frac{2}{k-1}(M_t - M'_t)$$

And p denotes the number of periods ahead you want to forecast. a_t is the level forecast. It consists of the normal MA forecast and the difference between the current forecast and the average of previous forecasts. That way, if there is an increase in forecasts, the new forecast will be corrected. The second component b_t is the trend correction. It adjusts the new forecasts with the difference in the new forecast and the previous forecasts. This also makes that the Double MA forecasting method can deal with increasing trend processes. A disadvantage of this method is that all previous observations within k have the same weight and there is no seasonality in the data.

In the graph below you can see an example of the forecasting methods described above for some random dataset. K is chosen to be three is this example. As you can see the Simple Average underestimates strongly. Because of the low value of k, the Moving Average forecast does underestimate, but not severely. These forecasting methods are used to estimate the demand of a certain product X in a number of weeks. Because the demand is increasing more than linearly, most of the methods underestimate the demand for the next period. Only the double moving average forecast is able to forecast the series quite accurate.



6.1.4 Exponential Smoothing

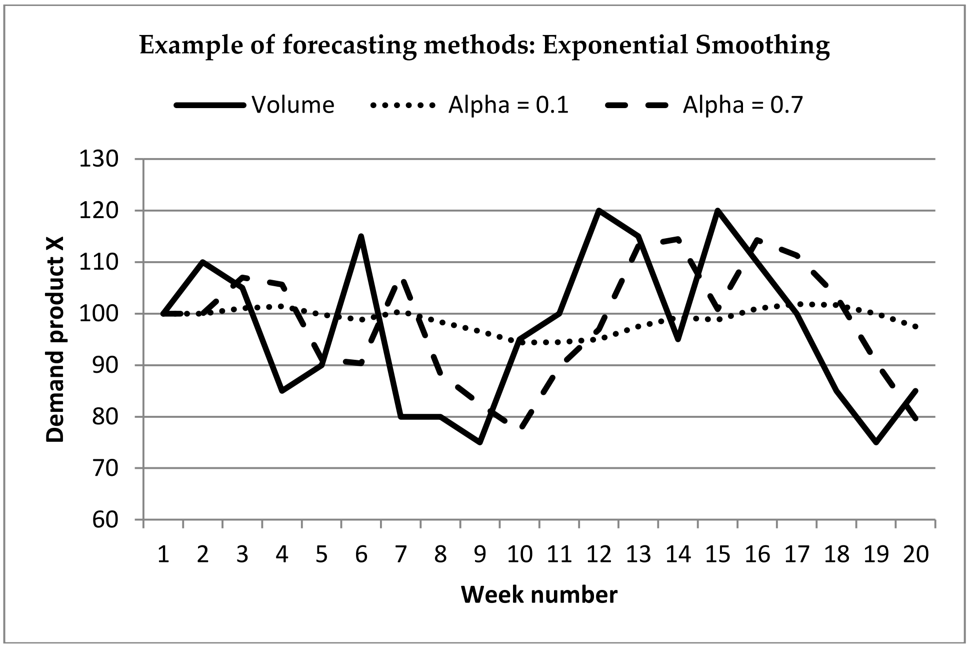
Another method to forecast time series is Exponential Smoothing. The equation is as follows:

$$\hat{Y}_{t+1} = \alpha \mathrm{Y_t} + (1-\alpha)\hat{\mathrm{Y}}_{\mathrm{t}} \tag{6.4}$$

So the next forecast depends on the previous value and the previous forecast. The parameter α has a value between 0 and 1. This forecasting method gives a recursive forecast, where the initial value Y_0 has to be set. If we set α to a large value, most of the weight will be put on the observation of the previous (most recent) period. A small α corresponds to a long 'memory' i.e. the forecasting process puts more weight on the history of the volume. This can be seen if we write out the equation:

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \alpha(1-\alpha)^3 Y_{t-3} + \cdots + (1-\alpha)^t Y_0$$

If α is large (say 0.7) most of the weight is on Y_t, however if α is small (say 0.1) more weight is put on observations from the past. In the first case the weights of the formula above are respectively (0.7;0.21;0.063;0.019;0.006 etc) and in the second case (0.1;0.09;0.081;0.073;0.066 etc). Hence you can see that the weight of past observations diminishes quickly if α is large. Also note that when the amount of data is small, a lot of weight can be on Y_0. This is because the sum of the weights should be equal to one, and with a large α it takes quite some periods before we approach one. In the example below you can see that a large α responds quicker to a change in the data and a small α smoothes the forecast more. In a rapidly changing environment the forecast using a large α is always a step behind the real volume as can be seen in the graph. In this example we choose $Y_0 = \hat{y}_0$, so the first estimate is equal to the first observation.



The best suited value for α can be chosen by calculating the MSE, MAE or MPE as described in chapter 3.5 and choosing the best performing α.

6.1.5 Exponential Smoothing adjusted for trend: Holt's Method

When data shows a clear trend, regular exponential smoothing will constantly underestimate the real volume. Therefore Holt (Holt, 1957) developed a method to deal with trend forecasting. Holt's procedure makes an estimate of the slope as well as the current level. Both the level and slope are smoothed by exponential smoothing with a different smoothing constant. The estimates for level and trend therefore change as more information becomes available. Because the smoothing constants for level and trend can be chosen, Holt's method is really flexible. The equations used for Holt's method are:

The level estimate:

$$\mathrm{L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1})} \qquad (6.5)$$

The trend estimate:

$$\mathrm{T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}} \qquad (6.6)$$

The forecast for p periods ahead:

$$\hat{Y}_{t+p} = L_t + pT_t \tag{6.7}$$

Where

L_t = The new smoothed value (level estimate)
α = The smoothing constant for the level $(0 < \alpha < 1)$
Y_t = The actual volume in period t
β = The smoothing constant for the trend $(0 < \beta < 1)$
T_t = The trend estimate
p = The number of periods to be forecasted ahead
Y_{t+p} = The forecasted volume for p periods ahead

The starting values L_0 and T_0 are set to Y_0 and 0 respectively. So the level estimate is set to the first observation and the trend is set to 0.

6.1.6 Exponential Smoothing adjusted for trend and seasonal variation: Holt-Winters' Method

To deal with data with a seasonal pattern, Winters developed an elaboration to the previously introduced Holts method. More details can also be found in (Winters, 1960) and the original paper (Holt, 1957). This forecasting method makes an estimate for the level, trend and seasonality. These three components are all weighted by their own smoothing constant. The four equations that define Holt-Winters' smoothing are:

The level estimate:

$$\mathrm{L_t} = \alpha \frac{\mathrm{Y_t}}{S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{6.8}$$

The trend estimate:

$$\mathrm{T_t} = \beta(\mathrm{L_t} - \mathrm{L_{t-1}}) + (1-\beta)\mathrm{T_{t-1}} \tag{6.9}$$

The seasonality estimate:

$$\mathrm{S_t} = \gamma \frac{\mathrm{Y_t}}{L_t} + (1-\gamma)S_{t-s} \tag{6.10}$$

The forecast for p periods ahead:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \tag{6.11}$$

Where

L_t = The new smoothed value (level estimate)
α = The smoothing constant for the level $(0 < \alpha < 1)$
Y_t = The actual volume in period t
β = The smoothing constant for the trend $(0 < \beta < 1)$
T_t = The trend estimate
γ = The smoothing constant for the seasonality estimate $(0 < \gamma < 1)$

S_t = The seasonality estimate
p = The number of periods to be forecasted ahead
s = The seasonality length
Y_{t+p} = The forecasted volume for p periods ahead

The initial conditions for this recursion are as follows:

$$L_{s+1} = Y_{s+1}$$
$$T_{s+1} = \frac{Y_{s+1} - Y_1}{s}$$
$$S_i = \frac{Y_i}{Y_1 + T_{s+1}(i-1)}, i = 1, \dots, s$$

Here again the values for α, β and γ should be determined by either minimising one of the error measures (MSE, MAE or MPE) or by using an optimisation algorithm. To start generating forecasts, we need to set the initial values for the level, L_t; the trend, T_t; and the seasonal indices, S_t. The formulas for this initial setting can be seen above. The level is set to the current level, the trend is set to the increase of the last s seasons and the seasonal indices are set to the ratio of the real value Y_i and the forecasted value based on trend and starting value. This method of forecasting will result in quite reliable initial values over time. It is possible to improve these initial values, but since we have enough historical data, we will not elaborate on this. The level, trend and seasonal indices will adapt itself if a cycle passes, and if we have enough cycles the influence of the initial values is really small. Therefore we will use this basic method of determining the initial values.

The multiplicative method assumes that the seasonal effects are proportional to the current volume. However, in some datasets the multiplicative Holt-Winters method might not fit the data well. The effect of level, trend and season might be independent of each other and therefore an additive model is better suited for the data. In the additive model it is assumed that the size of the seasonal effect is independent of the current volume. The four equations for the additive model are as follows.

The level estimate:

$$\mathrm{L_t} = \alpha(\mathrm{Y_t} - S_{t-s}) + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{6.12}$$

The trend estimate:

$$\mathrm{T_t} = \beta(\mathrm{L_t} - \mathrm{L_{t-1}}) + (1-\beta)\mathrm{T_{t-1}} \tag{6.13}$$

The seasonality estimate:

$$\mathrm{S_t} = \gamma(\mathrm{Y_t} - L_t) + (1-\gamma)S_{t-s} \tag{6.14}$$

The forecast for p periods ahead:

$$\hat{\mathrm{Y}}_{t+p} = L_t + T_t p + S_{t-s+p} \tag{6.15}$$

Where
L_t = The new smoothed value (level estimate)
α = The smoothing constant for the level $(0 < \alpha < 1)$
Y_t = The actual volume in period t
β = The smoothing constant for the trend $(0 < \beta < 1)$
T_t = The trend estimate
γ = The smoothing constant for the seasonality estimate $(0 < \gamma < 1)$
S_t = The seasonality estimate
p = The number of periods to be forecasted ahead
s = The seasonality length
Y_{t+p} = The forecasted volume for p periods ahead

The initial conditions for this recursion are as follows:

$$L_{s+1} = Y_{s+1}$$
$$T_{s+1} = \frac{Y_{s+1} - Y_1}{s}$$
$$S_i = Y_i - \left(Y_1 + T_{s+1}(i-1)\right), i = 1, \dots, s$$

So the first level estimate after the first complete period is set to the level at that time. The trend estimate is set to the increase (or decrease) per period over the last s periods. The initial values for the seasonality estimates are set for all observations until one cycle is completed. The seasonal estimate is set to the difference between the observed value and the value based on the trend estimate. In the additive model, it is easier to estimate the initial values than in the multiplicative method. This is because all effects are separate in the additive model. In the multiplicative model, the effects can enhance each other. Here again all smoothing estimates should be estimated by determining the best possible values.

6.1.7 Exponential Smoothing adjusted for trend and double seasonal variation.

If we use the method used above, we can only include one seasonal pattern. Several projects in our dataset only have one seasonal pattern, namely the week pattern. Call volume on Mondays is significantly higher than the rest of the week, and on Fridays the call volume is lower. However, there are also projects with multiple seasonal patterns. This is e.g. the case in project 437, which deals with invoices of unpaid subscription fees. If people don't pay the subscription fee, they get a reminder send by mail. Next to the week pattern, this process also shows a 4-week pattern, which is the frequency of sending the reminders. Furthermore projects can also have a year pattern. E.g. for magazine subscriptions it can be the case that people get a subscription for Christmas or Sinterklaas (5 December). To incorporate this, we look at an adaptation of the Holt-Winters model to incorporate multiple seasonal variations. This was first introduced by Taylor (Taylor, 2003). The multiplicative version of the double seasonal Holt-Winters exponential smoothing can be described as follows:

The level estimate:

$$\mathrm{L_t} = \alpha \frac{\mathrm{Y_t}}{W_{t-w}S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{6.16}$$

The trend estimate:

$$\mathrm{T_t} = \beta(\mathrm{L_t} - \mathrm{L_{t-1}}) + (1-\beta)\mathrm{T_{t-1}} \tag{6.17}$$

The (weekly) seasonality estimate:

$$\mathrm{W_t} = \delta \frac{\mathrm{Y_t}}{L_t S_{t-s}} + (1-\delta)W_{t-w} \tag{6.18}$$

The (monthly) seasonality estimate:

$$\mathrm{S_t} = \gamma \frac{\mathrm{Y_t}}{L_t W_{t-w}} + (1-\gamma)S_{t-s} \tag{6.19}$$

The forecast for p periods ahead:

$$\widehat{\mathrm{Y}}_{t+p} = (L_t + pT_t)S_{t-s+p}W_{t-w+p} \tag{6.20}$$

Where
L_t = The new smoothed value (level estimate)
α = The smoothing constant for the level $(0 < \alpha < 1)$
Y_t = The actual volume in period t
β = The smoothing constant for the trend $(0 < \beta < 1)$
T_t = The trend estimate
δ = The smoothing constant for the (weekly) seasonality estimate $(0 < \delta < 1)$
W_t = The (weekly) seasonality estimate
w = The (weekly) seasonality length
γ = The smoothing constant for the (monthly) seasonality estimate $(0 < \gamma < 1)$
S_t = The (monthly) seasonality estimate
p = The number of periods to be forecasted ahead
s = The (monthly) seasonality length
Y_{t+p} = The forecasted volume for p periods ahead

The initial conditions for this recursion are done in two steps. This method of determining the initial values was not described by (Taylor, 2003) so therefore we developed this method. First we define:

$$L_{s+1} = Y_{s+1}$$
$$T_{s+1} = \frac{Y_s - Y_{s-w}}{w}$$
$$W_{s-i} = \frac{Y_{s-i}}{L_{s+1} - T_{s+1}(i+1)}, i = 0 \dots w-1$$
$$W_i = W_{i+w}, i = s - w \dots 1$$

This first initial setting finds the values for the smallest seasonal period. It sets the level to the current level and sets the trend as the trend over the last (weekly) period. Then we set the seasonal estimates to be the division of the real observed value and the estimated value based on level and trend. This can be viewed as the deviation from the line based on level and trend. We therefore assume that the bigger seasonal period does not have a large impact on this small season and that the short term trend will deal with part of the big seasonal influence. After this first step, we continue (re)setting other variables

$$T_{s+1} = \frac{Y_s - Y_1}{s}$$

$$S_{s-i} = \frac{Y_{s-i}}{\big(L_{s+1} - T_{s+1}(i+1)\big) * W_{s-i}}, i = 0, \ldots s-1$$

Now we change the trend variable to the trend over the entire large season and then we estimate the large seasonal influence with the found values of W in the first step.

If we compare this model with the previously introduced model, there is not much difference between them. The basics of the model remain intact and only the extension for two seasonal influences is added. Note also that for clarity we added (weekly) and (monthly). This can however also be a different pattern, e.g. 4-weekly. In this model we also need an extra smoothing constant; this is a disadvantage since the values for these parameters should be determined. Chapter 6.3 deals with the problem of determining the parameter values.

Since a multiplicative model might not model the behaviour of the time series properly, we can also construct an additive version. This version is not described in (Taylor, 2003), but can be derived by following similar methods as before. The resulting model can be described as follows.

The level estimate:

$$\mathrm{L_t} = \alpha(\mathrm{Y_t} - W_{t-w} - S_{t-s}) + (1-\alpha)(L_{t-1} + T_{t-1}) \qquad (6.21)$$

The trend estimate:

$$\mathrm{T_t} = \beta(\mathrm{L_t} - \mathrm{L_{t-1}}) + (1-\beta)\mathrm{T_{t-1}} \qquad (6.22)$$

The (weekly) seasonality estimate:

$$\mathrm{W_t} = \delta(\mathrm{Y_t} - L_t - S_{t-s}) + (1-\delta)W_{t-w} \qquad (6.23)$$

The (monthly) seasonality estimate:

$$\mathrm{S_t} = \gamma(\mathrm{Y_t} - L_t - W_{t-w}) + (1-\gamma)S_{t-s} \qquad (6.24)$$

The forecast for p periods ahead:

$$\hat{\mathrm{Y}}_{t+p} = L_t + T_t p + W_{t-w+p} + S_{t-s+p} \qquad (6.25)$$

Where

L_t = The new smoothed value (level estimate)

α = The smoothing constant for the level ($0 < \alpha < 1$)

Y_t = The actual volume in period t
β = The smoothing constant for the trend $(0 < \beta < 1)$
T_t = The trend estimate
δ = The smoothing constant for the (weekly) seasonality estimate $(0 < \delta < 1)$
W_t = The (weekly) seasonality estimate
w = The (weekly) seasonality length
γ = The smoothing constant for the (monthly) seasonality estimate $(0 < \gamma < 1)$
S_t = The (monthly) seasonality estimate
p = The number of periods to be forecasted ahead
s = The (monthly) seasonality length
Y_{t+p} = The forecasted volume for p periods ahead

The initial conditions for this recursion are also done in two steps. This again was not introduced by Taylor, but follows similar logic as in the multiplicative case. First we define:

$$L_{s+1} = Y_{s+1}$$
$$T_{s+1} = \frac{Y_s - Y_{s-w}}{w}$$
$$W_{s-i} = Y_{s-i} - (L_{s+1} - T_{s+1}(i+1)), i = 0 \dots w-1$$
$$W_i = W_{i+w}, i = s - w \dots 1$$

This first initial setting finds the values for the smallest seasonal period. It sets the level to the current level and sets the trend as the trend over the last (weekly) period. Then we set the seasonal estimates to be the difference between the real observed value and the estimated value based on level and trend. We therefore assume that the bigger seasonal period does not have a large impact on this small season and that the short term trend will deal with part of the big seasonal influence. After this first step, we continue with setting the other variables

$$T_{s+1} = \frac{Y_s - Y_1}{s}$$
$$S_{s-i} = Y_{s-i} - (L_{s+1} - T_{s+1}(i+1) + W_{S-i}), i = 0, \dots s-1$$

Now we change the trend variable to the trend over the entire large season and then we estimate the large seasonal influence with the found values of W in the first step.

Both two last models yield good results in real world applications and by setting the initial values as described above, we can generate good forecasts after a few periods. You can generate forecasts if you only have one large period of data, but it is advised to have at least three periods of data to increase the accuracy of the seasonal estimates.

6.2 Forecasting methods: Regression models

In this sub chapter several forecasting methods will be discussed, including their advantages and disadvantages. We start with the most basic forecasting methods. For more reference about these forecasting methods see (Hanke & Wichern, 2009) or (Brockwell & Davis, 1996).

6.2.1 Autoregressive Models (AR)

Autoregressive models are used when there is autocorrelation in the data. Autocorrelation is when observations from a certain time period are dependent on observations from another time period. E.g. if you consider the weather forecast: if it is sunny today, it is likely that it will be sunny tomorrow (however a lot of other factors play a part). Also in call centre data we see autocorrelation. Typically we see that the call volume of today is dependent on the call volume of yesterday and on the call volume of a week ago. This second fact can be explained by the fact that there is a difference in call volumes per day. On Monday's the call volume is higher than in the rest of the week, therefore the call volume of a Monday depends on the previous Monday as well as on the day before (in this case the Friday). To incorporate autocorrelation in our models, Autoregressive Models (AR-models) have been developed. AR models are defined by the number of lags, so an AR(5) model incorporates 5 lags. The AR model is said to be of order p if it incorporates p lags. The AR(p) model is defined as follows:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ \dots + \beta_p Y_{t-p} + \varepsilon_t \qquad (6.26)$$

Where the error terms ε_t have conditional expectation 0, variance σ^2 and are independent and identically distributed (iid) random variables. Y_t is the dependent variable, $Y_{t-1}, \dots, Y_{t-p}$ are the explanatory variables and $\beta_0, \dots, \beta_p$ are the regression parameters (with β_0 the constant term). By means of e.g. Ordinary Least Squares (OLS) the model can be fitted to the data, which yields the following forecast equation:

$$\hat{Y}_t = b_0 + b_1 Y_{t-1} + \cdots + b_p Y_{t-p} \qquad (6.27)$$

6.2.2 Moving Average Models (MA)

The moving average models in this subchapter are different from the models introduced in the previous subchapter. We will first introduce the qth-order moving average model:

$$Y_t = \beta_0 + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ \dots + \beta_q \varepsilon_{t-q} \qquad (6.28)$$

Where the ε's are the explanatory variables and the β's are the parameters to be estimated. Again $\varepsilon_t \dots \varepsilon_{t-q}$ are iid random variables. This model is quite similar to the previous model, except that the forecast now depends on previous errors instead of previous observations. For some time series an upward shock (i.e. there is a positive error) will cause a downward shock in the next period. In that case the β_1 will be negative.

6.2.3 Autoregressive Moving Average Models (ARMA)

Both methods introduced above can be combined into a new model. This Autoregressive Moving Average Model, ARMA, uses the notation ARMA (p,q). Here p defines the order of the autoregressive part and q is the order of the moving average part. This yields the following definition:

$$Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \cdots + \gamma_q \varepsilon_{t-q} \tag{6.29}$$

Where again $\varepsilon_t \dots \varepsilon_{t-q}$ are iid random variables. These ARMA models can be used in a wide variety of stationary time series. Calculating values for the β's is computational difficult if the order is high. A standard regression model can easily be solved by applying the method of least squares. For the ARMA model that doesn't work. This is because the error terms are dependent on the rest of the model, so when we change the value of any of the β's, all error terms change as well. Algorithms like the 'Innovations Algorithm' (Brockwell & Davis, 1996) can be used to determine the parameter estimates. In most statistical packages like Stata, SPSS or SAS there is a method incorporated for estimating these models. Stata e.g. has a method which combines the BHHH algorithm (named after their inventors (Berndt, Hall, Hall, & Hausman, 1974)) and the BFGS algorithm by (Broyden, 1970), (Fletcher, 1970), (Goldfarb, 1970) and (Shanno, 1970). Both these algorithms are elaborations of the Newton's method for optimisation but follow similar logic. In Newton's method you approximate the function *f(x)* (in this case a function on the errors of the regression model) by a quadratic function. Then you take one step towards the maximum/minimum of that function (in our case the minimum). This step is repeated until convergence. Both these methods are optimisation algorithms and they are alternated such that a good solution can be found. Also for other methods discussed here with a moving average factor this optimisation method is used. Note that in case $q = 0$ the model can still be estimated as an autoregressive model, which can be solved by OLS.

6.2.4 Autoregressive Integrated Moving Average Models (ARIMA)

The ARMA method is designed for processes which are (weak sense) stationary or (covariance) stationary. Strict stationary processes are when the probability distribution does not change over time. Weak sense stationarity is a relaxation which holds when the first two moments of the distribution remain constant over time. This also means that the covariance only depends on the distance between two observations and not on the place of these observations. A non-stationary process can often be converted to a stationary process by differencing. This can be done by replacing the original series by a series of differences and these differences are then modelled as an ARMA process. Usually the first difference is enough to construct a covariance stationary process, but it is also possible to take the second difference or even higher numbers. An ARIMA process is defined by ARIMA (p,d,q), where p and q are respectively the order of the autoregressive part and the order of the moving average part. The third variable d is the order of the differences. In case of $d = 1$ the model is defined by:

$$\Delta Y_t = \mu + \beta_1 \Delta Y_{t-1} + \beta_2 \Delta Y_{t-2} + \cdots + \beta_p \Delta Y_{t-p} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \cdots + \gamma_q \varepsilon_{t-q} \tag{6.30}$$

Where Δ denotes the first difference. This method is used quite a lot in general forecasting, as well as in specific call centre forecasting situation.

6.2.5 Seasonal Autoregressive Integrated Moving Average Models (SARIMA)

If the data we would like to forecasts exhibits a seasonal pattern, we need to adapt the ARIMA model. If there is e.g. a clear correlation between observations of January this year with the observation of January last year, the seasonal period $S = 12$ in case of monthly data. If you wanted to incorporate this in the normal ARIMA model, you would have to set $p = 12$. This would increase the number of parameters and hence the calculation time. Furthermore, not all lags between 1 and 12 might have a significant effect and therefore should not be in the model. An adaptation to the ARIMA model is to add a seasonality factor. This seasonal factor is added by adding an extra differencing term. If you for example have a seasonality of 12, then the seasonal ARIMA(0,0,0)(0,1,0)$_{12}$ is can be written as:

$$Y_t = \hat{\mu} + Y_{t-12} + \varepsilon_t \tag{6.31}$$

Here we have a seasonality of 12 and we incorporate only the difference of this seasonality. The first set of inputs for this SARIMA model is similar to the ARIMA inputs. The second set of inputs for this seasonal ARIMA model shows which autoregressive, integrated and moving average factors should be incorporated in the model. Hence the model can be expanded enormously if we incorporate more terms. E.g. the ARIMA(1,1,0)(0,1,1)$_{12}$ can be written as:

$$\begin{aligned} Y_t = \hat{\mu} + Y_{t-1} + Y_{t-12} - Y_{t-13} + \beta_1(Y_{t-1} - Y_{t-2} - Y_{t-13} + Y_{t-14}) + \varepsilon_t \\ + \beta_2\varepsilon_{t-12} \end{aligned} \tag{6.32}$$

Which can be rewritten into:

$$W_t = \beta_1 W_{t-1} + \varepsilon_t + \beta_2\varepsilon_{t-12}$$

Where

$$W_t = \Delta\Delta_{12}Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$$

So here you can see how the seasonal ARIMA model works. First you apply both differencing procedures to obtain W_t and then you can construct the regression equation on W_t. If you then fill in W_t in the regression equation, you get the ARIMA model. This model can also be calculated in a statistical package like Stata.

Models that have a moving average term are rather hard to estimate. This is because the parameter estimates directly influence the error terms and therefore make the model different. Finding a good estimate to these models is therefore an optimisation model which takes considerable time to calculate. Several software packages like Stata can calculate these problems. However programming these models from scratch in the Anago software is rather difficult. Therefore, in the remainder of this thesis we will use the moving average models as benchmarks to estimate the performance of the Holt-Winters seasonal models. In chapter 8 we will compare the Holt-Winters model to the ARIMA models to see which model performs better.

6.3 Holt-Winters Parameter optimisation

A large problem of the Holt-Winters algorithm is finding the appropriate parameter values for the smoothing process. This problem only increases when we add the extra seasonality factor to the model. A common criticism is that this extension leads to 'over-parameterisation'. The big advantage of the algorithm is that finding forecasts is really easy once the parameters are found, but those parameters are really hard to estimate.

In several software packages there are algorithms implemented to find optimal values. These algorithms usually combine a local search optimisation with a global search. The local search optimisation will search for improvements in the parameters in the neighbourhood of the current parameter values. A problem then can be that this optimisation results in a local optimal solution. Therefore the global search method makes sure that more (locally) optimal values are found and hopefully it will also find the global optimum.

Getting a good value for the parameters can cost quite some calculation time, especially if you have to optimise four parameters. If you would choose to naively calculate every parameter setting where parameters can range from zero to one with two decimals, this would result in calculating 101^4 Holt-Winters forecasts. Therefore we will introduce a method which has a really fast calculation time, but yields good results. For this we will first introduce the concept of Latin Hypercube Designs.

6.3.1 Latin Hypercube Designs

Latin Hypercube Designs are mostly used when you want to estimate an unknown function with unknown parameters. E.g. you want to estimate a certain insurance risk for fire. Because this risk is complicated and a lot of factors contribute to the probability of fire, simulation is often used to calculate these probabilities (e.g. a Monte Carlo simulation). These simulations are done for the different values of the input parameters. These input parameters can be seen as the different scenarios which could take place. However, these simulations can be really time consuming and running the simulation for every possible parameter value can take ages. Therefore Latin Hypercube Designs are used to determine useful parameter values.

First we will illustrate the concept of Latin Hypercube Designs in two dimensions. If you take chess board (so size 8x8) and put 8 rooks on that board in such a way that they cannot attack each other. That way there is one rook on every row and column. An easy solution to this problem is to put all the rooks on the diagonal of the grid. Now we want to set the rooks in such a way that the minimum distance between two rooks is as large as possible. Determining the distance between two points can be done in several ways, but here we will only use the Euclidean distance. This is defined as

$$d(X,Y) = \left[\sum_{k=1}^{m} |x_k - y_k|^2\right]^{\frac{1}{2}} \tag{6.33}$$

To illustrate the use of this measure it is easiest to show an example of an optimal Latin Hypercube Design:

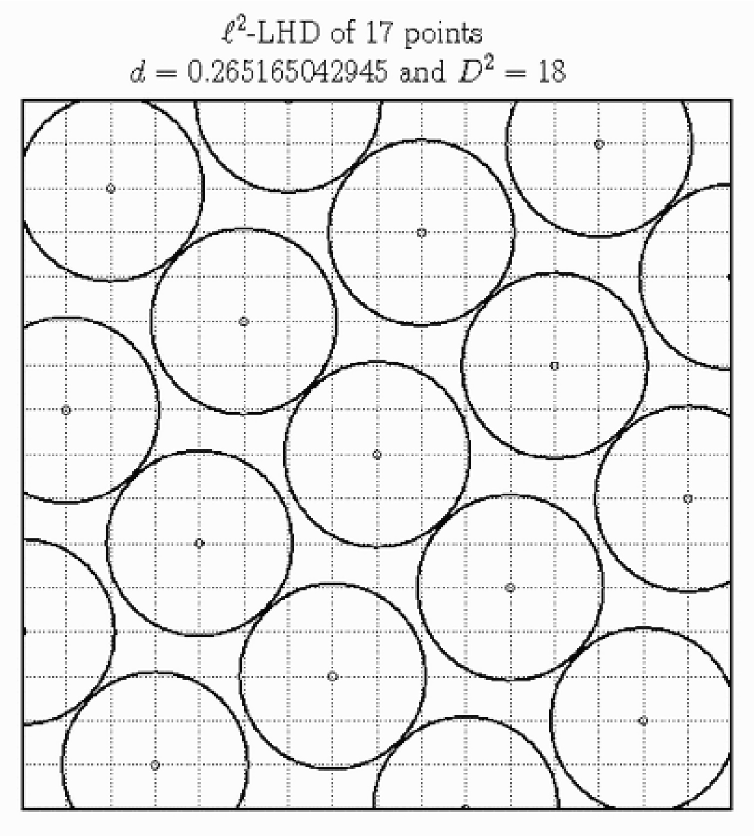


Figure 8: Latin Hypercube Design in two dimensions with 17 points and the Euclidean distance measure (taken from www.spacefillingdesigns.nl)

Here you can see that while using this distance measure, solving the LHD in two dimensions is basically packing circles into the grid. In three dimensions it can be seen as putting balls in a box. LHD are thus a distribution of point in a number of dimensions in such a way that the points are evenly spread over the entire area.

In two dimensions an optimal LHD can be found (see e.g. (van Dam, Husslage, den Hertog, & Mellissen, 2007)). However in higher dimensions it is hard to find optimal LHD. For some designs we know they are optimal, just by calculating all possible options, which is really time-consuming. For other problems, we can use algorithms to find good solutions. E.g. the algorithm proposed by (Jin, Chen, & Sudjianto, 2005) can be used. On the website www.spacefillingdesigns.nl the best LHD found so far can be seen.

Back to the example of the insurance risk. We needed parameter values for the simulation, and we wanted few simulation runs to decrease the calculation time. However, we also want to calculate scenarios (i.e. parameter settings) which represent the possible occurrences well. Because we don't have any knowledge of the optimal parameter setting, we want the parameters to be evenly distributed over their possible values. The LHD's can therefore be used, because in a LHD the points are evenly spread over the area.

For the Holt-Winters parameter search we can also use LHD's. The parameters $\alpha, \beta, \gamma\ and\ \delta$ can have a value between 0 and 1. To decrease the calculation time, we want the values to be optimal for two decimals. When we would calculate all possible outcomes this would result in calculating 101^4 possible forecasts. Now we take a LHD of 101 points in four dimensions and use this as the parameters. Of these 101 simulations, we calculate the values of the error measures introduced in equations (3.14), (3.15) and (3.16). These errors are the differences between the forecasts for the last ten weeks (so fifty observations ahead) which are done on the day before these weeks and the real observed number of calls. With these different error measures we have to determine which error measure is the best. This is done by calculating the relative performance to the best simulation. The best simulation gets value one and the errors of the other simulations are divided by the error of the best simulation. This is done for all three error measures. If we then add up these values, we can find the best settings for the parameters of these designs.

6.3.2 Local search optimisation

However, there might be improvements possible. Therefore, we take the five best parameter settings and try to decrease their errors. This is done by changing the values of the parameters one by one and see whether the errors become smaller. E.g. we decrease α by 0.01 and we see if the error measures decrease. It can be that some of the measures decrease and some don't that is why we calculate the following:

$$\frac{MSE_i}{minMSE} + \frac{MAE_i}{minMAE} + \frac{MPE_i}{minMPE}$$

Where $minMSE$ is the value of the MSE of the current design (before adapting α). If this sum is lower than three, the new design is considered an improvement. We will then look at increasing β. If it is no improvement we will increase α and recalculate the model.

This is continued until it is no longer possible to improve the parameter settings with these small adjustments. Of these five (locally) optimised designs the best one is chosen. Typically this process of improving the parameters takes about thirty steps per initial parameter setting in the data we tested. However it might take longer in different datasets. So in the entire process of optimising, we only need to calculate the forecast about 250 times, which is significantly lower than the naive 101^4 calculations.

6.3.3 Seasonal detection

Another important input for the Holt-Winters double seasonality model is the length of the season. In our data we typically see a seasonal length of 5 and a larger seasonal length of 20 (monthly patters) or 260 (year pattern, 52 weeks of five working days). Since there might also be other patterns in the data, which are not known in advance, we want to design an automatic seasonal detection. This detection is done by a straightforward regression model. First we estimate the following regression model:

$$Calls_i = \beta_0 + \beta_1 * Calls_{i-1} + \beta_2 * Calls_{i-2} + \cdots + \beta_{\left\lfloor \frac{n}{5} \right\rfloor} * Calls_{i-\left\lfloor \frac{n}{5} \right\rfloor} + \varepsilon_i$$

Where $\left\lfloor \frac{n}{5} \right\rfloor$ denotes the floor function of $\frac{n}{5}$ and n is the total number of observations. The subscript i indicates the day. This limitation is made because if we look at larger lag values, significance cannot be shown. We need at least five times the series to be certain of that seasonal influence. Furthermore the Holt-Winters model also needs at least three periods to get good seasonal parameters.

Once we calculated the regression model, we can determine which factors are significant and which are not. First we calculate the p-value of the parameters and we dispose the lag variable if the p-value is larger than 0.01. So with 99% confidence we can say that the remaining parameters are significant. However, after removing several lag variables, we need to recalculate the model. This is because the removed variables have had influence of the parameter values and variances. If we then estimate the new model, we might again find insignificant parameters. We can again remove these parameters from the model and repeat this process until no insignificant parameters are left in the model.

We now have a model with probably several significant parameters. We now have to determine which parameters indicate the seasonality of the process. First you could have really low significant values (e.g. a lag of one or two). The fact that the first lag variable is significant is that the process is dependent on its direct history. However, in the Holt-Winters forecasting model, the current level and trend are incorporated into the model. Therefore, we will not incorporate significance within the same week (i.e. with a lag of less than five days). In most datasets from the call centre world, we will see a highly significant factor for a lag of five days. This is the parameter that indicates a weekly seasonality. However, there might be other lag variables that are more significant. Therefore we look at the seasonality with the highest significance and set this as one of the seasons.

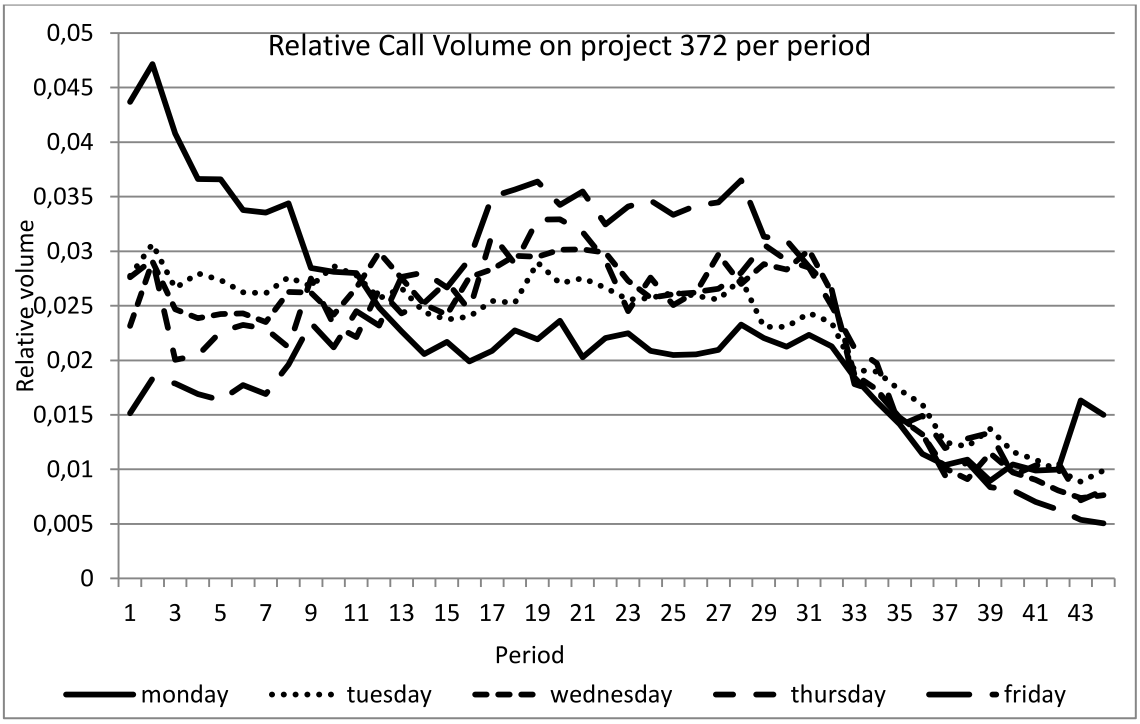
For the second season, we will look at the next parameter with a high significance. This lag parameter should also be at least five. When this is the case that second season will also be assigned.

It might be that there is still a significant seasonality factor left after the two seasons are appointed. This cannot be avoided and we also do not want to extend the model to more seasonality factors, since that will increase the problem of over-parameterisation even more. Obviously it is also possible to find zero or one significant seasonal parameter.

7 DAY PATTERNS

The call centre for which we forecast the incoming calls needs forecasts on quarter of an hour level. For each of those quarters the workload and the number of needed employees are calculated. In the previous chapter we saw that forecasts are done per day, so now we have to find a way to distribute the forecasts over the days.

Typically Monday has a higher call volume then the other days. If we look at the volume per quarter, we see that Mondays also have a different pattern then other weekdays. On Monday morning the volume is relatively high and lower in the rest of the day (we look at which percentage of the total day volume is in a certain period). Friday also exhibits a slightly different pattern. The call volume on Friday morning is lower and in the afternoon it makes up for this. See the figure below for these patterns.



For the forecasting process, we need a way to determine the patterns for the separate days. First we examine the pattern to see if there is a trend or dependent on previous observations. However, once we run a regression on the pattern with two lag variables as exogenous variables, no significant results show. Also, when a plot is made of the pattern, no trend can be discovered. This has been tested with several call centre datasets.

From the professionals from the call centre there was also the impression that holidays changed the patterns. If e.g. the call centre was closed on Monday because of Easter, they claimed it changed the pattern of Tuesday. However, after investigating this trait, we have not encountered any influence of these holidays.

So it seems that there is no significant change in the pattern over time. However, we still need to make an estimation of the day pattern. This will be done by calculating the average over previous periods and use this average as the forecasted day pattern for the next period(s). We will make this forecast for each day separate because of the different patterns for each day. To determine how far in the history we should look, we will again use the error measures MSE (3.14) and MAE (3.15) defined in chapter 3. The MPE (3.16) will not be used, since the difference between forecast and actual value should be divided by the actual value. But since the actual value can be zero, this leads to an error. With the remaining two methods we estimate the pattern by calculating the average over the past. This past can be of size one (so only one period is used to estimate the pattern) up until size thirty. The MSE and MAE are calculated for all these sizes of periods for comparison. See Appendix E for the results.

If we look at the results, we see that if we increase the number of periods, the error (generally) also becomes smaller. This is another indication that the day patterns are stable over time. The best result can be achieved to calculate the average over as many previous patterns as available to forecast the next pattern. However, calculation time is also important in the application of the forecasting procedure. When we look at the results again, we see that the error decreases rapidly if we increase the number of periods from one to two, but for higher numbers, the error decreases only slowly. To determine how long the period should be, we look at the difference between the lowest error and the highest error (usually when we use the previous pattern to estimate the next pattern). We determine the 'score' of a certain forecast i by the following formula:

$$\frac{MSE_i - \min_i(MSE_i)}{\max_i(MSE_i) - \min_i(MSE_i)}$$

So the worst forecast gets score 1 and the best gets score 0. A similar formula was used for the MAE measure. We consider the forecasts to be good if they have a score of at most 0.05. Given that we want to minimise the calculation time to determine the day patterns we want the length of the forecasting region to be as low as possible, while still generating good forecasts. The data suggests we need to use a period of length 17 (see Appendix E). So to forecast the pattern of the next Monday, we will calculate the average pattern over the last 17 Mondays and take this as the estimated pattern.

With these forecasted patterns, we can complete the forecasting process. Now scale down the daily forecasts to a quarter of an hour level. In the next chapter we will determine the quality of this forecasting method.

8 RESULTS

In this chapter the results will be discussed. This is the comparison between the extension to the Holt-Winters model and the ARIMA models developed in chapter 6. First we will start off with as short introduction on the projects that we tested.

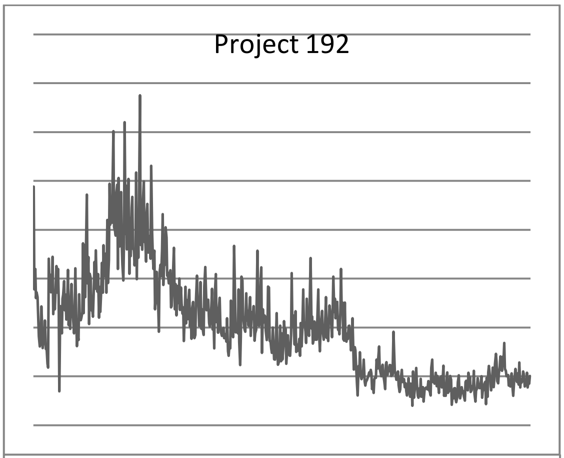
Project	Length in weeks	Average volume	standard deviation	small season	large season
192	131	199.15	109.06	5	10
194	131	244.67	191.09	5	20
242	116	145.60	39.64	5	15
437	64	106.00	104.73	20	60

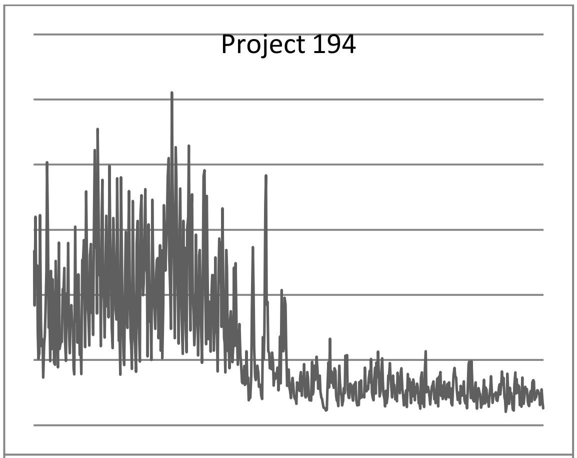
Several projects are tested in this chapter. More explanation on the project can be found in Appendix A. The seasons are determined by the method introduced in chapter 6. The projects 194 and 437 are concerned with sending invoices and reminders, so we would suspect them to have a regular pattern. In the table above we can see that they have a 4-week pattern (so 20 days). Furthermore almost all projects show a week pattern (i.e. a small season of length 5). We do not have the same amount of data for all projects, which is due to the fact that projects are terminated, renamed or combined. We can also see that the standard deviation of the call volume varies a lot. Project 242 is a really stable project, where project 194 and 437 are really volatile.

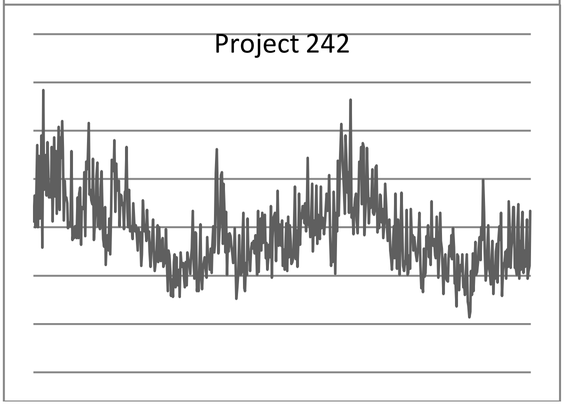
To determine the performance of the model we will forecast the last 50 observations (i.e. 10 weeks) of the series. We will then compare these forecasts with the actual value and calculate the different error measures (MSE (3.14), MAE(3.15) and MPE(3.16)). We will do this process for the four different Holt-Winters models to see which model performs best. That way we can also see if the second season is really helpful or not.

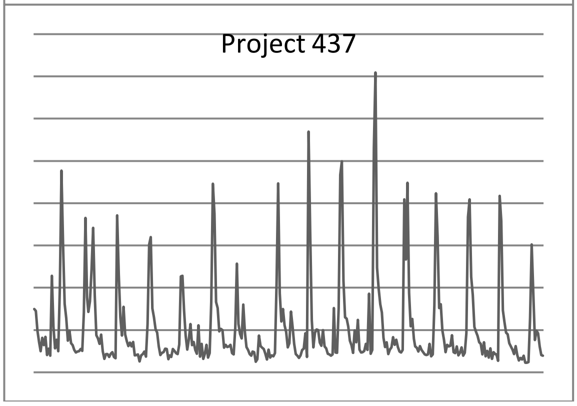
Before we start forecasting, we first have to adapt the raw data. This is done by detecting missing values (chapter 3) and outliers (chapter 4) and correcting them. With this corrected datasets we will start the forecasting process. To determine the parameters, we will apply the parameter optimisation process introduced in section 6.3. Here the parameters are optimised regarding to their performance over the last 50 observations. We thus optimise over observation 100 to 51 (seen from the end of the dataset) and then test the performance over observation 50 to 1 (also seen from the end).

First we will start of with a short introduction of the projects by means of showing the patterns.









Above you see the patterns for the different projects in this chapter. In the charts, there is no indication of level, or time frame, since that is not relevant for the current analysis. Project 192 is a project which has had a higher level in the past, but the number of calls has reduced and the volatility is much lower. We do keep this project, since we are looking for a forecasting method which can easily adapt to new situations. In the call centre, projects can be terminated, combined or a new impulse can be given to a certain magazine. Therefore volume can change and we want the system to be able to adapt to that. Project 194 has the same problem, in the start of the project the volatility was really high, and now the project has settled on a constant flow of calls. Because the difference in level and volume is really high, we will only look at the second part. Project 242 is a project which hints at a seasonal effect of about 350 days (note that this is far greater than a year, since it are working days). However there is not enough data to support this claim. Finally, project 437 has a strong seasonal pattern of length 20 with a flat trend.

These projects will be forecasted in the next subchapters. First we will forecast on basis of the Holt-Winters methods introduced before. Then we will forecast based on the ARIMA models (and related models). Finally a comparison will be made between the different forecasting methods.

8.1 Holt-Winters forecasting results.

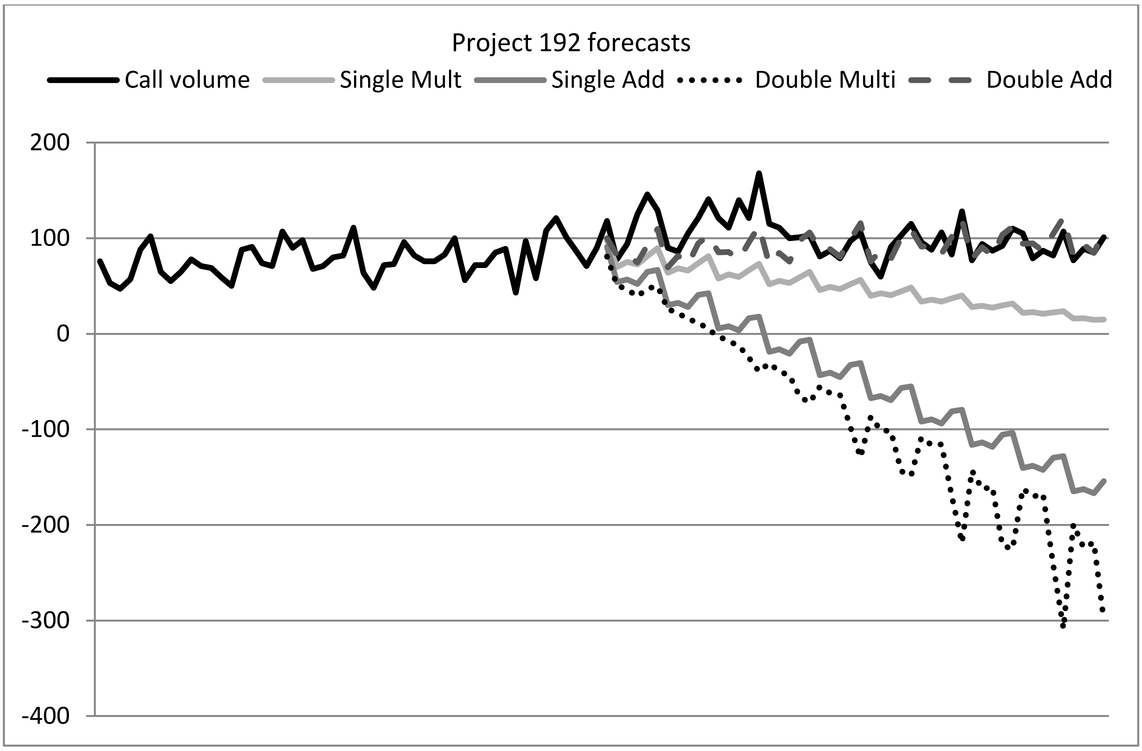
For the forecasting we first have to optimise the parameters. In the table below are the results:

Project 192	**single multi**	single add	double multi	double add
mse	**183,76**	188,62	208,44	269,26
mae	**10,24**	10,66	11,48	12,14
mpe	**0,14876912**	0,149623	0,15475798	0,17178285
Project 194	single multi	single add	**double multi**	double add
mse	1161,6	1232,16	**957,28**	997,54
mae	22,92	24,32	**21,36**	21,5
mpe	0,20518941	0,2200554	**0,203179325**	0,21440845
Project 242	**single multi**	single add	double multi	double add
mse	**557,72**	586,92	579,16	661,94
mae	**17,48**	17,52	17,44	20,42
mpe	**0,16955824**	0,1637617	0,184346352	0,20933731
Project 437	**single multi**	single add	double multi	double add
mse	**2685,72**	6680,24	3777,42	6369,96
mae	**31,68**	54,4	37,3	51,88
mpe	**0,31800078**	0,6070767	0,431071304	0,56730962

In this table you can see the results of the parameter optimisation process. The values of the parameters can be found in Appendix F. Above table shows the error measures of the optimisation process. So for project 192, the single multiplicative model yields the lowest errors based on the optimisation (so observation 100 to 51 from the end of the dataset). However, the best scoring measure on that data doesn't have to be the best forecasting method for the next 50 observations. Therefore we have to look more closely at the forecasts.

First project 192. Based on the forecast the single multiplicative model yields the best results, where the double additive model yields the worst results. If we then look at the forecasts of the next 50 observations, we get the following table and graph:

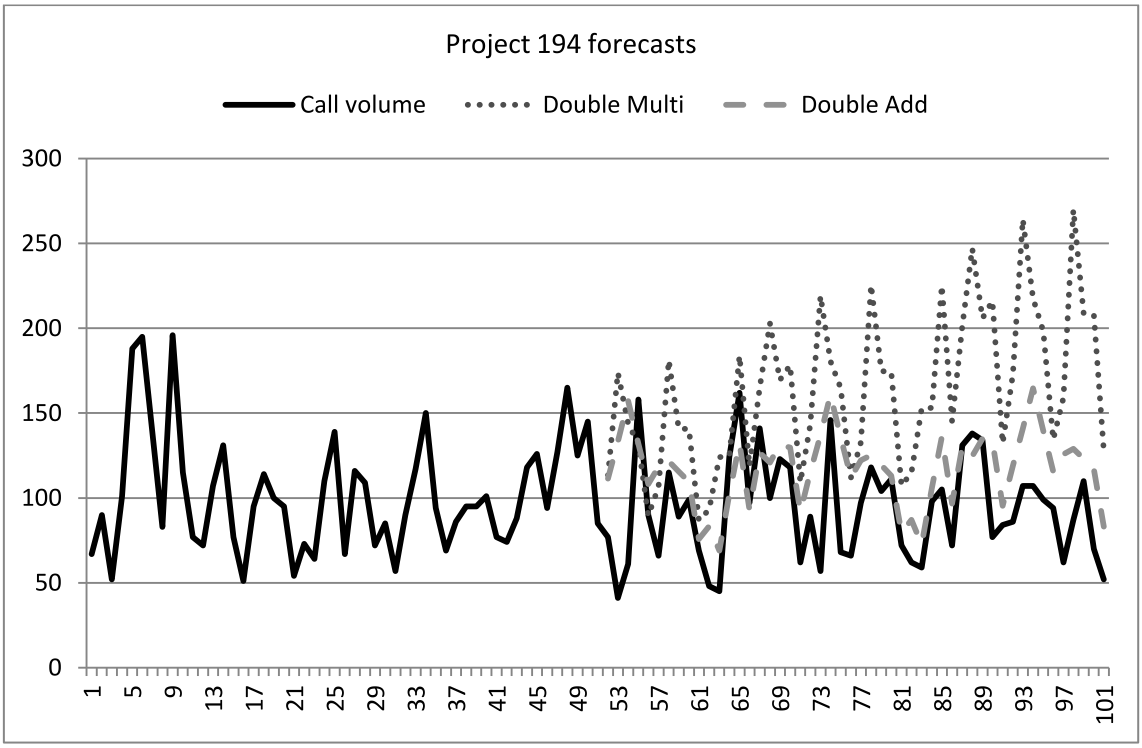
	Single multi	Single add	Double multi	**Double add**
mse	3260,42	25767,79	45857,40	**519,44**
mae	53,61	147,52	193,14	**16,30**
mpe	0,53	1,54	1,99	**0,15**



In the graph you see that the previously best method (single multiplicative) has a negative trend, which causes it to underestimate the real volume. The previously worst method (double additive) now yields the best results. In this dataset it becomes clear that it is not straightforward that the best method beforehand will also make good estimations afterwards. Therefore a good sense of logic is needed. If you would see these forecasts, you would quickly rule out the single additive and double multiplicative model. And after seeing the negative trend of the single multiplicative model you would also not choose that model, but the double additive instead.

Project 194 yields the following results:

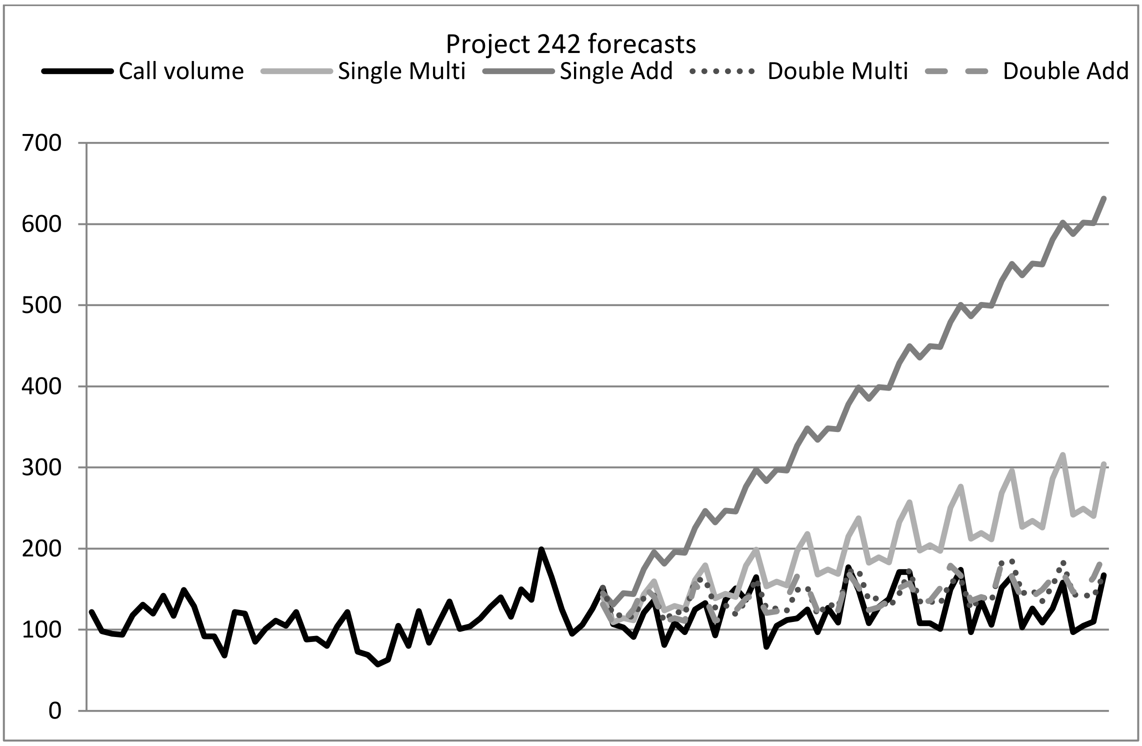
	single multi	single add	double multi	**double add**
mse	96635,53	224900,86	6843,45	**1327,05**
mae	266,85	411,23	71,46	**28,61**
mpe	3,00	4,73	0,90	**0,40**



In this graph only the double seasonality models are plotted. This is because both the single seasonality models yield very bad result and would decrease the usefulness of the graph. From the parameter optimisation process it has shown that the double multiplicative model yields the best result. In the forecasts this method also yields good results, but the double additive model yields better results. Here we see that there is as slight upward trend in the last 30 data points before we start forecasting, which causes the double multiplicative model to follow that trend.

Project 242 yields the following results:

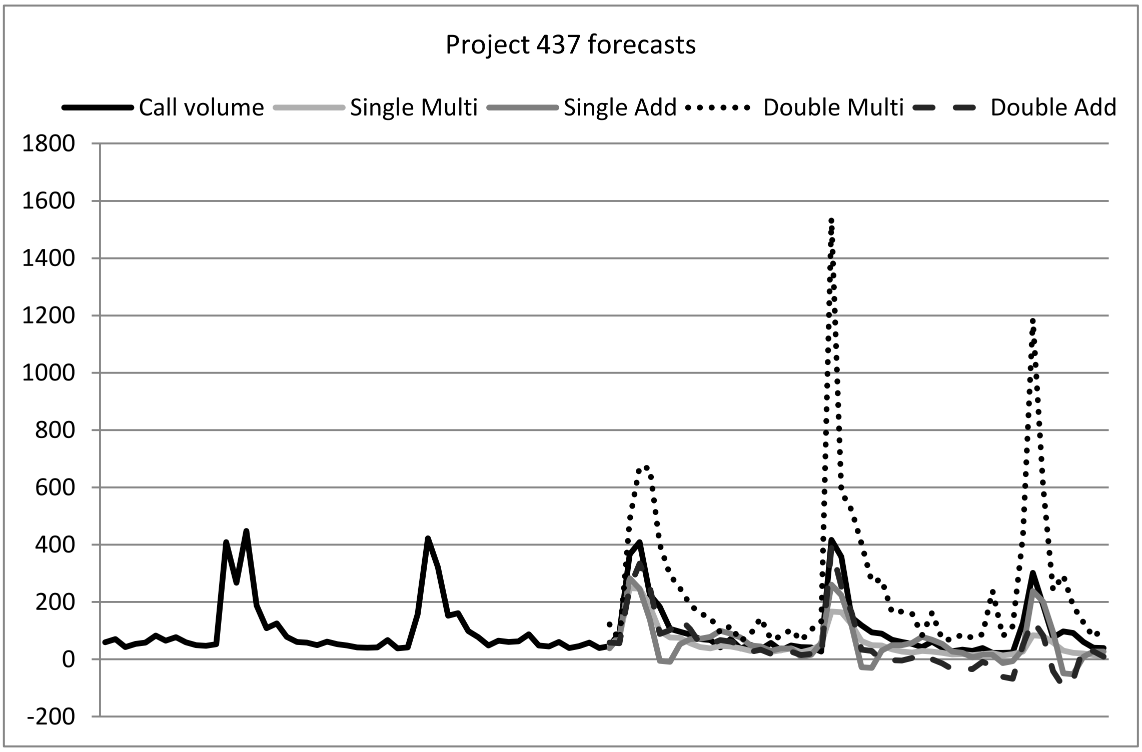
	single multi	single add	**double multi**	double add
mse	7197,005851	83519,70252	**594,072115**	712,8622859
mae	72,93580442	251,7509135	**21,13537478**	22,0158298
mpe	0,605385995	2,084100904	**0,188497705**	0,195256468



In the optimisation process the single multiplicative model was considered the best model, but in the real forecasts this model overestimates the real volume. This is again because of the trend in the last weeks. The second best model in the optimisation (double multiplicative) is here the best forecasting method.

The final results are below (project 437)

	single multi	single add	double multi	double add
mse	**5040,90794**	4983,063859	65805,81003	4921,077518
mae	**44,8031152**	48,38987164	159,1532102	54,80638082
mpe	**0,411101233**	0,545094631	1,733908911	0,842416477



Similarly as in the parameter optimisation, the single multiplicative Holt-Winters method yields the best results. Some other methods however also yield good results, but they forecast volumes below zero. Getting a forecast below zero is possible in the Holt-Winters models, since there is no probability distribution connected which keeps this from happening. Furthermore, we only see this happening in the additive models. In the multiplicative model, both seasonality influences cannot be negative and for this project the $\beta's$ from the optimisation process are both set to zero. This causes that the initial trend, which is >0, will not change over time. Therefore it is not possible for them to become smaller than zero. In the additive model, all influences have their own contribution and they do not affect each other. Therefore it can happen that all influences result in a negative forecast. There is also no way of keeping the forecasts above zero other than changing the forecasts afterwards. We cannot influence the separate factors (level, trend and both seasonal factors) such that the sum will remain positive.

Now that we have seen all the forecasting methods at work, we can conclude the following: Methods which are considered the best method in the optimisation procedure do not necessarily yield the best results. The best measure can better be achieved by looking at a graph of the forecast and choose the most logical graph. There we should carefully look at a strong negative or positive trend. If you e.g. look at project 192 we see that some Holt-Winters methods result in a negative call volume. These can easily be ruled out. Also the methods in project 242 can be ruled out because of the strong positive trend. A call centre planner should be able to see this quickly. However, we also saw that if we rule out the illogical forecasts, we can rely on the best method from the parameter optimisation process. This might not always yield the best result, but gives a good forecast. Furthermore, it is good practise to review the forecasts every week. Once new information becomes available, we can make more reliable forecasts.

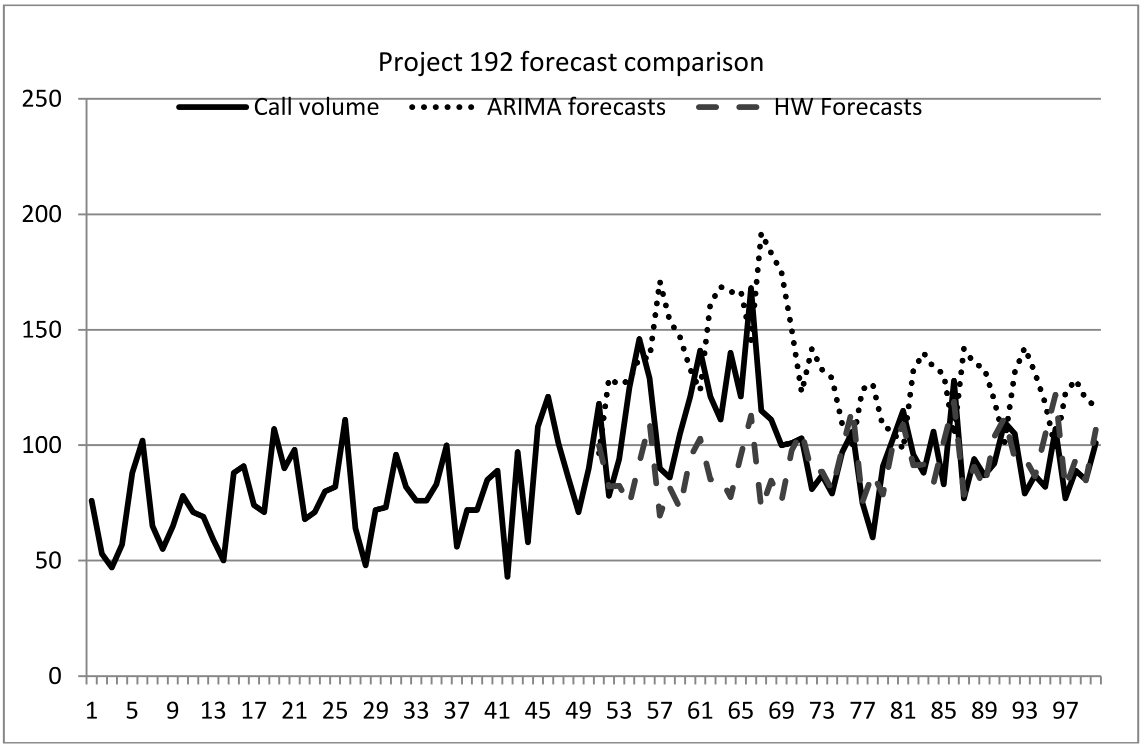
In the next section, we will discuss the ARIMA models for all of the projects to compare the performance of those methods with the best Holt-Winters model.

8.2 ARIMA forecasting results

In this chapter the results of the ARIMA models will be discussed. We will also include the best Holt-Winters forecast to make the comparison. For the ARIMA modelling we will use Stata to determine the best model. This can result in an ARMA, ARIMA or SARIMA model. In this chapter we will only discuss the model with the best fit. As you can see, none of the models results in a SARIMA model; this is because the optimisation procedure couldn't find an upward slope toward the best fit. First we will start off with project 192.

Project 192 shows autocorrelation and therefore we apply differencing for this project. To show this autocorrelation we use the Augmented Dickey-Fuller test (Said & Dickey, 1984). The test statistic before differencing is -2.003. The critical value at 5% is -2.860, hence we cannot reject the hypothesis that there is a unit root in the stochastic process, i.e. the process is non-stationary. We therefore apply differencing to the series to get a stationary process. After the first differencing the value of the ADF-test is -13.495 and hence we can reject the hypothesis that the process is non-stationary. The resulting model is an ARIMA (10,1,5). Increasing or decreasing any of the parameters results in worse performance. This model is achieved by starting out with a big model (i.e. lots of AR and MA parameters) and gradually dropping the insignificant parameters. The performance of the resulting ARIMA is as follows:

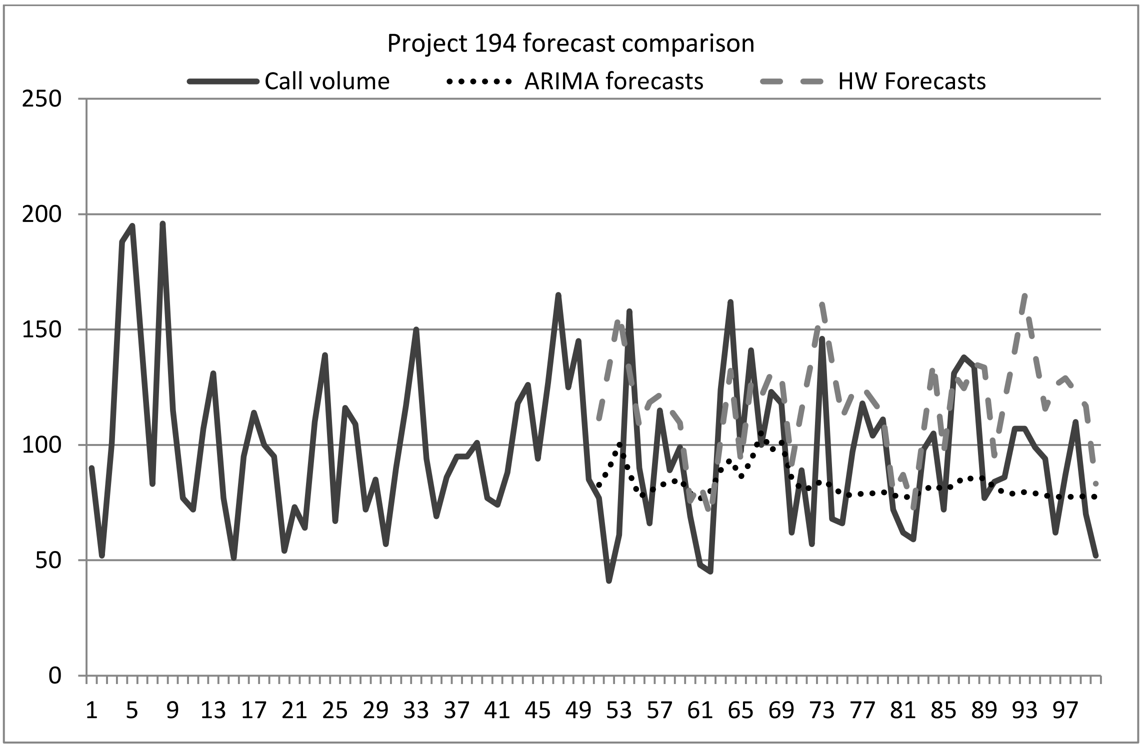
	ARIMA	HW
mse	1820,3	**519,44**
mae	36,96937	**16,30**
mpe	0,400536	**0,15**



As you can see in the plot, the ARIMA forecasts are slightly above the graph, where the HW forecasts are below the graph for the first part of forecasting. However the HW forecast outperforms the ARIMA model if we look at the error measures.

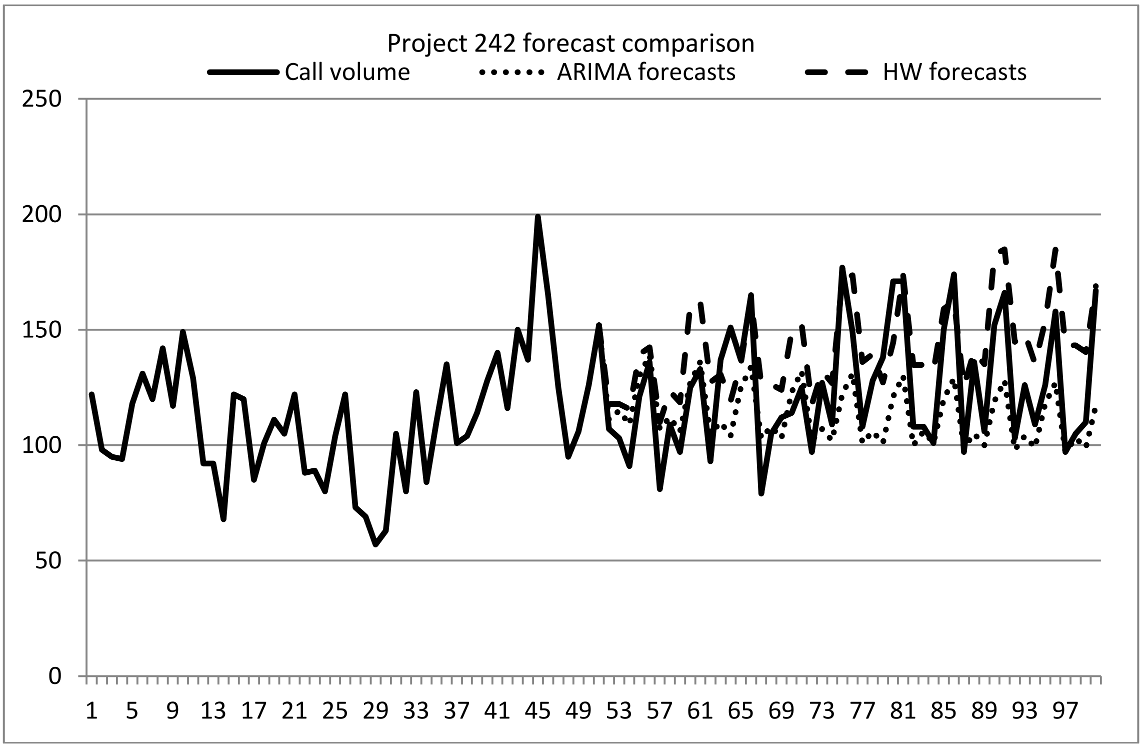
For project 194 (the last part) there is no significant autocorrelation (ADF test statistic of -6.842), therefore we do not use differencing. Furthermore only the 1st and 20th lag appear to be significant factors. In the graph below the results of the forecast can be found.

	ARIMA	HW
mse	882,9097	519,44
mae	24,44388	16,30
mpe	0,273273	0,15



As you can see in the chart the ARIMA forecast is a lot flatter than the real volume. This is probably due to the fact that there are just two significant factors. That way the process cannot incorporate all fluctuations. The performance is worse than the most of the Holt-Winters models, especially for the MAE and MPE measures. The errors from this model are:

Project 242 is also a project which is not stationary. Before differencing the Augmented Dickey Fuller statistic has a value of -2.765 and is hence we cannot reject the hypothesis that the process is non-stationary. After differencing the value of the ADF test is -12.646 and hence we reject the hypothesis and a differencing of one period. Furthermore, after estimating the model and fine tuning, it resulted in an ARIMA model with 1,2,3,4 and 5 as relevant Autoregressive factors and 1,5 and 6 as relevant Moving Average factors.

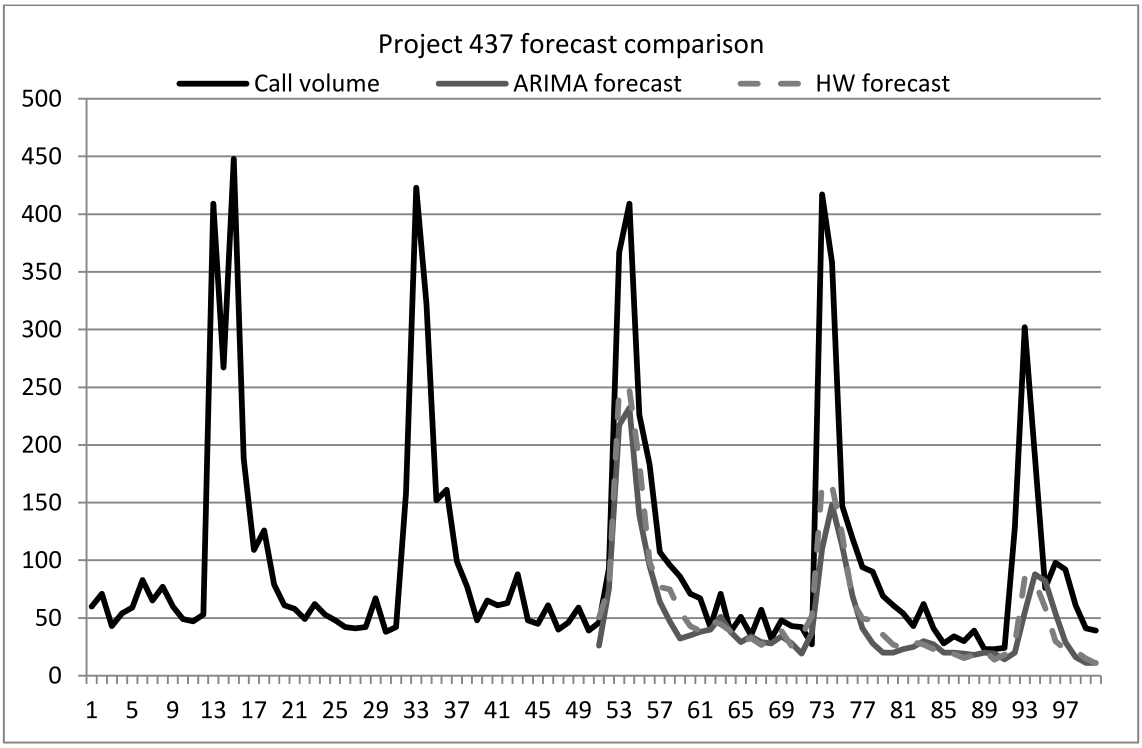


With errors:

	ARIMA	HW
mse	540,56	594,0721
mae	17,32	21,13537
mpe	0,128514	0,188498

If you compare these error measures with the best Holt-Winters method (double seasonality multiplicative model) you see that the ARIMA model outperforms the HW model slightly. Furthermore you see that the ARIMA model underestimates the true volume, where the HW model overestimates the volume.

For project 437 we see a much simpler model. Here only the 1,2 and 20 of the Autoregressive factor yield significant results. The moving average factors all showed to be insignificant. The error measures are:



	ARIMA	HW
mse	6605	5040,908
mae	50,68	44,80312
mpe	0,4439	0,411101

In the graph you can see the comparison between the best Holt-Winters model and the best ARIMA model. As you can see, both methods underestimate the peaks in the graph. As seen in the projects mentioned above, one of the proposed methods of the Holt-Winters forecasting model outperforms the ARIMA forecast, where the ARIMA forecast takes more time to calculate. However, it should be noted that not all Holt-Winters methods outperform the ARIMA forecast and that it is necessary to have a look at the forecasts to see whether they are logical. Another problem with this dataset is that variances can be a lot higher if the volume is higher. If a stochastic process has different variance over time, we call that process heteroskedastic.

9 CONCLUSIONS

In literature a lot of comparisons have been made between Holt-Winters based models and ARIMA models. These models are quite different because the ARIMA models assume an underlying distribution, where the Holt-Winters models are just a combination of separate effects. In this chapter we will make a comparison between these two methods and give some recommendations for future research.

In a large scale comparison of forecasting models (Granger & Newbold, 1973) it is shown that Box Jenkins models outperform (comparing MSE of the forecasts) Holt-Winters models in two third of the cases. The Box Jenkins procedure requires the user to identify an appropriate model in a general class of ARIMA models. This procedure therefore requires the user to make choices about the appropriate factors in the ARIMA model. The comparison made in that paper are therefore somewhat biased, since it is compared to the fully automatic Holt-Winters model. The paper by Chatfield (Chatfield, 1978) shows that the Holt-Winters forecasting procedure can also be improved when human interference is used. This paper was based on the same datasets used in Granger & Newbold. Simply having a user choose between additive and multiplicative Holt-Winters increases the performance. Furthermore clearing data from outliers improves the performance as well.

In our data we have seen that carefully choosing the right Holt-Winters method is important for the performance. In three of four projects the best Holt-Winters model outperforms the ARIMA model. However, the other Holt-Winters models are generally worse than the ARIMA model. Choosing a good forecasting method is therefore essential to a good performance. The best method in the optimisation process of the HW models is not necessarily the best method for the forecasting part. To achieve the best result you should look whether the trend in the forecast is logical (so not to steep or strongly negative). Holt-Winters models which forecast negative volumes can be ruled out immediately. It is harder to determine if a positive slope is too big. This resulting slope, after all, resulted from the model and is based on past observations. But since the Holt-Winters method constantly adapts itself to new observations, any new data should be put into the model to correct the forecasts. As a general rule it is good to look at the forecast of the error measure which performed best in the optimisation process (i.e. in forcasting the last 50 observations).

For future research several recommendations can be made. The missing data analysis and outlier detection relies heavily on the assumption that call arrivals are Poisson distributed. However, this assumption only holds on small intervals (i.e. neighbourhoods). More research can be done about the distribution of calls over the different days and within the days. With a better known distribution, the missing data imputation can be improved.
The Holt-Winters models developed in this thesis require human interference to get the best results. It is not certain that a procedure which yields good results in the optimisation procedure also yields good results when forecasting. Future research may be aimed at a method of automated choice of the Holt-Winters method.

Combining forecasting models might also yield good results. This could e.g. be a combination of Holt-Winters and ARIMA where both forecasts get a certain weight. Allan Timmermann has written quite come articles about combining forecasting methods, mostly in financial literature. See e.g. (Guidolin & Timmermann, 2009) or the book (G. Elliott, 2006)

The automatic seasonality detection method developed in this thesis is a straightforward method. Seasonality influences can only be detected when there is plenty of data. For smaller datasets, this method is therefore not suitable. A method can be developed to cope with this problem. We could e.g. use a bootstrapping method to generate more datasets and be able to draw conclusions about the sensitivity to small samples

Some datasets show signs of heteroskedasticity (i.e. different variance over time). To deal with this heteroskedasticity we can look at ARCH (Autoregressive Conditional Heteroskedasticity) models or GARCH (Generalised Autoregressive Conditional Heteroskedasticity) models. For these models see (Engle, 1982) and (Bollerslev, 1986) respectively.

Furthermore, the usefulness of SARIMA models in call centre data could be examined. In this thesis the estimation model of Stata could only find a sub-optimal SARIMA estimate in these samples, there might be other methods to find a better fit. Furthermore a SARIMA-GARCH model might give a good fit.

Also the Kalman Filter could be used in future research on call centre data. In this thesis the missing data detection and the outlier detection where done separately and are based on the Poisson distribution. The Kalman filter can work with data which contain random noise and produce estimates which are closer to the true mean and therefore might be more useful in forecasting volumes. In comparison; the Holt-Winter forecast gives an estimate based on future observations, without taking randomness into account.

Furthermore the forecasting could also be based on other factors than past call volumes. E.g. the influence of marketing campaigns or the number of magazine subscriptions sold is also a good indicator for call volume. Therefore we need more data and a more extensive model.

Forecasting of call volume will remain a difficult topic in research. It will always be hard to predict the behavior of callers and other external factors like the weather. Therefore this field of research will remain interesting in the future. It will always be a slippery slope to achieve good service while minimizing the costs. I would like to conclude with a quote by Andrew Lang about the fact that careful application of statistics is essential for making forecasts:

"An unsophisticated forecaster uses statistics as a drunken man uses lamp-posts, for support rather than for illumination."

Bibliography

Aldor-Noiman, S., Feigin, P. D., & Mandelbaum, A. (To Be Published). Workload forecasting for a call center: methodology and a case study. *Annals of Applied Statistics*.

Antipov, A., & Meade, N. (2002). Forecasting Call Frequency at a Financial Service Call Centre. *The Journal of the Operational Research Society, 53*(9), 953-960.

Atlason, J., Epelman, M. A., & Henderson, S. G. (2004). Call center staffing with simulation and cutting plane methods. *Annals of Operations Research, 1-4*(127), 333-358.

Azen, S., & Van Guilder, M. (1981). Conclusions regarding algorithms for handling incomplete data. *Proceedings of the Statistical Computing Section, American Statistical Association*(4), 53-56.

Berndt, E., Hall, B., Hall, R., & Hausman, J. (1974). Estimation and Inference in Nonlinear Structural Models. *Annals of Social Measurement, 3*, 653-665.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics, 31*, 307-327.

Brockwell, P. J., & Davis, R. A. (1996). *Introduction to Time Series and Forecasting.* New York: Springer Texts in Statistics.

Brown, L. D., & Zhao, L. H. (2001). *A new test for the Poisson distribution.* Philadelphia: Department of Statistics, University of Pennsylvania.

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., et al. (2002). *Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective.* Philadelphia: Department of Statistics, The Wharton School, University of Pennsylvania.

Broyden, C. (1970). The Convergence of a Class of Double-rank Mnimization Algorithms. *Journal of the Institute of Mathematics and Its Applications, 6*, 76-90.

Buck, S. (1960). A method for estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society B, 22*, 302-306.

Chatfield, C. (1978). The Holt-Winters Forecasting Procedure. *Journal of the Royal Statistical Society, 27*(3), 264-279.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*(Series B), 1-38.

Efron, B. (1994). Missing Data, imputation, and the bootstrap (with discussion). *Journal of the American Statistical Association*(89), 463-79.

Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. *Econometrica, 50*, 987-1008.

Fletcher, R. (1970). A New Approach to Variable Metric Algorithms. *Computer Journal, 13*, 317-322.

G. Elliott, C. G. (2006). *Handbook of economic forecasting* (Vol. 1). North Holland.

Gilson, K., & Khandelwal, D. (2005). Getting more from call centers. *The McKinsey Quarterly*.

Goldfarb, D. (1970). A Family of Variable Metric Updates Derived by Variational Means. *Mathematics of Computation, 24*, 23-26.

Granger, C., & Newbold, P. (1973). Some comments on the evaluation of economic forecasts. *Applied Economics, 5*, 35-47.

Guidolin, M., & Timmermann, A. (2009). Forecasts of US Short-term Interest Rates: A Flexible Forecast Combination Approach. *Journal of Econometrics, 150*, 297-311.

Hanke, J. E., & Wichern, D. W. (2009). *Business Forecasting* (Ninth Edition ed.). New Jersey: Pearson Education.

Hardy, M., & Bryman, A. (2004). *Handbook of data analysis.* London: SAGE Publications Ltd.

Holman, D., Batt, R., & Holtgrewe, U. (2007). *The global call center report: International perspectives on management and employment.* ILR Collection, Research Studies and Reports, Cornell University.

Holt, C. (1957). Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. *ONR Research Memorandum 52.* Pittsburgh: Carnegie Institute of Technology.

Jamshidian, M. (2004). Strategies for Analysis of Incomplete Data. In M. Hardy, & A. Brymann, *Handbook of data analysis* (2nd ed., pp. 113-130). London: SAGE Publications Ltd.

Jin, R., Chen, W., & Sudjianto, A. (2005). An Efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*(134), 268-287.

Kim, J., & Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods & Research*(6), 215-240.

Koole, G., & Mandelbaum, A. (2002). Queueing models of call centers: An introduction. *Annals of Operations Reserach, 113*, 41-59.

Little, R. J., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). New Jersey: John Wiley & Sons.

Little, R. J., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* New Jersey: John Wiley & Sons, Inc.

Mandelbaum, A. (2004). *Call Centers; Research Bibliography with Abstracts.* Haifa: Faculty of Industrial Engeneering and Management Technion - Israel Institute of Technology.

Rubin, D. (1976). Inference and missing data (with discussion). *Biometrika, 63*, 581-592.

Said, S. E., & Dickey, D. A. (1984). Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika, 71*(3), 599-607.

Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of Art. *Psychological Methods, 7*(2), 147-177.

Shanno, D. (1970). Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation, 24*, 647-656.

Shen, H., & Huang, J. Z. (2005). Analysis of call centre arrival data using singular value decomposition. *Applied Stochastic MOdels in Business and Industry*(21), 251-263.

Shen, H., & Huang, J. Z. (2008). Porecasting Time Series of Inhomogeneous Poisson Processes with Application to Call Center Workforce Management. *The Annals of Applied Statistics, 2*(2), 601-623.

Soyer, R., & Tarimcilar, M. (2008). Modeling and Analysis of Call Center Arrival Data: A Bayesian Approach. *Management Science, 54*(2), 266-278.

Taylor, J. W. (2003). Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. *The Journal of the Operational Research Society, 54*(8), 799-805.

Taylor, J. W. (2008). Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center. *Management Science, 54*(2), 253-265.

van Dam, E., Husslage, B., den Hertog, D., & Mellissen, J. (2007). Maximin Latin Hypercube Designs in two dimensions. *Operations Research, 55*(1), 158-169.

Winters, P. (1960). Sales by Exponentially Weighted Moving Averages. *Management Science, 6*(3), 324-342.

APPENDIX A

In this appendix all sub projects of the Sanoma workgroup are mentioned for reference. The workgroup number is 100090.

187: Sanoma nw abbo 0800 SAP

Data from week 7, 2008 until week 21, 2009. Handles questions about subscriptions of callers calling the 0800-number.

188: Sanoma abbo SAP

Data from week 31, 2008 until week 21, 2009. Handles questions about subscriptions.

189: Sanoma bezorgklachten SAP

Data from week 31, 2007 until week 21, 2009. Handles complaints about delivery. A similar project (372) started later on and took over the calls for this project. At the end of the data, the amount of calls for this project was really low.

190: Sanoma opzeggingen SAP

Data from week 44, 2007 until week 21, 2009. Handles cancellation of the subscriptions.

191: Sanoma wijzigingen overig SAP

Data from week 44, 2007 until week 21, 2009. Handles changes to subscription details. E.g. bank account number or delivery address.

192: Sanoma artikelen SAP

Data from week 31, 2007 until week 21, 2009. Handles with any kind of merchandise and other articles sold by Sanoma.

193: Sanoma Flex SAP

Low amount of data and terminated early. Not used in the study.

194: Sanoma maningen SAP

Data from week 31, 2007 until week 21, 2009. Handles with unpaid subscription fees.

195: Sanoma Disneyclub SAP

This project had reliable data until the end of 2008. After that it was terminated.

196: Sanoma geen keuze SAP

Data from week 31, 2007 until week 21, 2009. Handles with all people who do not make a choice in one of the automatic menus. Quite low volumes.

242: Sanoma betaling SAP

Data from week 35, 2007 until week 21, 2009. Handles with all questions related to payments.

244: Sanoma specials SAP

This project had reliable data until the end of 2008. After that it was terminated.

322: Sanoma doorverbonden recovery

This project was terminated

372: Sanoma uitval bezorgklachten

Data from week 46, 2007 until 21, 2009. In the first months of the project, the number of calls was really low. From week 6, 2008 the volume steadily increased. It took over the calls from project 189. This project handles complaints about delivery.

411: Sanoma acceptgiro SAP

Data from week 19, 2008 until week 21, 2009. It handles transaction forms for the subscriptions.

437: Sanoma maningen Disney SAP

Data from week 33, 2008 until week 21, 2009. It handles with unpaid subscription fees of the Disneyclub.

APPENDIX B

Scoring matrix for the neighbourhood estimation. The values in the matrix indicate the 'scores' for the different sizes of the neighbourhood. Lower scores mean better results.

Same time period observations

Same day observations

	1	2	3	4	5	6	7	8	9	10
1	300	289	255	195	176	86	56	63	107	135
2	297	277	237	180	151	71	48	48	87	124
3	294	282	241	206	180	120	80	77	109	129
4	290	276	243	208	183	133	98	84	108	125
5	283	245	211	150	103	52	39	27	38	64
6	264	234	205	137	99	50	34	24	26	51
7	268	245	218	175	147	81	60	44	59	85
8	277	256	227	196	170	122	99	78	97	105
9	269	255	225	199	180	137	113	93	108	111
10	260	235	220	187	157	105	81	62	73	87

APPENDIX C

```
. regress calls holiday

      Source |       SS       df       MS              Number of obs =     657
-------------+------------------------------           F(  1,   655) =    5.11
       Model |   15820.306     1   15820.306           Prob > F      =  0.0241
    Residual |  2027107.57   655  3094.82072           R-squared     =  0.0077
-------------+------------------------------           Adj R-squared =  0.0062
       Total |  2042927.88   656  3114.21932           Root MSE      =  55.631

------------------------------------------------------------------------------
       calls |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     holiday |  -16.43511   7.269138    -2.26   0.024    -30.70873   -2.161486
       _cons |   124.0659   2.286425    54.26   0.000     119.5763    128.5555
------------------------------------------------------------------------------

. regress calls l1.calls holiday

      Source |       SS       df       MS              Number of obs =     656
-------------+------------------------------           F(  2,   653) =  467.09
       Model |  1171894.84     2   585947.42           Prob > F      =  0.0000
    Residual |  819170.488   653  1254.47242           R-squared     =  0.5886
-------------+------------------------------           Adj R-squared =  0.5873
       Total |  1991065.33   655  3039.79439           Root MSE      =  35.419

------------------------------------------------------------------------------
       calls |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       calls |
         L1. |   .7543603   .0249337    30.25   0.000     .7054004    .8033201
     holiday |  -4.661855   4.688783    -0.99   0.320    -13.86877    4.545056
       _cons |   30.15822   3.428304     8.80   0.000     23.42639    36.89005
------------------------------------------------------------------------------

. regress calls l1.calls holiday monday tuesday wednesday thursday

      Source |       SS       df       MS              Number of obs =     656
-------------+------------------------------           F(  6,   649) =  282.31
       Model |  1439510.48     6  239918.414           Prob > F      =  0.0000
    Residual |  551554.844   649   849.85338           R-squared     =  0.7230
-------------+------------------------------           Adj R-squared =  0.7204
       Total |  1991065.33   655  3039.79439           Root MSE      =  29.152

------------------------------------------------------------------------------
       calls |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       calls |
         L1. |   .8259081   .0214104    38.58   0.000      .783866    .8679501
     holiday |  -2.811179   3.861218    -0.73   0.467    -10.39317    4.770809
      monday |   8.510937   3.624127     2.35   0.019     1.394508    15.62737
     tuesday |   -50.2588   3.718125   -13.52   0.000    -57.55981   -42.95779
   wednesday |  -15.32798   3.603767    -4.25   0.000    -22.40443   -8.251531
    thursday |  -25.11711   3.605836    -6.97   0.000    -32.19762   -18.03659
       _cons |   37.70506   3.479474    10.84   0.000     30.87268    44.53745
------------------------------------------------------------------------------
```

APPENDIX D

Results of Anomaly Detection. Arrival rates are estimated by the neighbourhood of all days (so estimate Monday rate by Tuesdays, Wednesdays etc.)

Total										
MSE	943.6	915.3	993.5	998.1	**900.3**	946.2	996.0	1040.6	1058.5	1041.6
MAE	22.39	22.42	23.26	22.81	**21.22**	21.67	22.44	23.03	23.10	22.72
MPE	0.182	0.186	0.191	0.186	**0.171**	0.174	0.180	0.185	0.185	0.182
Monday										
MSE	**1587.5**	1821.2	1943.3	2015.0	1632.0	1716.0	1871.2	1966.2	1992.9	1873.8
MAE	32.73	35.76	36.69	36.27	**30.80**	31.37	33.07	34.05	34.33	32.63
MPE	0.283	0.315	0.320	0.310	**0.248**	0.252	0.268	0.278	0.278	0.258
Tuesday										
MSE	945.2	582.6	516.0	575.6	**514.7**	562.4	542.2	529.1	553.8	528.3
MAE	24.18	18.54	**17.44**	18.75	18.05	18.84	18.61	18.30	18.56	18.05
MPE	0.184	0.155	**0.149**	0.155	0.152	0.157	0.158	0.156	0.157	0.155
Wednesday										
MSE	687.0	**628.2**	732.8	655.9	654.7	678.5	725.1	787.7	788.7	809.6
MAE	17.33	18.95	21.27	19.48	**18.86**	18.55	19.54	20.56	20.13	20.12
MPE	0.157	0.152	0.164	0.154	**0.151**	0.149	0.155	0.161	0.160	0.160
Thursday										
MSE	702.9	752.9	818.0	758.3	**651.9**	691.2	739.8	782.6	795.8	779.7
MAE	19.50	21.11	21.59	20.35	**18.58**	19.28	20.41	21.21	21.33	20.95
MPE	0.152	0.164	0.167	0.159	**0.151**	0.155	0.162	0.167	0.168	0.167
Friday										
MSE	**814.8**	821.3	987.1	1017.0	1067.5	1103.8	1126.2	1163.5	1187.5	1238.7
MAE	18.58	**18.23**	19.73	19.65	20.07	20.61	20.89	21.40	21.53	22.18
MPE	**0.140**	0.149	0.162	0.155	0.156	0.158	0.162	0.168	0.168	0.172

Results of Anomaly Detection. Arrival rates are estimated by the neighbourhood of the same days (so estimate Monday rates by other Mondays only).

Total										
MSE	1115	**1086**	1146	1188	1259	1314	1380	1399	1431	1484
MAE	22.94	**22.19**	22.82	23.29	24.20	25.09	25.89	26.24	26.78	27.29
MPE	0.188	**0.181**	0.183	0.183	0.191	0.199	0.205	0.208	0.212	0.217
Monday										
MSE	1223	**1189**	1325	1450	1540	1594	1702	1722	1751	1797
MAE	24.39	**24.00**	24.83	25.66	26.93	28.28	29.73	30.22	30.59	30.93
MPE	0.159	**0.156**	0.160	0.163	0.172	0.182	0.191	0.195	0.197	0.200
Tuesday										
MSE	751	625	636	**615**	652	703	760	793	825	865
MAE	20.79	19.44	18.95	**18.54**	18.65	19.76	20.43	20.94	21.58	22.01
MPE	0.186	0.181	0.175	**0.167**	0.167	0.179	0.184	0.187	0.192	0.196
Wednesday										
MSE	**1033**	1036	1096	1144	1226	1255	1274	1262	1248	1311
MAE	20.80	**20.71**	21.96	22.65	23.70	24.35	24.85	24.65	24.74	25.24
MPE	**0.172**	0.174	0.181	0.182	0.189	0.197	0.203	0.204	0.205	0.209
Thursday										
MSE	775	**749**	852	967	1076	1125	1171	1200	1252	1305
MAE	20.30	**18.87**	19.94	21.33	22.73	23.49	24.09	24.41	24.98	25.51
MPE	0.196	**0.175**	0.182	0.191	0.205	0.212	0.216	0.218	0.223	0.228
Friday										
MSE	1954	1829	1821	**1763**	1803	1891	1994	2016	2081	2142
MAE	28.42	**27.95**	28.44	28.27	28.97	29.55	30.36	30.99	31.98	32.77
MPE	0.227	0.217	0.216	**0.212**	0.222	0.224	0.231	0.236	0.244	0.251

Results of Anomaly Detection. Arrival rates are estimated by the neighbourhood of the same day for Mondays, and by the neighbourhood of the other weekdays (except Mondays) for Tuesday till Friday.

Total										
MSE	862.9	**800.6**	849.6	868.7	925.5	950.5	981.1	995.9	1034	1081
MAE	20.32	**19.08**	19.53	19.5	20.26	20.71	21.25	21.49	21.92	22.35
MPE	0.167	0.155	0.157	**0.155**	0.16	0.164	0.168	0.17	0.173	0.176
Monday										
MSE	1365	**1238**	1457	1573	1618	1688	1750	1767	1816	1924
MAE	24.84	**23.79**	25.39	26.56	27.45	28.66	29.91	30.37	30.69	31.49
MPE	0.161	**0.155**	0.162	0.166	0.172	0.181	0.19	0.195	0.198	0.203
Tuesday										
MSE	560.6	**399.5**	431.9	430.4	459.4	452.5	454.3	446	463	479.6
MAE	17.88	**14.85**	15.94	16.12	16.69	16.65	16.74	16.66	16.92	17.08
MPE	0.155	**0.139**	0.144	0.146	0.149	0.151	0.151	0.151	0.153	0.154
Wednesday										
MSE	680.4	672	**658.3**	670.1	719.6	745.9	775.9	808	828.3	877.6
MAE	17.13	17.95	17.29	**17**	17.46	17.81	18.22	18.74	19.05	19.73
MPE	0.155	0.146	0.144	**0.142**	0.148	0.148	0.153	0.157	0.16	0.163
Thursday										
MSE	700.8	**466.3**	495.1	470.6	545.4	547.3	587.9	603.9	655.5	672.8
MAE	19.44	15.94	15.89	**15.71**	16.69	17.09	17.71	17.96	18.55	18.68
MPE	0.151	0.135	**0.133**	0.133	0.139	0.145	0.148	0.151	0.154	0.156
Friday										
MSE	1143	**1109**	1217	1213	1298	1332	1352	1369	1422	1470
MAE	23.48	**21.94**	23.22	22.24	23.17	23.56	23.89	23.93	24.59	24.99
MPE	0.218	0.195	0.202	**0.185**	0.193	0.195	0.198	0.195	0.201	0.204

APPENDIX E

Results to determining the weekday patterns for project 372

	Monday		Tuesday		Wednesday	
# periods	MSE	MAE	MSE	MAE	MSE	MAE
1	0.000178	0.010452	0.000295	0.013405	0.000376	0.014889
2	0.000143	0.009153	0.000233	0.011871	0.000299	0.013553
3	0.000127	0.008645	0.000201	0.011079	0.000265	0.012778
4	0.000120	0.008388	0.000192	0.010779	0.000243	0.012105
5	0.000117	0.008305	0.000189	0.010690	0.000230	0.011832
6	0.000111	0.008093	0.000182	0.010520	0.000221	0.011607
7	0.000105	0.007896	0.000178	0.010412	0.000216	0.011531
8	0.000104	0.007849	0.000178	0.010425	0.000216	0.011525
9	0.000104	0.007868	0.000175	0.010391	0.000215	0.011451
10	0.000102	0.007755	0.000175	0.010405	0.000213	0.011427
11	0.000102	0.007748	0.000175	0.010414	0.000210	0.011399
12	0.000103	0.007835	0.000171	0.010306	0.000206	0.011269
13	0.000103	0.007800	0.000171	0.010256	0.000207	0.011282
14	0.000102	0.007775	0.000168	0.010168	0.000206	0.011238
15	0.000099	0.007739	0.000167	0.010134	0.000204	0.011191
16	0.000098	0.007674	0.000167	0.010116	0.000203	0.011169
17	0.000098	0.007674	0.000166	0.010116	0.000201	0.011107
18	0.000099	0.007728	0.000165	0.010110	0.000201	0.011092
19	0.000099	0.007721	0.000166	0.010132	0.000199	0.011045
20	0.000100	0.007741	0.000165	0.010114	0.000198	0.010992
21	0.000100	0.007762	0.000163	0.010067	0.000199	0.011008
22	0.000101	0.007799	0.000164	0.010074	0.000199	0.011011
23	0.000099	0.007766	0.000164	0.010056	0.000199	0.011007
24	0.000099	0.007751	0.000163	0.010019	0.000198	0.010976
25	0.000099	0.007725	0.000162	0.009998	0.000198	0.010992
26	0.000098	0.007690	0.000161	0.010004	0.000199	0.011032
27	0.000098	0.007711	0.000161	0.009976	0.000198	0.011017
28	0.000098	0.007700	0.000161	0.009971	0.000198	0.011012
29	0.000099	0.007715	0.000160	0.009953	0.000198	0.011003
30	0.000098	0.007726	0.000160	0.009966	0.000198	0.010982

Results to determining the weekday patterns for project 372

	Thursday		Friday	
# periods	MSE	MAE	MSE	MAE
1	0.000377	0.015065	0.000206	0.011016
2	0.000290	0.013022	0.000165	0.009887
3	0.000251	0.012210	0.000142	0.009206
4	0.000235	0.011757	0.000135	0.009057
5	0.000230	0.011630	0.000134	0.009082
6	0.000219	0.011374	0.000133	0.009081
7	0.000209	0.011120	0.000129	0.008879
8	0.000204	0.010983	0.000127	0.008838
9	0.000202	0.010959	0.000127	0.008796
10	0.000200	0.010913	0.000124	0.008746
11	0.000200	0.010904	0.000122	0.008708
12	0.000200	0.010897	0.000121	0.008608
13	0.000196	0.010841	0.000121	0.008580
14	0.000196	0.010805	0.000122	0.008612
15	0.000195	0.010806	0.000122	0.008610
16	0.000195	0.010802	0.000123	0.008657
17	0.000195	0.010776	0.000123	0.008645
18	0.000198	0.010800	0.000124	0.008691
19	0.000198	0.010770	0.000124	0.008702
20	0.000197	0.010754	0.000123	0.008721
21	0.000198	0.010783	0.000123	0.008708
22	0.000197	0.010784	0.000122	0.008672
23	0.000197	0.010778	0.000123	0.008699
24	0.000199	0.010847	0.000122	0.008659
25	0.000198	0.010761	0.000123	0.008691
26	0.000197	0.010765	0.000124	0.008709
27	0.000197	0.010773	0.000124	0.008682
28	0.000197	0.010785	0.000123	0.008642
29	0.000197	0.010751	0.000123	0.008646
30	0.000197	0.010724	0.000123	0.008635

APPENDIX F

Results of forecasting optimisation.

192	**single multi**	single add	double multi	double add
alpha	**0,82**	0,77	0,71	0,01
beta	**0,23**	0,36	0,41	0,29
delta	**0,18**	0,36	0,46	0,27
gamma	**0,16**	0,42	0,04	0,03
mse	**183,76**	188,62	208,44	269,26
mae	**10,24**	10,66	11,48	12,14
mpe	**0,14876912**	0,149623	0,15475798	0,17178285
194	single multi	single add	**double multi**	double add
alpha	0,91	0,95	**0,03**	0,68
beta	0,64	0,57	**0,44**	0,03
delta	0,75	0,66	**0,68**	0
gamma	0,57	0,95	**0,1**	0,63
mse	1161,6	1232,16	**957,28**	997,54
mae	22,92	24,32	**21,36**	21,5
mpe	0,20518941	0,2200554	**0,203179325**	0,21440845
242	**single multi**	single add	double multi	double add
alpha	**0,71**	0,71	0,92	0,43
beta	**0,93**	0,93	0	0
delta	**0,23**	0,23	0,79	0,48
gamma	**0,71**	0,71	0,68	0,42
mse	**557,72**	586,92	579,16	661,94
mae	**17,48**	17,52	17,44	20,42
mpe	**0,16955824**	0,1637617	0,184346352	0,20933731
437	**single multi**	single add	double multi	double add
alpha	**0,05**	0,42	0,13	0,29
beta	**0**	0	0,05	0,06
delta	**0,24**	0,26	0,77	0,05
gamma	**0,32**	0,32	0,55	0,27
mse	**2685,72**	6680,24	3777,42	6369,96
mae	**31,68**	54,4	37,3	51,88
mpe	**0,31800078**	0,6070767	0,431071304	0,56730962