

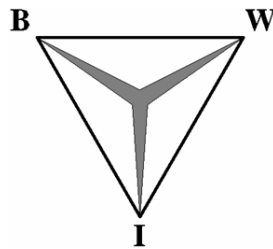
Predicting Arrivals in Call Centers

Koen van den Bergh



Predicting Call Arrivals in Call Centers

Koen van den Bergh



BMI-Paper

Supervisor: drs. A. Pot

Vrije Universiteit
Faculty of Sciences
discipline Business Mathematics & Informatics
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands

August 2006

Executive Summary

Even though there is an enormous amount written about forecasting, the number of articles about call center forecasting is not very impressive. The articles capture a reasonable amount of mathematical forecasting models, but the many undiscovered models can be applied to enhance the accuracy of forecasting. Great-scale research will be needed to broaden the knowledge about this subject.

In this paper a complete literature overview of different forecasting methods is provided, with a particular focus on the forecasting methods that are expounded and utilized for predicting daily call frequencies in call centers.

After the explanation of these methods, two of them are tested using a real call center database. The results and all what enters into the forecast will be discussed in this paper.

In conclusion a comparison will be given of the differences and the advantages / disadvantages of the different methods.

Preface

One of the graduate courses of my study Business Mathematics & Informatics is the course BMI-paper. In this course one can chose a subject to gain more in-depth knowledge. As the name of my study indicates, it captures three components: economics, mathematics, and Informatics. In my paper all three components are included.

I have tried to setup a clear and complete overview regarding the existing mid-term call center call frequency forecasting techniques. If there are any notes or suggestions regarding this paper, I sincerely like to hear from you.

I enjoyed working on this project. Although the subject discussed in this paper has not been treated during my study, by working on this paper I did get the opportunity to go deeply into this matter.

Besides the examination of the possible forecast methods, some of the forecast methods were actually done. Most fascinating was to see that the discussed models transformed a database into good forecasts.

It is clear that the prediction of call center arrivals is still in its infancy and that there is a lot more to discover. Hereby I would like to stimulate other people that are interested in this matter to further investigate the forecasting of call center arrivals, so that more knowledge can be gained.

To conclude I would like to thank my supervisor Auke Pot for the effort and time spent. I enjoyed our cooperation and he helped me in many aspects.

I sincerely hope that you will enjoy reading this paper as much as I enjoyed writing it.

With kind regards,

Koen van den Bergh

Contents

1	Introduction	3
2	Forecasting Models	5
2.1	ARIMA Model	5
2.2	Dynamic Regression Models	7
2.3	Exponential Smoothing	12
2.3.1	Single Exponential Smoothing	12
2.3.2	Holt's Linear Model	13
2.3.3	Holt-Winters' Trend and Seasonality Model	14
2.4	Regression Analysis	15
3	Applying Techniques	19
3.1	Data	19
3.2	Software	21
3.3	Measurements of Forecasting Accuracy	23
3.4	Forecasting Results	23
3.4.1	ARIMA Model Predictions	25
3.4.2	Holt-Winters' Trend and Seasonality Model Predictions	27
3.4.3	Summary of Results	28
3.5	Suggestions for Improvement	29
4	Conclusions and Further Research	31
5	Appendix	35
5.1	ARIMA Functions R-code	35
5.2	Holt-Winters Functions R-code	36
5.3	Error R-code	37
5.4	Charts of the Predictions	39

Chapter 1

Introduction

The main purpose of this paper is to provide a complete literature overview of different forecasting methods, with a particular focus on the forecasting methods that are expounded and utilized for predicting call frequencies in call centers. Before proceeding, a short expound will be given of a call center.

A call center is a centralized office within a company that both answers incoming and makes outgoing telephone calls to clients (telemarketing). An important task of call centers is staffing. If too many staff members are scheduled at the same time, this will cause low efficiency and high costs. On the other hand, the benefit is that all incoming calls can be dealt with, and thus there will be fewer discontented clients. The result of not scheduling enough call center workers will be low costs and highly efficient telephone clerks. A negative outcome, however, may be that the number of incoming calls exceed the number of available clerks. Many calls will remain unanswered, which shall cause dissatisfaction and eventually even loss of clients. The result of such a scenario will be a loss of sales and thus a loss of income.

The main issue is thus to find a balance. This means scheduling enough staff members so that every call can be handled, except in very rare circumstances. This way, the operational costs will be kept as low as possible and the clients are satisfied simultaneously. In order to compute the right number of staff members, a database of daily call volumes must be available. With the help of this database and a mathematical forecasting method it is possible to forecast the daily expected calls. Self-evidently, the number of telephone clerks needed, can be computed by means of these daily expected calls.

As one might understand commercial firms have great interest in forecasting the right number of personnel, seeing the advantages of investing little money and still reaching high efficiency.

Even though there is an enormous amount written about forecasting, the number of articles about call center forecasting is not very impressive. In an assessment of future directions

for call center research, Gans, Koole, and Mandelbaum [8] even write that the practice of time series forecasting at call centers is "still in its infancy".

A distinction can be made between the articles written about forecasting call frequencies. There are three periods of time that are subject to the forecasting. One can forecast on the long term, the mid term, and on the short term. Longer term forecasts such as yearly and monthly predictions, are used for budgeting and staff planning, planning operational changes, training, and scheduling vacations. Mid term forecasts, such as weekly and daily forecasts, are needed for workforce staffing and scheduling. Short term forecasts, such as hourly, measure how well a call center is staffed for the current day. Traditionally, short term forecasts have been defined as making predictions for periods of less than three months. Medium forecasts span the three months to two year time frame and long term forecasts deal with any period of time longer than two years. However, in call center applications, these time frames are often difficult to predict and may not be very useful.

The focus in this paper is forecasting daily call volumes. This means that several methods used in previous papers on the topic of mid term forecasting will be expounded.

After the explanation of these methods, two of them will be tested using a real call center database. The results and all what enters into the forecast will be discussed in the chapter Applying Techniques.

In the last chapter the conclusions of this research will be given together with some directions for further research.

Chapter 2

Forecasting Models

Previous scholarly texts have shown that there are several methods to be used for forecasting time series. However, not all of the existing forecasting methods are applied to forecast call frequencies at call centers. In this section, an overview of the models used will be presented. For each method, the model will be given together with the accompanying literature in which the models were used. In none of the articles about forecasts for call centers the specific model applied was shown. Therefore, it is only possible to give a general model. Subsequently, the following models are discussed:

- ARIMA Model
- Dynamic Regression Models
- Exponential Smoothing
- Regression Analysis

2.1 ARIMA Model

ARIMA is the abbreviation for AutoRegressive Integrated Moving Average. The ARIMA model is a widely used forecasting model invented by Box and Jenkins [5]. The basis of the ARIMA model is the ARMA model, which consists of two sorts of terms: the autoregressive terms (AR) and the moving average (MA) terms.

Autoregressive (AR) terms are lagged values of the dependent variable, and serve as independent variables in the model. The General autoregressive model is given by

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad t = p + 1, p + 2, \dots, p + n. \quad (2.1)$$

Here ϕ_1, \dots, ϕ_p and α are unknown parameters. The process ε_t is white noise with the property that $E[\varepsilon_t y_{t-k}] = 0$ for all $k \geq 1$. So the regressors y_{t-k} are exogeneous with

$k = 1, \dots, p$. As the time series y_t is observed for $t = 1, \dots, n$, the p -lagged explanatory variable is available only from time $t = p+1$ onwards. This model is called an autoregressive model of order p , also written as $AR(p)$.

Moving Average (MA) terms are lagged values of the errors between past actual values and their predicted values who also serve as independent variables. A general Moving average process is given by the next formula

$$y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (2.2)$$

where ε_t is white noise. Here α and $\theta_1, \dots, \theta_q$ are unknown parameters. This process is always stationary, with mean $\mu = E[y_t] = \alpha$, variance $\gamma_0 = \sigma^2(1 + \sum_{j=1}^q \theta_j^2)$, and covariances $\gamma_k = \sigma^2(\theta_k + \sum_{j=k+1}^q \theta_j \theta_{j-k})$ for $k \leq q$ and $\gamma_k = 0$ for $k > q$. This model is also known as a moving average model of order q , which can be written as $MA(q)$.

Combining both models gives the *autoregressive moving average model*, also known as $ARMA(p, q)$

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (2.3)$$

To denote this formula in a more concise way, lag operators are used. Applying lag operator (denoted L) once, the index is moved back one time unit and if this is applied k times, the index is moved back k units.

$$\begin{aligned} Ly_t &= y_{t-1}, \\ L^2 y_t &= y_{t-2}, \\ &\vdots, \\ L^k y_t &= y_{t-k}, \end{aligned}$$

The lag operator is distributive over the addition operator, i.e.

$$L(\varepsilon_t + y_t) = \varepsilon_{t-1} + y_{t-1}.$$

By using lag operators it is possible to rewrite the ARMA models using lag operators, which is done in equation (2.4).

$$\begin{aligned} AR(p) : (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t &= \varepsilon_t, \\ MA(q) : y_t &= (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t. \end{aligned} \quad (2.4)$$

With the lag polynomials

$$\begin{aligned}\phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p, \\ \theta(L) &= 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q,\end{aligned}$$

it is possible to rewrite an ARMA process in a more compact way:

$$\begin{aligned}AR &: \phi(L)y_t = \varepsilon_t, \\ MA &: y_t = \theta(L)\varepsilon_t, \\ ARMA &: \phi(L)y_t = \theta(L)\varepsilon_t.\end{aligned}\tag{2.5}$$

The I in ARIMA indicates integrated and refers to the practice of differencing, which transforms a time series by subtracting past values from itself. The ARIMA(p,d,q) is given by

$$\phi(L)(1 - L)^d y_t = \alpha + \theta(L)\varepsilon_t,\tag{2.6}$$

where d is the indication of the amount of past values that is used to subtract itself. So far the description of the ARIMA model. For a comprehensive expound of the ARIMA model the reader is pointed to the book of Heij et al [9].

Forecasting practitioners have long realized the potential of the ARIMA methodology for modeling time series data with strong seasonal patterns. In fact, most articles on forecasting call volumes in call centers refer to the ARIMA model. Nijdam [14] illustrated the effectiveness of ARIMA models for predicting monthly telephone traffic in the presence of a persistent seasonal pattern. In the most recent article, W. Xu [16], a Forecasting Specialist in the Forecasting/Modeling Group in Worldwide Customer Service Strategic Planning & Analysis at FedEx explained that ARIMA is one of the forecasting models used in SAS, their major forecasting software.

2.2 Dynamic Regression Models

The ARIMA model discussed in the previous section deals with single time series and does not allow the inclusion of other information in the models and forecasts. However, frequently other information may be used to aid in forecasting time series. Information about holidays, strikes, changes in the law, or other external variables may be of use in assisting the development of more accurate forecasts.

An example of this type of useful information is the influence of advertising campaigns. Often the effect such of information does not show up in the forecast variable (y_t) immediately, but is divided across several time periods. For instance, the effect of an advertising campaign persists for some time after the end of the campaign. In this case, monthly sales figures (y_t) may be modeled as a function of the advertising costs in each of the previous

few months, that is $x_t, x_{t-1}, x_{t-2}, \dots$. So, the output series, y_t is affected by the input series, x_t . In fact, the input series x_t influences the output series y_t over several future time periods.

The general dynamic regression model is give by the next equation

$$y_t = a + \nu_0 x_t + \nu_1 x_{t-1} + \nu_2 x_{t-2} + \dots + \nu_k x_{t-k} + n_t. \quad (2.7)$$

Here y_t denotes the output time series and $x_t, x_{t-1}, \dots, x_{t-k}$ denotes the input time series. n_t is an ARIMA process as described in the previous section. The values ν_0, \dots, ν_k present the so-called impulse response weights (also transfer function weights). These coefficients are measures of how y_t is affected by x_{t-i} . The general dynamic regression model equation can be rewritten to make the formula shorter. This is done in the following way

$$\begin{aligned} y_t &= a + \nu_0 x_t + \nu_1 x_{t-1} + \nu_2 x_{t-2} + \dots + \nu_k x_{t-k} + n_t, \\ &= a + (\nu_0 + \nu_1 L + \nu_2 L^2 + \dots + \nu_k L^k) x_t + n_t, \\ &= a + \nu(L) x_t + n_t. \end{aligned} \quad (2.8)$$

In this shorthand notation $\nu(L)$ is called the transfer function since it describes how a change in the explanatory variable x_t is transferred to y_t . As can be seen, the lag operator, which was explained in the previous subsection, is also being applied here. Owing to this, the Equation (2.8) can be written in a more compact way.

The order of the transfer function is k , which is the longest lag in x that is used. This order can sometimes be very large. For this reason, the dynamic regression model will be written in a more parsimonious form. This rewritten form is as follows

$$y_t = a + \frac{\omega(L)}{\delta(L)} x_{t-l} + n_t, \quad (2.9)$$

where

$$\begin{aligned} \omega(L) &= \omega_0 - \omega_1 L - \omega_2 L^2 - \dots - \omega_s L^s, \\ \delta(L) &= 1 - \delta_1 L - \delta_2 L^2 - \dots - \delta_r L^r, \end{aligned}$$

and r , s , and b constants.

In this model the two polynomials, $\omega(L)$ and $\delta(L)$, replace the polynomial $\nu(L)$, which reduces the number of parameters to estimate, making more efficient use of the data and so producing more accurate forecasts. Note that the subscript for x is $(t-l)$. This means that there is a delay of l periods before x begins to influence y . So x_{t-l} influences y_t first. When $l > 0$, x is often called a "leading indicator" since x_t is leading y_t by l periods.

The reason why this rewritten form is more parsimonious is that the values of r and s are usually going to be much smaller than the value of k . This will be illustrated by the next example. Suppose $\omega(L) = 1.2 - 0.5L$ and $\delta(L) = 1 - 0.8L$. Then

$$\begin{aligned} \frac{\omega(L)}{\delta(L)} &= \frac{1.2 - 0.5L}{1 - 0.8L}, \\ &= (1.2 - 0.5L)(1 - 0.8L)^{-1}, \\ &= (1.2 - 0.5L)(1 + 0.8L + 0.8^2L^2 + 0.8^3L^3 + \dots), \\ &= 1.2 + 0.46L + 0.368L^2 + 0.294L^3 + 0.236L^4 + \dots, \\ &= \nu(L). \end{aligned} \tag{2.10}$$

In this case, $\nu(L)$ would have had an immensely large order number k , while $\omega(L)$ and $\delta(L)$ would have had respectively an order of $r = 1$ and $s = 1$. Therefore, the restated model is more parsimonious.

It is straightforward to extend the dynamic regression model to include several explanatory variables, such that solutions can be tractable. In case of an additive relation between the explanatory variables the formula is given by Equation (2.11)

$$\begin{aligned} y_t &= a + \sum_{i=1}^m \frac{L^i \omega_i(L)}{\delta_i(L)} x_{i,t} + n_t, \\ &= a + \sum_{i=1}^m \frac{\omega_i(L)}{\delta_i(L)} x_{i,t-l_i} + n_t, \end{aligned} \tag{2.11}$$

where

$$\begin{aligned} \omega_i(L) &= \omega_{i,0} - \omega_{i,1}L - \dots - \omega_{i,s_i}L^{s_i}, \\ \delta_i(L) &= 1 - \delta_{i,1}L - \dots - \delta_{i,r_i}L^{r_i}, \end{aligned}$$

where n_t is an ARIMA process. For further information about dynamic regression models, one may consult Makradadis et al [13], in which the models are elaborated extensively.

Models of the stated form above are called dynamic regression models because they involve a dynamic relationship between the response and explanatory variables. According to the terminology of Box and Jenkins [5], the model is sometimes also referred to as a transfer function model.

Intervention Analysis Intervention analysis is a special case of the dynamic regression model. Intervention analysis has become well-known by the article "*Intervention analysis with applications to economic and environmental problems*" of Box and Tiao [6]. The explanatory variable in the model represents an intervention, which is a one-off event which impacts the time series of interest. The influence of this one-off event may be immediate

or it may be spread over a period of time. Nevertheless, the intervention is assumed to occur only once.

Thus intervention analysis allows to have possible changes in the mechanism generating a time series, which causes it to have different properties over different time intervals. This is very useful, because during the period for which a time series is observed, it is sometimes the case that a change occurs. This change can affect the level of the series. For example, a change in the tax laws may have a persistent effect on the daily closing prices of shares on the stock market.

There are four different forms of intervention: step functions, delayed response, pulse functions, and decayed response. The simplest form is the step function. In the step function the expectation of the influence of the intervention will be a sudden and lasting decline or increase in the forecast output series. Suppose the intervention occurred at time u . The dummy variable can be defined by

$$x_t = s \begin{cases} 0 & t < u \\ 1 & t \geq u \end{cases},$$

which is zero before the intervention and one after the intervention. This is called a "step intervention" because the graph of x_t against t resembles a step, which can be seen in Figure 2.1. The intervention analysis model will then be given by

$$y_t = a + \omega x_t + n_t. \quad (2.12)$$

Here y_t denotes the output time series, x_t denotes the input time series and n_t is an ARIMA process. The value of ω represents the size or drop in the forecasting variable y_t .

The other graphs of x_t against t represent the remaining three types of intervention. These graphs are shown in Figure 2.1. The names of the different types of intervention are chosen properly. They indicate the form of the graph. An extensive elaboration of the aforementioned types of intervention can be found in the book written by S. Makridakis, S.C. Wheelwright, and R.J. Hyndman [13].

Dynamic regression models (Transfer function models) are used in the article of Andrews and Cunningham [1]. They used it to model the number of daily calls for orders (buying merchandise) and inquiries (e.g. checking order status) at L.L. Bean. In addition to the day of the week, covariates include the presence of holidays, catalogue mailings, as well as forecasts for orders that are independently produced by the company's marketing department. These periodic disruptions were modeled using intervention analysis along with other transfer functions to capture the effects of more regular influences.

Bianchi, Jarrett, and Hanumara [3] report in their article, that AT&T Bell Laboratories

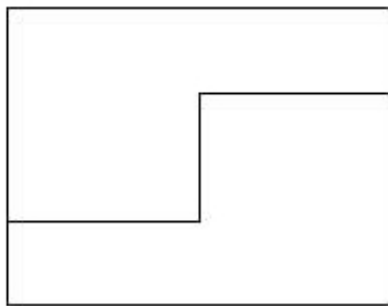


Figure 1: Step Function

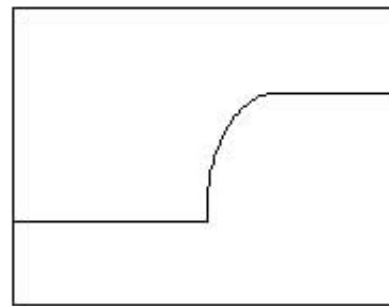


Figure 2: Delayed Function

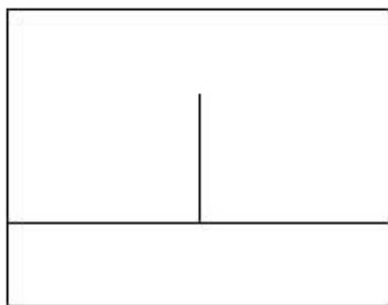


Figure 3: Pulse Function

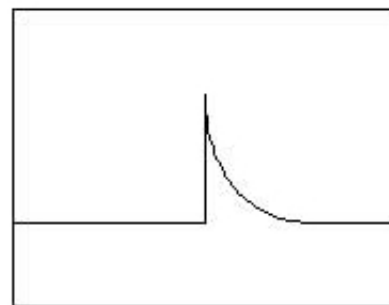


Figure 4: Decayed Function

Figure 2.1: Different types of intervention

used an adaptation of the Holt-Winters forecasting model with its telemarketing scheduling system, called NAMES for forecasting incoming calls in telemarketing centers. In their study they try to evaluate the current use of the Holt-Winters model for forecasting as done by the NAMES system and indicate whether improvement is possible through the use of ARIMA time series modeling. The data they examined, revealed the presence of outliers, which was a big problem for the current NAMES software system. Intervention analysis, the way in which ARIMA models can be used to account for outliers, performed significantly better than either of the Holt-Winters models in more than 50% of the time series studies. Five years later, Bianchi, Jarrett, and Hanumara [4] performed a similar study to analyze and improve existing methods for forecasting calls to telemarketing centers for the purpose of planning and budgeting. The use of additive and multiplicative versions of Holt-Winters exponentially weighted moving average models was analyzed and compared to Box-Jenkins (ARIMA) modeling with intervention analysis. Once more, they found that ARIMA models with intervention analysis performed better for the time series studied.

2.3 Exponential Smoothing

In the late 1950s exponential smoothing methods were developed for the first time by operations researchers. It is not clear whether Holt [10], Brown [7] or maybe even Magee [12] was the first to bring exponential smoothing. Since the development of this concept, it became a widely used method, partly due to its simplicity and low costs.

There are several exponential smoothing methods. The major ones used are Single Exponential Smoothing, Holt's Linear Model (1957) and Holt-Winters' Trend and Seasonality Model. Subsequently, these methods will be described. An extensive elaboration on these models may be found in Makradadis et al [13].

2.3.1 Single Exponential Smoothing

The simplest form of exponential smoothing is single exponential smoothing, and can only be used for data without any systematic trend or seasonal components. Given such a time series, a logical approach is to take a weighted average of past values. So for a series y_1, y_2, \dots, y_{t-1} , the estimate of the value of y_t , given the information available up to time t , is

$$\begin{aligned}\hat{y}_t &= w_0 y_{t-1} + w_1 y_{t-2} + w_2 y_{t-3} + \dots, \\ &= \sum_{i=0}^{\infty} w_i y_{t-i+1},\end{aligned}\tag{2.13}$$

where w_i are the weights given to the past values of the series and sum to one. Since the most recent observations of the series are also the most relevant, it is logical that these

observations should be given more weight than the observations further in the past. This is done by giving declining weights to the series. These decrease by a constant ratio and are of the form:

$$w_i = \alpha(1 - \alpha)^i,$$

where $i = 0, 1, 2, \dots$ and α is the smoothing constant in the range $0 < \alpha < 1$. For example, if α is set to 0.5, the weights will be:

$$\begin{aligned} w_0 &= 0.5, \\ w_1 &= 0.25, \\ w_2 &= 0.125, \\ &\vdots \end{aligned}$$

The equation for the estimate of y_t now becomes

$$\hat{y}_t = \alpha y_{t-1} + \alpha(1 - \alpha)y_{t-2} + \alpha(1 - \alpha)^2 y_{t-3} + \dots, \quad (2.14)$$

since

$$\hat{y}_t = \alpha y_{t-1} + (1 - \alpha)(\alpha y_{t-2} + \alpha(1 - \alpha)y_{t-3} + \dots), \quad (2.15)$$

it can be seen that

$$\hat{y}_t = \alpha y_{t-1} + (1 - \alpha)\hat{y}_{t-1}. \quad (2.16)$$

2.3.2 Holt's Linear Model

Holt's linear model is an extension of single exponential smoothing. Holt designed this method to allow forecasting of data with trends. The method is as follows

For a time series y_1, y_2, \dots, y_{t-1} , the estimate of the value of y_{t+u} , is given by the next formula

$$\hat{y}_{t+u} = m_t + ub_t \quad u = 1, 2, \dots, \quad (2.17)$$

where m_t denotes an estimate of the level of the series at time t and b_t denotes an estimate of the slope of the series at time t . The one step ahead formula is equal to

$$\hat{y}_t = m_{t-1} + b_{t-1}, \quad (2.18)$$

with the updating equations m_t and b_t as

$$\begin{aligned} m_t &= \alpha_0 y_t + (1 - \alpha_0)(m_{t-1} + b_{t-1}), \\ b_t &= \alpha_1(m_t - m_{t-1}) + (1 - \alpha_1)b_{t-1}, \end{aligned} \quad (2.19)$$

with $0 < \alpha_0 < 1$ and $0 < \alpha_1 < 1$.

2.3.3 Holt-Winters' Trend and Seasonality Model

The exponential smoothing methods examined thus far can deal with almost any type of data as long as they are non-seasonal. Nevertheless, in case of seasonality these methods are not very useful. Holt-winters' trend and seasonality model can actually manage seasonality.

Holt's linear model was extended by Winters[15] to include seasonality. The Holt-Winter's method is based on three smoothing equations: one for the level, one for the trend, and one for seasonality. It is similar to Holt's model. The only difference is that an equation is added to deal with seasonality. In fact there are two different Holt-Winters' methods. It depends on whether seasonality is modeled in an additive or multiplicative way. The choice of which to use depends on the characteristics of the specific time series. First the multiplicative model will be discussed.

Multiplicative Model The next formula gives the Holt-Winter's multiplicative function

$$\hat{y}_{t+u} = (m_t + ub_t)c_{t-s+u}, \quad (2.20)$$

where m_t denotes the level of the time series at time t and b_t represents the trend at time t . In this equation, the seasonal component is given by c_{t-s+u} , with s as the length of the seasonal period, and \hat{y}_{t+u} the forecast for u periods ahead. For monthly data ($s = 12$), the next formula is obtained

$$\hat{y}_{t+1} = (m_t + b_t)c_{t-11}. \quad (2.21)$$

Because there are three components to exponential smoothing, three separate smoothing constants are required. The first two constants were already known: α_0 for the level, α_1 for the slope. Now a third constant α_2 is added for the seasonal component. The updating equations of m_t , b_t and c_t are respectively

$$\begin{aligned} m_t &= \alpha_0 \frac{y_t}{c_{t-s}} + (1 - \alpha_0)(m_{t-1} + b_{t-1}), \\ b_t &= \alpha_1(m_t - m_{t-1}) + (1 - \alpha_1)b_{t-1}, \\ c_t &= \alpha_2 \frac{y_t}{m_t} + (1 - \alpha_2)c_{t-s}, \end{aligned} \quad (2.22)$$

with α_0, α_1 and α_2 between 0 and 1.

Additive Model The additive model is slightly different. Instead of multiplying the Holt model by the seasonal component, the seasonal factor is added. The formula will then be given by

$$\hat{y}_{t+u} = m_t + ub_t + c_{t-s+u}. \quad (2.23)$$

Besides the difference mentioned before, the smoothing equations of m_t and c_t also differ from those used in the multiplicative model. For perfection all the update equations are shown beneath.

$$\begin{aligned} m_t &= \alpha_0(y_t - c_{t-s}) + (1 - \alpha_0)(m_{t-1} + b_{t-1}), \\ b_t &= \alpha_1(m_t - m_{t-1}) + (1 - \alpha_1)b_{t-1}, \\ c_t &= \alpha_2(y_t - m_t) + (1 - \alpha_2)c_{t-s}. \end{aligned} \tag{2.24}$$

In only a few articles about forecasting call volumes at call centers, the exponential smoothing methods are discussed. In the article of W. Xu [16], she declared that exponential smoothing was one of the forecasting techniques implemented in Fedex's major forecasting software system (SAS). As described before Bianchi, Jarrett, and Hanumara [3] report in their article, that AT&T Bell Laboratories used an adaptation of Holt-Winters forecasting model with its telemarketing scheduling system, called NAMES, for forecasting incoming calls to telemarketing centers. Nevertheless, in their study to evaluate the current use of the Holt-Winters model for forecasting as done by the NAMES system and indicate whether improvement is possible through the use of ARIMA time series modeling. They found ARIMA models with intervention analysis to perform better for the time series studied. In their subsequent article, which was also discussed in the previous section, Bianchi, Jarrett, and Hanumara [4] performed a similar study to analyze existing and improved methods of forecasting calls to telemarketing centers for the purposes of planning and budgeting. The use of additive and multiplicative versions of Holt-Winters exponentially weighted moving average models was analyzed and compared to Box-Jenkins (ARIMA) modeling with intervention analysis. Once more, they found ARIMA models with intervention analysis to perform more effectively for the time series studied.

2.4 Regression Analysis

Another method used for forecasting call frequencies in call centers is regression analysis. A major advantage of the regression analysis is it is easier to understand, as will become clear below. A short description of regression analysis shall be presented. A complete overview of regression analysis can be found in the book of Heij et al [9].

A regression analysis of time series is often categorized into three effects: the seasonal effect, the trend effect and the random effect.

When time series are measured per quarter, per month, or even per day, they may contain seasonal variation. The seasonal component in a time series refers to patterns that are repeated over a one-year period and that average out in the long run. The patterns that do not average out are included in the constant and trend components of the model. Whereas

the trend is of dominant importance in long-term forecasting, the seasonal component is very significant in short-term forecasting as it is often the main source of short-run fluctuations. Frequently, seasonal effects can be detected from plots of the time series, and also from plots of the seasonal series that consist of the observations in the same month or quarter over different years.

The components can be added-up or multiplied. The models obtained are respectively called the additive model and the multiplicative model. Both models are stated beneath

- **multiplicative**

$$y_t = \alpha_{s_k} * \beta_{t_k} * \gamma_{r_k},$$

- **additive**

$$y_t = \alpha_{s_k} + \beta_{t_k} + \gamma_{r_k},$$

where α_{s_k} denotes the factor determined by the influence of the season s_k at time k , β_{t_k} designates the factor determined by the influence of the trend t_k at time k and γ_{r_k} designates the factor determined by the effect of the random component r_k at time k . It is possible to extend the equation. For instance, it is possible to include the daily effects, such as say the day of the week effect.

Because time series are divided into several components, textual documentation on the subject often refer to it as time series decomposition. Of these pre-mentioned models the multiplicative model is the most widely applied, and often gives better predicting values than the additive one.

It is possible to adopt the mentioned models in combination with linear regression. Before this model is given, first the theory of linear regression will be expounded concisely.

The most simple form of linear regression is simple linear regression. Simple linear regression is used in situations to evaluate the linear relationship between two variables. One example is the relationship between bank wages and education. In other words, simple linear regression is used to develop an equation by which we can predict or estimate a dependent variable (Bank wages), given an independent variable (education). The simple regression equation is given by

$$y = a + bx, \tag{2.25}$$

with y as the dependent variable, and a as the y intercept. The variable b is the gradient or slope of the line and x is the independent variable.

Sometimes it is needed to include more than one independent variable. For example, one might find out that if the variables education and initial salary are both included as

independent variables, the prediction of the dependent variable bank wages is much better. The term multiple regression denotes a number of independent variables that are included. The general formula of multiple regression is as follows

$$y = a + \sum_{i=1}^n b_i x_i, \quad (2.26)$$

with y as the dependent variable, and a as the y intercept and x_i for $i = 1, \dots, n$ the independent variables.

As mentioned before there is a possibility to apply decomposition of time series in combination with linear regression. If an additive time series decomposition is combined with multiple regression the result is given by the next equation

$$y_t = s_t + t_t + r_t + a + \sum_{i=1}^n b_i x_i. \quad (2.27)$$

Articles about the prediction of call center call volumes show that regression analysis is not frequently implemented method. In 1998, Klungle [11] tried to forecast the number of incoming calls for the Emergency Road Service. They found this number to vary at different times of the day significantly during winter and spring seasons. They also tried to model the incoming calls with the Holt-Winter's method and a neural network method, but found that the regression analysis method performed best. Antipov and Meade [2] developed a forecasting model for the number of daily applications for loans at a financial service telephone call center. They built a regression analysis model with a dynamic level, multiplicative calendar effects and a multiplicative advertising response. It was shown to be more effective than an ARIMA model, which was used as a benchmark.

Chapter 3

Applying Techniques

It is now time to try one of the techniques described in the previous section. Because of the fact that the ARIMA model is the most widely applied model in the previous call center forecasting articles, this will be the first method adopted. The second method which will be adopted, is the Holt-winters' trend and seasonality model. This method is chosen because of its simplicity and its speed.

Data is needed to forecast the call frequencies. Without data it is impossible to fit a model, let alone to predict. In the first subsection a description will be given of the data set used in this examination.

The prediction can be done using a software program, which can perform the computations relatively fast, which are needed for the forecast. The software program including the add-ins are discussed in the subsection software.

To assess the results some measurements of forecasting accuracy must be determined. These measurements are explained shortly in the third subsection measurements of forecasting accuracy.

The actual forecasting results will be placed in the fourth subsection results. Even though these results have been gained with great devotion, there will always be possibilities for improvement and further studies. Suggestions for possible improvements are provided in the last section of this chapter.

3.1 Data

The data needed for the forecasts have been supplied by an insurance firm, which sells its products to individuals as well as firms. The insurances they sell range from travel insurances to life insurance policies.

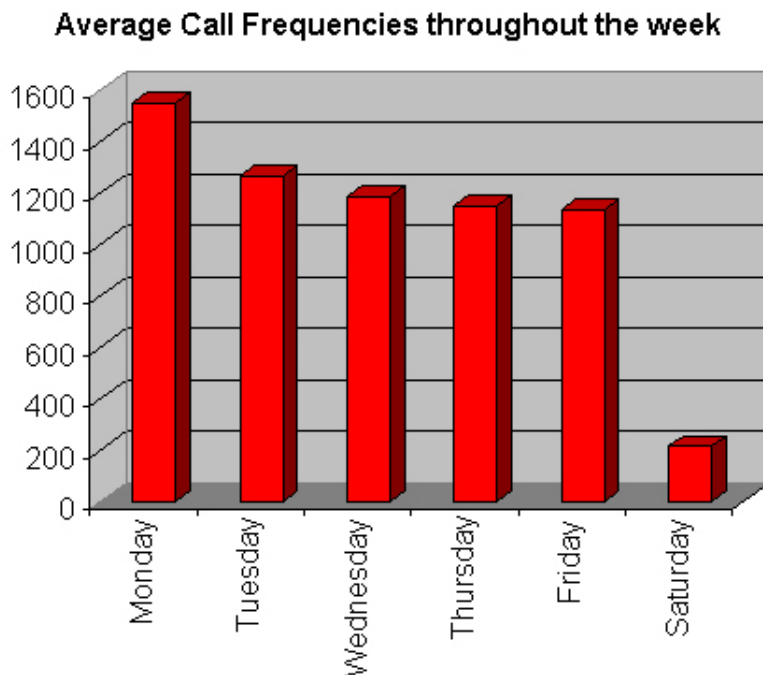


Figure 3.1: Average week of car damage call volumes

As any other insurance company, it has a call center, which is used for both selling purposes and insurance claims. The examined database contained car damage call center frequencies. In the database, which was an Microsoft Excel file, the amount of calls were recorded for every day in the past three years. The data ranges from Monday the 6th of January in 2003 till Saturday the 22th of April in 2006.

Because the insurance company closes its call center on Sundays, the database only captures call amounts from Monday till Saturday. Sometimes, call amounts of particular days were missing too. Most of these missing values are ascribed to National holidays, like Queensday, Christmas, Easter, etc. Because these missing data could cause the predicting techniques to fail, these data had to be filled in. The manner of assigning values to the gaps will be explained later on.

The data showed a strong seasonal effect within the week. Figure 3.1 displays the average week of car damage call volumes. As can be seen from the figure, most calls come in on Monday. From Tuesday till Friday the amount of calls descend steadily. On Saturday the graph drops to its lowest point of the week, which is around 200 calls.

As mentioned before, the missing values had to be compensated. As the figure displayed, the particular day of the week is a very important factor. Thus, the missing values were to be appointed with care. The manner to compensate for these missing values, was to take

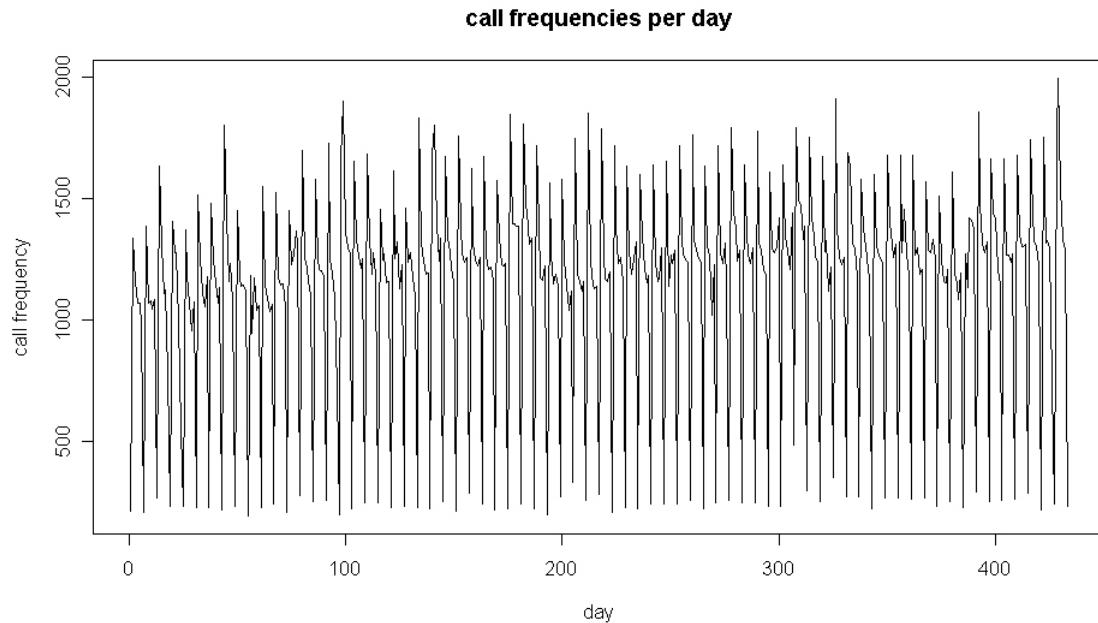


Figure 3.2: Chart of the supplied database

the average of call amounts of a specific day in the week, of the two weeks before and of the two weeks after.

So if the missing day was a Saturday, the value was calculated by taking the average of the amount of the two Saturdays before and two Saturdays after.

Figure 3.2 shows the amount of calls per day. This is just a part of the whole database supplied by an insurance company.

3.2 Software

As mentioned earlier, if one wants to make predictions, a software program is indispensable. There are a lot of software programs out there which can be used. Programs such as SPSS, S-PLUS, R are just a few examples of the many programs.

Because R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License, R was the program to be used. Another advantage is that R compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

R is a language and environment for statistical computing and graphics. R provides a wide

variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. More information about R can be found at the site

<http://cran.r-project.org>.

Within R it is possible to write your own extensions and functions. Due to this and the fact that R is Open Source, a lot of extensions and functions written by others can be found on the internet.

The forecasts made in this examination are made using self-written functions. These self-written functions were based partly based on an extension and partly based on built-in functions.

In R extensions are called packages. The used package for the forecasts of the ARIMA method, the first predicting technique, is written by Rob J. Hyndman and is given the name *forecast*. It includes all kinds of convenient forecast functions. If one is interested, information about the package and the package itself can be downloaded from the site

<http://www-personal.buseco.monash.edu.au/hyndman/Rlibrary/forecast/>.

The functions of the forecast package, which were used for forecasting the call frequencies with the ARIMA method, were *best.arima()* and *forecast()*.

As previously told in the ARIMA model section, the order of the model(p,d,q) has to be selected before the predicting phase can begin. The selection of the right order is mainly based on the AIC (Akaike Information Criterion). This AIC should be minimized. The model which has the minimal AIC is considered the best model. Searching for this minimal AIC value is mainly a trial and error process and can go on forever.

The function *best.arima()* automates this exhaustive search, it looks as the name indicates for the best ARIMA model. At the moment *best.arima()* found the best model, the function *forecast()* can begin producing the forecasts using this model. It should be mentioned though, this whole process can take hours. Naturally, this is partly dependent on the computer's built-in CPU. The average run in this examination took at least ten hours.

The functions used for the second predicting technique, Holt-winters' trend and seasonality model, were already in R. These functions were *ts()*, *HoltWinters()*, and *predict()*.

ts() is a function, used for mutating the database into time series. The function *HoltWinters()* fits the model to the data and *predict()* is used for calculating the actual predictions using the model returned by the *HoltWinters()* function.

If one is interested, one can find all the self-written R-code in the Appendix.

3.3 Measurements of Forecasting Accuracy

Before the forecasting results can be given, some measurements of forecasting accuracy must be determined. This section captures the equations of the most widely applied measurement methods. The following list of methods shall be utilized for assessing the accuracy of forecasts

- Error Total

$$ET = \sum_{t=1}^n e_t,$$

- Mean Percentage Error

$$MPE = \frac{1}{n} \sum_{t=1}^n 100 * \frac{e_t}{y_t},$$

- Mean Absolute Percentage Error

$$MAPE = \frac{1}{n} \sum_{t=1}^n |100 * \frac{e_t}{y_t}|,$$

- Mean Squared Error

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2,$$

- Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2},$$

where e_t is the error at time t or

$$e_t = y_t - \hat{y}_t.$$

In this equation and the equations above the variable y_t means the actual call frequency outcome at day t and \hat{y}_t the predicted call frequency at day t .

A forecast is a good forecast if the ET, MPE, and MAPE values are close to zero and the MSE and RMSE measures are small. The Smaller the MSE and RMSE values are, the better the forecasts.

3.4 Forecasting Results

Now the actual results of the two predicting techniques utilized will be given. Before these results will be given it is necessary to first present the difficulties and the choices made within the research.

The first choice made, was to predict the daily call frequencies using 630 data points. This means that both predicting techniques would be fitted to 630 daily call volume data. The choice to fit to 630 data points was made, due to the fact that 630 call frequencies make exactly two years of data. Two years is just good enough to find a good model, and at the same time it is not too long. This way the history will not have too great an influence. Eventually, this will improve the quality of the forecasts.

To compute the precise forecast accuracy predictions have to be made over a long period. The longer this period is, the more accurate the forecast errors are. In this examination, the forecasts are 312 days long, which is long enough to provide precise forecast accuracy numbers.

The forecasts could be done in different ways, which means there were several variables that could be adapted. Within the ARIMA predicting method, one of those adaptations was the amount of times `best.arima()` got the assignment to calculate. For example, the choice could be made that `best.arima()` computed the best order ARIMA model once in a week and that this model would be exploited self-evidently for the predictions of one week. However, it was also possible for `best.arima()` to calculate the best model each day, which would then be exploited for the prediction of only one day.

The same was found for the Holt-Winters' trend and seasonality prediction model. Only within the Holt-Winters' method the function `Holt-Winters()` was called for computing the best fitting model. So, decisions had to be made about the amount of times the function `HoltWinters()` was called.

Another variable was the moment of time the prediction was made. For example, it is possible to produce today the forecasts for the next month. Another option is to produce the forecasts today for tomorrow. Because the main purpose of the prediction is to schedule the right amount of personnel, the moment of time is taken into consideration. The highest prediction accuracy would probably be reached by predicting one day in advance. But call center employees will certainly not accept this. Maybe the flexible student has some interest in working in this way, but the average worker wants to have some security in his/her live and wants to know when he/she has to work weeks in advance. The assumption is made that schedules will be accepted by employees if they are presented to them at least two weeks in advance.

As previously mentioned, forecasting with the ARIMA method is time consuming. Thus it was necessary to make some choices about the pre-mentioned variables and the amount of forecasts. There has been chosen to perform three forecasts for each prediction method. These forecasts are listed below.

- 1 best fitted model (# of times `best.arima()` / `HoltWinters()`) is calculated once a week

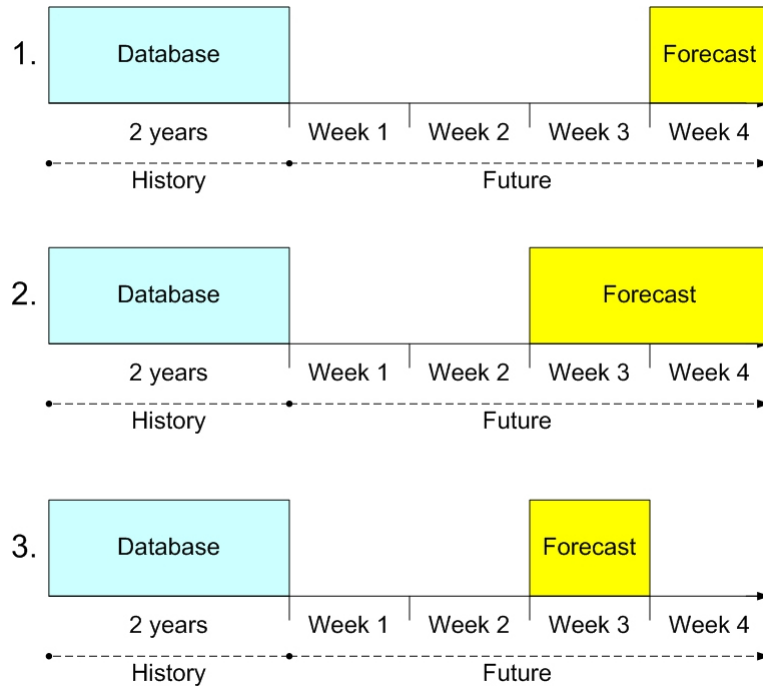


Figure 3.3: Graphical presentation of produced forecasts

and the prediction is made for the fourth week

- 2 best fitted model ($\#$ of times `best.arma()` / `HoltWinters()`) is calculated once in two weeks and the prediction is made for the third and fourth week
- 3 best fitted model ($\#$ of times `best.arma()` / `HoltWinters()`) is calculated once a week and the prediction is made for the third week

These forecasts can also be presented graphically, shown in Figure 3.3. Subsequently, the ARIMA and Holt-Winters' prediction results will be discussed.

3.4.1 ARIMA Model Predictions

Prediction method 1 As noted before, the first prediction method lets the function `best.arma()` calculate the best fitting ARIMA model once a week. This model will then be exploited to forecast the fourth week. The result of this method is graphically displayed in Figure 3.4. The forecast accuracy measurements (FAM) of this prediction method are stated in Table 3.1.

Prediction method 2 In the second prediction method the function `best.arma()` computes the best fitting ARIMA model once in two weeks. This resulting best fitted model will be used to forecast the third and fourth week. The forecast accuracy measurements (FAM) of this prediction method are printed in Table 3.2.

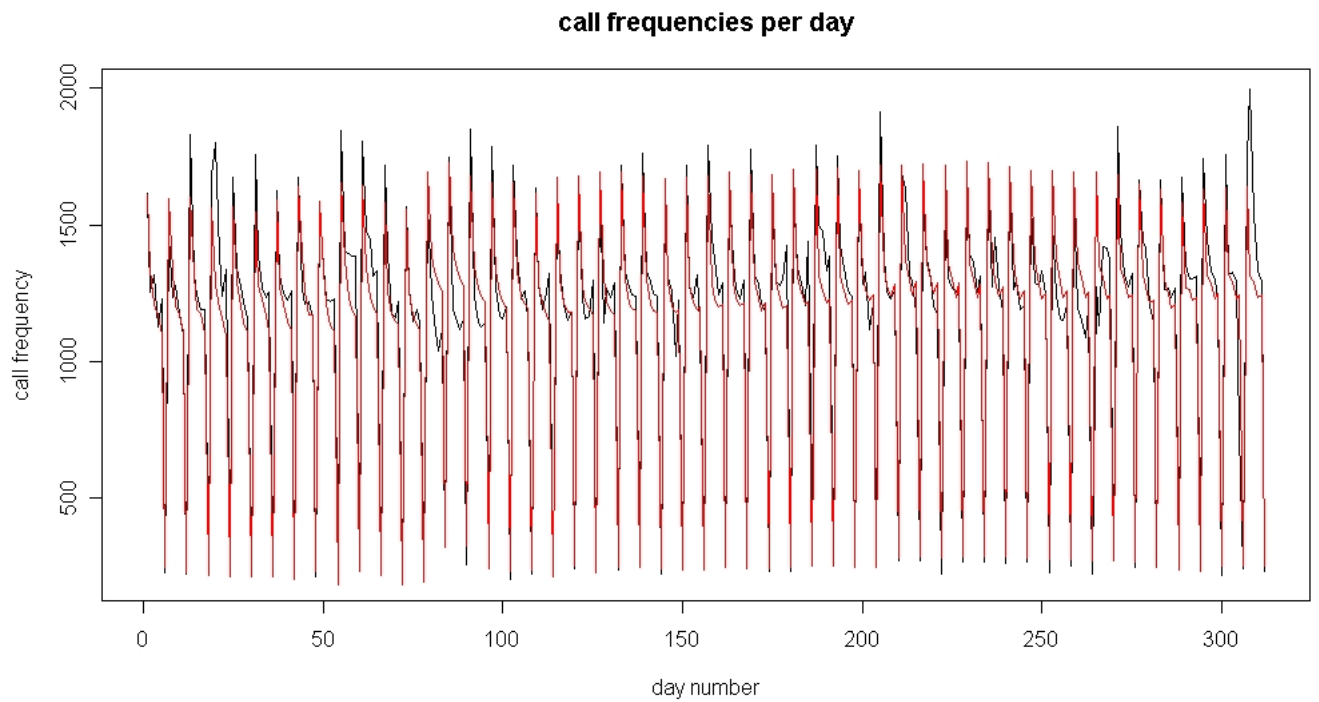


Figure 3.4: ARIMA model: prediction method 1

FAM	Value
ET	-7011.069
MPE	-1.190
MAPE	7.124
MSE	12455.994
RMSE	111.606

Table 3.1: FAM of the ARIMA model: prediction method 1

FAM	Value
ET	-5742.256
MPE	-0.687
MAPE	7.066
MSE	12578.818
RMSE	112.155

Table 3.2: FAM of the ARIMA model: prediction method 2

Prediction method 3 The last method of forecasting is nearly the same method as the first prediction method. The `best.arima()` function calculates the best fitting ARIMA model once a week, which will then be used for forecasting. Instead of forecasting the fourth week, which was the case in the first prediction method, the third week will be predicted. The forecast accuracy measurements (FAM) of this last prediction method are stated in Table 3.3.

FAM	Value
ET	-5870.608
MPE	-0.760
MAPE	7.045
MSE	12570.011
RMSE	112.116

Table 3.3: FAM of the ARIMA model: prediction method 3

For brevity, only the chart of the first prediction method was given. The other charts can be found in the Appendix.

3.4.2 Holt-Winters' Trend and Seasonality Model Predictions

Prediction method 1 Stated above, the first prediction method lets the function `HoltWinters()` calculate the best fitting Holt-Winters' model once a week. This model will then be exploited to forecast the fourth week. The forecast accuracy measurements (FAM) of this prediction method are stated in table 3.4.

FAM	Value
ET	-1165.442
MPE	0.982
MAPE	7.792
MSE	12288.889
RMSE	110.855

Table 3.4: FAM of the Holt-Winters' model: prediction method 1

Prediction method 2 In the second prediction method the function `HoltWinters()` computes the best fitting Holt-Winters' model once in two weeks. This resulting best fitted model will be utilized to forecast the third and fourth week. The forecast accuracy measurements (FAM) of this prediction method are printed in Table 3.5.

Prediction method 3 The last method of forecasting is nearly the same method as the first prediction method. The HoltWinters() function calculates the best fitting Holt-Winters' model once a week, which will then be used for forecasting. Instead of forecasting the fourth week, which was the case in the first prediction method, the third week will be predicted. The forecast accuracy measurements (FAM) of this last prediction method are stated in Table 3.6.

The reader that is interested in the charts belonging to the various prediction methods, is referred to the Appendix.

3.4.3 Summary of Results

To compare all the gained results the easiest way is to compare them right next to each other. Table 3.7 contains the accuracy measurements of all the performed forecasts. Obviously, ARIMA stands for the ARIMA forecasts and HW for the Holt-Winters' forecasts. As can be seen, the Mean Absolute Percentage Errors (MAPE) are printed in bold. This is because the MAPE values are considered as the most important. These values actually

FAM	Value
ET	-667.064
MPE	1.1289
MAPE	7.946
MSE	12861.868
RMSE	113.410

Table 3.5: FAM of the Holt-Winters' model: prediction method 2

FAM	Value
ET	-1254.951
MPE	0.834
MAPE	7.420
MSE	12366.934
RMSE	111.207

Table 3.6: FAM of the Holt-Winters' model: prediction method 3

FAM	ARIMA 1	ARIMA 2	ARIMA 3	HW 1	HW 2	HW 3
ET	-7011.069	-5742.256	-5870.608	-1165.442	-667.064	-1254.951
MPE	-1.190	-0.687	-0.760	0.982	1.1289	0.834
MAPE	7.124	7.066	7.045	7.792	7.946	7.420
MSE	12455.994	12578.818	12570.011	12288.889	12861.868	12366.934
RMSE	111.606	112.155	112.116	110.855	113.410	111.207

Table 3.7: Summary of results

tell the deviation of the predictions in relation to the actual outcome values. According to the table, the third prediction with the ARIMA model is the most accurate, albeit there is not a big difference with the other ARIMA predictions. The MAPE values of ARIMA are much lower than the values of Holt-Winters'.

A big difference can be seen in The Error Total (ET) measurements between the ARIMA method forecasts and the Holt-Winters' forecasts. The ET values of the Holt-Winters' forecasts are significantly lower than the ARIMA method forecasts. This means that the daily call frequency predictions of the ARIMA method are often below the actual outcomes. Thus, the call frequencies are predicted too low. The ET values of the Holt-Winter's forecasts are closer to zero, which is somewhat better.

In the remaining results there was no other clear difference between the ARIMA and Holt-Winters' methods. The Mean Percentage Error (MPE) values are all close to zero, which is the value it should have. The Mean Squared Error (MSE) is continuously around the 12,500 and the Root Mean Squared Error (RMSE), which is deducted from the MSE is around the 112.

As mentioned earlier, the MAPE value is the leading measurement. According to the MAPE value the third forecast made with ARIMA is the most accurate one, notwithstanding the fact that there is no big difference with the other ARIMA predictions. Striking is the fact that the ARIMA forecasts all have better MAPE values than the Holt-Winters' forecasts. On the other hand, it is not that surprising. The average computation time needed for the ARIMA forecasts was about ten hours, while the calculations of all the forecasts with the Holt-Winters' model were done within 1 minute. According to the computing effort and the time invested, the ARIMA model should give better predictions.

3.5 Suggestions for Improvement

As has become clear in the previous sections, some choices needed to be made. This was mainly because of the fact that `best.arma()` required a lot of time to find the best ARIMA model. In spite of all the efforts made and the time that has been contributed to this study, there are some possibilities for improvement. First, one could adjust the time to prediction. In this examination the predictions were made two or three weeks in advance. For example, it is a possibility to make the predictions just one week in advance. In other words, one could predict the second week instead of the third and fourth, which were done in this examination. Predicting the second week will by all odds result in better prediction values.

Second, one could compute the best fitted ARIMA order model every day. In this study, `best.arma()` calculated the best model only once a week or once in two weeks. In this case it is not clear if this suggestion will actually be an improvement, but it is certainly worth trying. A warning should be made to the reader; this calculation would probably take you

60 hours!

Third, the attributes given to the function `best.arima()` could be changed. The ARIMA forecasts produced in this examination all used the next call

```
> best.arima("database", max.p = 10, stationary=TRUE, max.q = 10,  
  max.order=30)
```

This means that `best.arima()` searched for the best fitted ARIMA(p,d,q) model, while satisfying the requirements to search only for models, which satisfy the equations $p \leq 10$, $d \leq 10$, and $q \leq 10$. Possibly, changing these requirements to $p \leq 15$, $d \leq 15$, and $q \leq 15$ would give more accurate forecasts, but then again the computing time would increase exponentially.

Once again, the reader who wishes to forecast and make these possible improvements, should have a computer with an immense computing speed or an enormous load of free time.

Chapter 4

Conclusions and Further Research

The question is which of the aforementioned models to use for forecasting. Actually there is no single best forecasting model. Each model may be best fitted into a specific situation. In this section the advantages/disadvantages will be described of each model.

ARIMA Model The short-term forecasts of the ARIMA model are more reliable than the long-term forecasts. This is because of the fact that the emphasis in forecasts is on short-term forecasting. ARIMA models namely place heavy emphasis on the recent past rather than the distant past.

One of the most important disadvantages for the ARIMA modeling approach is the necessity of a large amount of data. Without a large amount of call center call frequency recordings ARIMA does not function well.

Another disadvantage of this modeling approach is that it is a complex technique, which requires a great deal of experience and although it often produces satisfactory results, those results depend on the researcher's level of expertise, for ARIMA model building is an empirically driven methodology of systematically identifying, estimating, diagnosing, and forecasting time series. This cycle of model building continuously requires the presence and judgement of expert analysts, which can be very expensive.

The computing time needed for the ARIMA model is longer than most of the aforementioned forecasting techniques. For purposes, which need a quick and reliable forecast, ARIMA is not the appropriate model to exploit.

Though these disadvantages might give the feeling ARIMA is not the most ideal forecasting model, ARIMA is capable of producing outstandingly good predictions, which is certainly worth trying.

Dynamic Regression Models Dynamic regression models are actually extensions of the ARIMA model and deal with the shortcomings of the ARIMA model. So the same ad-

vantages / disadvantages described in the previous paragraph bear on dynamic regression models.

The ARIMA model discussed in the previous section deals with single time series and does not allow the inclusion of other information in the models and forecasts. However, frequently other information may be used to aid in forecasting time series. Information about holidays, strikes, changes in the law, or other external variables may be of use in assisting the development of more accurate forecasts. Dynamic regression models offer the possibility to include this information, which is a clear advantage.

Exponential Smoothing Also with exponential smoothing models heavy emphasis is placed on the recent past rather than the distant past. So short-term forecasts of exponential smoothing models are more reliable than the long-term forecasts.

The major advantages of widely used smoothing methods are their simplicity and low cost. Its forecasting equations are easily interpreted and understood by management. So it is popular in call center call volume forecasting systems.

The origin of low costs lies in the fact that exponential smoothing methods are simple to use and can be applied almost automatically. The attendance of expert analysts is thus kept to a minimum. Besides this, the computing time needed to calculate the forecasts is negligible.

A major disadvantage of exponential smoothing methods, is its sensibility to outliers. Outliers can have dramatic effects on forecasts, and should be eliminated.

Self-evidently the disadvantage of single exponential smoothing is that it only works for no trend and no seasonality patterns. Holt's linear models' shortcoming is its disability to manage seasonality. Nevertheless, Holt-Winters' trend and seasonality model can manage trends and seasonality and thus solves the inadequacies of the other exponential smoothing methods.

It is possible that better accuracy can be obtained using more sophisticated methods, such as ARIMA. However, when forecasts are needed for an incredibly amount of items, as is the case in many inventory systems, smoothing methods are often the only methods that are fast enough for acceptable implementation.

Regression Analysis The main disadvantage of regression analysis is that it is hard to implement in an automatic modeling system. It may face multicollinearity problems, serial correlation problems, heteroscedasticity problems, etc. which need bunches of tests with human judgments. Through the inference of expert analysts the costs shall increase, by which this method will not belong to the most economical ones.

The process of building the right forecast model is the most time swallowing part. Once a model is obtained, the computation efforts will be negligible.

Though this description might give a negative impression, this model can give outstandingly good predictions.

Comparison of Models As seen from different forecasting model analysis and comparisons, we can classify the models into different patterns. This overview is given in Table 4.1.

Method	Seasonality	Trend	Random	Interventions
ARIMA Model	yes	yes	yes	no
Dynamic Regression Models	yes	yes	yes	yes
Exponential Smoothing models				
Single Exponential Smoothing	no	no	yes	no
Holt's Linear	no	yes	yes	no
Holt-Winters Trend and Seasonality	yes	yes	yes	no
Regression Analysis	yes	yes	yes	yes

Table 4.1: Comparison of Models

In the table there has been indicated which models can respectively deal with seasonality, trend, random, and intervention. If models can not manage seasonality, or trend, they at least can handle the random part. For example, single exponential smoothing is not sophisticated enough to recognize seasonality or trend patterns. The only thing it can do is handle the random part, which in the exponential smoothing case consists of giving a weighted moving average of lagged values of the dependent variable. In other words, the least a prediction method can do is predicting the random part.

As described in the section Dynamic Regression Models, an intervention is a one-off event which impacts the time series of interest. The influence of this one-off event may be instantaneous or it may be spread over a period of time. A change in the tax laws may, for example, have a continuing effect on the daily closing prices of shares on the stock market. Similarly, the construction of a dam on a river may have a dramatic effect on the time series of streamflow below the dam. In the column interventions, the forecasting methods are selected which can handle the interventions described in the section.

Further Research As shown in the previous subsections, many different methods exist for predicting call volumes in call centers. Nevertheless, these are just a few of the many

possible forecasting methods. As correctly stated by Gans, Koole and Mandelbaum [8] the practice of time series forecasting at call centers is "still in its infancy". There are many methods for making quantitative forecasts which have not been applied to call center data. For instance, it is possible to extend ARIMA to include seasonal effects (SARIMA). SARIMA has already been applied successfully to predict the consumption of goods by civilians and the urban freeway traffic flow. Another possibility is to use the State Space models methodology, which is also applied to predict urban freeway traffic flow.

Neural Networks are increasingly used for quantitative prediction purposes. The applications of this artificial forecasting method are endless. For example, it is useful in determining rainfall expectations, the performance of the stock prices, and the short-term load of electrical engineering.

These models form merely a fraction of the models that can possibly be used for forecasting call volumes in call centers. Extensive research shall point out if these models can be utilized successfully.

Chapter 5

Appendix

5.1 ARIMA Functions R-code

The self-written ARIMA functions which utilized the `best.arima()` and `forecast()` functions. The following code is used for all the ARIMA prediction methods, though some changes were required in the variables to obtain the different prediction methods. The next code is used for forecasting method 1.

```
predictArima <- function(x) {
  predictionsOfOneYear <- numeric(312)
  initialDay <- 73
  weekSizeToPredict <- 6
  index <- 1

  while(!initialDay >= 385) {
    bestModel <- best.arima(x[initialDay:(initialDay+629)],
                           max.p = 10,stationary=TRUE, max.q = 10, max.order=30)
    initialDay <- initialDay + weekSizeToPredict
    temp <- predictArimaSub(bestModel)

    for(i in 1:weekSizeToPredict) {
      predictionsOfOneYear[index] <- temp[i]
      index <- index + 1
    }
  }
  predictionsOfOneYear
}

predictArimaSub <- function(x) {
  weekSizeToPredict <- 6
```

```

predictionsOfOneWeek <- numeric(6)

tempArray <- forecast(x,h=24)

for(j in 1:weekSizeToPredict) {
  predictionsOfOneWeek[j] <- tempArray$mean[18+j]
}
predictionsOfOneWeek
}

```

The computations were called by typing the following line

```
forecastArima <- predictArima("database time series")
```

5.2 Holt-Winters Functions R-code

Below the self-written Holt-Winters function is stated, which utilized the `ts()`, `HoltWinters()` and `predict()` functions. The following code is used for the Holt-Winters prediction methods. In this case also the variables could be changed to obtain the different prediction methods. The next code is used for forecasting method 1.

```

HW <- function(x) {
  predictionsOfOneYear <- numeric(312)
  initialDay <- 73
  weekSizeToPredict <- 6
  index <- 1

  while(!initialDay >= 385) {
    convertToTimeseries <- ts(x[initialDay:(initialDay+629)],start=1,freq=6)
    temp <- predict(HoltWinters(convertToTimeseries),n.ahead=24)
    initialDay <- initialDay + weekSizeToPredict

    for(i in 1:weekSizeToPredict) {
      predictionsOfOneYear[index] <- temp[18+i]
      index <- index + 1
    }
  }
  predictionsOfOneYear
}

```

The computation of predictions started after entering the next line

```
forecastHoltWinters <- HW("database time series")
```

5.3 Error R-code

The following code reflects the computations of forecasting accuracy measurements described in the subsection Measurements of Forecasting Accuracy.

```
returnErrorArray <- function(x,y) {
  errorArray <- numeric(5)

  error <- makeErrorArray(x,y)
  et <- computeET(error)
  mpe <- computeMPE(error,y)
  mape <- computeMAPE(error,y)
  mse <- computeMSE(error)
  rmse <- computeRMSE(error)

  errorArray[1] <- et
  errorArray[2] <- mpe
  errorArray[3] <- mape
  errorArray[4] <- mse
  errorArray[5] <- rmse

  errorArray
}

maakErrorArray <- function(x,y) {
  error <- numeric(length(x))

  for(i in 1:length(x)) {
    error[i] <- x[i]-y[i]
  }
  error
}

computeET <- function(x) {
  et <- numeric(1)

  for(i in 1:length(x)) {
    et <- et + x[i]
  }
  et
}

computeMPE <- function(x,y) {
```

```
mpe <- numeric(1)

for(i in 1:length(x)) {
  mpe <- mpe + (100*(x[i]/y[i]))
}
mpe <- mpe/length(x)
mpe
}

computeMAPE <- function(x,y) {
  mape <- numeric(1)

  for(i in 1:length(x)) {
    mape <- mape + abs(100*(x[i]/y[i]))
  }
  mape <- mape/length(x)
  mape
}

computeMSE <- function(x) {
  mse <- numeric(1)

  for(i in 1:length(x)) {
    mse <- mse + x[i]^2
  }
  mse <- mse/length(x)
  mse
}

computeRMSE <- function(x) {
  sqrt(computeMSE(x))
}
```

The computations were called by typing the following line

```
measurementValues <- returnErrorArray("predicting values","actual
outcome values")
```

5.4 Charts of the Predictions

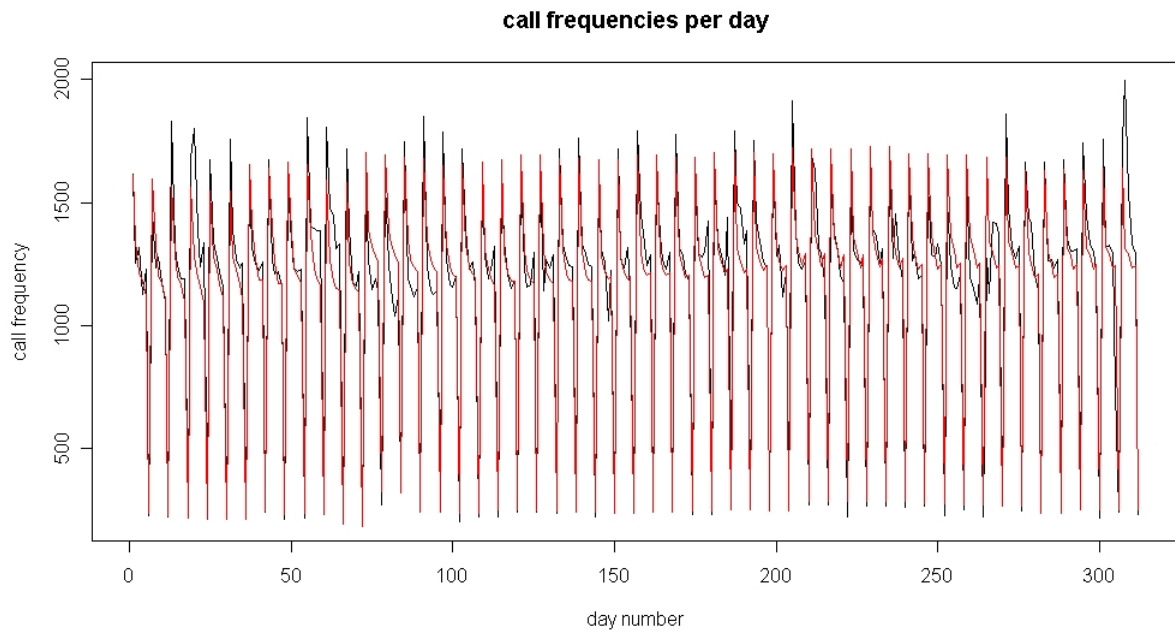


Figure 5.1: ARIMA model: prediction method 2

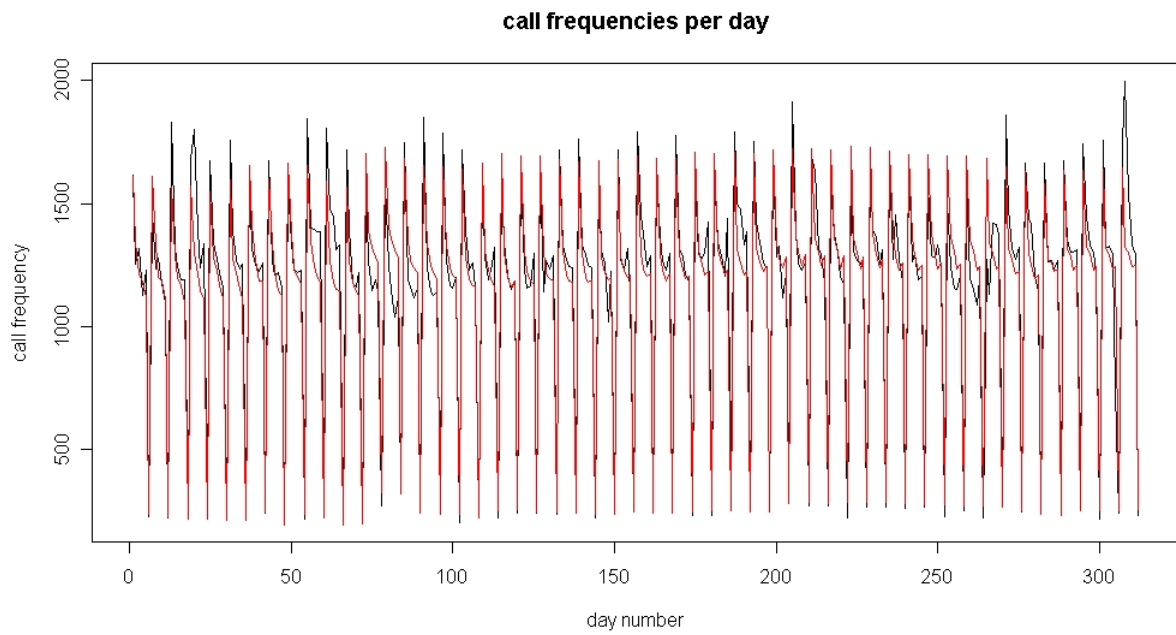


Figure 5.2: ARIMA model: prediction method 3

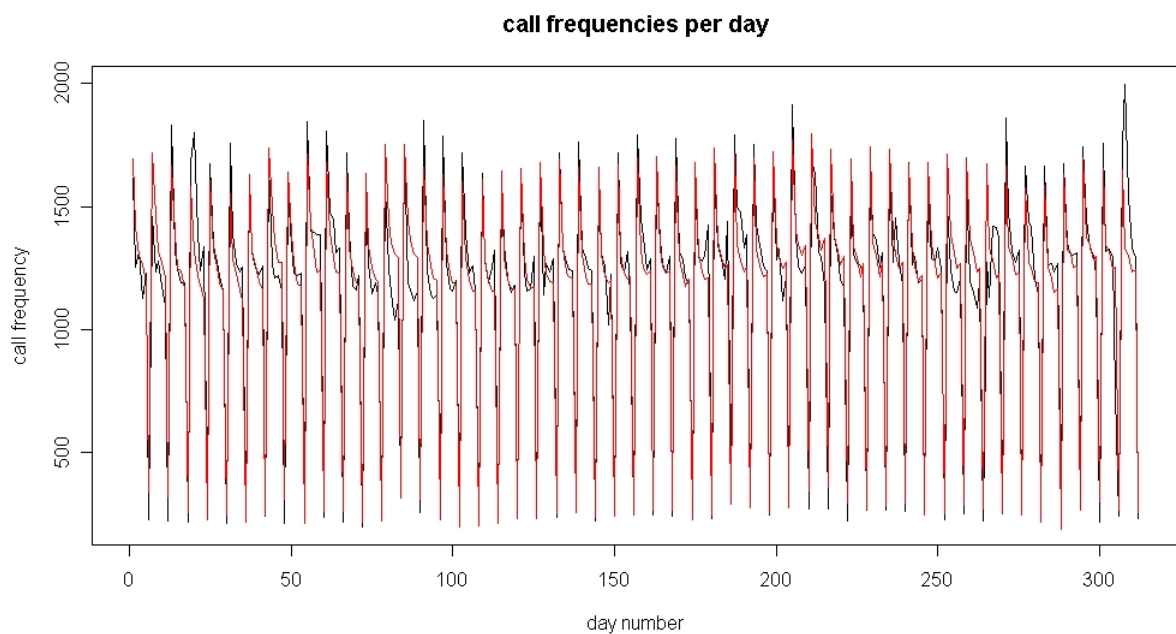


Figure 5.3: Holt-Winters' trend and seasonality model: prediction method 1

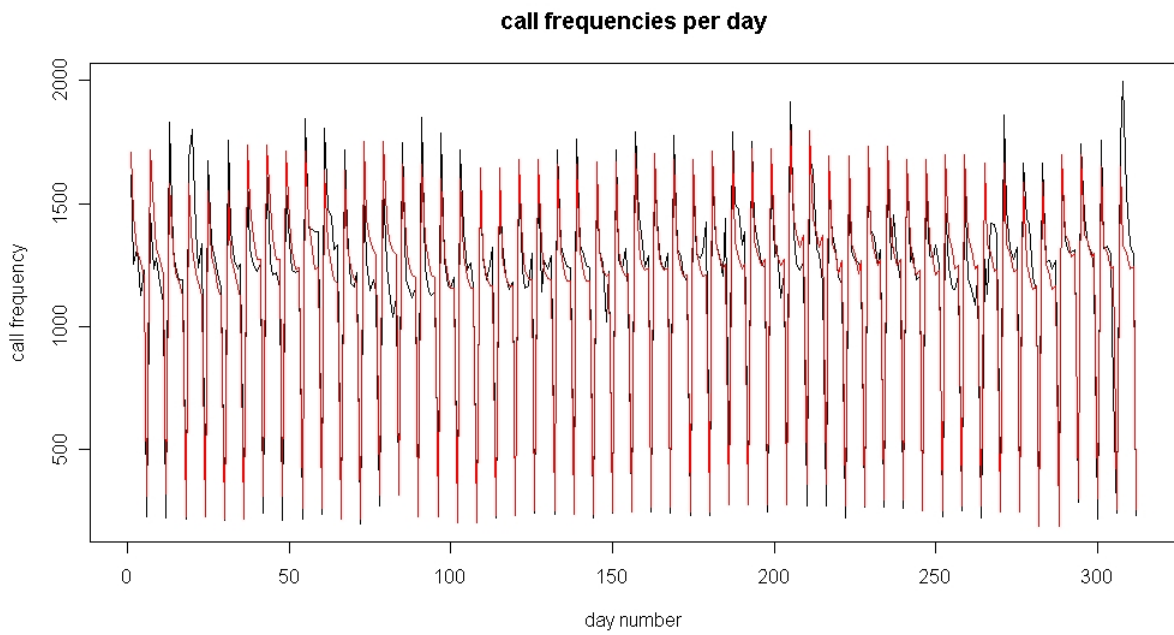


Figure 5.4: Holt-Winters' trend and seasonality model: prediction method 2

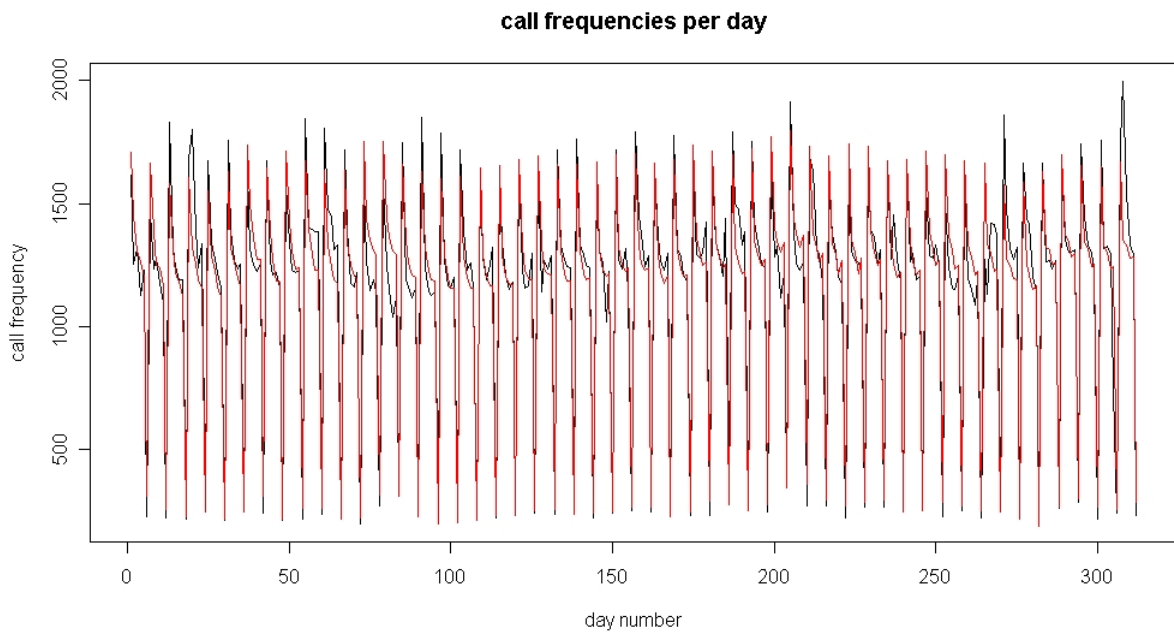


Figure 5.5: Holt-Winters' trend and seasonality model: prediction method 3

Bibliography

- [1] B. H. Andrews and S. M. Cunningham. L.L. Bean improves call-center forecasting. *Interfaces*, 25(6):1–13, 1995.
- [2] A. Antipov and N. Meade. Forecasting call frequency at a financial services call centre. *Journal of the Operational Research Society*, 53(9):953–960, 2002.
- [3] L. Bianchi, J.E. Jarrett, and C.R. Hanumara. Forecasting incoming calls to telemarketing centers. *The Journal of Business Forecasting Methods & Systems*, 12(2):3–12, 1993.
- [4] L. Bianchi, J.E. Jarrett, and C.R. Hanumara. Improving forecasting for telemarketing centers by ARIMA modeling with intervention. *International Journal of Forecasting*, 14(4):497–504, 1998.
- [5] G.E.P. Box and G.M. Jenkins. *Time Series Analysis Forecasting and Control*. Holden-Day, San Francisco, California, 1970.
- [6] G.E.P. Box and G.C. Tiao. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70(349):70–79, 1975.
- [7] R.G. Brown. Exponential smoothing for predicting demand. presented at the tenth national meeting of the Operations Research Society of America, San Fransisco, 16 November 1956, 1956.
- [8] N. Gans, G. Koole, and A. Mandelbaum. Telephone Call Centers: Tutorial, Review, and Research Projects. *Manufacturing & Service Operations Management*, 5:79–141, 2003.
- [9] C. Heij, P.M.C. de Boer, P.H. Franses, T. Kloek, and H.K. van Dijk. *Econometric Methods with Applications in Business and Economics*. Oxford University Press, 2004.
- [10] C.C. Holt. Forecasting seasonals and trend by exponentially weighted moving averages. Technical report, Carnegie Institute of Technology, Pittsburgh, PA, 1957.
- [11] R. Klungle and J. Maluchnik. Call center forecasting at AAA Michigan. *The Journal of Business Forecasting*, 20(4):8–13, 1998.

- [12] J.F. Magee. *Production Planning and Inventory Control*. McGraw-Hill, New York.
- [13] S. Makridakis, S.C. Wheelwright, and R.J. Hyndman. *Forecasting: Methods and Applications*. John Wiley and Sons, Inc., 1998.
- [14] J. Nijdam. Forecasting telecommunications services using Box-Jenkins (ARIMA) models. *Telecommunication Journal of Australia*, 40(1):31–37.
- [15] P.R. Winters. Forecasting sales by exponentially weighed moving averages. *Management Science*, 6:324–342, 1960.
- [16] Weidong Xu. Longe range planning for call centers at FedEx. *The Journal of Business Forecasting Methods and Systems*, 18(4):7–11, Winter 1999/2000.

