Daily sales forecasting in foodservice: Developing a model for application in an expert system

BPJ 420 – Final Report

28 September 2016

H.F. Versluis

11300389

Final Year

B.Eng. (Industrial)

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

DEPARTEMENT BEDRYFS- EN SISTEEMINGENIEURSWESE
DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING

VOORBLAD VIR INDIVIDUELE WERKOPDRAGTE - 2016 FRONT PAGE FOR INDIVIDUAL ASSIGNMENTS - 2016

Persoonlike besonderhede / Personal details	
Studentenommer **Student number**	11300389
Voorletters en van **Initials and surname**	H.F Versluis
Titel Title	Mr.
Selnommer **Cell number**	084 645 5250
Werkopdrag / Assignment	
Modulekode **Module Code**	BPJ 420
Werkopdragnommer **Assignment number**	Final Report
Onderwerp **Subject**	Daily sales forecasting in foodservice: Developing a model for application in an expert system
Dosent **Lecturer**	Prof. Yadavalli
Datum **Date**	28/09/2016

Verklaring / Declaration

1. Ek begryp wat plagiaat is en is bewus van Universiteitsbeleid in hierdie verband
2. Ek verklaar dat hierdie ny eie oorspronklike werk is
3. Waar iemand anders se werk gebruik is (hetsy uit 'n gedrukte bron, die internet of enige ander bron), is dit behoorlik erken en die verwysings ooreenkomstig departementele vereistes gedoen
4. Ek het nie 'n ander student se vorige werk gebruik en as my eie ingedien nie
5. Ek het niemand toegelaat en sal niemand toelaat om my werk te kopieer met die doel om dit as sy of haar eie werk voor te hu nie

1. I understand what plagiarism is and I am aware of the University's policy in this regard.
2. I declare that this is my own original work
3. Where other people's work has been used (either from a printed source, internet or any other source) this has been carefully acknowledged and referenced in accordance with departmental requirements
4. I have not used another student's past work to hand in as my own
5. I have not allowed and will not allow, anyone to copy my work with the intention of handing it in as his/her own work

Handtekening **Signature**	

Executive Summary

The process of forecasting is one that has been applied by many and has evolved significantly over the past several decades. In the foodservice industry, effective and profitable operation relies heavily on accurate and reliable sales forecasts. This project was conducted with the aim of determining if it is possible to develop a forecasting tool consisting of a model or selection of models capable of forecasting daily sales for different restaurants with similar operating environments which can subsequently be developed into an Expert System. This Expert System should replicate the expertise, comprehension, intuition, and intelligence of the forecast knowledge expert in the process of forecasting daily sales in the foodservice industry and can then be used by individuals who has limited or no experience/understanding of the process of forecasting.

This document includes a background on the company for which the model will be developed. It also highlights the problem that were addressed by this project as well as the approach and scope that will were followed towards finding a solution to this problem. A literature review explaining the methodology and application of potential methods that were considered in the selection and development of the final model(s) are included. Data analysis were carried out on the data that were acquired and similarities and observations documented. Finally the appropriate methods were tested and evaluated against one another.

The results obtained from this project indicated that it is indeed possible to develop a forecasting tool capable of forecasting daily sales for different restaurants with similar operating environments. It was also determined that an Expert System can be developed based on the results obtained. The methodology that the Expert System will follow in producing forecasts are also included in this document.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIC - Akaike's Information Criterion

ARIMA – Autoregressive Integrated Moving Average

CCS – Copper Canyon Spur

MAPE – Mean Absolute Percentage Error

MS-ARIMA – Multiple-Seasonal Autoregressive Moving Average

MS-ARIMAX – Multiple-Seasonal Autoregressive Moving Average with External Regressors

MS-TBATS – Multiple-Seasonal, Trigonometric, Box-Cox Transform, ARMA Errors, Trend and Seasonal Factors

TBATS – Trigonometric, Box-Cox Transform, ARMA Errors, Trend and Seasonal Factors

WB – Wimpy Bloemfontein

WM – Wimpy Midrand

WR – Wimpy Roodepoort

1 Introduction and Background

1.1 The Company

Copper Canyon Spur is a family restaurant franchise situated in Lambton Gardens, Johannesburg. The restaurant was opened in 2003 and has since then grown and expanded to cater for customers of every age, race, and ethnicity. The restaurant forms part of the Spur Group (Pty) Ltd which has 505 outlets worldwide of which 318 are Spur Steak Ranches (Spurcorporation.com, 2016). The restaurant is open every day of the year and caters for various types of events such as birthday parties, year-end functions etc. Copper Canyon Spur has a gross revenue of approximately R1-million per month.

The restaurant changed ownership in August 2015 and is therefore currently still in the process of adapting to new management principles and methodologies. Although the new stakeholders have extensive experience in the restaurant franchise industry, this is the first time they are joining the Spur Group. New challenges will inevitably arise when entering a new environment and with it, new approaches, methods and solutions for overcoming these challenges will be developed. One of the challenges that has been identified is the forecasting of daily sales.

1.2 Problem Statement

As stated in section 1.1, one of the challenges faced by the company is daily sales forecasting. Currently, an Ad Hoc forecasting method is being used to predict future daily sales. This method has proven to be quite effective when used for forecasting daily sales of three other restaurants (which are all part of the same but different franchisor) but has shown a decrease in accuracy when it was applied at Copper Canyon Spur. This is due to the fact that applying this type of forecasting method requires experience and expertise in a particular restaurant franchise environment to be effective.

The stakeholders expressed the need for a forecasting model that can be used to forecast daily sales not only for Copper Canyon Spur but also for other restaurants with a similar operating environment. This model can then later be developed into an expert system model that can be implemented at various other restaurants. When implemented into an expert system, the process of producing forecasts should be automated and produce forecasts to be used by individuals without experience in forecasting.

2 Project Aim

The aim of this project will be to create an accurate and reliable forecasting model that can be used to forecast daily sales of different restaurants with similar operating environments. The model should be as user-friendly as possible and allow for continuous development and improvement even after the project has ended. The model should also be adaptable to allow for the future development of an expert system model which would replicate the expertise, comprehension, intuition, and intelligence of the forecast knowledge expert in the process of forecasting daily sales in the foodservice industry.

3 Project Approach, Scope & Deliverables

3.1 Approach & Scope

The first part of the project entailed conducting a thorough, in-depth literature study of existing forecasting methods and approaches. Potential forecasting methods will be identified and the methodology behind the application of these methods will be explained. Related works will also be studied.

A thorough analysis of the daily sales data gathered will then be conducted to get an understanding of the behavior and characteristics of the data and to aid in the process of deciding which forecasting methods should, and can be used. Based on the results obtained from the data analysis the different methods identified will then be applied and different forecasting models developed.

These models will then be tested, evaluated and compared to determine which model/models should be used for application in an expert system.

4 Literature review and problem investigation

4.1 Importance of Forecasting in the Foodservice Industry

Planning is essential for proper and effective management, and forecasting is an important subset of the planning function (Choi, 1999). Rahmlow and Klimberg (2002) identified some of the most important decision areas as well as the impact that forecasting has on these areas within an organization, the results are displayed in Table 1.

Table 4-1: Areas within the Organization Used by Forecasting (Rahmlow and Klimberg, 2002)

Area	%
Budgeting	85
Operations Decisions	67
Financial Decisions	65
Staffing	50
Contingency Planning	47
Investment Decisions	41
New Product	39
Ordering	31

In the food retail industry, a major contributing factor to successful operation and optimal stock management is forecasting (Arunraj and Ahrens, 2015). Kokkinou (2013) states that, as restaurant operators deal with highly perishable products, overestimation of sales can lead to unnecessary labor costs and stock wastage. Underestimation of sales can lead to unsatisfactory customer service and loss of revenue due to stock-outs and insufficient labor capacity. An accurate and reliable forecasting method can reduce wastage of stock and labor, improve customer satisfaction and could ultimately lead to an increase in revenue. It will also provide restaurant operators with information that can be used towards better planning and decision making to ensure effective and profitable operation.

4.2 Forecasting Methods

Forecasting methods can be divided into two general categories: qualitative and quantitative forecasting methods. Qualitative forecasting methods are often applied when limited data is available or when time is insufficient. These forecasting methods usually rely on the judgment of experts within a certain field. They usually take less time to construct and can be relatively inexpensive and easy to understand. A major disadvantage of qualitative forecasting is that it can be largely opinionated and as a result be subjective.

Quantitative forecasting methods rely on historical data to predict the future by finding trends and relationships in the historical data. Quantitative methods can further be classified into time series and causal methods. Time series methods are based on the assumption that past occurrences and behavior has some relevance in the future. They do not focus on what caused this behavior but rather assume that whatever caused this behavior will continue doing so in the future. Predictions are made by determining the impact of trends and seasonal factors on past data and extrapolating this behavior into the future.

Causal methods investigate the impact of principal factors influencing the behavior of the historical data and analyses their effect on the variables under investigation. The relationship between a dependent and independent variable (or variables) are determined and used to create a forecast.

Table 2 includes some examples of qualitative and quantitative forecasting methods.

Table 4-2: List of Forecasting Techniques (Lawrence et al., 2009)

Qualitative Methods	Quantitative Methods	
	Time Series Methods	Causal Methods
Judgment	Moving Average	Regression
Historical Analogy	Exponential Smoothing	Econometric
Focus Group	Trend Analysis	Input-Output
Market Research	Decomposition	Disaggregated
Diffusion	Advanced Time Series methods	Neural nets
Markovian	Box-Jenkins (ARIMA)	

In addition to this, combinations of these methods can also be used. These models are known as Hybrid Models. Hybrid Models are used to improve forecast accuracy by combining two or more forecasting methods with alternative capabilities to accommodate for the limitations that may be present when only one of the methods is used (Arunraj and Ahrens, 2015).

4.3 Selecting Forecasting Methods

4.3.1 Methodology

Selecting the appropriate forecasting method(s) is a crucial part of the forecasting process. As made clear in the preceding section, a wide variety of methods is available, each with its own limitations and capabilities. Armstrong (2001) identified some key principles and factors to consider in the selection of appropriate forecasting methods.

The principles are listed as follows:

- Use forecasting methods that contain methodical and detailed steps that can be explained and replicated.
- If sufficient data is available, use quantitative instead of qualitative methods.
- If large changes in the forecasts can be expected, use causal methods instead of time-series methods.
- Unless considerable proof is present that a complex method will improve forecasts, use simple forecasting methods.

Some factors to consider during the decision process of selecting an appropriate method are:

- Data availability: Is sufficient data available to study past behavior of the variable to be forecast?
- Type of data: Is the data cross-sectional data (obtained at a single point in time) or time series data (obtained at regular intervals in time)?
- Is sufficient knowledge available on the impact of principal factors influencing the behavior of the historical data?
- Is sufficient domain knowledge available on the subject of the forecast?

4.3.2 Application

At least 10 years of daily sales data for Copper Canyon Spur has been acquired, along with daily sales data acquired from three other restaurants forming part of the same but different franchisor (i.e. not part of the Spur Corporation) varying between 4 - 8 years. It can, therefore, be concluded that sufficient time series data is available.

At this point limited knowledge on the principal factors influencing the behavior of the historical data is available, but according to the literature, several factors do exist and should not necessarily be excluded from the study. An example of this is in a study conducted by Arunraj and Ahrens (2015), where they found that there are several factors influencing the daily sales of bananas that are worth investigating. Some of the factors are events, such as regular holidays, festivals and school vacations. It is also not clear if large changes in the data are likely to occur, but it may be safer to assume that they may, or may not, due to factors such as those mentioned above.

As the author has no experience and limited knowledge in the domain of forecasting daily sales in the foodservice industry, it can be concluded that insufficient domain knowledge is available.

In consideration of the above-mentioned factors and principals, the potential methods to be used in this study are identified as time series (or extrapolation) methods as well as regression analysis. Combinations of these methods can also be considered given that substantial evidence exists that it will improve forecasts.

Note to the reader:

The following sections **(sections 4.4.1 - 4.6.3)** are largely based on the textbook *Forecasting: Principles and Practice* by Hyndman and Athanasopoulos, (2013) to ensure that the notation used to describe the various forecasting methods that are included in this study remain consistent. Another contributing factor is the fact that throughout the textbook the authors explain how to apply these methods by using the statistical computing software package known as *R*, which is potentially the main computer package to be used in this project. Other sources may also be included and will be referenced appropriately.

4.4 Exponential smoothing time series forecasting

4.4.1 Time series components

Before introducing the potential time series methods that may be used in this project, we need to define the basic patterns describing the behavior of the data.

- A trend is present when there is a continuous increase or decrease in the data over a long period of time. It can either be linear or non-linear.
- Seasonality is present when there are patterns that repeat at regular fixed intervals such as annual, monthly, or weekly intervals.

Cyclic behavior is present when the data shows signs of increases and decreases but not at fixed intervals. It is usually observed over a period of at least 2 years.

Figure 4-1 shows 4 time series with different or no combinations of trend, seasonality, and cyclic behavior.

- The top left graph displays strong annual seasonality along with some cyclic behavior over a period of 6-10 years.
- The top right graph displays no seasonality but a clear downward trend.
- The bottom left graph displays strong seasonality along with an upward trend.
- The bottom right graph displays no obvious trend, seasonality or cyclic behavior.

(Hyndman and Athanasopoulos, 2013)

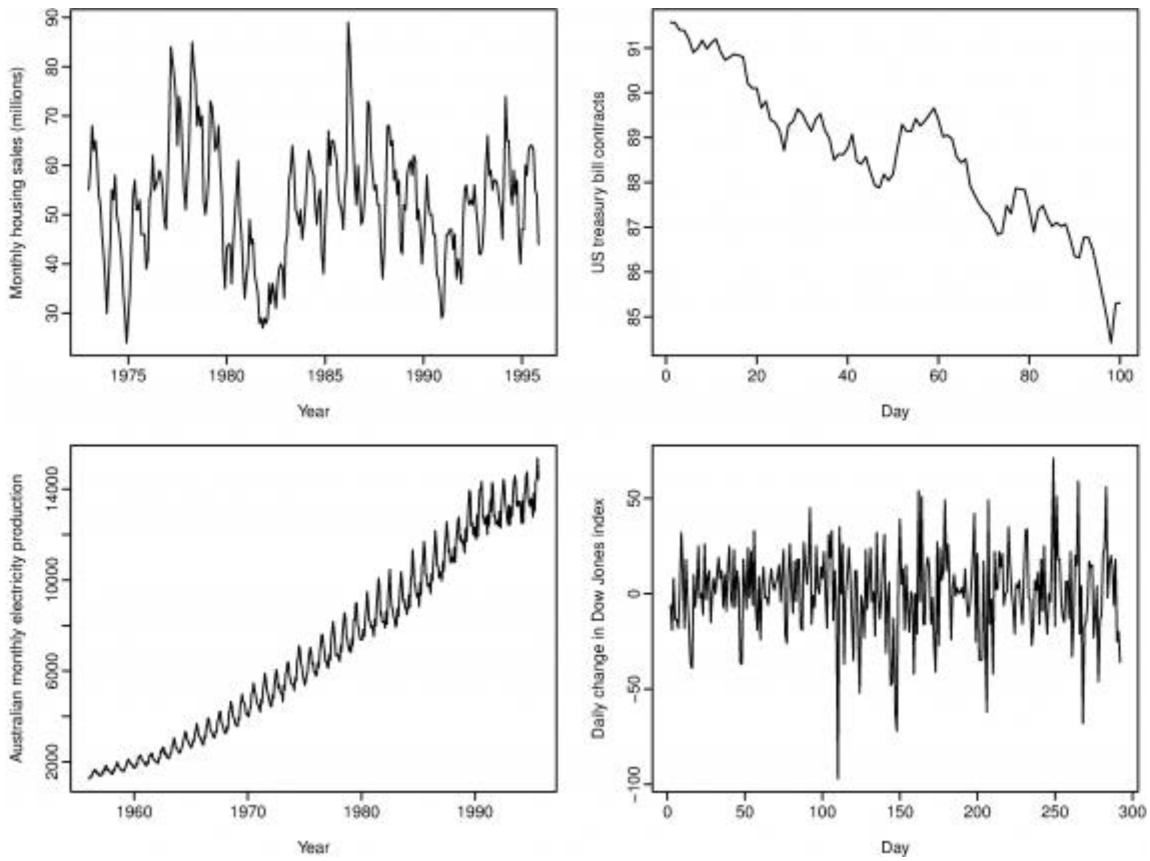

Figure 4-1: Four time series exhibiting different types of time series patterns (Hyndman and Athanasopoulos, 2013).

4.4.2 Naïve and moving average methods

The most basic form of forecasting is known as the naïve method. This forecasting method entails simply observing the last value of a time series and using this as a forecast for the next values of the series, given by:

$$\hat{y}_{t+1|t} = y_t$$

Where $\hat{y}_{t+1|t}$ is the forecast for time $t+1$, at the end of time t. The forecast can be regarded as a weighted average with the entirety of the weight assigned to the last value of the series.

An alternative to this method and one of the most widely used forecasting methods is the moving average method. This method produces forecasts by simply using the average of the last N values of a series, given by:

$$\hat{y}_{t+1|t} = \frac{1}{N} \sum_{i=t-N+1}^{t} y_i$$

Where N is a given parameter. This can be regarded as a weighted average with equal weights assigned to the last N observations (Hyndman and Athanasopoulos, 2013) (Winston, 2004).

4.4.3 Simple exponential smoothing

A combination of the above-mentioned methods is known as simple exponential smoothing. This method generates forecasts by assigning exponentially decreasing weights to past observations with the most recent observation given the largest weight and the earliest observation the smallest, given by:

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots$$

With α a smoothing parameter, $0 \leq \alpha \geq 1$. This method can be used when there is no trend or seasonality present and the data fluctuates around a constant base level (Hyndman and Athanasopoulos, 2013), (Winston, 2004).

4.4.4 Exponential smoothing with trend

Holt (1957) expanded the simple exponential smoothing method to accommodate series where a trend is present. This method is known as Holt's method. The method generates forecasts by combining two smoothing equations, one estimating the level (or base) of the data, and one estimating the trend:

$$\begin{aligned} \hat{y}_{t+h|t} &= \ell_t + hb_t \\ \ell_t &= \alpha + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \end{aligned}$$

Where $\hat{y}_{t+h|t}$ is the forecast for time $t+h$, $(h = 1,2,3, \ldots)$ at the end of time t, ℓ_t is an approximation of the base level of the series at time t, b_t is an approximation of the trend of the series at time t, α is a smoothing parameter for the base of the series, $0 \leq \alpha \geq 1$ and β is a smoothing parameter for the trend, $0 \leq \beta \geq 1$. To initialize the forecast we can choose $\ell_0 = y_1$ and $b_0 = y_2 - y_1$ (Hyndman and Athanasopoulos, 2013).

4.4.5 Exponential smoothing with trend and seasonality

Holt's method can further be expanded into a method that can be used with time series where trend and seasonality are present. This is known as the Holt-Winters method. This method uses a smoothing equation for the level component, another for the trend component, and a third one for the seasonal component. It can further be classified into additive and multiplicative Holt-Winters methods. When throughout the series seasonal patterns are relatively constant, the additive method produces better forecasts:

$$\begin{aligned} \hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t-m+h_m^+} \\ \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \end{aligned}$$

We define m as the period of seasonality, where for example $m = 12$ for monthly data and $m = 4$ for quarterly data. We define ℓ_t, b_t, α, and β the same as with Holt's method. Additionally we have s_t which is an estimate of the seasonal component of the series with γ a smoothing parameter, $0 \leq \gamma \geq 1 - \alpha$. With $h_m^+ = \lfloor (h-1) \bmod m \rfloor + 1$, we ensure that the approximations of the seasonal components come from the last year of the series. To initialize the forecast we can choose $\ell_0 = \frac{1}{m}(y_1 + \cdots + y_m)$, $b_0 = \frac{1}{m}[\frac{y_{m+1} - y_1}{m} + \cdots + \frac{y_{m+m} - y_m}{m}]$, $s_0 = y_m - \ell_0$, and $s_{-1} = y_1 - \ell_0$.

When the seasonal patterns appear to change relative to the level of the series, the multiplicative method produces better forecasts:

$$\hat{y}_{t+h|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$$
$$\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$$
$$s_t = \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}$$

To initialize the forecast we can choose $\ell_0 = \frac{1}{m}(y_1 + \cdots + y_m)$, $b_0 = \frac{1}{m}[\frac{y_{m+1} - y_1}{m} + \cdots + \frac{y_{m+m} - y_m}{m}]$, $s_0 = \frac{y_m}{\ell_0}$, and $s_{-1} = \frac{y_1}{\ell_0}$. It should be noted that there are numerous methods of initializing a forecast each of which depend on the amount and type of data available. The initialization methods proposed here are only recommendations that could prove to be useful (Hyndman and Athanasopoulos, 2013).

4.5 ARIMA modelling

Another method of time series forecasting is known as ARIMA modelling. While the focus of exponential smoothing methods is on the trend and seasonality of the time series, ARIMA modelling is based on finding autocorrelations in the series.

4.5.1 Autocorrelation and ACF plots

Autocorrelation is a method of determining linear relationships between lagged values of a series. A correlation coefficient r_k with a lag length of $k = 1$ will measure the relationship between y_t and y_{t-1}, r_2 will measure the relationship between y_t and y_{t-2} and so forth, r_k can be expressed as:

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

With T the length of the series. The correlation coefficient r_k always lies between -1 and 1, where a positive value indicates a positive relationship and a negative value a negative relationship. Plotting the correlation coefficients normally forms the autocorrelation function (or ACF). ***Figure 4-2*** shows an example of an ACF plot. ACF plots are useful in identifying significant correlations between lagged values of a series. A positive spike indicates a positive correlation and a negative spike a negative correlation (Hyndman and Athanasopoulos, 2013).

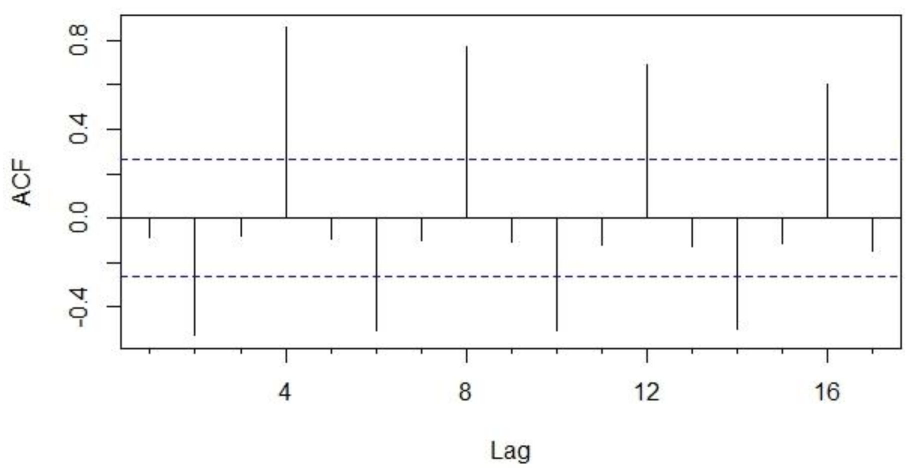

Figure 4-2: Autocorrelation function generated with *RStudio*.

4.5.2 Stationary time series

We define a stationary time series as a series with characteristics that does not depend on the time at which observations are made. Thus, a time series with seasonal or trend components are not stationary as the value of the series change in accordance to the time at which observations are made. A time plot of a stationary series will generally contain unpredictable patterns and indicate a relatively constant horizontal mean with the variance roughly constant throughout the plot. An example of a stationary time series is a white noise series, displayed in ***Figure 4-3*** with its resulting ACF plot displayed in ***Figure 4-4***.

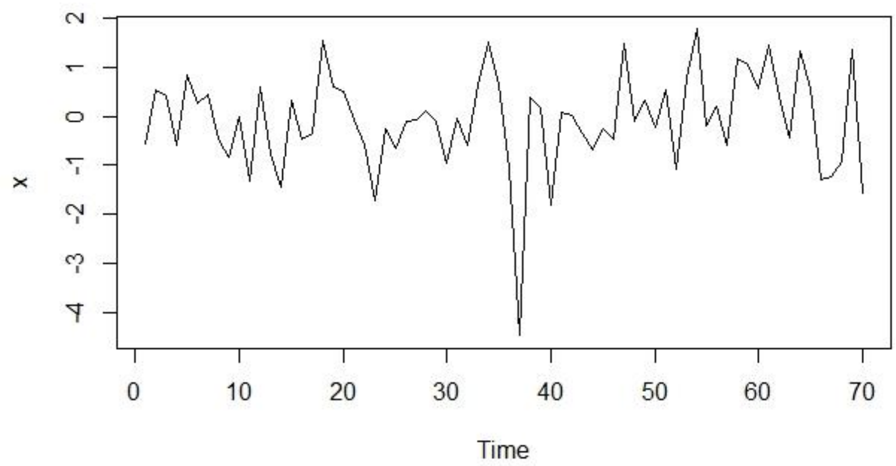

Figure 4-3: Time plot of a random white noise series generated with *RStudio*.

Each autocorrelation of a white noise series are expected to be near zero and 95% of the spikes are expected to be within $\pm 2/\sqrt{T}$, with T the length of the series, these bounds can be seen in ***Figure 4-4*** displayed by the blue dotted lines. When more than 5% of the spikes, or if one or more large spike lie outside these bounds, the series is likely not to be a white noise series. Time series that has no seasonal or trend components, but displays cyclic behavior, are also stationary as the length of the cycles are not fixed and therefore not observable over specific intervals.

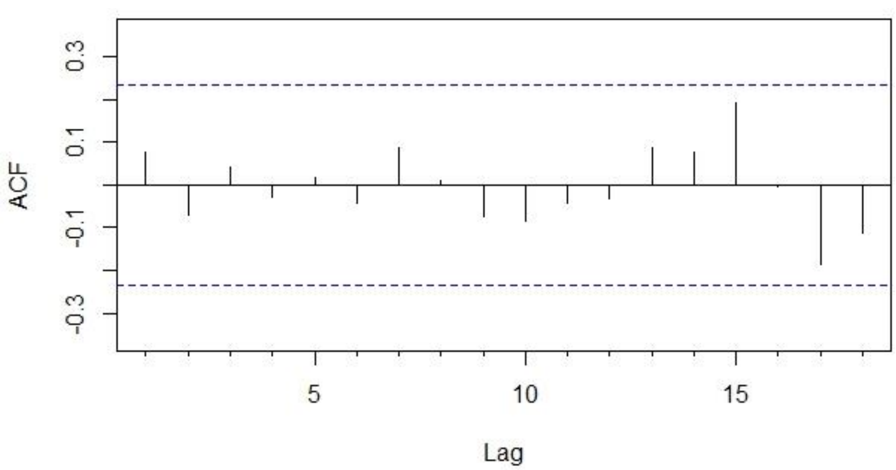

Figure 4-4: Autocorrelation function for the white noise series displayed in Figure 4-3.

When studying the ACF plot of a non-stationary time series, we often see a large positive spike at lag 1 and a slow decrease in the autocorrelations towards zero, while for a stationary time series we usually observe a relatively quick drop in the autocorrelations. An example of this is the ACF plot of 5 years of daily sales for Copper Canyon Spur from January 2009 till December 2013 displayed in ***Figure 4-5***. This series is clearly non-stationary as the plot has a large positive value at lag 1 (the x-axis starts by default at zero when plotting an ACF with *R* but the first autocorrelation is at lag 1) with a slow decrease in the autocorrelations towards zero on the positive side as well as on the negative side of the graph (Hyndman and Athanasopoulos, 2013).

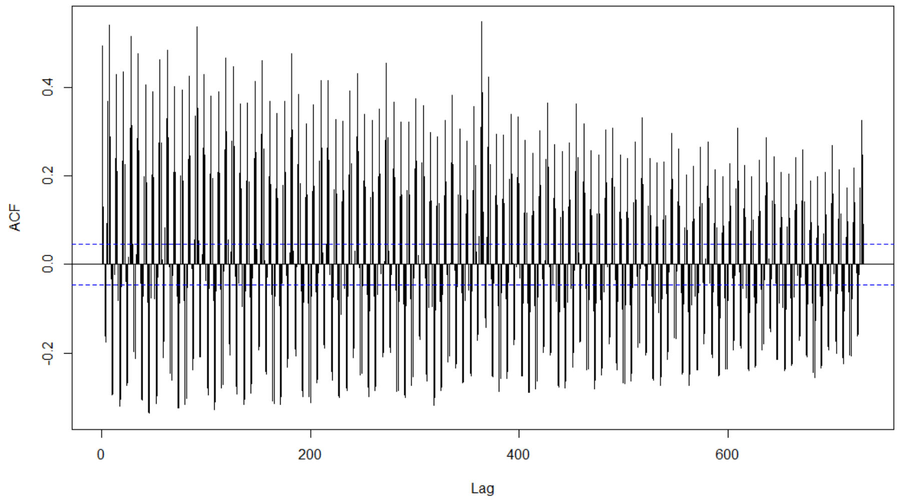

Figure 4-5: ACF plot of daily sales from 2009-2013 for Copper Canyon Spur.

4.5.3 Differencing

Differencing is a method that can be used to transform a time series by attempting to remove the seasonal and trend components from the series to make the series stationary. By calculating the differences between consecutive observations we can obtain a stable mean and eliminate changes in the level of the series (which accounts for trend and seasonality). The differenced series can then be expressed as:

$$y_t' = y_t - y_{t-1}$$

The series will then have $T - 1$ values as we cannot calculate a difference y_1' for the first value of the series. In some cases, a second-order difference may be required to make the series stationary. We then calculate the differences between consecutive values of the first-order differenced series:

$$\begin{aligned} y_t'' &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \end{aligned}$$

Which will then contain $T - 2$ values. Seasonal differencing is done by calculating the differences between observations and their corresponding observations during the previous season:

$$y_t' = y_t - y_{t-m}$$

With m the number of seasons. For monthly data we will calculate the difference between y_t and y_{t-12}. Occasionally we may need to do a first difference as well as a seasonal difference to obtain a stationary series. We will obtain the same results if we start with either the first or the seasonal difference, but the literature suggests that for series with strong seasonal patterns, starting with the seasonal difference may often lead to a stationary series and eliminate the need for an additional first difference (Hyndman and Athanasopoulos, 2013).

4.5.4 Backshift notation

The backward shift operator B can be used when dealing with time series lags in order to simplify complex equations:

$$By_t = y_{t-1}$$
$$B(By_t) = B^2 y_t = y_{t-2}$$

It can also be used for equations involving differencing:

$$y_t' = y_t - y_{t-1} = y_t - By_t = (1-B)y_t$$
$$y_t'' = (1-B)^2 y_t$$

And $(1-B)^d y_t$ a dth order difference of a series (Hyndman and Athanasopoulos, 2013).

4.5.5 Autoregressive models

The term autoregression can be defined as a method of finding linear relationships between a variable and past values of the same variable. We can then forecast this variable using a linear combination of its past values which is known as autoregressive forecasting:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t$$

With p the order of the model which equals the number of past values included in the model, c is a constant and e_t is white noise. This is known as an **AR(p) model**. Different values for the paramaters $\phi_1, \dots, \phi_p$ results in different patterns in the time series. These paramaters are calculated with a computer package such as *R*. Forecasting with autoregressive models generally requires the time series to be stationary (Hyndman and Athanasopoulos, 2013).

4.5.6 Moving average models

A moving average model uses a linear combination of past forecast errors to produce forecasts:

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

With e_t white noise. We refer to this as an **MA(q) model.** Different values for the paramaters $\theta_1, \dots, \theta_q$ results in different patterns in the time series which are also calculated with a computer package (Hyndman and Athanasopoulos, 2013).

4.5.7 Non-seasonal ARIMA models

We obtain a non-seasonal ARIMA model by combining differencing, autoregression and a moving average model. ARIMA is an acronym for AutoRegressive Integrated Moving Average. This model can be given as:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t$$

With y'_t the differenced series that may have been differenced more than once. This is known as an **ARIMA(p, d, q) model**, whith p the order of the autorogressive part of the model, d is the degree of first differencing of the series, and q is the order of the moving average part of the model. By using backshift notation we can rewrite this equation as:

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q) e_t$$

Selecting values for p, d, and q can easily be done by use of a computer package. It can also be done manually, but requires experience and expertise in the field of ARIMA modelling. To estimate the paramaters $c, \phi_1, \dots, \phi_p$, and $\theta_1, \dots, \theta_q$ we use maximum likelyhood estimation (MLE) which is a method applied by a computer package that finds the paramaters that leads to the highest propability that the observed data comes from the selected model. This is similar to the method of least squares estimation discribed in section **4.6** (Hyndman and Athanasopoulos, 2013).

4.5.8 Seasonal ARIMA models

By adding additional seasonal terms to an ARIMA model we obtain a seasonal ARIMA model which can be used to forecast a wide range of seasonal time series, given by:

$$\text{ARIMA } (p, d, q)\ (P, D, Q)_m$$

Where (p, d, q) is the non-seasonal part of the model and $(P, D, Q)_m$ the seasonal part of the model. With m the number of periods in a season. The seasonal part of the model contains similair terms than that of the non-seasonal part of the model, but they involve backshifts of the seasonal period. This is best described with backshift notation:

$$\phi_p(B)\Phi_P(B^m)(1 - B)^d(1 - B^m)^D y_t = c + \theta_q(B)\Theta_Q(B^m) e_t$$

With the non-seasonal components:

$$\text{AR: } \phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

$$\text{MA: } \theta_q(B) = 1 + \theta_1 B + \cdots + B^q$$

$$\text{Non-seasonal difference: } (1 - B)^d$$

And the seasonal components:

$$\text{AR: } \Phi_P(B^m) = 1 - \Phi_1 B^m - \cdots - \Phi_P B^{Pm}$$

$$\text{MA: } \Theta_Q(B^m) = 1 + \Theta_1 B^m + \cdots + \Theta_Q B^{Qm}$$

$$\text{Seasonal difference: } (1 - B^m)^D$$

(Hyndman and Athanasopoulos, 2013)

4.6 Regression analysis

4.6.1 Simple linear regression

If we believe that the value of a variable (defined as the dependent variable) depends linearly on the value of another variable (defined as the independent variable) we can use simple linear regression to estimate this relationship:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Where y is the dependent variable and x the independent variable. The parameter β_0 determines the intercept and parameter β_1 the slope of the linear line $y = \beta_0 + \beta_1 x$, ε is an error term indicating that the value of y may not always fall on this line and captures the effect of anything other than x. The assumptions we make for ε is that the errors have a mean of zero, are not autocorrelated, and are not related to the predictor variable. We obtain estimates for β_0 and β_1, known as least squares estimates, by minimising the sum of the squared errors:

$$\sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_i)^2$$

Which leads to:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

With N the total number of dependent variables or errors, $\bar{y}$ the average of all the y values, and $\bar{x}$ the average of all the x values, the estimated line $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ is known as the regression line. Forecasts are obtained from $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, for $i = 1, \dots, N$, (Hyndman and Athanasopoulos, 2013).

4.6.2 Non-linear relationships

Often the dependent variable and independent variable are related in a non-linear fashion. The easiest way of estimating this relationship is by transforming the variables x and y so that they are linearly related and then estimating a regression line using the transformed variables (Hyndman and Athanasopoulos, 2013).

4.6.3 Multiple regression

We use multiple regression to estimate the relationship between a dependent variable and several independent variables. The relationship can be expressed as:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + e_i$$

With y_i the dependent variable to be forecast and $x_{1,i}, \dots, x_{k,i}$ the k predictor (or independent) variables. The predictor variables should all be numerical. The coefficients $\beta_1, \dots, \beta_k$ determines the impact that each predictor has on the independent variable after considering the impact of all other predictors. We make the same assumptions for the errors as with simple linear regression. We estimate the values of $\beta_1, \ldots, \beta_k$ by minimizing the sum of the squared errors as with simple linear regression:

$$\sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{1,i} - \cdots - \beta_k x_{k,i})^2$$

This is usually done with a computer package such as R, Excel, or MATLAB. The least squares estimates $\hat{\beta}_1, \ldots, \hat{\beta}_k$ are then used to produce forecasts with:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \cdots + \hat{\beta}_k x_{k,i}$$

for $i = 1, \ldots, N$, (Hyndman and Athanasopoulos, 2013).

4.7 Related Works

Arunraj and Ahrens (2015) did a study on daily sales of bananas in a food retail store in Germany. By creating a hybrid seasonal autoregressive integrated moving average and quantile regression forecasting model they found that there are several factors influencing the demand for the bananas that are worth investigating when forecasting in the food sales industry. Some of the factors are Events, such as regular holidays, festivals and school vacations. Weather, such as temperature and snow conditions. Seasonality, such as day of the week, day of the month, month of the year, season of the year and quarter of the year. Price, product characteristics, promotions, discounts and more. Many of these factors have also proven to have an effect on sales within the foodservice industry and are worth investigating.

Lee and Kim (2015) did a study on Pizza sales forecasting in South Korea using big data Analysis. By using data crawling they investigated the effects of time, weather, news, economic indices, trends, sports events and past sales volume on the sales of pizza. They developed three forecasting models, one using the simple average method, another using big data in regression analysis and one using a multilayer perceptron method. The study found that forecasting accuracy can be improved significantly by using big data.

Sanchez (1994) developed an expert system capable forecasting entrée items for three service lines at a university dining center. The system was able to improve forecasting accuracy by up to 20% and displayed the expertise and judgment of a forecasting knowledge expert at the disposal of any person who may not even be educated in the concepts of forecasting.

5 The Data

Daily sales of three different restaurants were obtained for the development of the model. These restaurants include Copper Canyon Spur and three Wimpy restaurants situated in Roodepoort Johannesburg, Midrand Johannesburg and in Central Bloemfontein respectively.

For all the restaurants four years of daily sales spanning over different time periods was used to develop the different forecasting models, with one year of daily sales used to validate the forecasts. Using 4 years of data to develop the models has proven to produce the most accurate forecasts for all four restaurants. This can be ascribed to the fact that the daily sales of the restaurants are extremely dynamic and the franchisors continuously update their menus and promotions, which has an effect on sales patterns. Other factors such as new restaurants and shops opening up nearby also has a significant effect on the sales of the restaurant and therefore the patterns and trends of more than five years in the past may not be as applicable when forecasting sales for the present. Alternatively using less than 4 years of data for development of the models decreased forecasting accuracy as this does not provide enough data for the models to adequately calculate trend and seasonal factors.

Through consultation with the stakeholders, it was determined that sales forecasts for one year are sufficient as short term planning is the main focus for management. Forecasting accuracy also declines as forecasting horizons increase due to the fact that the forecasts may not accommodate changing trends from year to year.

5.1 The Restaurants

5.1.1 Copper Canyon Spur - CCS

Copper Canyon Spur is open on all the days of the year and is situated in an Industrial area. Initially daily sales from January 2011 to December 2015 were used for developing and testing the different models as well as sales from January 2010 to December 2014 to verify certain findings that will be discussed in subsequent sections. At the beginning stages of this project, sales for 2016 were not yet available against which forecasts could be validated but the data has been acquired and forecasts were produced for the period from September 2015 to August 2016. In the subsequent sections, the daily sales from January 2011 to December 2015 will mainly be used to compare forecasting methods and results between the different restaurants.

5.1.2 Wimpy Midrand - WM

Wimpy Midrand is one of the few Wimpy restaurants with a drive-through section and is open on most days of the year but closed on certain public holidays such as *Good Friday* as well as *Christmas and Boxing Day*. Sales from August 2012 to July 2016 were used for the development and validation of the different models.

5.1.3 Wimpy Roodepoort - WR

Wimpy Roodepoort is situated in a shopping mall and only closes on *New-Years-Day, Christmas Day and Boxing Day*. Sales from January 2010 to December 2014 were used for development and validation of models due to the fact that the restaurant changed ownership early 2015.

5.1.4 Wimpy Bloemfontein - WB

Wimpy Bloemfontein was situated in a shopping mall but has closed permanently in 2015. The restaurant was open from Mondays to Saturdays and closed on Sundays. In June 2012 the restaurant started opening on Sundays as well. Due to this change in the weekly business cycle, daily sales from June 2007 to May 2012 were used for development and testing because fitting the model over a period where sales patterns changed so drastically produced highly inaccurate results. This is also an opportunity to test whether the final model or combination of models will be able to forecast sales for a restaurant that is closed on certain days of the week.

5.2 Understanding the Data

In his paper titled *Lessons Learned and Challenges Encountered in Retail Sales Forecast*, Song (2015) states that it is critical to understand multiple seasonalities in the data. As retail sales data often contain strong annual and weekly seasonalities with some cases including monthly as well as quarterly seasonalities. This is evident by examining the ACF plots for all four restaurants.

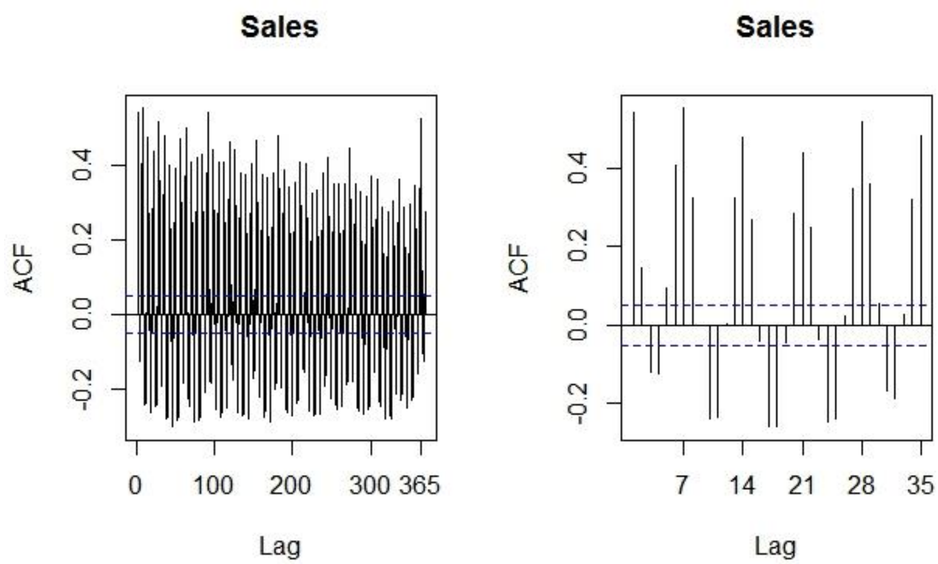

Figure 5-1: ACF plot of Copper Canyon Spur sales data

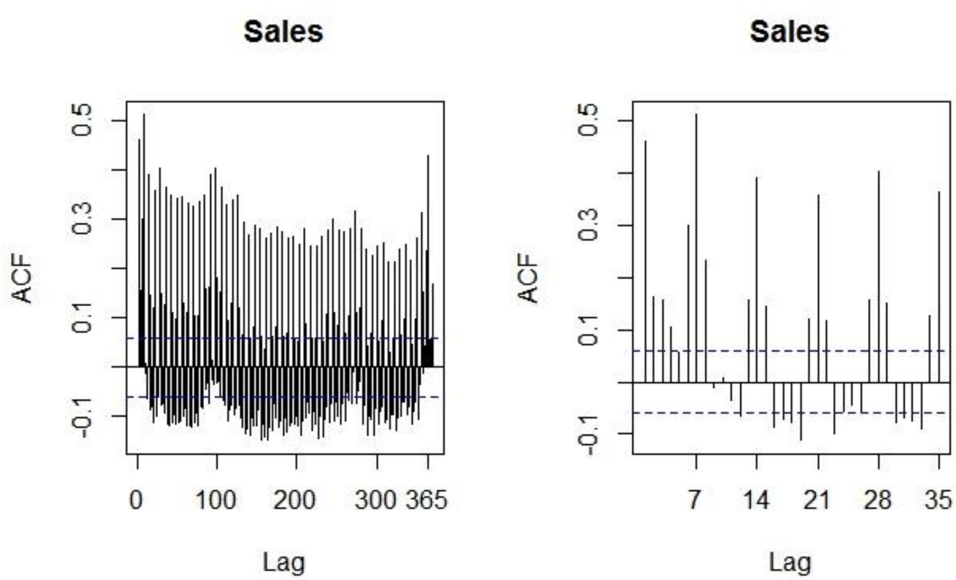

Figure 5-2: ACF plot of Wimpy Midrand sales data

The ACF plots are displayed for lag 1 to 365 as well as for lag 1 to 35 for all four restaurants. As can be seen in ***Figure 5-1***, ***Figure 5-2***, ***Figure 5-3***, and ***Figure 5-4*** there are significant spikes at lag 365, indicating an annual seasonality, as well as significant spikes at lag 7 and multiples thereof, indicating strong weekly seasonality. As will later be shown in Section 6.3 the data also contains a monthly seasonality, but the ACF plot fails to capture this due to the fact that all months do not contain the same number of days and the ACF cannot identify a correlation between a combination of 28, 30, and 31 lagged values. It is, therefore, essential that the potential model should be capable of incorporating multiple seasonalities in order to produce the most accurate and realistic forecasts.

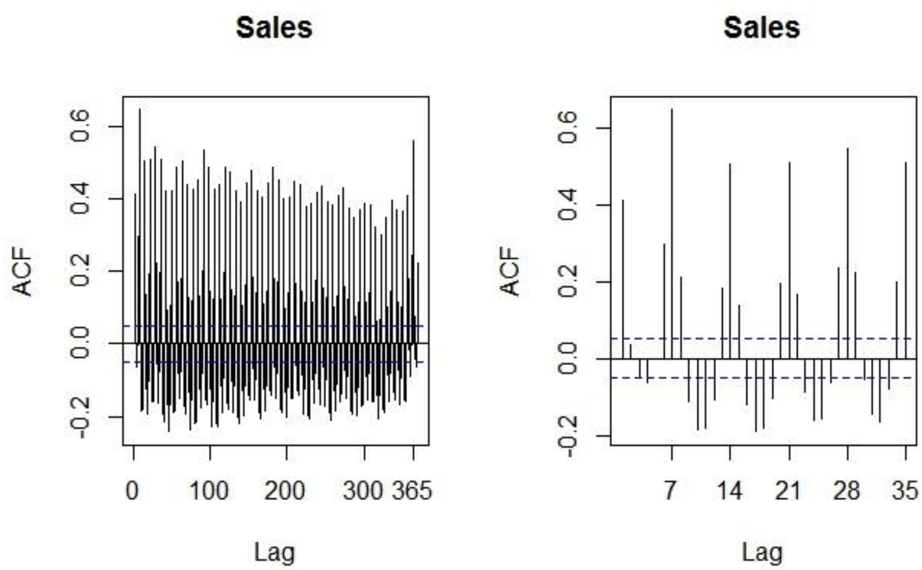

Figure 5-3: ACF plot of Wimpy Roodepoort sales data

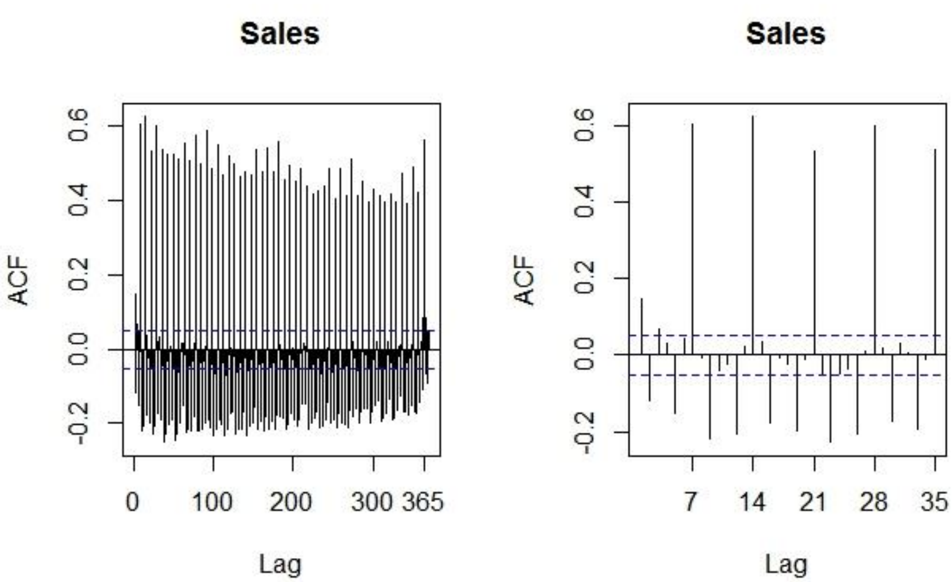

Figure 5-4: ACF of Wimpy Bloemfontein sales data

6 Testing the models

By considering the findings in the previous section, it is clear that only the models identified thus far containing seasonal components should be used, which eliminates the Naïve, Moving Average, Simple Exponential Smoothing and Exponential Smoothing with Trend models. The remaining models will be tested and compared to determine which models can be used in developing the final model. All models will be fitted and tested with the statistical computational software *R*.

6.1 Forecasting Accuracy

The most commonly used method of evaluating forecast accuracy is the Mean Absolute Percentage Error (MAPE):

$$MAPE = mean(\left|\frac{y_i - \widehat{y_i}}{y_i} \times 100\right|)$$

With y_i the *i*th observation (or actual value) and $\widehat{y_i}$ the forecast of y_i. This method is useful in comparing forecasting accuracy of different models as it is expressed as a percentage of the average of all forecasting errors made by a specific model. The forecasting methods that will be tested in the subsequent functions will be evaluated by comparing their respective MAPE values.

6.2 Exponential Smoothing with Trend and Seasonality (Holt-Winters)

The Seasonal Exponential Smoothing models are generally designed to handle short seasonal periods such as 12 for monthly data or 4 for quarterly data (Hyndman, 2016). The Holt-Winters model in *R* can only handle seasonal periods up to 24, which therefore eliminates the possibility of incorporating the Annual or Monthly seasonality into the model. This is due to the fact that the amount of parameters to be estimated for the initial seasonal states depends on the length of the seasonal period, for long seasonal periods, this becomes almost impossible. The weekly seasonality was therefore used when testing this model. The results are displayed in ***Table 6-1*** below. As can be seen by the results, this method does not produce consistently accurate results for all the restaurants but produces a surprisingly low MAPE for Wimpy Midrand.

Table 6-1 Comparison of MAPE of the Holt-Winters model for all restaurants

	CCS	WM	WR	WB
MAPE	41.76%	19.63%	28.89%	31.48%

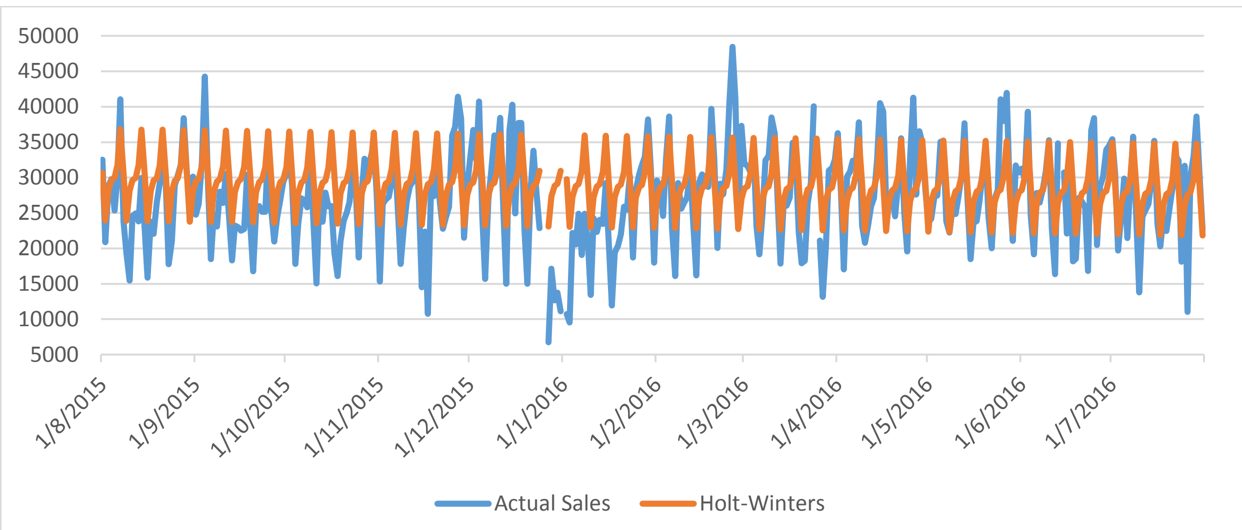

Figure 6-0 Wimpy Midrand Sales forecasts by Holt-Winters model

When investigating the plot of forecasts against actual sales for Wimpy Midrand displayed by ***Figure 6-0***, we see that the model captured the weekly seasonality quite accurately and that this appears to be the strongest seasonality in the model. It fails, however, to capture the annual and monthly seasonalities as can be seen by the period from 1/1/2016 to 1/2/2016 where the daily sales generally take a big dip and gradually increases.

When investigating the plot of the Copper Canyon Spur forecasts displayed by ***Figure 6-1***, we see that the model picked up a downward trend at the end of 2014 (possibly the same downward trend we see at the end of December 2015) which is actually part of the annual seasonal pattern and which looks similar for all four restaurants, and forecasted all the sales for 2015 with a downward trend, eventually producing sales forecasts that are negative and which can realistically never occur.

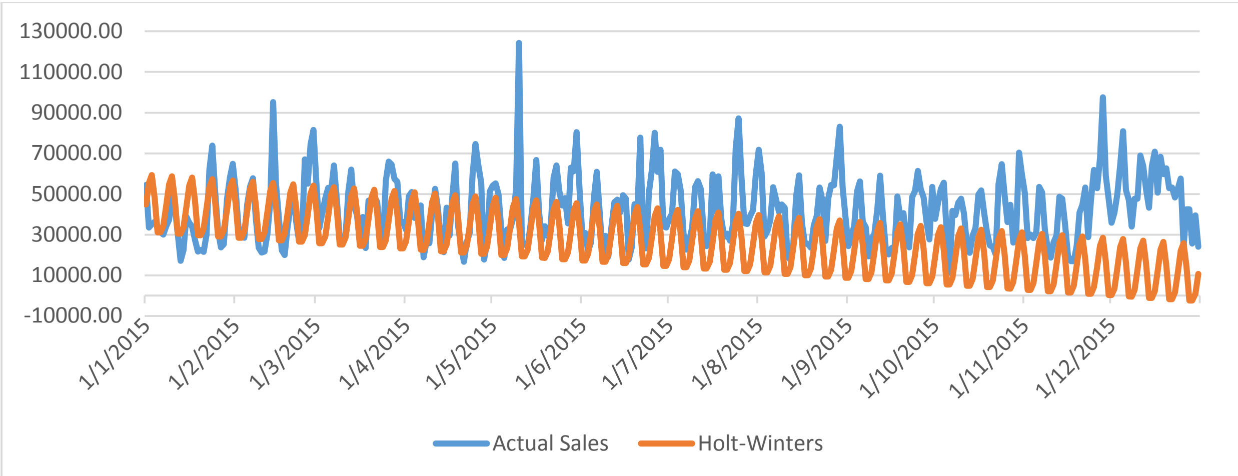

Figure 6-1 Copper Canyon Spur Sales forecasts by Holt-Winters model

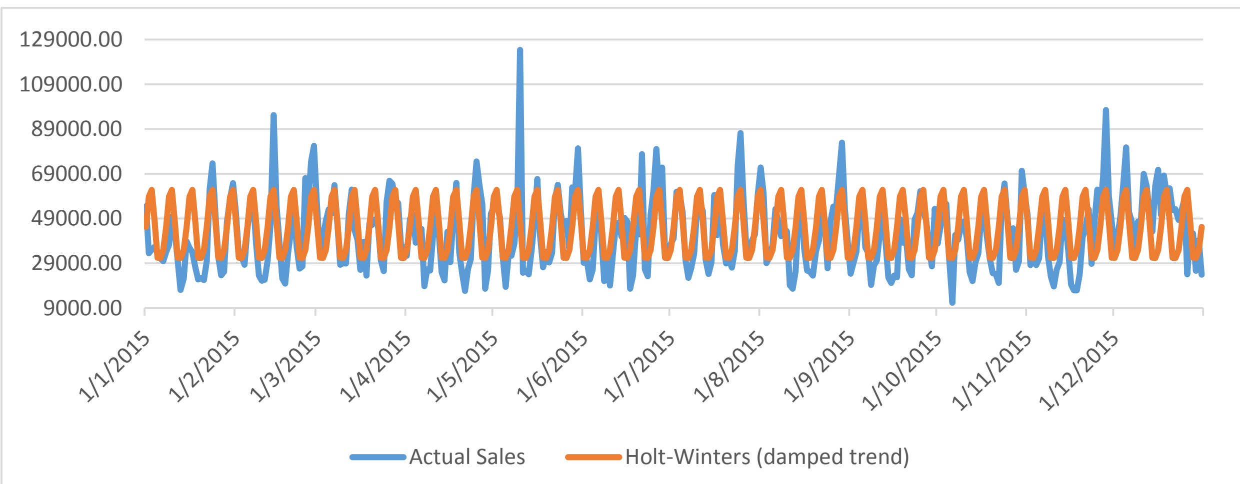

Figure 6-2 Copper Canyon Spur sales forecasts by Holt-Winters model with damped trend

When the model was refitted with the trend damped, the MAPE showed a significant decrease from 41.76% to 27.32% and the plot displays a much more realistic forecast pattern as can be seen in ***Figure 6-2***. Damping the trend also improved the MAPE for Wimpy Roodepoort from 28.89% to 23.94% but did not improve forecasting accuracy for Wimpy Midrand and Wimpy Bloemfontein. The updated table is displayed below.

Table 6-2 MAPE of the Holt-Winters model with damped trend for CCS and WR

	CCS	**WM**	**WR**	**WB**
MAPE	27.32%	19.63%	23.94%	31.48%

6.3 TBATS model

After considering the results obtained from the preceding sections, it became clear that a model capable of incorporating multiple seasonalities is needed in order to produce accurate and reliable forecasts. Subsequently, the author researched possible methods that accommodate this important feature. One such method that was discovered after the literature review was conducted is an innovations state space model named TBATS. TBATS is a forecasting model developed by De Livera et al. (2012) capable of forecasting time series with complex seasonal patterns as well as multiple seasonalities. TBATS is an acronym for the most important characteristics of the model: Trigonometric Fourier representations, Box-Cox transformations, ARMA errors, Trend, and Seasonal components.

It builds on the Holt-Winters model by including a Box-Cox transformation (which is a method of transforming the data to stabilize the variance), replacing the error term with an $ARMA(p, q)$ process, and replacing the seasonal component of the model with a trigonometric representation based on Fourier series that incorporates multiple seasonalities. Because of the trigonometric functions, the model is capable of incorporating non-integer seasonalities.

The model was fitted to the sales data for all four restaurants with the results displayed in ***Table 6-3***. To illustrate the effect that the multiple seasonalities have on the data, the model was first fitted with only a single seasonal period of 7 for the weekly seasonality, then with seasonal periods of 7 and 365.25 for the weekly and annual seasonalities respectively, and finally with seasonal periods of 7, 30.4375, and 365.25 for weekly, monthly, and annual seasonalities respectively. The values of 365.25 and 30.4375 are used to accommodate for leap years and varying days within different months.

Table 6-3 MAPE of the TBATS model with different seasonal periods for all restaurants

Seasonal period(s)	CCS	WM	WR	WB
7	28.20%	17.69%	20.80%	30.79%
7; 365.25	24.40%	14.29%	17.79%	26.45%
7; 30.4375; 365.25	19.42%	14.07%	15.24%	23.33%

As can be seen from the table, forecasting accuracy from the TBATS model for all four restaurants improved significantly as additional seasonalities are incorporated. This also shows that there is indeed a monthly seasonality present in the data as for Copper Canyon Spur, Wimpy Roodepoort, and Wimpy Bloemfontein the MAPE was lowered significantly with the addition of the monthly seasonality resulting in a 20.41%, 14.33%, and 11.8% improvement of their forecasting accuracies respectively. Even though the MAPE for Wimpy Midrand decreased only slightly with the addition of the monthly seasonal component resulting in a 1.54% improvement in forecasting accuracy, this is still an indication that the monthly seasonality is present, albeit its effect may not be as strong as on the other restaurants.

When we investigate the time plot of the forecasts for Wimpy Roodepoort by the TBATS model with weekly and annual seasonalities incorporated, displayed by ***Figure 6-3***, we see how this model now captures not only the weekly seasonality of the data but also the annual seasonality as can be seen by the gradual wave-like fluctuations of the mean of the forecasts throughout the year.

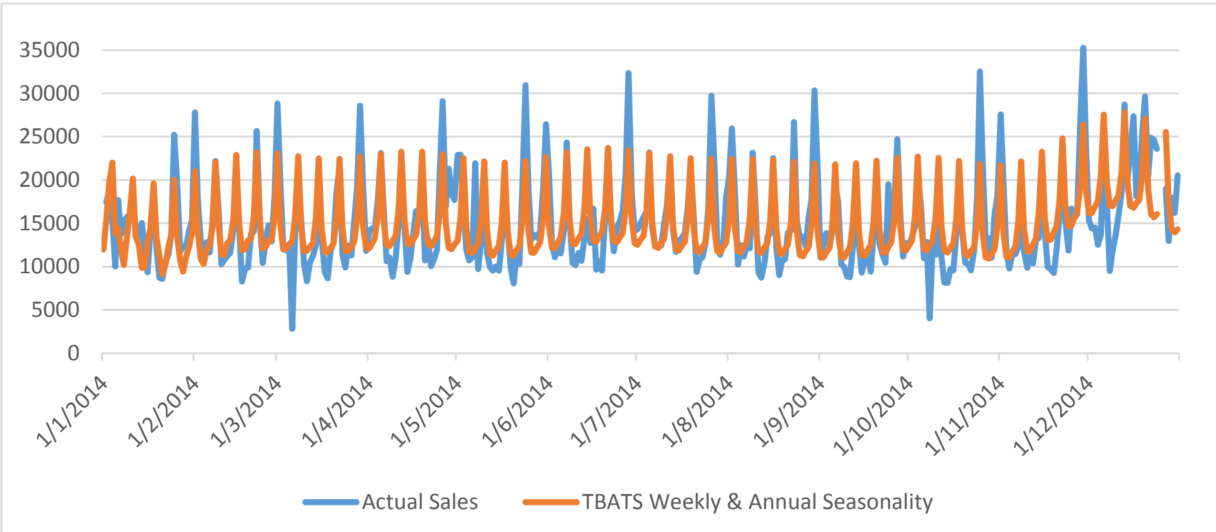

Figure 6-3 Forecast for Wimpy Roodepoort by TBATS model with weekly & annual seasonalities

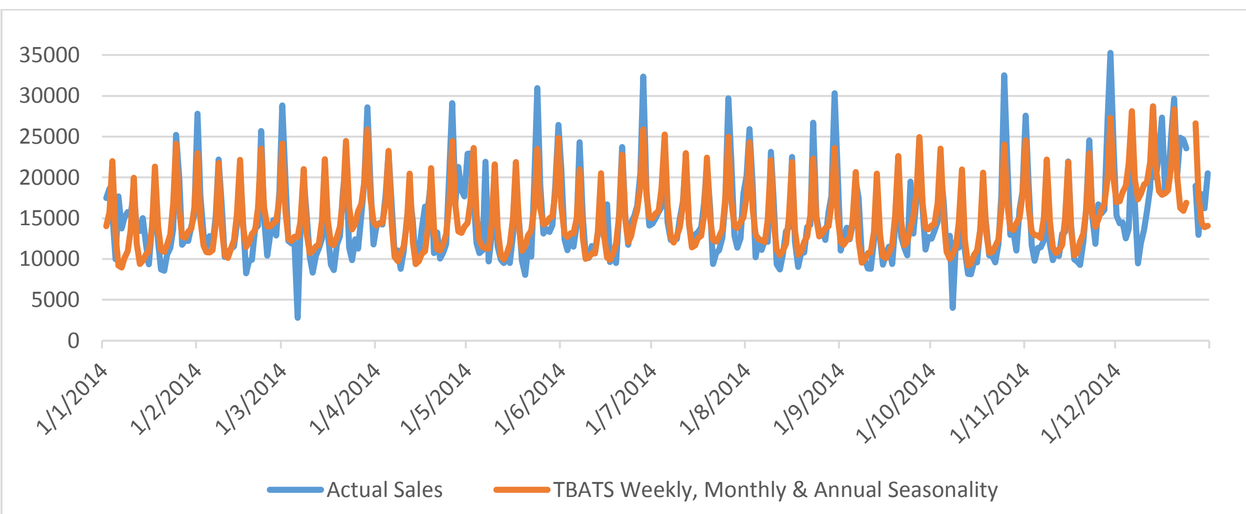

Figure 6-4 Forecast for Wimpy Roodepoort by TBATS model with weekly, monthly & annual seasonalities

When we investigate the plot of the forecasts for Wimpy Roodepoort by the TBATS model with weekly, monthly, and annual seasonalities incorporated, displayed by ***Figure 6-4,*** we see how the model now captures the weekly and annual seasonality, and additionally also the monthly seasonality that can be seen by the local maxima at the beginning and end of each month and the local minima near the middle of the month. The exception is during the December month where the sales start increasing from the start of the month to reach a local maximum near the middle of the month and then decrease towards the end of the month, but this is due to the annual seasonality and not the monthly seasonality, which makes logical sense, as middle December is the time when shopping malls and restaurants are the busiest during the year as most people are on holiday and/or doing pre-Christmas shopping.

Figure 6-5 displays the observations that we made regarding the seasonality of the data quite clearly. This is a decomposition of the daily sales from January 2010 to December 2014 for Wimpy Roodepoort that was used to fit the model into its trend and seasonal components. At the top of the graph, we can see the observed data that was used to fit the model. The *level* component can be seen as a baseline for the trend, and the *slope* the change per unit time. *Season1* is the weekly seasonality, *Season2* is the monthly seasonality, and *Season3* the annual seasonality.

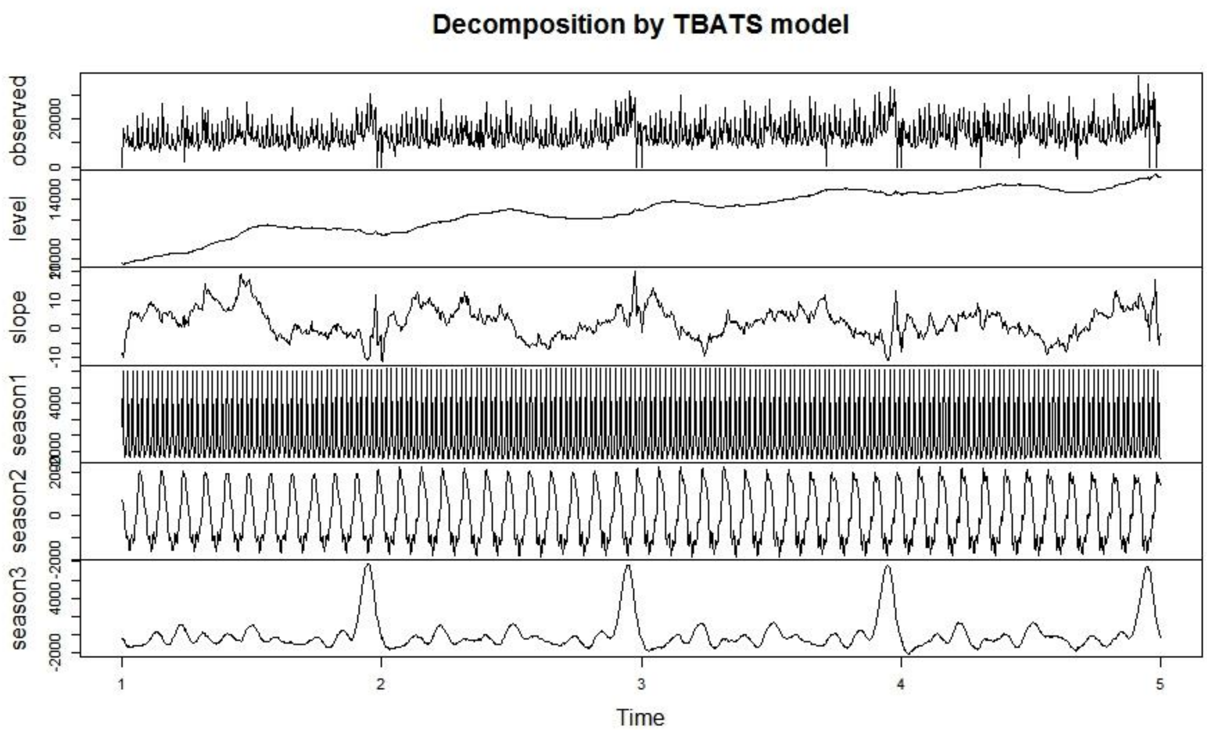

Figure 6-5 Decomposition of the trend and seasonal components of daily sales for Wimpy Roodepoort from Jan 2010 to Dec 2014

6.4 Multiple-Seasonal ARIMA model

As is evidently clear from the preceding sections, fitting a model without incorporating the multiple seasonalities present in the data will not produce better forecasts than what we have already achieved by the TBATS model. Therefore a non-seasonal ARIMA will not be adequate.

The design of seasonal ARIMA models faces the same problem as with the Seasonal Exponential Smoothing models mentioned in ***Section 6.2***, they usually only accommodate seasonal periods up to 12. The ARIMA function in *R* can handle any seasonal period up to 350 but generally runs out of memory when the seasonal period is greater than 200, the reason for this is not yet clear to the developers of this function. Seasonal ARIMA models are based on seasonal differencing to find autocorrelations between a value and the previous value one seasonal period back. This poses a problem for annual seasonalities, as this means that we have to compare what happened today to what happened one year ago and there is no guarantee that the seasonal pattern is smooth, with leap years also posing a problem. With monthly seasonalities, the problem is the varying number of days within a month as discussed in ***Section 5.5***. Additionally, seasonal ARIMA models can also not incorporate multiple seasonalities (Hyndman, 2016).

Alternatively, a method of producing forecasts with ARIMA modelling containing multiple non-integer seasonalities is to model the seasonal pattern by using Fourier terms. The resulting model will then be:

$$y_t = a + \sum_{k=1}^{K}[\alpha_k \sin\left(\frac{2\pi kt}{m}\right) + \beta_k \cos\left(\frac{2\pi kt}{m}\right)] + N_t$$

Where N_t is an $ARIMA(p, d, q)$ process, a is a constant, α_k and β_k are smoothing parameters, and K is the number of Fourier terms to include for each seasonal period = m. The value of K can be chosen by minimizing the AIC of the model, which is known as Akaike's Information Criterion, a measure of the likelyhood that the fitted model came from the data that was used in the fitting process. Small values of K produces a more smooth seasonal pattern, whereas increasing the value of K allows for more wiggly seasonal patterns. For all four restuarants a value $K = 3, K = 10$, and $K = 43$ for the weekly, monthly, and annual seasonalities respectively proved to provide the lowest AIC value. The (p, d, q) order of the $ARIMA$ function is also chosen by minimizing the AIC.

Table 6-4: MAPE of fitted multiple-seasonal ARIMA models and their orders for all restaurants

	CCS	**WM**	**WR**	**WB**
$\boldsymbol{ARIMA(p,d,q)}$**:**	(1,1,1)	(0,1,4)	(5,1,0)	(0,1,3)
MAPE	22.83%	13.71%	13.57%	21.66%

After obtaining the correct order of the ARIMA models and optimal K values for the seasonalities, the models was used to produce forecasts for all restaurants. The results are displayed in ***Table 6-4***. In comparing the MAPE of these models with the ones obtained for the fitted multiple-seasonal TBATS models, we see an improvement of 2.56% for Wimpy Midrand, an improvement of 10.95% for Wimpy Roodepoort, and an improvement of 7.16% in forecasting accuracy for Wimpy Bloemfontein. For Copper Canyon Spur however, we see a deteriation of 14.93% in forecasting accuracy of the ARIMA model in comparison with the TBATS model. The possible reasons for these observations will be discussed in the subsequent section.

6.5 Multiple-Seasonal ARIMA versus Multiple-Seasonal TBATS models

The fact that the multiple-seasonal ARIMA models (hereafter referred to as MS-ARIMA models) produced better forecasts than the multiple-seasonal TBATS models (hereafter referred to as MS-TBATS models) for three out of the four restaurants could be due to various factors. The TBATS model is based on Exponential Smoothing models, and it is a well-known fact that some exponential smoothing methods and some ARIMA models are equivalent. There are however many ARIMA models with no Exponential Smoothing counterparts and vice versa. All the ARIMA orders used for the forecasts of the four restaurants has no Exponential Smoothing counterpart and this could explain why the ARIMA models were better capable of handling the data whereas Exponential Smoothing methods might utilize a "second-best" option. Another reason may also be that some of the restaurants contain missing data for some days of the year. Because ARIMA models are capable of incorporating external regressor variables, the ARIMA function in *R* automatically detects the missing values as a dummy variable and then removes it's effect from the data. The TBATS model does not allow for external regressors and these values had to be estimated or forecasted instead, which might not have been accurate enough and additionally affect the forecasts negatively. These values are however only a small fraction of the total amount of data, except for Wimpy Bloemfontein, which is closed on Sundays, when fitting the MS-TBATS model for this restaurant the values on Sundays were kept at zero, as it is a reoccurring value every seven days throughout the year and should, therefore, be incorporated into the weekly seasonality.

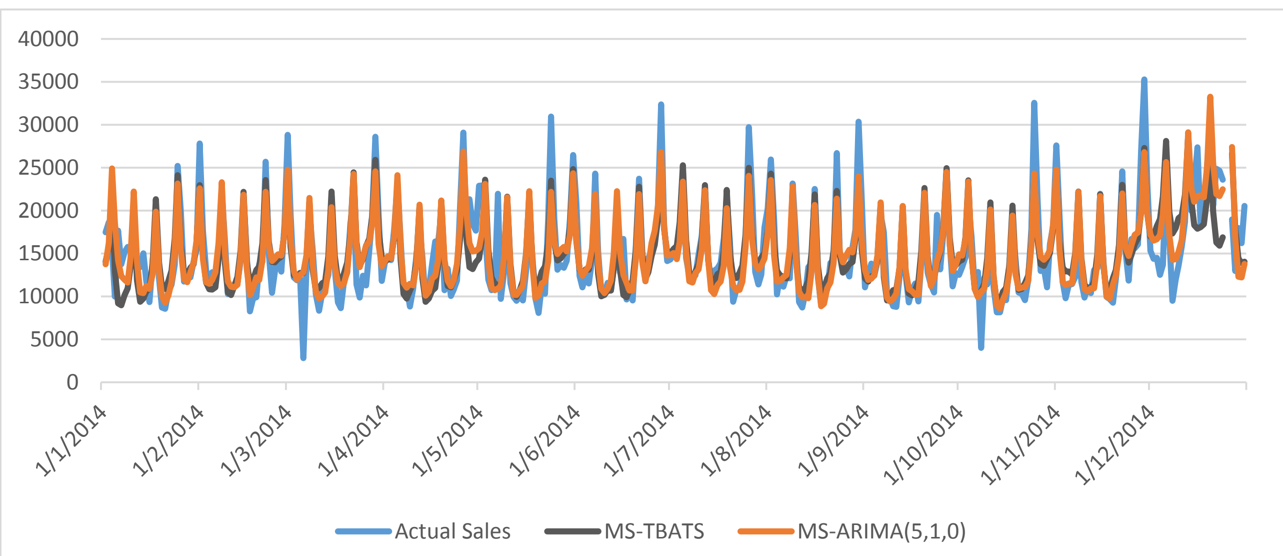

Figure 6-6 Multiple-seasonal ARIMA and TBATS forecasts for Wimpy Roodepoort

When comparing the forecasts of the MS-ARIMA(5,1,0) model with those of the MS-TBATS model for Wimpy Roodepoort displayed by ***Figure 6-6***, we see that they are very similar, with both capturing the multiple seasonalities present. The MS-ARIMA model does, however, fit the troughs and peaks due to the weekly seasonality better.

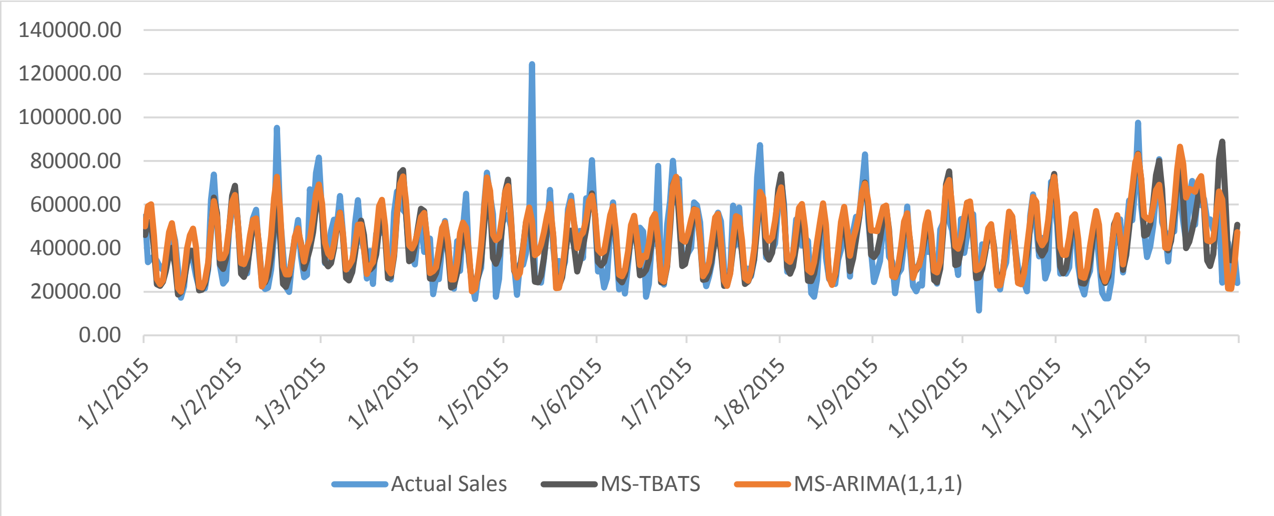

Figure 6-7 Multiple-Seasonal ARIMA and TBATS forecasts for Copper Canyon Spur

As can be seen in ***Figure 6-7,*** the MS-TBATS model captured the seasonal pattern of 2015 for Copper Canyon Spur more accurately than the MS-ARIMA model did. ***Table 6-5*** displays the comparison of the MAPE produced by the MS-ARIMA model with that of the MS-TBATS model for forecasts over three periods, each a year in length, stretching over approximately three years from January 2014 to August 2016. We see how the MS-ARIMA model produces the most accurate forecasts for the period January 2014 to December 2014, but thereafter the forecasting accuracy of the MS-TBATS model starts outperforming that of the MS-ARIMA model for the subsequent periods. We also observe how forecasting accuracy deteriorates through these three periods for both models.

Table 6-5 Comparison of the MAPE produced by the MS-ARIMA and MS-TBATS models for Copper Canyon Spur over three different periods

Forecasted Period	MS-ARIMA	MS-TBATS
Jan 2014 – Dec 2014	17.99%	18.39%
Jan 2015 – Dec 2015	22.83%	19.42%
Sep 2015 – Aug 2016	24.12%	22.74%

The main advantage that MS-TBATS models have over MS-ARIMA models is that the seasonal patterns are allowed to change slowly over time, while for the MS-ARIMA models the seasonal patterns are required to remain constant. This is usually not an issue as seasonality generally remains remarkably constant. When we investigate the plots of daily sales of approximately four years from January 2013 until August 2016 for Copper Canyon Spur displayed by ***Figure 6-8***, ***Figure 6-9***, ***Figure 6-10***, and ***Figure 6-11*** on the next page, we can see that the seasonal pattern does change slightly from 2013 to 2014 with the variance slightly increasing and the data becoming more volatile. We can also observe changes in seasonal patterns from 2014 to 2015, as well as from 2015 to 2016. It may, therefore, be possible that the MS-TBATS model incorporated this slight change in seasonality (observed during the later periods of the data used to fit the models) into the forecasts while the MS-ARIMA model kept the initial seasonal pattern (observed during the earlier periods of the data used to fit the models) constant in producing forecasts. Subsequently, the MS-TBATS model captured the most recent seasonal patterns more accurately than the MS-ARIMA model did.

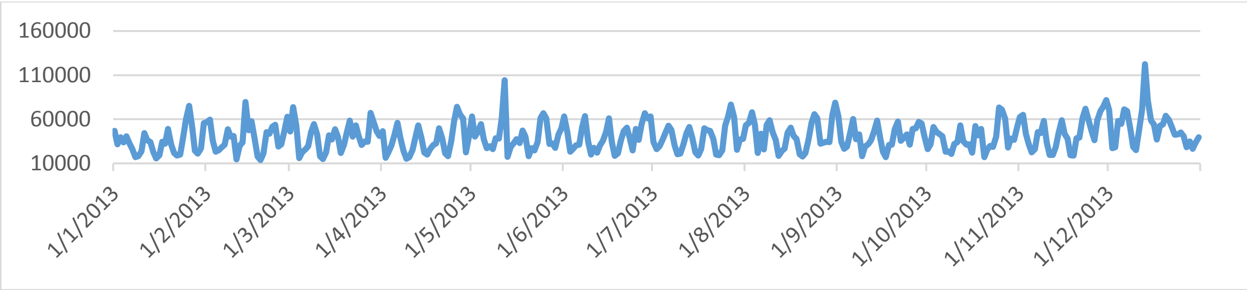

Figure 6-8 Copper Canyon Spur daily sales for 2013

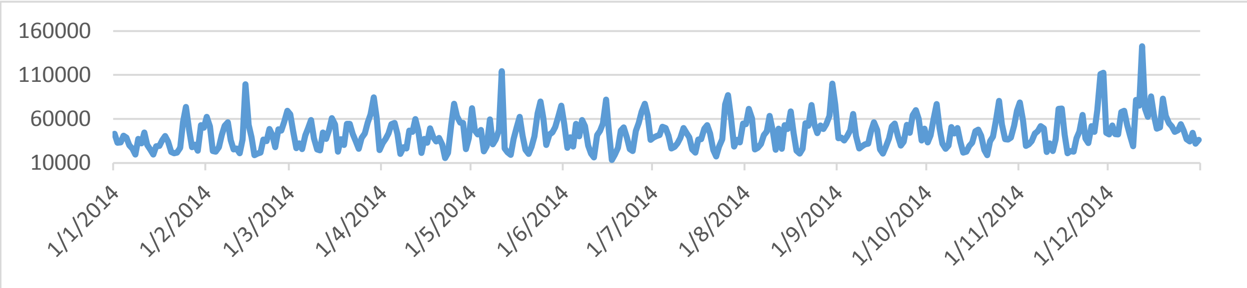

Figure 6-9 Copper Canyon Spur daily sales for 2014

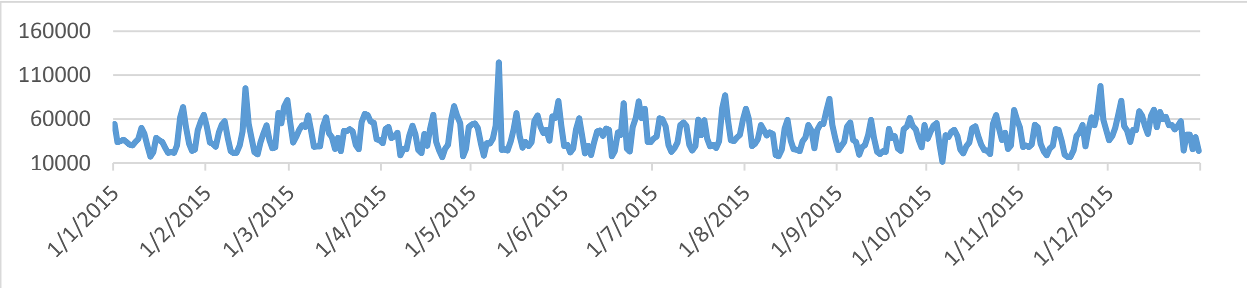

Figure 6-10 Copper Canyon Spur daily sales for 2015

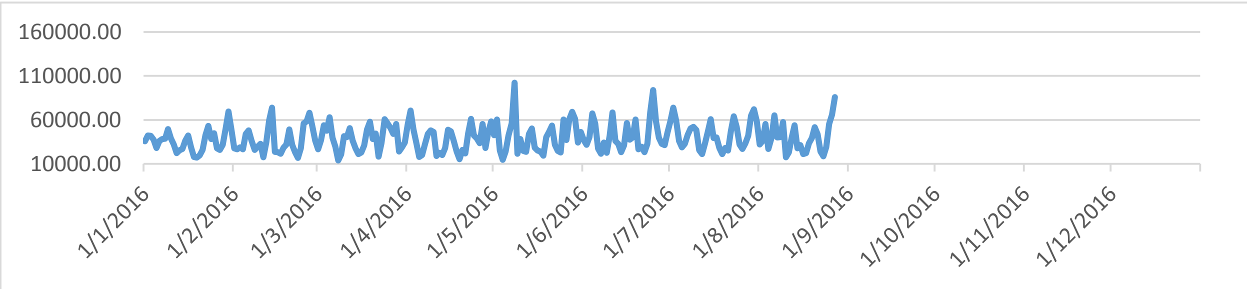

Figure 6-11 Copper Canyon Spur daily sales for 2016

6.6 Multiple-Seasonal ARIMA model with External Regressors

As can be seen in the plots of the forecasts produced in the previous sections, the models are now capable of capturing the trends and multiple seasonalities in the data, but there are still some outliers present that are not forecasted accurately. An example of this can be seen in ***Figure 6-10*** between 1/5/2015 and 1/6/2015 where there is a significant spike, producing a daily sales value of more than R120 000, the highest daily sales value recorded for the year. This spike can also be seen occurring in 2013, 2014, and 2016, displayed by ***Figure 6-8***, ***Figure 6-9***, and ***Figure 6-11***, each time during the month of May. With further investigation, it was discovered that this spike occurred each year on *Mother's Day*. Other events and public holidays also proved to have an effect on the data, such as *Valentine's Day*, *Human Rights Day*, *Youth Day*, *Reconciliation Day* and more.

The advantage of ARIMA models is that they can include external regressor variables and measure their effect on the data after the seasonal and trend components has been extracted and then forecast the combined effect of the variable along with the trend and seasonal components. This can be done by using a dummy variable that takes a value of one on certain days of the year and a value of zero on all other days of the year. The dummy variable takes the form of a vector that is the same length as the data being used to fit the model. Multiple dummy variables can also be combined into a Matrix with the multiple vectors of these variables arranged into rows and the columns of the matrix corresponding to the length of the data being used in fitting the model.

This was done for all four restaurants. Only public holidays that occurs during weekdays were used as the public holidays falling on weekends did not prove to have an effect on the sales. For Copper Canyon Spur non-public holidays which are special events such as *Mother's Day*, *Valentine's Day,* and *Father's Day,* were also used. Other events used for all the restaurants were *Human Rights Day, Good Friday, Family Day, Freedom Day, Workers Day, Youth Day, Woman's Day*, *Heritage Day*, *Reconciliation Day*, *Election Day* (if there was an election day in the sales period used), as well as normal public holidays. Additionally for Copper Canyon Spur days before and after a public holiday also proved to have an effect on sales. Also included as external regressors for Wimpy Midrand and Wimpy Roodepoort is days where load shedding occurred, as this has proven to produce substantially lower sales on these specific days. This is of course with the assumption that information on load shedding schedules will be made available beforehand in order to incorporate this into the forecasts. School holidays were included only for Wimpy Roodepoort.

Table 6-6: MAPE for MS-ARIMA and MS-ARIMAX forecasts for all restaurants

	CCS	**WM**	**WR**	**WB**
MS-ARIMA	22.83%	13.71%	13.57%	21.66%
MS-ARIMAX	21.17%	11.86%	11.38%	20.95%

The results from the forecasts produced with the Multiple-Seasonal ARIMA with External Regressors model (hereafter referred to as MS-ARIMAX model) are displayed in ***Table 6-6*** along with the results from the MS-ARIMA model for comparison.

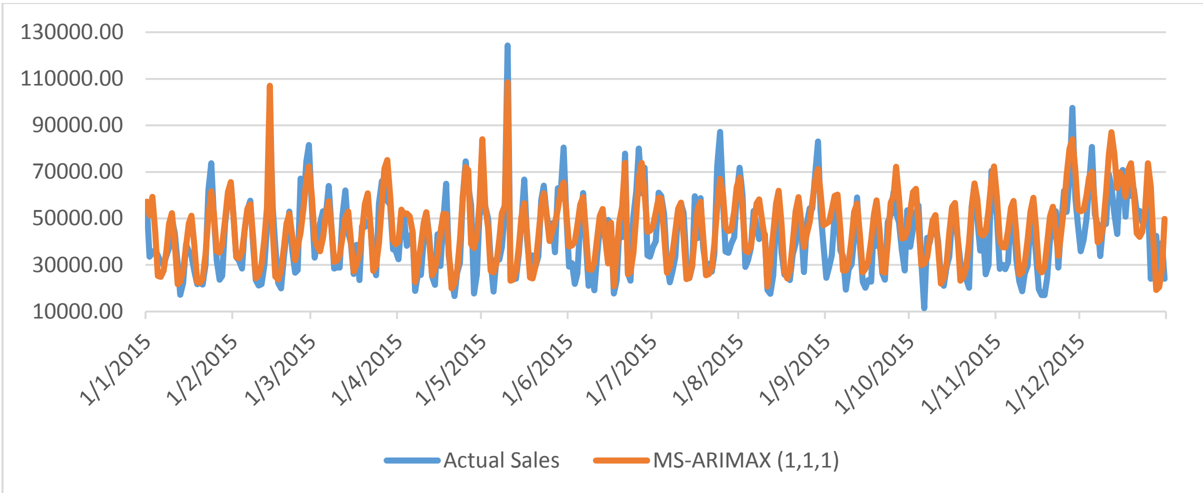

Figure 6-12: MS-ARIMAX sales forecasts for Copper Canyon Spur

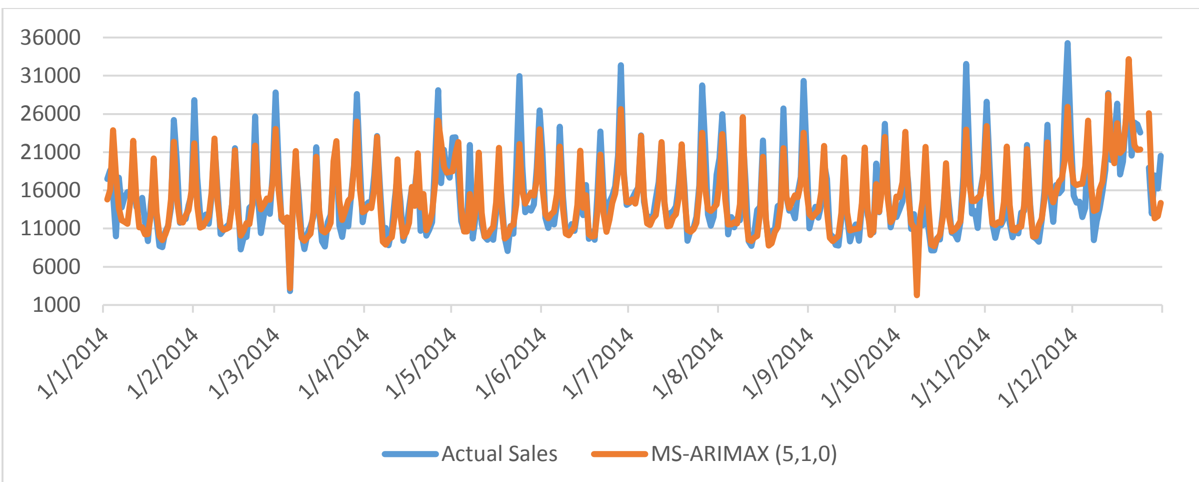

Figure 6-13: MS-ARIMAX sales forecasts for Wimpy Midrand

As can be seen from the results in ***Table 6-6***, MS-ARIMAX models improved forecasting accuracy for Copper Canyon Spur by 7.27%, for Wimpy Midrand by 13.49%, for Wimpy Roodepoort by 12.82% and for Wimpy Bloemfontein by 3.28% in comparison with the MS-ARIMA models. As can be seen in ***Figure 6-12***, the MS-ARIMAX model now captures the effect of *Valentine's Day* observed between 1/2/2015 and 1/3/2015 as well as the effect of *Mother's Day* observed between 1/5/2015 and 1/6/2015 for Copper Canyon Spur forecasts. In ***Figure 6-13*** we see how the forecasts now incorporate the effect of load shedding observed between 1/3/2014 and 1/4/2014 as well as between 1/10/2014 and 1/11/2014 for Wimpy Midrand.

7 Selecting the best model/models

We now compare the results obtained from the previous sections in order to determine which model or combination of models will be best suited to produce forecasts for restaurants with similar operating environments. The results are summarized in ***Table 7-1***

Table 7-1: Comparison of MAPE for all the models tested

	CCS (2014)	CCS (2015)	CCS (2015-2016)	WM	WR	WB
Holt-Winters	29.47%	27.32%	43.48%	19.63%	23.94%	31.48%
MS-TBATS	18.39%	19.42%	22.74%	14.07%	15.24%	23.33%
MS-ARIMA	17.99%	22.83%	24.12%	13.71%	13.57%	21.66%
MS-ARIMAX	16.01%	21.17%	21.76%	11.86%	11.38%	20.95%

As can be seen in from the results displayed in the table, the MS-ARIMAX and MS-TBATS models are identified as the best models to be considered in the selection of models to be used by the Expert System. Evidently, the MS-ARIMAX model is the most accurate and reliable forecasting model as for five out of six cases it produced better forecasts than the MS-TBATS model with most of these substantially better than those of the MS-TBATS model (15.07%, 25.33%, 10.2%, better forecasting accuracy for WM, WR, WB, respectively and 12.94% for CCS in 2014) whereas when the MS-TBATS model outperformed the MS-ARIMAX model it only improved forecasting accuracy by 8.27%.

However, seeing as the MS-TBATS is still capable of outperforming the MS-ARIMAX model in some cases, even if it does not incorporate the factors of public holidays and special events, it would be wise to include this model in the final selection of models to be incorporated into the Expert System. The effects of public holidays are essential to be included in the forecasts of daily sales, because as we have seen for Copper Canyon Spur, if a model forecasts a value of approximately R60 000 for *Mother's Day* and the resulting sales turns out to be more than double the forecasted value, this could be a very difficult situation for the restaurant which they would not at all be prepared for. To incorporate these days into the MS-TBATS forecasts, all the values for 2015 where a Public Holiday factor was included in the MS-ARIMAX model where extracted and inserted on the same days for the MS-TBATS model, replacing these values of the MS-TBATS model. This lowered the MAPE of the forecasts from the original MS-TBATS model from 19.42% to 18.05%, a 7.05% improvement. The resulting forecasts can be seen in ***Figure 7-1.***

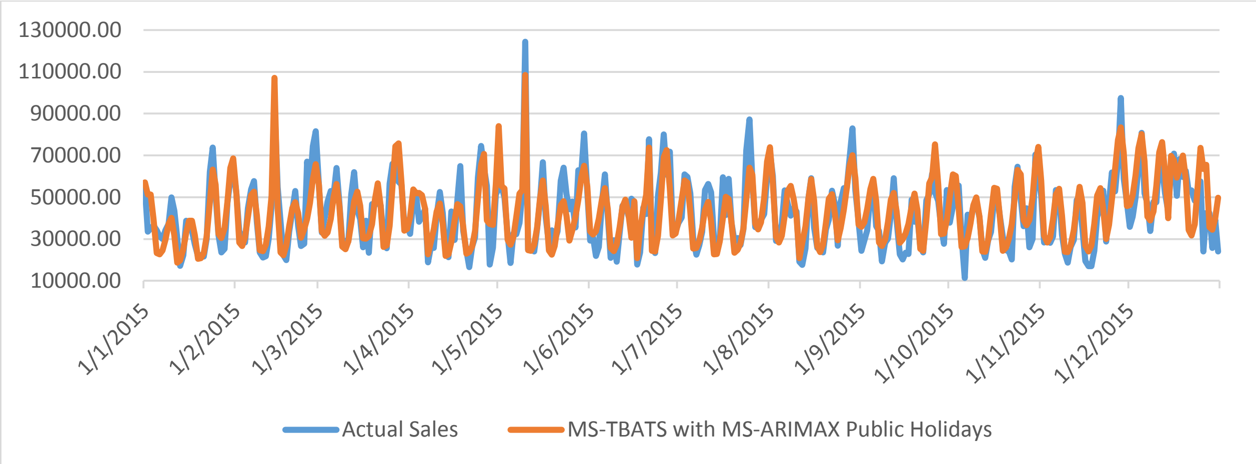

Figure 7-1: MS-TBATS sales forecasts with MS-ARIMAX public holidays included

8 Proposed implementation

As we have now identified the best models to be incorporated in the expert system as the MS-ARIMA and MS-TBATS models we can now discuss how this will be implemented. The process flow displayed by ***Figure 8-1*** on the next page illustrate the methodology and decision process that the Expert System will follow in order to produce the most accurate and reliable forecasts.

Most of the functions that are included are already automated functions in software packages such as *R* and other forecasting tools and it will, therefore, be possible to automate the entire process. It should be noted that prior to entering data into the expert system, the data should first be inspected to ensure that it is up to the required standard. Long periods with missing values will affect the forecasts negatively and should, therefore, be estimated or forecasted with earlier data before it can be used in producing forecasts.

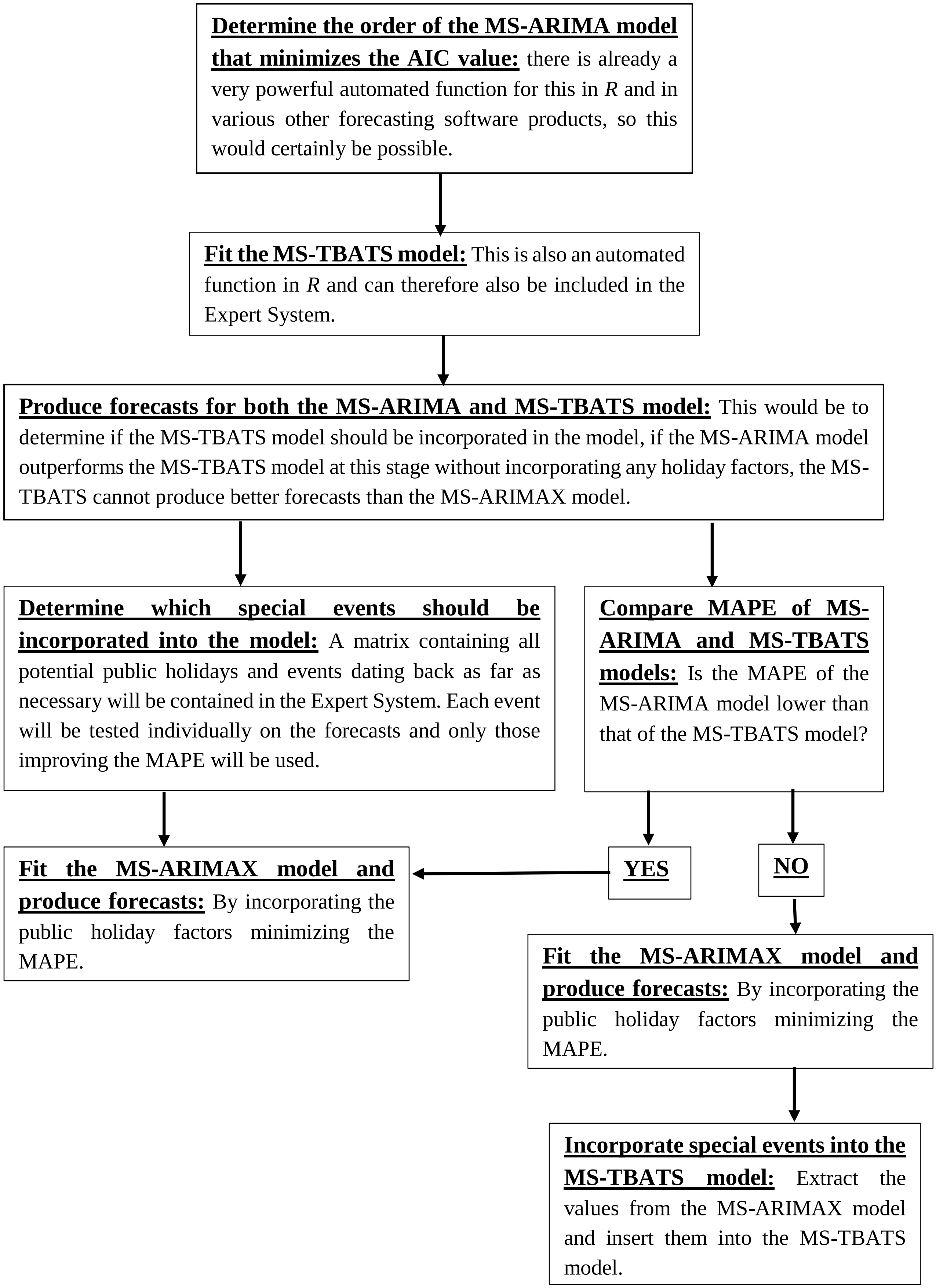

Figure 8-1 Process flow of the methodology that the Expert System will follow

9 Conclusion and Recommendations

This project was started with the aim of determining if it would be possible to develop a forecasting tool capable of forecasting daily sales for different restaurants with similar operating environments. After considering all possible methods that were identified, eliminating methods based on the data analysis and testing the remaining methods, it can be concluded that it is indeed possible to develop a forecasting tool capable of producing accurate and reliable forecasts for different restaurants. Throughout the data analysis and model testing phases, we observed how the data for all four restaurants displayed remarkably similar traits in terms of seasonal patterns and how they reacted to the different methods that were tested.

It should be noted that the forecasts produced are only as good as the data that are used. Forecasting accuracy depends on various factors and as can be seen from the results, equally accurate forecasts for all four restaurants are not possible. Forecasting accuracy for Copper Canyon Spur decreased significantly from 2014 to 2016. The main reason for this is due to the fact that the restaurant changed ownership in 2015, this should logically have an effect on sales patterns as different management principles and promotion strategies are now being used. Another factor is the fact that the previous owner held various events at the restaurant, such as birthday parties, corporate events, year-end functions etc. This has a substantial effect on sales for these specific days, causing sales patterns to become more volatile without an observable trend for these outliers. Unfortunately, no data were, however, available indicating on which days these events were as the previous owner did not keep a record of the dates of these events. If the data were, however, available, these days could be incorporated as external regressors just as with public holidays. It is therefore recommended that restaurant owners planning on using this system in the future keep a detailed record of all events and observations made that may seem to have an effect on sales, in order to incorporate this into future forecasts.

References

Alysha M. De Livera, Rob J. Hyndman & Ralph D. Snyder (2011) *Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing*, Journal of the American Statistical Association.

Armstrong, J. (2001). *Principles of forecasting*. Boston, MA: Kluwer Academic.

Arunraj, N. and Ahrens, D. (2015*). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting*. International Journal of Production Economics, 170, pp.321-335.

Choi, J. (1999). *The restaurant industry: business cycles, strategic financial practices, economic indicators, and forecasting*. Ph.D. Virginia Polytechnic Institute and State University.

Hyndman, R. and Athanasopoulos, G. (2013). *Forecasting: Principles and Practice*.

Kokkinou, A. (2013). *Forecasting for the Start-Up Restaurant Owner*. Journal of Foodservice Business Research, 16(1), pp.101-112.

Lawrence, K., Klimberg, R. and Lawrence, S. (2009). *Fundamentals of forecasting using Excel*. New York, N.Y.: Industrial Press.

Lee, D. and Kim, J. (2015). *Pizza sales forecasting using big data analysis*. Information, 18(5(A), pp.1577-1584.

Rahmlow, H. and R. Klimberg. '*Forecasting Practices of MBAs*." Advances in Business and Management Forecasting 3 (2002): 113-123.

Robjhyndman.com. (2016). *Forecasting with long seasonal periods*. [online] Available at: http://robjhyndman.com/hyndsight/longseasonality/ [Accessed 30 Aug. 2016].

Sanchez, N. (1994). *Applying expert systems technology to the implementation of a forecasting model in foodservice*. Undergraduate. Kansas State University.

Spurcorporation.com. (2016). *Group profile / Spur Corporation*. [online] Available at: http://www.spurcorporation.com/business-profile/group-profile/ [Accessed 21 Mar. 2016].

Winston, W. (2004). *Introduction to probability models*. Belmont, CA: Brooks/Cole-Thomson Learning.

10 Appendices

10.1 Appendix A: Signed Industry Sponsorship Form

Department of Industrial & Systems Engineering
Final Year Projects
Identification and Responsibility of Project Sponsors

Final Year Projects may be published by the University of Pretoria on *UPSpace* and may thus be freely available on the Internet. These publications portray the quality of education at the University, but they have the potential of exposing sensitive company information. It is important that both students and company representatives or sponsors are aware of such implications.

Key responsibilities of Project Sponsors:

A project sponsor is the key contact person within the company. This person should thus be able to provide guidance to the student throughout the project. The sponsor is also very likely to gain from the success of the project. The project sponsor has the following important responsibilities:

1. Confirm his/her role as project sponsor, duly authorised by the company. Multiple sponsors can be appointed, but this is not advised. The duly completed form will considered as acceptance of sponsor role.
2. Review and approve the Project Proposal, ensuring that it clearly defines the problem to be investigated by the student and that the project aim, scope, deliverables and approach is acceptable from the company's perspective.
3. Review the Final Project Report (delivered during the second semester), ensuring that information is accur te and that the solution addresses the problems and/or design requirements of the defined project.
4. Acknowledges the intended publication of the Project Report on UP Space.
5. Ensures that any sensitive, confidential information or intellectual property of the company is not disclosed in the Final Project Report.

Project Sponsor Details:

Company:	**Copper Canyon Spur**
Project Description:	**Daily sales forecasting in foodservice: Developing a model for application in an expert system**
Student Name:	**H.F Versluis**
Student number:	**11300389**
Student Signature:	
Sponsor Name:	CASPER VERSLUIS.
Designation:	
E-mail:	CASPER VERSLUIS@GMAIL.COM

Tel No:	
Cell No:	0725030228
Fax No:	
Sponsor Signature:	