Time-Series Forecasting Methods In Call Centers

Carl van Roon (2554086)
Supervisor: Prof. Dr. Ger Koole

Vrije Universiteit Amsterdam
Research Paper Business Analytics
March 9, 2019

1 Summary

Forecasting is a method that helps companies to predict the future. In the case of Call Centers it is possible to predict the incoming calls on different kinds of levels. For example, a forecaster can forecast on a yearly, quarterly, monthly, weekly, daily and intra-daily level. A well-executed forecast can help both customers and companies alike. Reducing both costs for the company and waiting times for the customers. This paper sets out to investigate several traditional methods of forecasting. Furthermore, weekly fractions are introduced. Weekly fractions help expand the scope of a one-day ahead forecast to a one-week ahead forecast. This one-week ahead forecast is still on a daily level and still uses the traditional methods. Last, a simple linear regression method is introduced to compare to the aforementioned methods. The performance of each method is measured via the WAPE. The result of this paper is that the weekly fractions help reducing the WAPE significantly, while still using simple methods. Linear regression is very suitable for this kind of data by using dummy variables. More research needs to be done on linear regression to see how much can be improved upon.

Contents

2 Introduction

The 21st century is nicknamed the Information age, where the economy is based on information technology. Companies have stored huge amounts of data for years. Yet, the analysis of this data is a recent phenomenon. For example: Exastax, a Big Data Solution company, noted that in 2017 most airlines did not take advantage of big data technology [1].

Call centre Helper, a leading online Contact Centre and Customer Service magazine, found that about half of all contact centres in 2016 used manual forecasting [2]. In Figure 1 other methods of forecasting can be seen, some of which are discussed later.

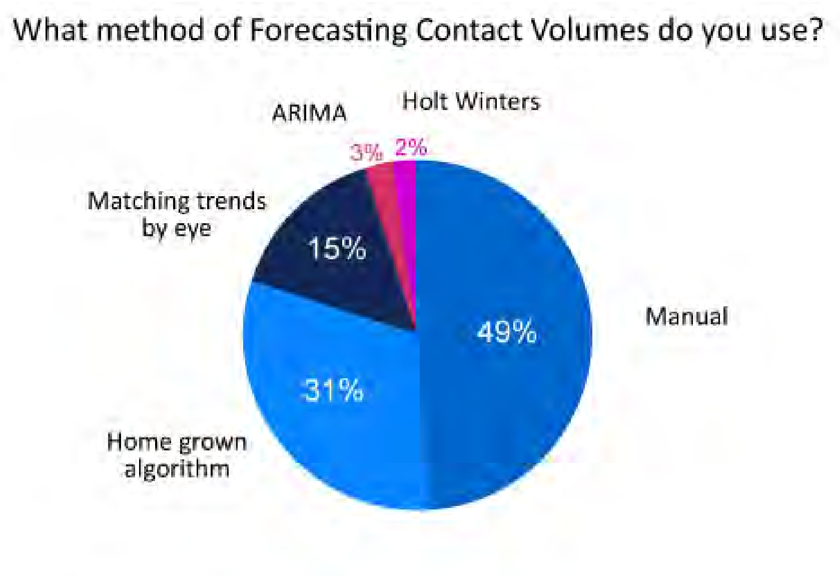


Figure 1: Usage of Forecasting methods

The 50% of manual forecasting shows that no formulas or algorithms are used, rather there is an educated guess. However, forecasting can be a tedious and repetitive task. Humans tend to cut corners on repetitive tasks leading to errors. Furthermore, humans are biased and tend to look at data that favors them. A study done by Briony D. Pulford [3] shows that humans tend to be overconfident in situations with positive outcomes than for those with negative outcomes. Mathematical models are able to cope with all kinds of data and they can do it without bias or getting bored and tired. Hence, a more mathematical approach to forecasting is advisable. Indeed, Davenport and Harris [4] argue and show that data-driven insight via predictive modelling and statistical analysis generate impressive business results.

This paper shows which methods can be used and on what scope of time. It compares the traditionally known methods like moving average and Holts' method to a method that incorporates weekly fractions. Last a comparison is drawn between the aforementioned methods and a simple linear regression with dummy variables.

In order to not make the scope of this paper too broad, the focus of this paper lies on the forecasting of incoming calls. Namely:

- Which forecasting methods can be used?
- Which forecasting methods are used in practice?
- How do these methods perform on real data?

Furthermore, the paper incorporates a method to forecast on a longer time span while still using classical methods. This method aims to help these methods and lower the overall error.

This paper starts with the application in practice for call centers in section 3. Section 4 shows which methods can be used and which methods are researched in this paper. This section also discusses the new method. Moreover, this section looks at other literature as well. Section 5 will go in depth into the data and the struggles when working with this data. Furthermore this section discusses how the best method is determined in this paper. Section 6 will show the results found in the research, together with an analysis of the results. Section 7 will cover the conclusion found via the previous section. Last, section 8 will cover a discussion about the algorithm used and future endeavors on forecasting.

3 Application in Practice

Since call centers are usually the first line of contact between customers and companies, a fast handling of calls can be crucial for a company. In most call centers the capacity costs in general, and human resource costs in particular, account for 60–70% of operating expenses [5]. Hence, a good staffing level can drastically reduce costs. The fast handling of calls can be done with enough staffed call center employees. In order to staff the right amount of people a forecast for the number of calls is necessary. This whole cycle from forecasting calls to staffing the employees is commonly referred to as *Work Force Management* (WFM). The definition of WFM, according to Koole [6], is: "The common name of the planning cycle that results in the schedules of the call center agents, usually a few weeks before the period for which the schedule is made. As input it uses historic call center data on traffic loads and information on agent availability". Figure 2 shows the idea of WFM where each column represents a stage of WFM and the rows represent the time span in which each process of each stage happens. The process range from a year in advance (Strategic) to a day in advance (Operational).

As mentioned in the introduction, the scope of this paper lies in the use of historic traffic loads of call centers to predict future traffic loads. Namely, a daily forecast with one day and one week in advance. This paper does not research the influences of events like commercials or advertisements. Nor does this paper research long-term forecasting.

3.1 In-house forecasting

While a forecast is easily made, the hardest part of forecasting is to be accurate. The forecaster is required to have considerable knowledge about the data and the methods used in forecasting. Furthermore, working with data follows the "Garbage In, Garbage Out" principle, where bad input of data results in bad forecasting. A bad forecast can result in overstaffing or understaffing. When a call center is overstaffed agents are idle for a lot of time. This means that the call center could have had less agents in service and thus reducing costs. On the other hand if the call center is understaffed, all agents will be occupied and not able to help arriving calls. If a customer has to wait for too long before being assisted by a call center worker, the customer is prone to hanging up the phone. This, in turn, results a dissatisfied customers.

3.2 Outsourced forecasting

A good forecast is not only crucial for in-house call centers. Whenever a company decides to outsource their call center, a contract is made where an agreement on a number of handled calls is made. When the handled number of calls is exceeded, the company usually has to pay a fee for the extra number of calls the outsourced company has made. Additionally, when a company expects that they are not able to handle the traffic load with the current amount of

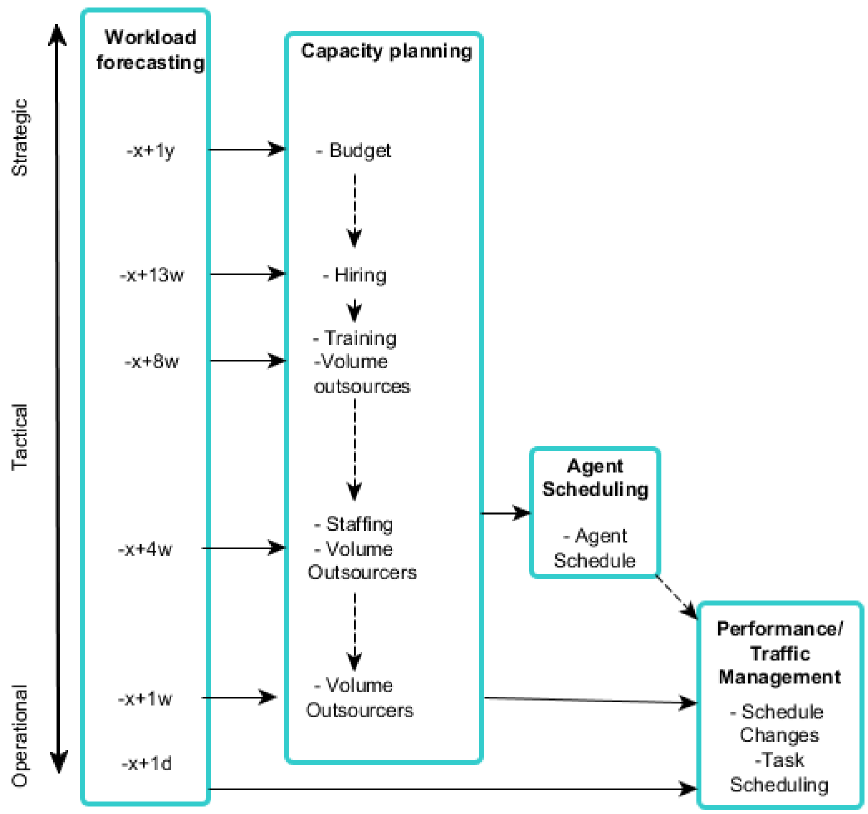


Figure 2: The WFM Processes courtesy of CCMath

staff, a percentage of calls can be redirected to the outsourced company. Hence, forecasting is crucial for both when the call-center is outsourced or in-house.

3.3 Effectiveness of methods in practice

The effectiveness of each of the methods discussed in the next section depends on the kind of data presented. According to Makridakis, Wheelwright and Hyndman, four different of data patterns can be distinguished. These four are: horizontal, seasonal, cyclical and trend.

First, a *horizontal* pattern occurs whenever the data fluctuates around a certain mean. One can imagine this happening whenever data does not increase or decrease over time. One example can be a product which sales stay constant.

Second, a *seasonal* pattern occurs whenever a time series is influenced by certain seasonal factors. These seasonal patterns can occur intra-weekly, per quarter of the year, every month and so on. An example of this is the increase in sales of kids toys around December or the increase in consumption of ice cream in the summer.

Third, a *cyclical* pattern happens whenever data rises or falls without a fixed period. The difference between a cyclical pattern and a seasonal pattern is the fact that the former does not happen at regular periodic basis and tends to vary in length. An example of cyclical patterns is an economic time series that changes due to economic fluctuations.

Last, a *trend* pattern exists whenever there is a long-term increase or decrease in data. For instance, a company that grows usually exhibits an increase in sales over time.

These four methods are not mutually exclusive. A time series can both exhibit a trend and a seasonality. The aforementioned company that grows over time can still have a lower number of sales in a certain period, but in the long run have an increase of sales. The existence of these four patterns in the data all influence how effective each method is. For example, a method not equipped to deal with seasonality will fare far worse than a method that is able to handle that seasonality. The next section will mention what every method is equipped to deal with besides the explanation of each method.

4 Methods and focus

Several methods for forecasting exists, ranging from naive to more complex mathematical structures. Several methods used for forecasting are:

- **Mean**
- **Moving Average**
- **Single Exponential Smoothing**
- **Adaptive Single Exponential Smoothing**
- **Holt's Linear Method**
- **Holt-Winters' Method**
- **Weekly fractions**
- **Linear Regression**
- ARIMA models
- Neural Networks

The methods in bold are used in this research. Either these methods are used in practice (e.g. Linear regression or Holt-Winters') or are well known forecasting methods. These methods are also found in the book Forecasting: Methods and Applications [7]. The methods not in bold will be discussed in the Discussion section of this paper.

Note that many variables in the next sections have a subscript denoting a certain time (t, $t-1$, etc). The context of time depends on the data provided by the call center, more of which will be explained in Section 5. For the sake of clarity and to avoid repetition, equal variables with different subscripts are mentioned once, but all note to a different point in time.

4.1 Mean

This method takes the mean over a certain time span $i-1,\ldots,t$ and uses this mean as the forecast for the next time $i=t+1$. It is possible to take one general mean and use it for all future forecasts. For a certain time t the mathematical formula is:

$$F_{t+1} = \frac{1}{t}\sum_{i=1}^{t} Y_i$$

Where F_{t+1} is the forecast for time $t+1$ and Y_i is the observed number of calls for time $i = 1,...,t$

Since more data becomes available at each time step, the mean is able to change per day as well. Instead of using the data of all previous days, the mean can also be calculated recursively via:

$$F_{t+1} = \frac{(t-1)F_t + Y_t}{t}$$

A big drawback of this method is that after a certain amount of time, the forecast is prone to rigidity, as new observations have less impact on the mean. Furthermore, this method suffers from the so-called "Flaw of averages", where plans based on average are usually wrong. This is due to the fact that the average does not take peaks and valleys into account. Hence, the mean would only be useful when the data has a horizontal pattern. However, as section 5 will show, this is not the case.

4.2 Moving Average (MA)

The moving average expands on the idea of the mean, but only looks at a certain time frame rather than the whole time. This means that after an observation has happened far enough in the past, the observation is discarded for the calculations, while newer observations are taken into account. The idea behind this is that old observations provide less value than newer observations. A common practice is to take the time span of a season. For example, with daily data the time span would cover a whole week, meaning that only the last 7 observations are taken into account. This results in the following formula for the k-order Moving Average (MA(k)):

$$F_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^{t} Y_i$$

Where F_{t+1} is the forecast for time $t+1$ and Y_i is the observed number of calls for time $i = 1, ..., t$ and k is the order of the moving average.

The advantage of this can be found in the fact that a forecast made with a Moving Average is not as rigid as the general mean, due the accountability of only the latest k observations. However, big changes in the k observations are still prone to the Flaw of averages. This means that data that is horizontal will work for this method. While not entirely capable of handling trend, it does so better than the mean.

Note that the first k observations cannot be forecast, as i will be earlier than the earliest observation.

4.3 Single Exponential Smoothing (SES)

Single Exponential Smoothing (SES) is an extension on the Moving Average. However, rather than discarding the older observations, SES assigns exponentially decreasing weights to older observations. In particular, this method uses the previous forecast and its forecasting error to forecast the next period, namely:

$$F_{t+1} = F_t + \alpha(Y_t - F_t)$$

Where F_t is the forecast for time t, Y_t is the observed number of calls for time t and α is a constant between 0 and 1. The higher α, the higher the "adjustment-factor" of the forecasting error of time t is. When $\alpha = 1$, SES is taking the latest observation as its forecast-i.e., naive forecasting. The value of α can be found by calculating the value on which the total error of all forecasting is the lowest, either by grid search or non-linear optimization.

Seeing as F_1 is not know, as it needs F_0 and Y_0, one should take $F_1 = Y_1$. SES is able to handle horizontal patterns and tends to lag behind on trend patterns. This lag is due to the fact that the method can only adjust the next forecast for some percentage of the most recent error. Seasonal patterns and cyclical patterns are still not recognized by SES.

4.4 Adaptive Single Exponential Smoothing (ASES)

Adaptive Single Exponential Smoothing (ASES) was developed by Trigg and Leach [8] and takes the idea of SES but does not use a fixed α. Rather, α can be modified as time passes and changes in the data occur. The motivation is that some time periods require low adjustments and some require high adjustments. ASES requires an α at every time t to use a formula similar to SES. The α_t is calculated via the smoothed estimate of forecast error and an absolute smoothed estimate of forecast error:

$$F_{t+1} = \alpha_t Y_t + (1 - \alpha_t)F_t$$

where

$$\alpha_{t+1} = |\frac{A_t}{M_t}|$$
$$A_t = \beta E_t + (1 - \beta)A_{t-1}$$
$$M_t = \beta|E_t| + (1 - \beta)M_{t-1}$$
$$E_t = Y_t - F_t$$

Here α_t is the adjustment constant, F_t is the forecast and Y_t is the observed number of calls, all at time t. The A_t is the aforementioned smoothed estimate of the forecast error, while M_t is the absolute estimate and E_t is the error at time t. Last the β is a constant between 0 and 1, defined by the user. Note that the α calculated is calculated for one time step in future $t+1$. This α_{t+1} is preferred as ASES is often too responsive to changes, hence a small lag of one period is incorporated.

Just as with SES, there is a problem at the initialization (i.e. $t = 1$). When calculating A_1 and M_1, A_0 and M_0 are required but are not defined, leading to an unknown F_2. Hence to intialize it is possible to take:

$$F_2 = Y_1,$$

$$\alpha_2 = \alpha_3 = \alpha_4 = \beta,$$
$$A_1 = M_1 = 0$$

Different starting values for α will return different α_t over time. Furthermore the value of β determines the strictness of changes to α. It is once again possible to find the best β to the data via non-linear optimization or grid search by taking the lowest overall error.

The advantage of ASES is that the adjustment constant changes over time, thus adapting to the data. However, there might be some lag between the data changing and the α_t to catch up. Hence, ASES is able to handle similar data patterns as SES. Furthermore, it might be the case that SES with an optimal α results in a lower overall error than ASES. Examples of this can be found in Gardner and Dannenbring [9].

4.5 Holt's Method

Holt's Method (Holt's) is an extension of SES. Holt's allows forecasting for data with a horizontal and trend pattern. Holt's method uses two smoothing constants, namely α and β. These constants are used to calculate the level estimate of the forecast and the slope of the forecast respectively. This results in:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$$
$$F_{t+m} = L_t + b_t m$$

Where L_t is the level estimate, Y_t is the observed number of calls and b_t is the trend estimate at time t. The forecast F_{t+m} is made for time $t+m$, where m is the number of time steps ahead the forecast needs to be made for. Furthermore, α and β are the adjustment constants between 0 and 1 for level and trend, respectively. Note that, unlike ASES, L_t and b_t need to be calculated before the forecast is made. Namely, the level is adjusted first via the observation at time t and the predicted observation at time $t-1$. The strictness of the adjustment depends on the value of α. Next, the trend is adjusted via the difference in levels and the trend of one time stamp earlier. Last, the forecast is calculated by taking the calculated level and adding the trend, possibly multiplied by the number of m time steps the users wants to predict. Hence, a one-day ahead forecast will be just the trained level L_t plus the trained trend b_t.

Once again the initialization is of importance here. At time $t = 1$ several variables are not defined at time $t = 0$, hence estimates for L_1 and b_1 are necessary. It is possible to set $L_1 = Y_1$ and $b_1 = Y_2 - Y_1$. Here one should optimize α and β to reduce the overall error via grid search or non-linear optimization.

Holt's method works especially well on data that has a trend and tends to outperform SES. However, once the data shows seasonality, Holt's has a hard time adjusting due to irregular movement of the data.

4.6 Holt-Winters' Method

All aforementioned methods have trouble adjusting to seasonality. However, the Holt-Winters' Method is able to handle seasonality, as well as horizontal and trend patterns by introducing a seasonality factor. More on seasonality is explained in Section 5. The extension of Holts' method was done by Peter Winters and encapsulates two different types of data series: Additive and multiplicative. An example of an additive and multiplicative time series can be found in Figure 3.

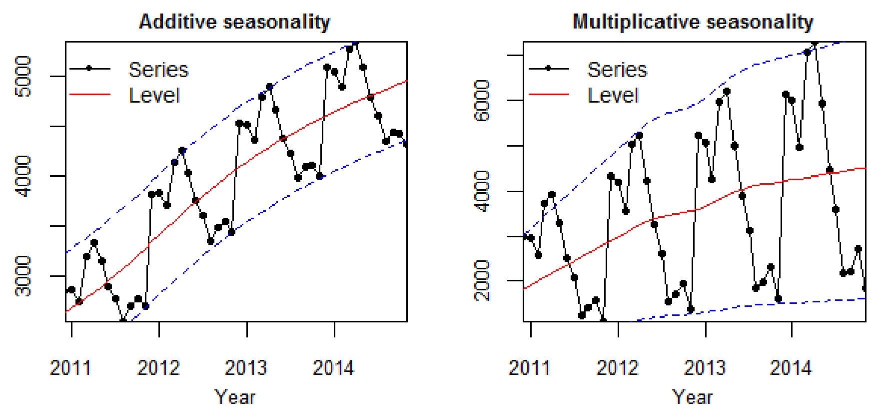


Figure 3: Plot showcasing additive versus multiplicative time series

This figure shows as time progresses the additive series' spread is constant, while the multiplicative spread becomes bigger. Depending on how the time series progresses the forecaster has to decide whether to use the multiplicative Holt-Winters' method or the additive Holt-Winters' method. First the multiplicative method:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$$

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

$$F_{t+m} = (L_t + b_t m)S_{t-s+m}$$

Here L_t is the level estimate, Y_t is the observed number of calls, S_t is the seasonal factor and b_t is the trend factor at time t. Furthermore, s is the length of the seasonality present in the data and m is the number of time steps ahead the forecast needs to be made for. Last, α, β and γ are the adjustment constants between 0 and 1 for the level, trend and seasonality respectively. The value for these constants can once again be found via grid search or non-linear optimization by minimizing the overall error. Holt-Winters' is very similar in its calculations to Holt's method, but the level takes the seasonality factor into account. The seasonality factor is a ratio between the observed number of calls and the level estimate. In case the actual calls are higher than the smoothed estimate, the ratio is higher than 1 and lower than 1 when the actual calls are lower than the smoothed level estimate. This ratio is then adjusted for via the γ and the seasonal factor of s periods previous. For the level estimate, it can be seen that there is an adjustment of the observed calls via the seasonality factor. Furthermore, the seasonality factor is once again used for the actual forecast.

While less common, the additive Holt-Winters' method is as follows:

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$$

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

$$F_{t+m} = L_t + b_t m + S_{t-s+m}$$

The formulas are very similar. The main difference is that the seasonal factors are subtracted and added, rather than divided.

For initialization, at least 2 seasons are required in order for Holt-Winters' to work, hence some data is already required. The level is initialized at time $t = s$ and is the average of the first season:

$$L_s = \frac{Y_1 + Y_2 + ... + Y_s}{s}$$

To initialize the trend, an average over all trends is taken:

$$b_s = \frac{1}{s}\left[\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + ... + \frac{Y_{s+s} - Y_s}{s}\right]$$

Last, the season factors for the multiplicative method are calculated via the ratio of the data values of the first season to the mean of the first season:

$$S_1 = \frac{Y_1}{L_s}, \quad S_2 = \frac{Y_2}{L_s}, \quad ... \quad S_s = \frac{Y_s}{L_s}$$

While the additive method subtracts the level estimate from the observed calls:

$$S_1 = Y_1 - L_s, \quad S_2 = Y_2 - L_s, \quad ... \quad S_s = Y_s - L_s$$

The main benefit of Holt-Winters is the fact it can handle seasonality very well, compared to the other methods. A time series can consist of a repetitive pattern where certain parts of the time series are consistently lower or higher than expected. One can imagine this for a store that is closed on the weekends: the next Monday likely has more customers than any other day of the week.

A drawback is that Holt-Winters' is unable to handle data with zero's. This means that in case a company has a day where a data point is 0 (for a call center meaning no incoming calls), Holt-Winters will crash. This is due to the fact that observations used for the initialization results in at least one zero for S_1 to S_s. Whenever this zero is used for the L_t calculation a division by zero will occur. For non-initialization observations the same happens when two observations that are one season are apart are zero. When both Y_{t-s} and Y_t are zero and $\gamma = 1$, the $S_{t-s} = 0$ and L_t will give a division by zero. Furthermore, since the multiplicative Holt-Winters uses fractions, the method will become quite unstable. This can be counteracted by setting the adjustment constants α, β, γ to 0 for that day. This means that the actual observations are not taken into account and the variables used in the past will get the full weight assigned. Last, Holt-Winters' method requires that some data is already present (the two periods for initialization). While this might be less of a problem for already existing companies, seeing as they are likely to collect data already, fresh starters without data are unable to use this method. Furthermore, the longer the seasonality, the longer the initialization will be. This can be seen for the weekly fractions.

4.7 Weekly fraction

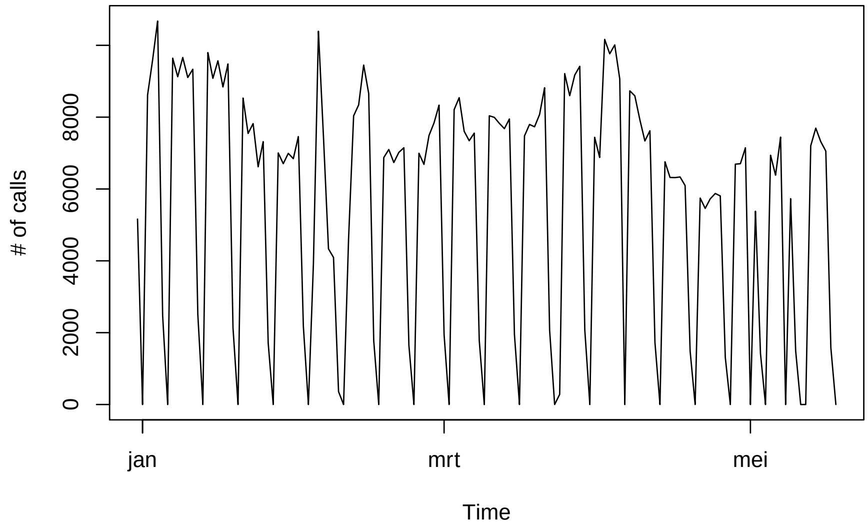
Weekly fractions are used in conjunction with the previous methods. Weekly fraction requires already present data as well. However, this method is able to forecast further into the future, with an arguably simpler methodology than Holt-Winters while still being able to handle seasonality. This method takes the idea that different time stamps in a season all are a fraction of the entire season. For example, if the season spans a week, every day of the week is a certain fraction of the total of that week. Hence, it is possible to sum up the days for one week, reducing the instability of the patterns and increasing the use of the more simpler methods to forecast. An example of this can be found in Figure 4. The result of summing up the days in a week removes the intra-week seasonality pattern and helps the methods that are able to handle horizontal and trend patterns. However, as seen in the Data section, the intra-year seasonality still remains.

Since the data is smoother, the more simpler methods of forecasting like Moving Average, Single Exponential Smoothing and Holt's method can be used to forecast one week ahead. Once this forecast for one season is made, this forecast is multiplied by the fractions in order to return a one season ahead forecast.

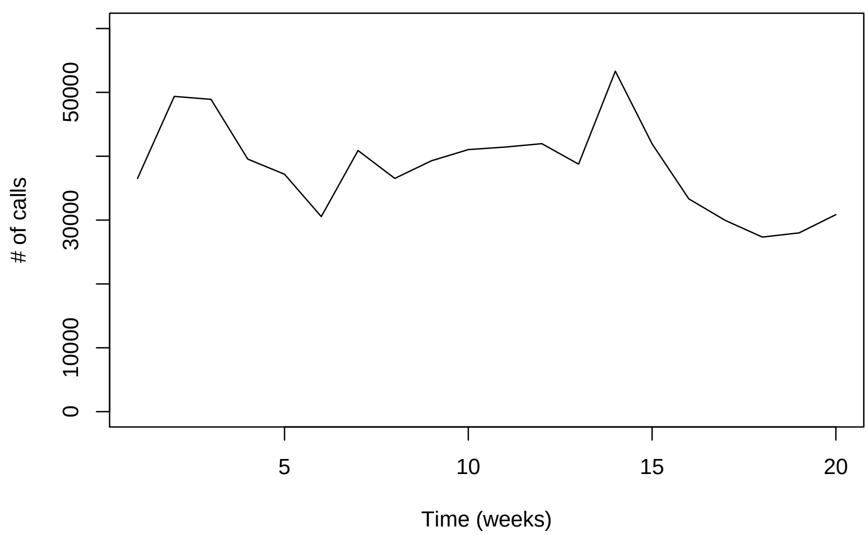
In order to retrieve the fractions in which to split up the forecasted season, some data of several seasons is required. The fractions can be calculated by summing up the fraction of the seasonality for multiple season and dividing by the sum of all data points. For clarification: if the seasonality is a whole week, split up in days, one should sum up all Mondays and divide by the total number of observations. In mathematical terms:

$$p_j = \frac{\sum_{h=1}^{k} Y_{h,j}}{\sum_{i=1}^{t} Y_i}$$

Where p_j is the fraction of part j of the seasonality, $Y_{h,j}$ is the observed



(a) Daily data for half a year



(b) Weekly data for half a year

Figure 4: Plots showcasing different scopes of data

number of calls on the part j in season $h = 1, \dots, k$ and Y_i is the observed number of calls on time $i = 1, ..., t$.

Next the forecast for one season is, as mentioned before, made by multiplying the fractions by the one season ahead forecast H_k. In mathematical terms for $j = 1, ..., s$:

$$F_{k,j} = p_j H_k$$

Where $F_{k,j}$ is the forecast for season k on part j of that season, p_j is the fraction and H_k is the one season ahead forecast for season k.

Note that this method changes the seasonality from weekly to yearly. With methods like Moving Average and Holt-Winters and the error measurements discussed later in this paper might give a skewed image on the performance of these methods. More on this can be found in the results for Moving Average and Holt-Winters respectively.

4.8 Linear Regression

The last and, according to Wout Bakker of CCMath, most commonly used method in forecasting in call centers nowadays is linear regression. Linear regression is usually done by "*Explanatory variables*" which will predict one variable. In mathematical terms for forecasting:

$$F_t = f_0 + f_1 X_{1,t} + ... + f_k X_{k,t} + e_t$$

Where F_t is the forecast for time t; f_0 is the intercept,$f_1, ..., f_k$ are estimates of coefficients, $X_{1,t}, ..., X_{k,t}$ the t-th observation of each explanatory variable and e_t is the estimated error at time t.

An example of this is the *Consumer Report* ratings of 77 cereals in the "*Healthy Breakfast*" dataset (found here) in R [10]. This dataset contains the number of grams of fat and sugar per serving. It can be possible to calculate the ratings of each cereal via the number of grams of fat and sugar in the cereal. In this case the regression would result in $Rating = 61.1 - 3.07 * Fat - 2.21 * Sugars$. Where -3.07 is the coefficient per gram of fat and -2.21 is the coefficient per gram of sugar. These coefficients are found by minimizing the sum of squared errors: $S = \sum_{i=1}^{n} e_i^2$. More information on this topic can be found in the book of Makridakis, Wheelwright and Hyndman[7].

However, time series data, like most call centers possess, do not have clear explanatory variables. Rather, the data is just the number of calls at a certain time. Hence, the introduction of so-called "dummy variables" is necessary. Dummy variables are either 0 or 1 depending on the context. This means that for daily data, the following 6 dummy variables can be created:

- $D_1 = 1$ if the day is Monday, zero otherwise
- $D_2 = 1$ if the day is Tuesday, zero otherwise

$\vdots$

- $D_6 = 1$ if the day is Saturday, zero otherwise

Note that the Sunday is absent. Linear regression requires 1 less dummy variable than possible to avoid multicollinearity. Multicollinearity is the phenomenon where one explanatory variable can be predicted from the others, hence some statistics about the individual variables might be skewed. For example, if the Sunday was present in the aforementioned lists, it could be predicted via the other days of the week as well-i.e., if all days except Sunday are zero, the Sunday has to be one. Therefore, Sunday is omitted.

For the daily data (5) without Sundays, 5 dummy variables are created for each day of the week. Note that the sixth day is absent, due to the aforementioned multicollinearity. Furthermore, at the end of the year a spike in calls can be seen as well, due to holidays and people contacting call centers. This suggest that dummy variables for the week of each year is helpful as well. With on average 52 weeks and 1 day per year, there are a total of 53 weeks per year, resulting in 52 dummy variables.

A linear regression should be able to handle horizontal, seasonal, cyclical and trend patterns. However, how well linear regression is able to handle these patterns depends on the skill of the forecaster. Furthermore, statistical knowledge is required in order to use this method correctly.

5 The Data

In order to forecast, data needs to be collected. The forecaster also needs to have some understanding about the data as numerous problems can arise when a forecaster blindly forecasts. A call center has systems in place in order to collect data: from when the call arrived to who called [5]. A plot of the whole data can be found in Figure 6, while a small excerpt of the data can be found in Figure 5 starting on the first of January 2009. This data was provided by an Indian call center.

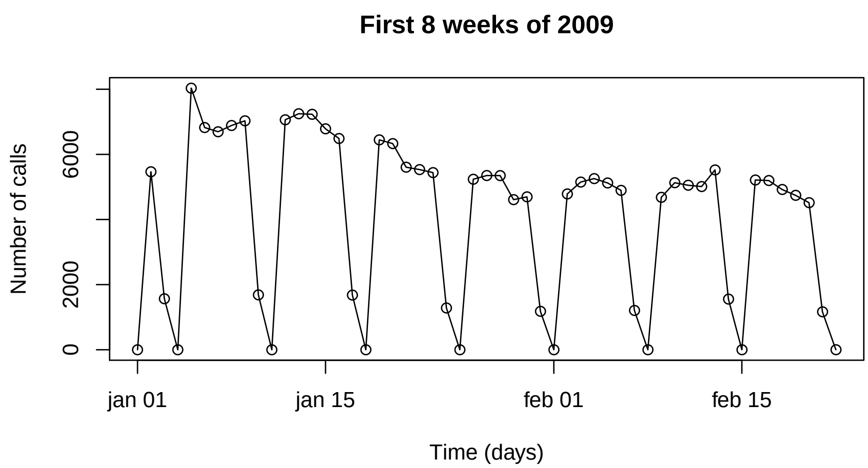


Figure 5: Plot of the first few weeks of 2009

As can be seen in Figure 5 a weekly pattern occurs, hence a seasonality of a week ($s = 7$) is present, while when looking at Figure 6 a yearly trend can be seen as well, where close to New Year and Christmas the calls spike. Next, there is no clear pattern seen on weekdays. Furthermore, as seen in Figure 8, the Saturdays show a big dip and there are no calls on Sunday. In order to avoid problems with the 0 calls (as can happen with Holt-Winters), it was decided to remove all Sundays from the data, resulting in Figure 7 and thus reducing the seasonality ($s = 6$). With the sum of all calls on a certain day it is possible to calculate the fractions, with the results found in Table 1.

The plot for the summed days into weeks can be found in Figure 9. The blue vertical lines are the week where a new year starts. This figure shows the intra-year period as well. Close to the new year, the calls spike. On the other hand, the calls seem to dip in the middle of the year.

Last, the plot without Sundays still shows days where no calls are coming in. This is due to the fact that these days fall together with special holidays. Hence, it can be assumed the call center was closed. However, unlike the Sundays it is inadvisable to remove these holidays from the data. First, as removing these

	Fractions
Monday	0.1803
Tuesday	0.1970
Wednesday	0.1983
Thursday	0.1877
Friday	0.1883
Saturday	0.0483

Table 1: Fractions for each day of the week

holidays can make the seasonality per week very inconsistent as there might be 4 or 5 days compared to 6 days in a normal week. In addition, some holidays fall on different days per year and thus can be hard to take into account. It can be expected from the data, that the methods mainly relying on trend and predictable data will have a hard time to successfully forecast.

A forecast can almost never be 100% accurate. Hence, in order to see how accurate each method is, a measure of error needs to be established. The error in forecasting is the difference between the forecasted value and the actual observed value (e.g. someone forecasts 1000 calls, but the actual was 990, there is an error of 10 calls). One of the most used measures of accuracy is Mean Squared Error (MSE), which can be calculated by squaring all errors of each observation and dividing by the number of observations. However, the result of MSE can be very hard to interpret (e.g. an MSE of 1187212.5 is very ambiguous). Furthermore, a big error will be amplified due to the squaring, hence the MSE is prone to big outliers. According to Koole [6] a better measure of error would be the Weighted Absolute Percentage Error (WAPE). The WAPE puts higher weights on days with higher call volumes and gives clearer measure of error as it returns a percentage. The WAPE can be calculated as follows:

$$WAPE = \frac{\sum_{i=1}^{n} |Y_i - F_i|}{\sum_{i=1}^{n} Y_i}$$

The WAPE is the measure of error to optimize the aforementioned α, β and γ for SES, Holt's and Holt-Winters' method. The goal in determining whether a method works lies in the WAPE. The lower the WAPE, the lower the error and thus the better the method is. However, other factors, like data requirements and ease of use will play a role as well.

6 Results

Table 2 shows the WAPE of every method for both the traditional way and the weekly fractions as well as Linear Regression. In order to make sure the WAPE is calculated fairly, days with 0 calls are not taken into the calculations. Furthermore, the table also shows which figures in the appendix correspond to which method. The figures are taken on the time period ranging from the 19th of August to the 14th of November 2008 with no Sundays included. Note that the Moving Average and Holt-Winters' method are marked with an $*$, this is due to the data required for this method. Hence the graphs span a time period of about one year later. More on this is explained per method.

	Traditional		Weekly fractions	
	WAPE	**Figure**	**WAPE**	**Figure**
Mean	0.2637	10	0.1512	16
Moving Average	0.2689	11	0.0776*	17
Single Exponential Smoothing	0.2857	12	0.1096	18
Adaptive Single Exponential Smoothing	0.2690	13	0.1099	19
Holt's Linear Method	0.3020	14	0.1119	20
Holt-Winters' Method	0.0951	15	0.0651*	21
Linear Regression	0.1208		22	

Table 2: Results of forecasting

It can be seen from the table that the weekly fractions vastly decrease the WAPE for every method. The table also shows that the fractional Holt-Winters' method has the lowest WAPE, with fractional MA a close second and traditional Holt-Winters' a close third. Even more, fractional SES is a very close fourth with just one percentage point higher.

One common remark that all methods have in common is the fact that special days or holidays are still not accounted for, as can be seen in the graphs below. Most graphs have a steep rise in the beginning of the graph due to 2 consecutive days without calls, resulting in the methods trying to correct themselves. Even more, the last part of the graphs shows 2 days without calls, namely the first and eleventh of November. This results in very erratic behaviour from the methods.

6.1 Mean

The result of the high WAPE of the traditional method can be clearly seen from Figure 10. As mentioned in the method section 4, the data does not change much when new observations are added. Furthermore, the plot shows the Flaw of Averages very well: The deep valleys are not accounted for leading to continuous overstaffing and thus high costs.

The weekly fractions immediately shows the effectiveness of incorporating the seasonality in the calculations. However, it still tends to overestimate the actuals. This is due the fact that when looking at the entire plot, there is a small downward trend until the end of 2008. This downward trend leads to the average being slightly higher than what the actuals would be.

6.2 Moving Average

Like the Mean, the Moving Average overestimates the more quiet days, albeit less than the Mean. However, the normal weekdays tend to be more underestimated than the normal Mean, due to the fact that only the last $k = 6$ observations are taken into account.

Since that data changed from daily to weekly the seasonality changed as well: from $k = 6$ to $k = 52$. This means that at least 52 weekly observations are required for this method, which might make it infeasible. Furthermore, this will skew the WAPE seeing as less observations are taken into the calculation. Hence, caution should be taken when comparing this method. Nonetheless, this method performs relatively well, due to the fact there are less higher peaks and lower dips, which benefits MA.

6.3 Single Exponential Smoothing

Once again, the forecasts are dragged down by the low volumes on Saturday. Furthermore, the clear dips after every low day are due to the weight assigned to the most recent observation. The best weight, the one with the lowest WAPE, is 0.05. This means that the data will be quite rigid, as not much adjustment is done. This can be seen clearly seen for the day after a Saturday with low calls. While the dip in calls is quite significant, the forecasting does not dip much. The reason for such a low alpha is due to the nature of the error measurement WAPE. A higher α would make SES more erratic, as it just takes the previous day as the forecast. This would mean that the forecast from Saturday to Monday, would lead to a really big WAPE, leading to a more smoothed SES.

The best α for weekly fractions SES is 0.7, resulting in a much more adapting forecast than the traditional method. The likely reason for the higher α is the fact the weekly data is less erratic. Hence, the difference between the weeks is less. This means that a higher α will not affect the WAPE as much as with the traditional method.

6.4 Adaptive Single Exponential Smoothing

From the plots can be seen that ASES does a better job than SES when it comes to following weekdays. However, the fact that Saturdays are overestimated and the small dips on the Mondays still show that ASES lags behind the actuals. With a $\beta = 0.05$, small changes in α_t will occur. These small changes are very consistent with SES, seeing as large changes will increase the WAPE in the same way as SES.

For ASES with fractions the best β is 0.45, resulting in moderate changes to α_t. The comments that applied to fractional SES also applies here. The fact that the data is less erratic helps to better the forecast.

6.5 Holt's Linear Method

Holt's Linear method clearly shows the inability to handle seasonality. Very notable in the Figure is the steady climb at the start of the figure. This is due to the aforementioned fact that a few days earlier 2 consecutive days with zero days were present resulting in a big dip. With $\alpha = 0.15$ and $\beta = 0.2$, the lowest WAPE is achieved for the traditional method.

For the weekly fractions the best α and β are 0.65 and 0.1 respectively, This shows that the level needs more adjustment in the fractional method than the traditional method, while the trend needs less adjustment. This is due to the fact that while the level changes over the time, the general trend stays the same. Surprisingly, the WAPE is higher than the WAPE for SES and ASES, albeit very small. This is likely due to the fact that grid search and not non-linear optimization was used to find the best parameters.

6.6 Holt-Winters' Method

Holt-Winters' immediately shows its strength for the traditional method. The fact that the seasonality component is added, drastically lowers the WAPE. Figure 15, shows that the low call volumes of the Saturdays are taken into account. The optimal WAPE is achieved with $\alpha = 0.45$, $\beta = 0.05$ and $\gamma = 0.25$

The fractional method of Holt-Winters' results in the lowest WAPE of every method. However, caution should be taken when looking at this method. This is due to the fact that, just like the Moving Average, one whole seasonality of 52 weeks is not used in the calculations of the WAPE. Furthermore, at least 2 seasons are required to use the weekly fractions method, meaning that at least 2 years of data is required. However, the data provided covers a little more than two and a half years. With the first year already not forecasted on due to initialization, leads to little room for error measurement. Furthermore, data of the second year is used for the initialization and is forecasted on. Hence the error measurement might be skewed. This means that the calculated WAPE should be looked at cautiously due to the two aforementioned points. Nonetheless, the best parameters for this method are: $\alpha = 0.45$, $\beta = 0.25$ and $\gamma = 0.75$.

6.7 Linear Regression

Last, linear regression has a low WAPE compared to its simplicity. All 57 coefficients and the intercept can be found in table 3. Every coefficient shows the effect of that certain day or that certain week in the month. The intercept shows the "normal" level of calls without any day or week effects. The table shows that weeks 2 and 3 and weeks 47 to 52 increase the call volume, likely due to holidays. Some assumptions need to be checked, before this model is valid. More on these validations can be found in the book of Makridakis, Wheelwright and Hyndman[7] . First, the variables need to be significant- i.e., the coefficient is not 0 or the coefficient is significantly different than 0. However, since this model deals with dummy variables rather than traditional explanatory variables, the assumption is made that once a single dummy variable is significant, all are significant. In this case an example of a significant variable for the Weeknumber would be week 33, with a p-value of 0.000245. In case of the weekday, Saturday would be a statistical significant day with a p-value of less than $2e - 16$. Hence, all variables can be taken into account. Besides the significance of variables, the assumption of uncorrelated variables and uncorrelated, normally distributed errors. Since all variables are dummy variables and the presence of multicollinearity has been taken care of by taking one less number of dummy variables than possible the uncorrelated variables are taken care of. In order to show the normality of the errors a Q-Q plot for the residuals versus the Normal Distribution can be used and can be found in Figure 23. The plot shows skewness on the left and rightmost part of the plot. After further investigation these residuals seem to stem from days with seemingly inexplicable low or high calls (e.g. a Monday with just 200 calls). Nonetheless, the Q-Q plot looks good enough to assume a normal distribution.

7 Conclusion

This research aims to research the use of simple and more complex forecasting techniques as well as the impact of weekly fractions and the use of linear regression.

The more traditional standard forecasting methods rely on non-erratic data with only a trend and not a seasonality. This is why the traditional way of forecasting has a high error measure and are therefore not advised to use.

From the results it can be seen that Holt-Winters works really well on call center data. The main benefit of this method is the implementation of seasonality, which other methods lack. This explains the effectiveness of the weekly fractions. The seasonality is incorporated in the calculations and the data can adapt. An additional benefit of weekly fractions is the longer term forecasting for a week rather than a day, while still preserving a low error. Furthermore, the general methods stay simple, but with a slight modification for the weekly to daily forecasts. This longer term forecasting gives opportunity for longer term scheduling.

While linear regression falls on a different category than both a traditional method and weekly fractions it is still very useful. The linear method used in this paper is very basic. This is due to the fact that the regression consists only of dummy variables to check which day of the week and which week of the year the forecast is. This means that the linear regression can be expanded to include special days like holidays. More on this can be found in Section 8.

8 Discussion

This paper focuses on the relatively short-term forecasting of incoming calls in a call center. This is done to make the scope of this paper not too broad. This also means that research can be done on middle to long term forecasting. While the previous methods might not all be suitable for this type of forecasting, linear regression can definitely be further tweaked to suit these needs. Next, the linear model used in this paper is still simple. Other factors can be used to better predict the incoming calls. Even more, the effect of events such as: Commercials, holidays, world events and news are not taken considered in this research. This research is a subject in and of itself. Researches can try to transform the data as well, for example by using log-transformations, which might benefit the forecasting methods.

Furthermore, other measures of error can be used and might yield different results. Another way to find the optimal parameters for α, β and γ like non-linear optimization might yield similar but somewhat different results as well.

Future studies could also combine the forecasting of calls with the other aspects of WFM like the number of agents to use and the scheduling of these agents.

Last, not all methods of forecasting are discussed in this paper, once again in the interest of the scope of the paper. Two of such methods are ARIMA-models and Neural Networks. Autoregressive Integrated Moving Average (ARIMA) models are more sophisticated models and hence more complex. ARIMA-models require a lot of attention from the forecaster to use the correct parameters and estimations and results in iterations of ARIMA-models.

Practice shows that Linear Regression is still favored according to CCMath, a company that excels in WFM. Further research should focus on the use of Linear Regression in forecasting as it seems the most promising. Especially the handling of special holidays where the call center is either closed or takes less calls tends to throw of most algorithms.

Other sources like the M-competitions, organized by Spyros Makridakis, have evaluated and compared different forecasting methods as well [11]. The most recent competition, called the M4-competition, tried to replicate the results of the previous three with an increasing number of time series and included the use of Machine Learning. Namely, the use of Neural Networks. However, it was discovered that the pure use of Machine Learning methods were inferior in accuracy and higher in computational requirements [12]. This reinforces the idea that Machine Learning methods should not be used in a vacuum, but are rather suited in conjunction with statistical methods.

References

[1] Exastax. How airlines are using Big Data; 2017. Available from: https://www.exastax.com/big-data/how-airlines-are-using-big-data/.

[2] Helper CC. A Guide to Workforce Forecasting in the Contact Centre; 2018. Available from: https://www.callcentrehelper.com/workforce-forecasting-57254.htm.

[3] Pulford BD. Overconfidence in human judgement. University of Leicester; 1996.

[4] Davenport TH, Harris JG. Competing on Analytics. Harvard Business Review Press; 2007.

[5] Gans N, Koole G, Mandelbaum A. Telephone Call Centers: Tutorial, Review, and Research Prospects. Manufacturing & Service Operations Management. 2003;5(2):79–141.

[6] Koole G. Call Center Optimization. Amsterdam: MG Books; 2013.

[7] Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting: Methods and applications. Wiley; 1998.

[8] Trigg DW, Leach AG. Exponential Smoothing with an Adaptive Response Rate. Journal of the Operational Research Society. 1967 Mar;18(1):53–59. Available from: https://doi.org/10.1057/jors.1967.5.

[9] Gardner E, G Dannenbring D. Forecasting with exponential smoothing: Some guidelines for model selection. Decision Sciences. 2007 06;11:370 – 383.

[10] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2008. ISBN 3-900051-07-0. Available from: http://www.R-project.org.

[11] of Nicosia U. M4-Competition; 2019. Available from: https://www.mcompetitions.unic.ac.cy/.

[12] Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLOS ONE. 2018 03;13(3):1–26. Available from: https://doi.org/10.1371/journal.pone.0194889.

9 Appendix

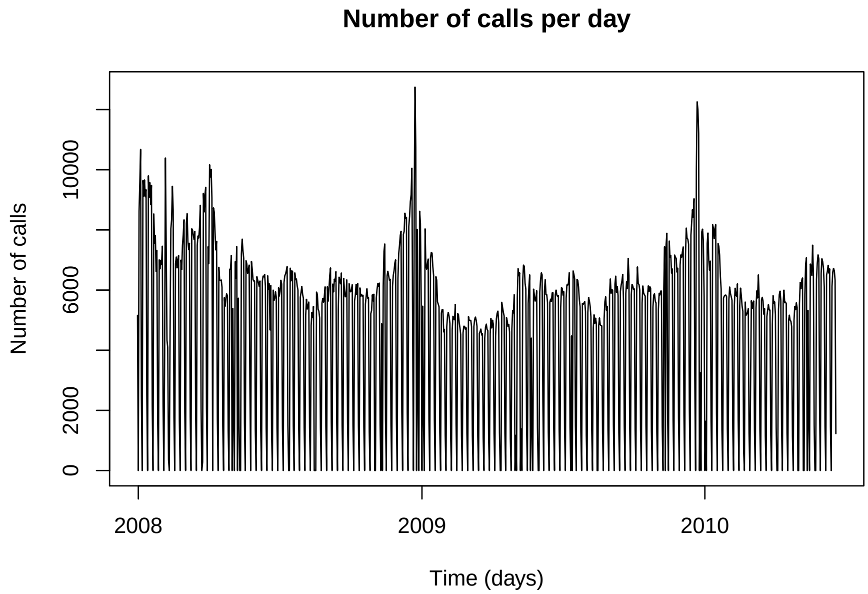


Figure 6: Plot of the whole data set

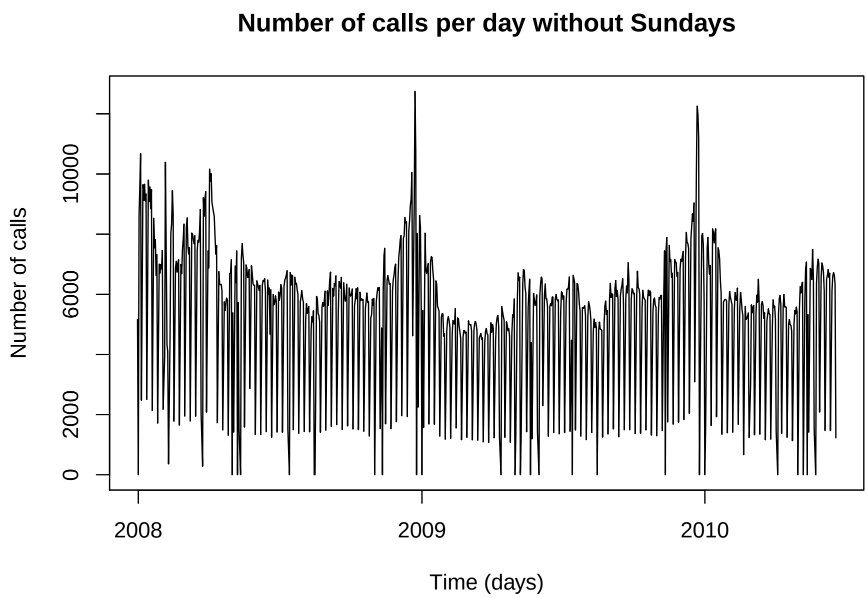


Figure 7: Plot of the whole data set with Sundays removed

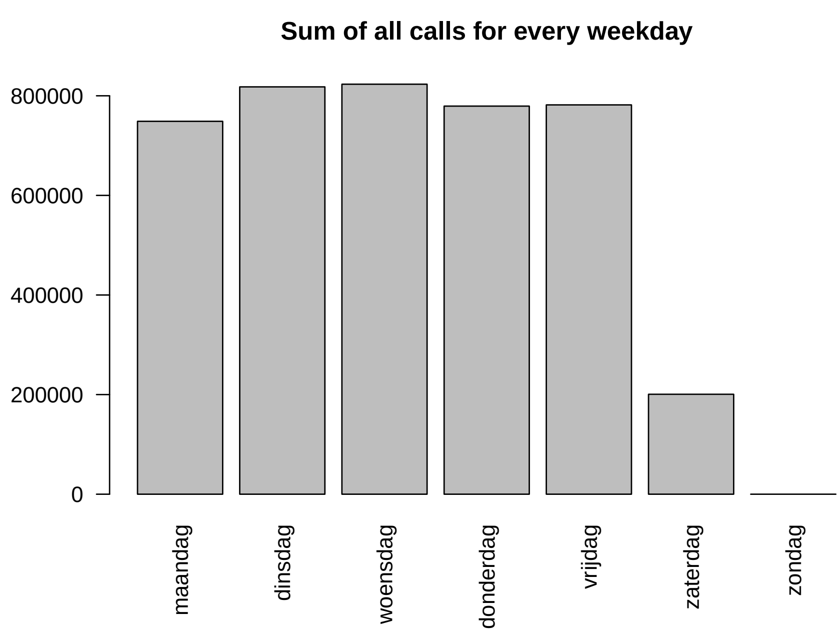


Figure 8: Sum of all calls for every weekday

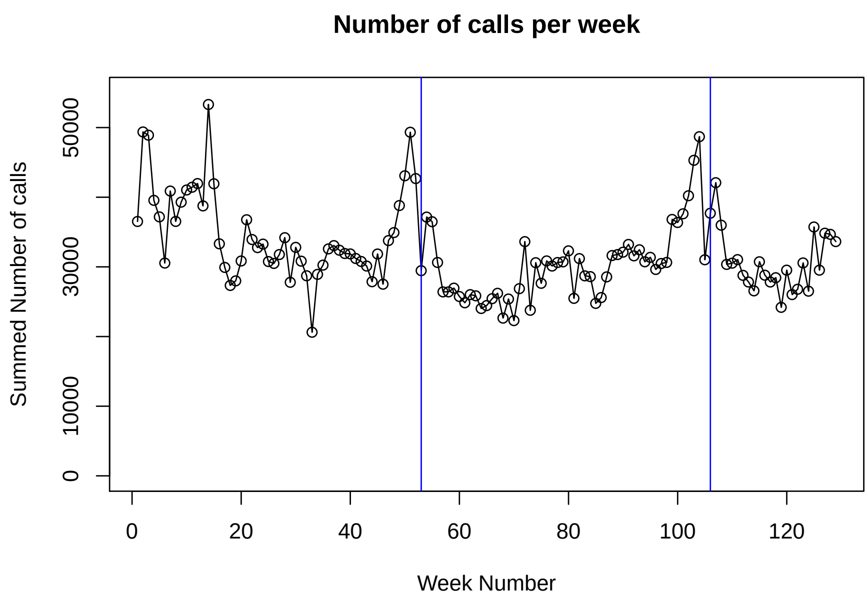


Figure 9: Summed up calls. The blue lines indicate the week in which the new year starts

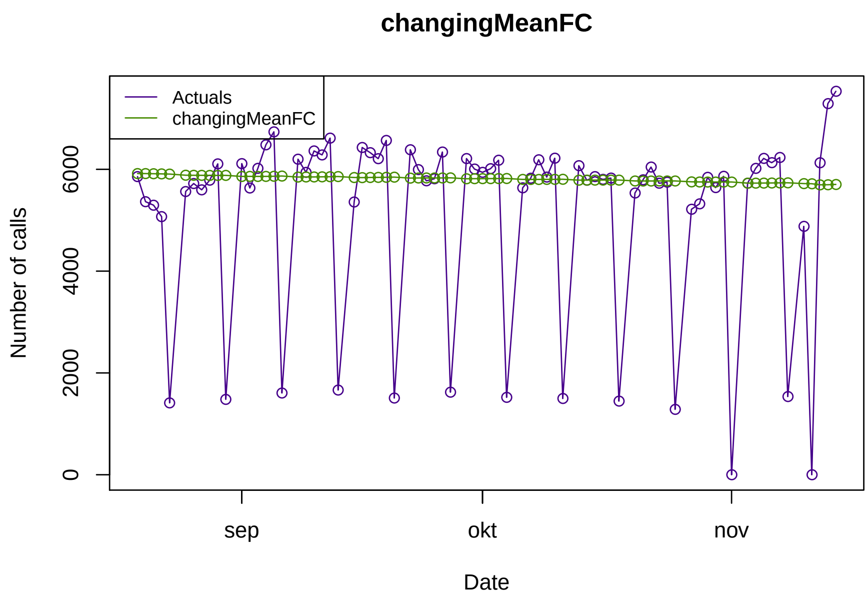


Figure 10: Traditional forecast for the Mean

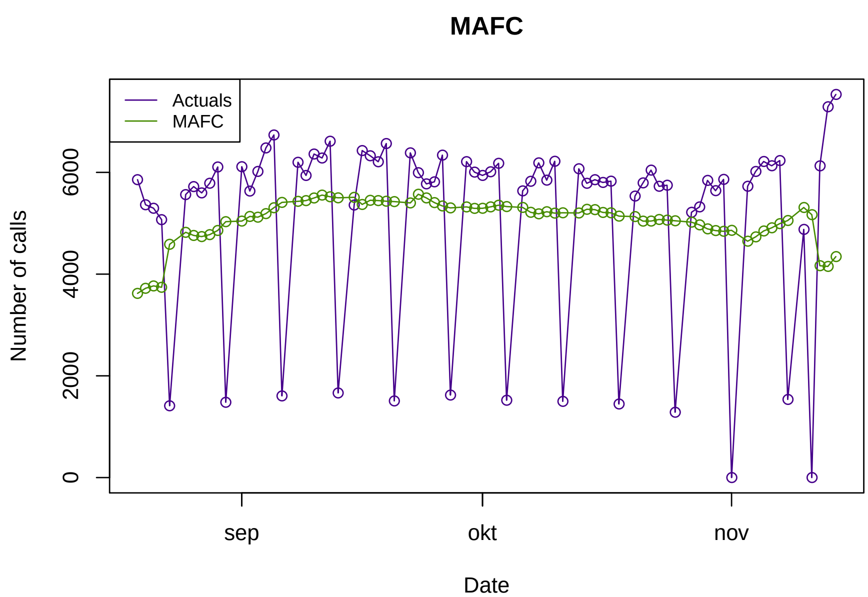


Figure 11: Traditional forecast for the Moving Average

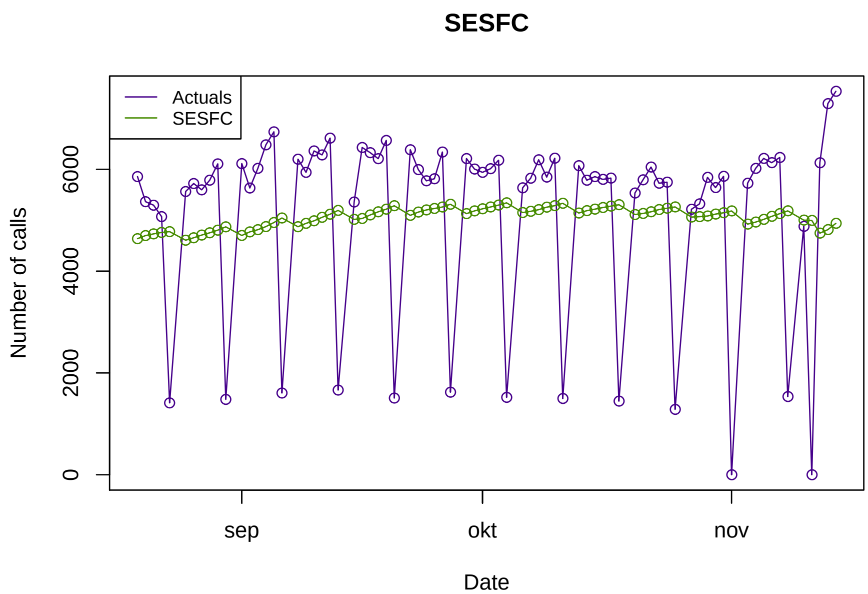


Figure 12: Traditional forecast for Single Exponential Smoothing

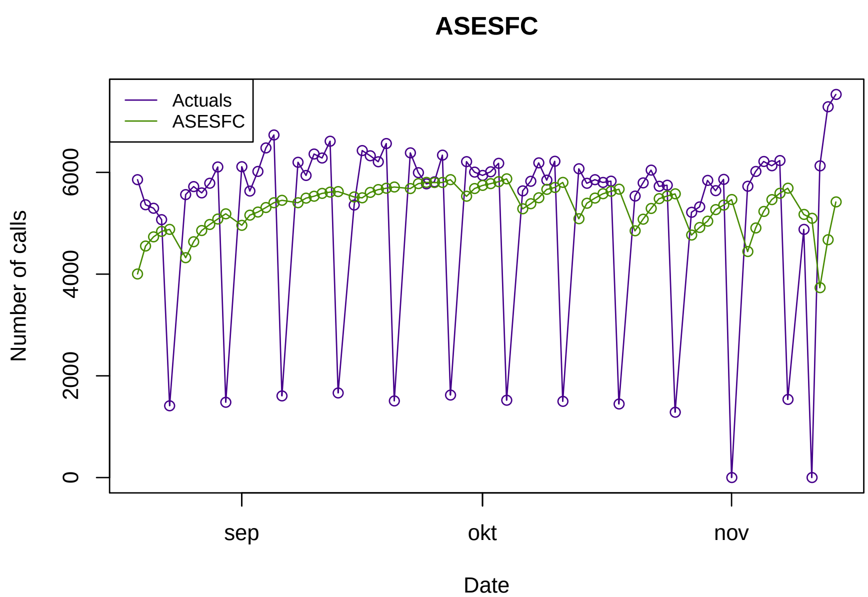


Figure 13: Traditional forecast for Adaptive Single Exponential Smoothing

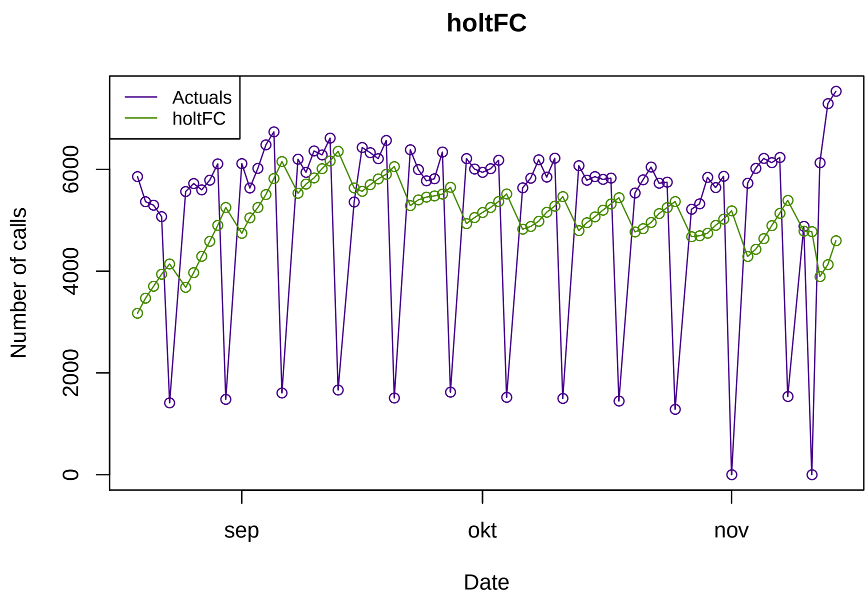


Figure 14: Traditional forecast for Holt's Linear Method

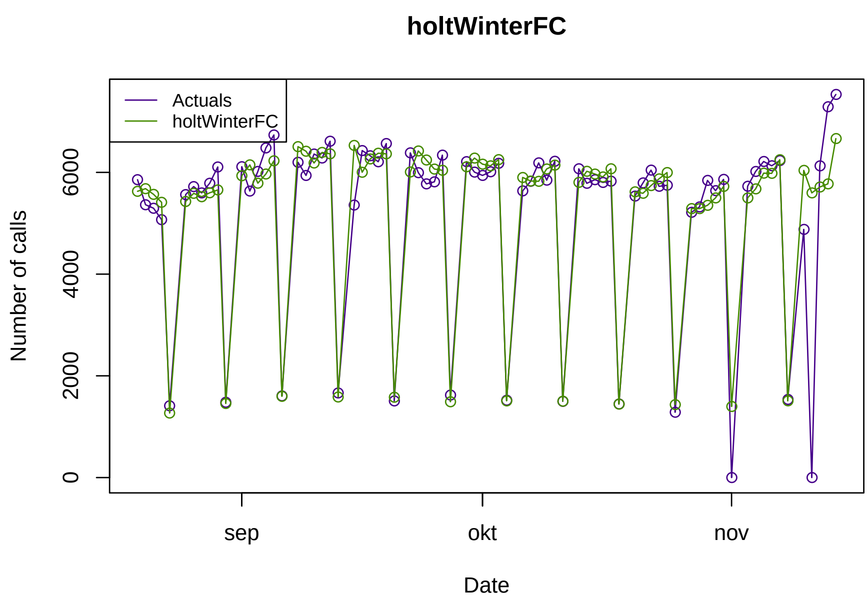


Figure 15: Traditional forecast for Holt-Winters' Method

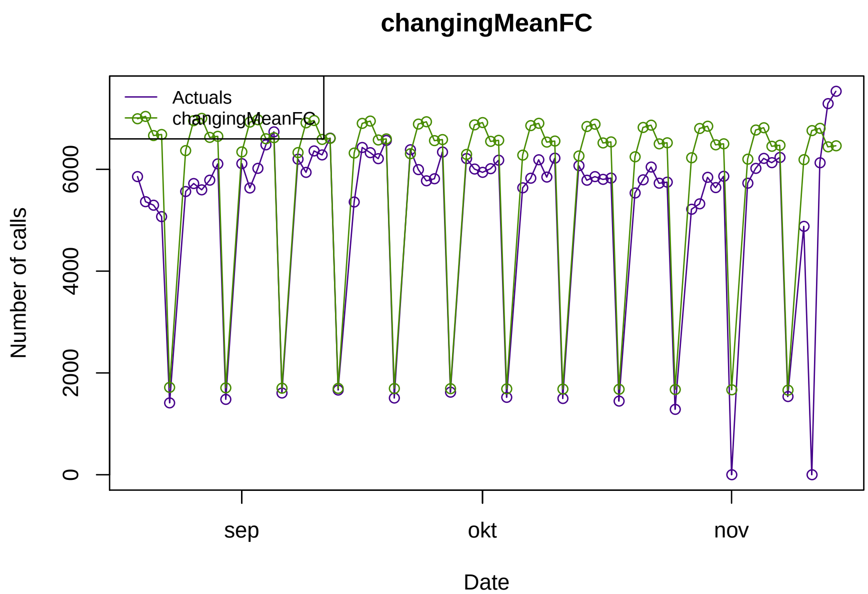


Figure 16: Fraction forecast for the Mean

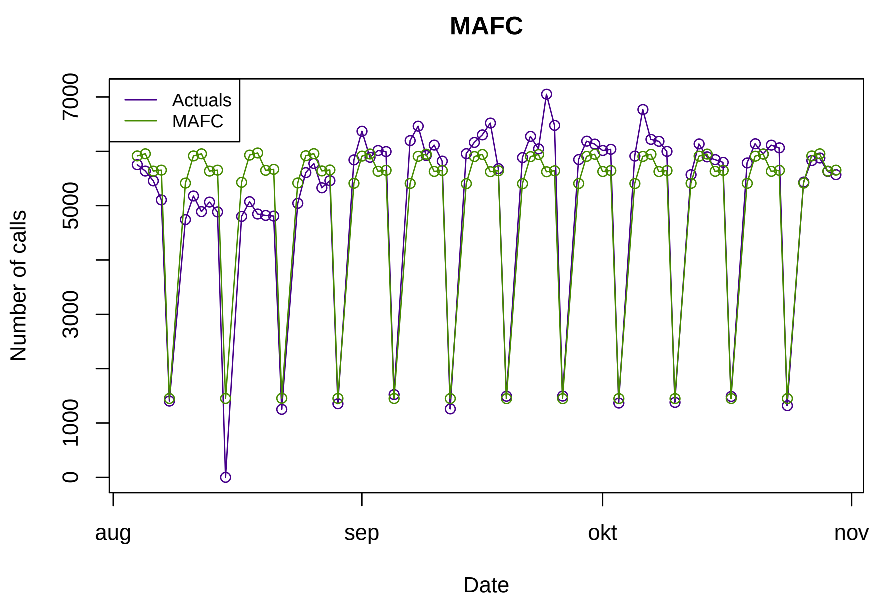


Figure 17: Fraction forecast for the Moving Average

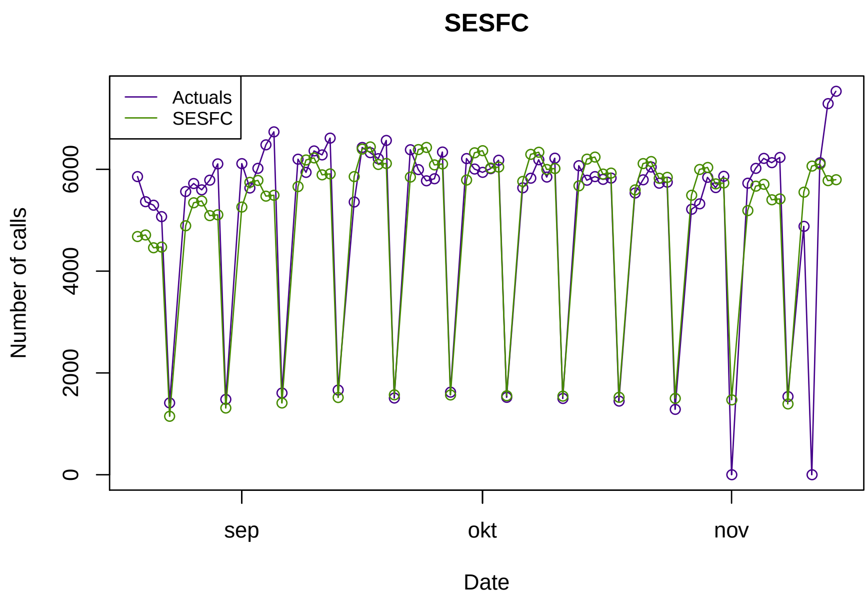


Figure 18: Fraction forecast for Single Exponential Smoothing

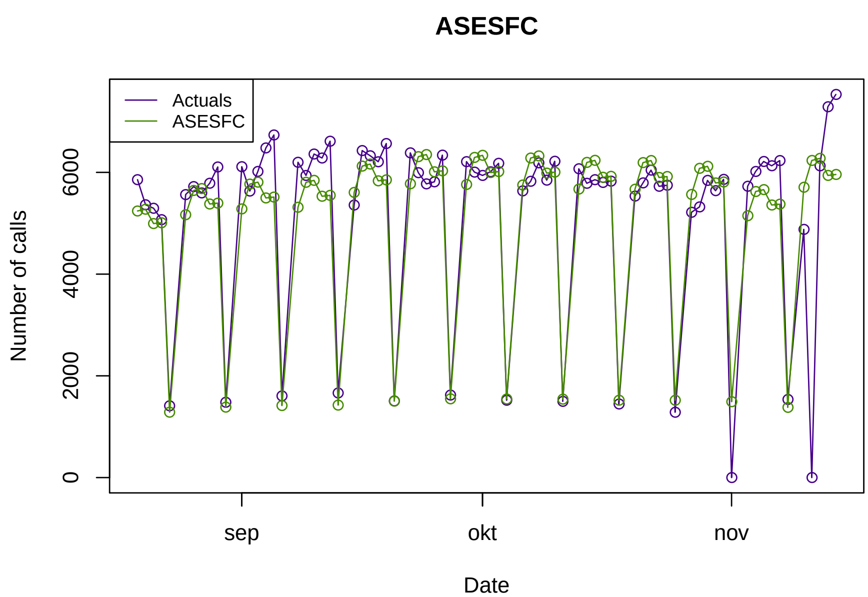


Figure 19: Fraction forecast for Adaptive Single Exponential Smoothing

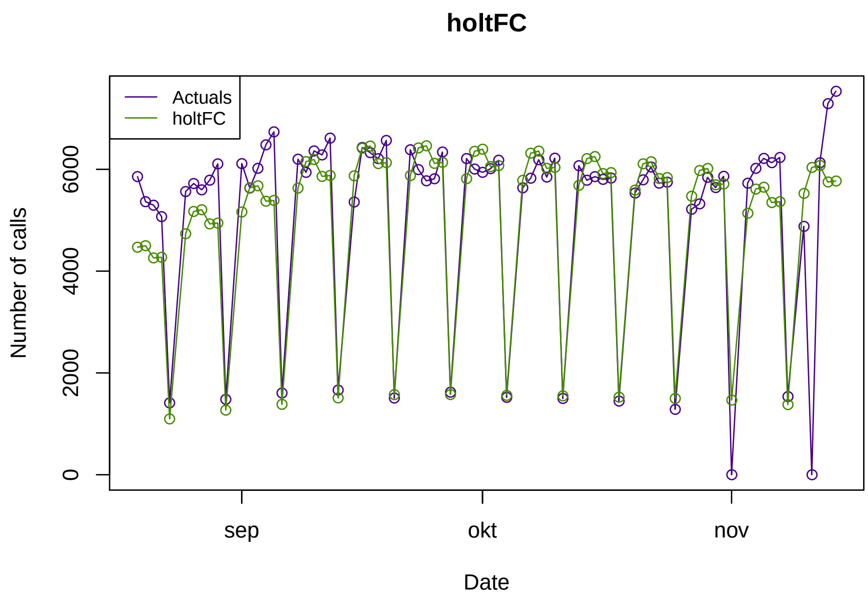


Figure 20: Fraction forecast for Holt's Method

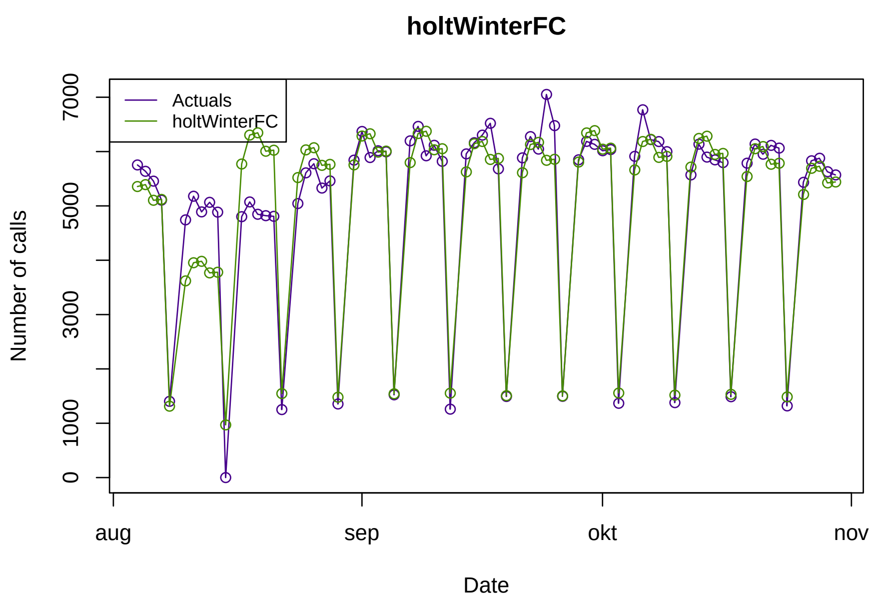


Figure 21: Fraction forecast for Holt-Winters Method

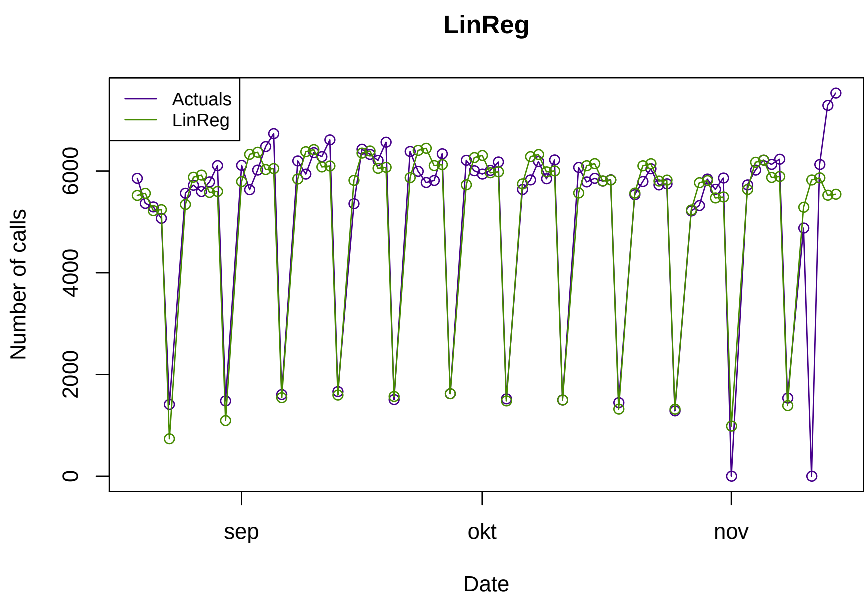


Figure 22: Forecast using Linear Regression

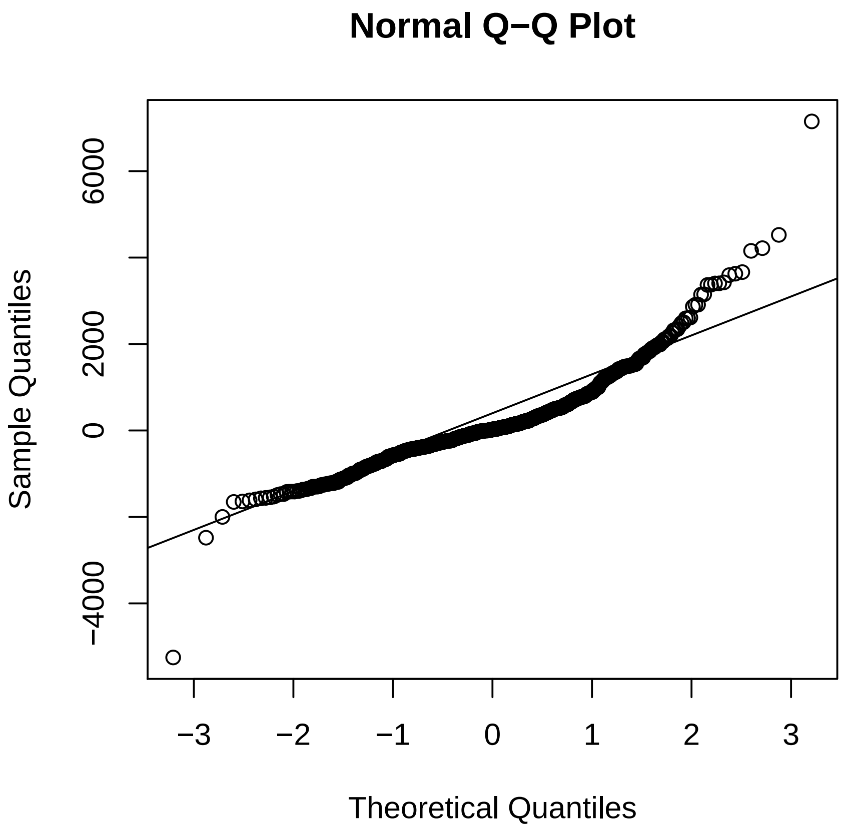


Figure 23: Q-Q plot of the Normal Distribution vs the residuals of linear Regression

Variables	Coefficients	Variables	Coefficients
Weeknumber 2	1384.89	**Weeknumber 31**	-798.75
Weeknumber 3	983.22	**Weeknumber 32**	-982.58
Weeknumber 4	-174.67	**Weeknumber 33**	-1978.75
Weeknumber 5	-530.83	**Weeknumber 34**	-1216.33
Weeknumber 6	-871.06	**Weeknumber 35**	-858.50
Weeknumber 7	-392.44	**Weeknumber 36**	-408.83
Weeknumber 8	-754.50	**Weeknumber 37**	-357.67
Weeknumber 9	-721.22	**Weeknumber 38**	-384.67
Weeknumber 10	-327.11	**Weeknumber 39**	-331.17
Weeknumber 11	-421.78	**Weeknumber 40**	-472.92
Weeknumber 12	-547.83	**Weeknumber 41**	-472.92
Weeknumber 13	-668.33	**Weeknumber 42**	-633.75
Weeknumber 14	-40.89	**Weeknumber 43**	-635.58
Weeknumber 15	-333.72	**Weeknumber 44**	-967.83
Weeknumber 16	-1205.67	**Weeknumber 45**	-565.25
Weeknumber 17	-1198.89	**Weeknumber 46**	-913.58
Weeknumber 18	-1304.94	**Weeknumber 47**	122.58
Weeknumber 19	-1238.67	**Weeknumber 48**	181.75
Weeknumber 20	-193.11	**Weeknumber 49**	609.25
Weeknumber 21	-756.44	**Weeknumber 50**	1181.33
Weeknumber 22	-240.06	**Weeknumber 51**	2126.17
Weeknumber 23	-477.67	**Weeknumber 52**	1854.50
Weeknumber 24	-328.94	**Weeknumber 53**	-589.33
Weeknumber 25	-685.92	**Monday**	-535.84
Weeknumber 26	-666.50	**Wednesday**	41.68
Weeknumber 27	-552.42	**Thursday**	-299.38
Weeknumber 28	-218.58	**Friday**	-280.26
Weeknumber 29	-1321.00	**Saturday**	-4784.52
Weeknumber 30	-427.08	**Intercept**	6737.05

Table 3: Variables and their coefficients for linear regression