



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Forecasting Chargeable Hours at a Consulting Engineering Firm

Applicability of Classical Decomposition, Holt-Winters & ARIMA

Bachelor's Thesis in
Industrial and Financial Management

School of Business, Economics and Law
at University of Gothenburg
Spring term 2012

Tutor: Taylan Mavruk

Authors:
Johan Agneman
Roger Lindqvist

Date of birth:
1985-06-09
1985-07-05

Acknowledgements

We would like to thank everyone that has supported us during the writing of this thesis. Without the support from others it would not have been possible to reach the full potential of this thesis.

First and foremost, our deepest gratitude goes to our supervisor Taylan Mavruk, assistant professor at the University of Gothenburg, for valuable guidance and advice. His willingness to motivate us contributed greatly to this thesis.

We specially thank CEO David Hellström, CFO Jan Gustavsson, and Business Unit Director Mathias Thorsson at Reinertsen. We are honestly grateful for their enthusiasm on helping us with this work.

Also, we extend our thanks to Per Lidström, associate professor at Lund University, for being a valuable sounding board during the final completion of this thesis.

Finally, we would like to thank Elizabeth Wedenberg and Linda Agneman for their language support.

Johan Agneman & Roger Lindqvist

29th of May 2012, Göteborg

Abstract

Reinertsen, a Swedish consulting engineering firm, is dissatisfied with the accuracy of its qualitative forecast of chargeable hours. This thesis investigates whether classical decomposition, Holt-Winters, or ARIMA can perform more accurate forecasts of chargeable hours than the qualitative method currently used at Reinertsen. This thesis also attempts to explain why or why not these forecasting methods improve the forecasting accuracy at Reinertsen. The purpose of this thesis is twofold: (1) to identify a suitable manpower forecasting method for Reinertsen; and (2) to contribute to previous literature on forecasting by further assessing the performance and the applicability of the chosen forecasting methods.

The data applied was monthly numbers of chargeable hours which covered the period between 2007 and 2011. The first 48 monthly observations were used to generate the forecasts while the remaining 12 monthly observations were used to evaluate the forecasts. The data contains trend and strong monthly fluctuations.

The results indicate that ARIMA and classical decomposition are inappropriate forecasting methods to forecast chargeable hours at Reinertsen. The relatively poor performance of classical decomposition and ARIMA is believed to be attributable to these methods inability to forecast varying fluctuations. The results also show that Holt-Winters yield the most accurate forecasts amongst the evaluated forecasting methods. The forecasted time series fluctuates much and the Holt-Winters method, which focuses on recent observations, might be better suited to capture these fluctuations. Consequently, the Holt-Winters method has the potential to improve the forecasting of chargeable hours at Reinertsen.

Table of contents

1. Introduction	1
1.1 Problem discussion.....	2
1.2 Research questions	3
1.3 Purpose	3
1.4 Limitations.....	3
2. Theory and literature review.....	3
3. Method and data	7
3.1 Research process.....	7
3.2 Research design.....	7
3.3 Reliability and validity of research.....	8
3.4 Data	8
3.4.1 Data analysis.....	9
3.5 Forecasting methods.....	10
3.5.1 Classical decomposition.....	10
3.5.2 Holt-Winters	11
3.5.3 ARIMA.....	12
3.6 Accuracy measures.....	15
4. Empirical results.....	15
4.1 Characteristics of data.....	15
4.2 ARIMA parameters	18
4.3 Forecasting performance	21
5. Analysis	24
6. Conclusion.....	27
Bibliography	i
Appendix 1: Forecasted values of chargeable hours	iv
Appendix 2: STATA commands.....	v

List of tables

1. Descriptive statistics.....	16
2. Seasonal indices	17
3. Runs test for binary randomness	18
4. Augmented Dickey-Fuller test of chargeable hours	18
5. Augmented Dickey-Fuller test (ADF) of first order differencing of chargeable hours	19
6. AC, PAC and Q of first order differencing of chargeable hours	20
7. AC, PAC and Q for residuals of ARIMA(0,1,10), ARIMA(0,1,11) and ARIMA(0,1,12)...	21
8. MAPE.....	23
9. MSE	23
10. Mean error	24
11. Forecasted values.....	iv
12. STATA commands.....	v

List of figures

1a: Chargeable hours over time	16
1b: Monthly fluctuation of chargeable hours	17
2: First order differencing of chargeable hours over time	19
3: AC and PAC of first order differencing of chargeable hours	20
4: AC and PAC for residuals of ARIMA(0,1,11)	21
5a: Current forecasting method	22
5b: Multiplicative decomposition	22
5c: Additive decomposition	22
5d: Multiplicative Holt-Winters	22
5e: Additive Holt-Winters	22
5f: ARIMA	22

1. Introduction

Rational decision making is a fundamental term in every organization. Berridge (2003) suggests that a rational decision is a decision that maximizes utility, and Simon (1979) states that the task of rational decision making is to select efficient options that are directed towards organizational goals. However, much research has revealed irrational tendencies in how people take decisions. Ito, Pynadath and Marsella (2010) concluded that human beliefs influence and thereby bias decisions, and Makridakis (1981) proposes that humans have difficulties to process all available information necessary for making rational decisions.

For many organizations, including firms, agencies and governments, forecasting is a tool that support the decision making process. Forecasting includes predicting the future, and the outcome affects decision making on economic policies, investment strategies, manpower planning and several other issues organizations deal with on daily, weekly, monthly and yearly basis (Pollack-Johnson, 1995). Yet, dealing with the future involves much uncertainty, and forecasting does not yield perfect prediction of the future. However, the purpose of forecasting is not to eliminate uncertainty but to reduce it, thus enable decision makers to take more rational decisions (Makridakis, 1981).

The literature on managerial forecasting is extensive and many models are available. Empirical studies have shown that the performance of different forecasting models varies with the characteristics of the data available (Meade, 2000; Hyndman and Kostenko, 2007), the time horizon to be forecasted (Bechet and Maki, 2002; Gooijer and Hyndman, 2006), and the area of application (Clarke and Wheelwright, 1976; Mahmoud, 1984). Consequently, the choice of forecasting model is crucial to generate accurate forecasts that reduce uncertainty and enable managers to take informed decisions.

A forecasting method could be qualitative or quantitative. Qualitative refers to human involvements while quantitative refers to systematic procedures. Quantitative methods include time series methods, which capture historical data to predict the future, and explanatory methods, which involve identifying independent variables to predict future movements in dependent variables (Makridakis and Wheelwright, 1980). Time series forecasting is a common forecasting practice when empirical data is available. However, Gooijer and Hyndman (2006), state that there is no consensus among researcher which method to use. Each situation is unique and the choice of forecasting method depends much on the characteristics of data available (Meade, 2000). Thus, an interesting area of research is to investigate the applicability of time series forecasting in a situation where the method is not currently used.

1.1 Problem discussion

The managers of Reinertsen, a Swedish consulting engineering firm, are dissatisfied with the accuracy of their manpower forecasts, leading to a utilization rate¹ that is considered too low. Indeed, good manpower forecasting can lead to a profitable advantage of every business (Bassett, 1973) and it is a key activity to be successful in today's business environment (Bechet and Maki, 2002). According to Bassett (1973), good manpower planning begins with a sales forecast, and proceeds directly to estimation of manpower needed to produce in accordance with the forecast. Reinertsen sells consultant hours, and more accurate forecasting of sold hours, or chargeable hours, could enable Reinertsen to improve its manpower planning. This in turn can lead to an improvement of utilization rate.

Reinertsen is divided into three different business units: Infrastructure; Energy and Industry; and Oil and Gas. Each business unit is then divided into more specialized departments, which in turn, consist of small specialized engineering groups. Reinertsen has been structured similarly since January 2008. Currently, a forecast for each month, based on qualitative estimations, is made on a yearly basis. The qualitative estimations, which are made for each business area separately, are reviewed to ensure that they are realistic. The forecast of each business area is then combined at company level to reach an overall forecast for each month the next coming year.

Makridakis and Wheelwright (1980) argue that a qualitative forecasting method depends much on the skills and experience of the people involved in the decision procedure. Makridakis (1981) also suggests that a major drawback of qualitative methods is the inconsistency associated with human involvement. This irrationality accounts for a large proportion of human forecasting error. The inconsistency that is due to human involvements can, however, be reduced by implementing a quantitative forecasting method. Consequently, it is interesting to see whether time series forecasting can produce more accurate forecasts of chargeable hours than the qualitative forecasting currently employed at Reinertsen.

According to Gooijer and Hyndman (2006), three commonly used times series forecasting methods are classical decomposition², Holt-Winters³ and ARIMA⁴. These methods vary in complexity where the former is considered the simplest method and the latter the most complex method. Each of these methods will be applied to the historical data of chargeable hours provided by Reinertsen to investigate whether they can minimize the firm's forecasting error. The result will reveal the appropriateness of each method to forecast chargeable hours at Reinertsen as well as provide explanations to why or why not these methods are applicable to forecast time series that is similar to that of chargeable hours at Reinertsen.

¹ Utilization rate is defined as the ratio of chargeable hours to the total number of manpower hours

² An approach that breaks down time series into several factors including trend factor and seasonal factor

³ An exponential smoothing method designed to handle trend and seasonality

⁴ A statistic approach that predicts the future by examining the characteristics of the data.

1.2 Research questions

This thesis will target the following research questions:

- Can classical decomposition, Holt-Winters, or ARIMA perform more accurate forecasts of chargeable hours than the current forecasting method used at Reinertsen?
- Why or why not are these methods appropriate to forecast time series that is similar to that of chargeable hours at Reinertsen?

1.3 Purpose

The purpose of this thesis is twofold: (1) to identify a suitable manpower forecasting method for Reinertsen; and (2) to contribute to previous literature on forecasting by further assessing the performance and the applicability of three commonly used time series forecasting methods.

1.4 Limitations

This thesis will focus on three different time series forecasting methods; classical decomposition, Holt-Winters, and ARIMA. These methods are selected because they are widely used among forecasters when historical data contains both trend and seasonality. Each method can, however, represent more than one model; both classical decomposition and Holt-Winters can represent an additive and a multiplicative model, respectively, and ARIMA can be modeled into several different models depending on the characteristics of the data available. Each method will be fitted to one time series and evaluated and compared to one qualitative method. The time series and the qualitative method are provided by Reinertsen. The accuracy of each forecast will be assessed with three different accuracy measures; mean error, mean square error, and mean absolute percentage error. These accuracy measures are chosen because they are frequently used among researchers and because they are good complement to each other.

2. Theory and literature review

Forecasting has been used to predict future outcomes since ancient Egypt but it was not until the Keynesian revolution that more systematic models were developed. Scandinavian countries began reporting official macroeconomic forecasts soon after the second world war and this was followed by other developed countries in the 1950's and in the 1960's (Hawkins, 2005). Business forecasting was long considered an easy practice. The stable growth after the second world war made forecasting straightforward and future outcomes were assumed to follow established trends. However, several macroeconomic crises in the 1970's highlighted the need for more sophisticated models (Makridakis, 1981). Also, global economy, international competition, and the changing business environment during

the past decades have impelled continuous improvements in the area of forecasting. The International Institute of Forecaster was established in 1982 and forecasting has since then been studied extensively, and, as a result, numerous of methods have been developed. Still, there are areas of forecasting where research has been limited (Gooijer and Hyndman, 2006).

Forecasting methods are divided into two main categories: qualitative methods and quantitative methods. Qualitative methods include judgmental predictions of sales forces, executives, experts or panels. Also surveys and iterative processes are included in this category (Pollack-Johnson, 1995). As mention earlier, Makridakis (1981) states that a major drawback of qualitative methods is the inconsistency associated with human decision-making. This irrationality accounts for a large proportion of human forecasting error. However, Armstrong (2006) argues that qualitative methods can be improved by more standardized procedures; an example is the Delphi-method where experts independently adjust each other' forecasts iteratively until a satisfactory forecast is obtained. Also, Bunn (1989), Pollack-Johnson (1995), and Armstrong (2006) suggest that qualitative forecasting methods can yield better forecasts if combined with quantitative forecasting methods. In general, qualitative methods are useful when empirical data is difficult to obtain and when independent shocks impose permanent structural changes (Pollack-Johnson, 1995). Quantitative methods include time series methods and explanatory methods. Time series methods capture historical data to predict the future while explanatory methods identify independent variables to predict future movements in dependent variables (Clarke and Wheelwright, 1976). Both time series methods and explanatory methods are considered objective, although, Pollack-Johnson (1995) concludes that also these methods involve human decision-making as to the choice of model. Also, according to Armstrong (2006), a problem with quantitative business forecasting methods is to obtain a sufficient number of observations that can statistically qualify the methods.

Although several empirical studies have attempted to examine the relative performance of quantitative forecasting and qualitative forecasting no consensus has been reached among researchers. Mahmoud (1984) concludes that many studies have indicated that forecasts based on qualitative methods have provided less accurate forecasts than quantitative methods. However, Makridakis and Wheelwright (1980) states that is difficult to compare the relative performance of qualitative forecasting because these methods are not standardized and depend much on the skills and experience of the forecasters. However, other studies indicate that qualitative, or at least partly qualitative methods, produce comparable or even better forecasts than quantitative methods do. For instance, Pollack-Johnson (1995) argues that qualitative forecasting in many situations outperforms quantitative forecasting and that a combination of the two methods is efficient. Meade (2000) argues, instead, that there appears to be no single model that yields the most accurate forecast in every situation; the performance varies across studies and depends on the characteristics of the data available.

Time series forecasting is a common forecasting practice when historical data is available. The practice is based on a sequence of evenly spaced data points which are extrapolated to predict future outcomes (Pollack-Johnson, 1995). Time series can be either stationary or non-stationary in nature.

Stationarity refers to a constant trend; that is, the time series has a constant mean and variance. Non-stationarity refers to longer upward, or downward, movements in the data; that is, the time series is trending. Whether the data contain a trend or not is an important factor to consider when selecting time series forecasting model (Box, Jenkins and Reinsel, 2011). A quick way to get an overview of the trend is to plot the data against time (Kalekar, 2004). Another method is the augmented Dickey-Fuller (ADF) test, which is a test that statistically examines whether the time series is stationary or non-stationary (Wong, Chan and Chiang, 2011). Time series can follow a discernible pattern or a random pattern. A random pattern is due to independent shocks that impose unpredictable variation in the data, and much randomness can skew time series forecasts. However, some methods capture randomness in data better than others (Hyndman and Kostenko, 2007). Consequently, random variance must be identified to enable an appropriate selection of forecasting method. A statistical test to check the amount of random variation in the data, suggested by Babbage (2004), is the Runs tests for binary randomness. Time series data can also contain seasonality, which refers to annual, recurrent, fluctuations in the data. Identification of seasonality facilitates selection of a suitable forecasting method, and a common method to detect seasonality includes computation of seasonal indices (Kalekar, 2004).

According to Gooijer and Hyndman (2006), the oldest methods to manage seasonality in the data is classical decomposition, which is an approach that breaks down the time series into several factors including trend factor and seasonal factor. Makridakis and Wheelwright (1980) describe how classical decomposition could be performed either as a multiplicative or an additive model. The multiplicative model yields more accurate forecasts if the seasonal fluctuations follow the trend; that is, the magnitude of the fluctuation is proportional to the underlying trend. On the contrary, the additive model yields more accurate forecasts if the seasonal fluctuations are independent of the trend; that is, the magnitude of the fluctuations is constant (Kalekar, 2004). In the classical decomposition method, each historical observation is equally important; that is, each observation is equally weighted when generating the forecast. Some researchers argue that this leads to an underlying inertia of the method, thus making it inappropriate for most time series. Other researchers argue that classical decomposition is a tool for data analysis rather than a forecasting method (Makridakis et al, 1998).

A method designed to handle both trend and seasonality is Holt-Winters triple exponential smoothing method (Gooijer and Hyndman, 2006). Exponential smoothing is a method that repeatedly revises a forecast, where recent observations get more weights and older observations less weights (Kalekar, 2004). Chen (1997), Gooijer and Hyndman (2006), and Gelper, Fried and Croux (2008) claim that Holt-Winters is a useful method when the data shows both trend and seasonality, and Chatfield and Yar (1998), suggest that Holt-Winters is a relatively simple forecasting method that yields good forecasts. However, Chatfield and Yar (1998) also state that a common drawback associated with Holt-Winters is the absence of helpful literature when it comes to some practical issues. For instance, there is no standardized method to generate starting values at the beginning of the time series. Partly as a result, widely different approaches can produce substantially different forecasts for what is apparently the same method (Chatfield and Yar, 1998). Also Holt-Winters can be divided into a multiplicative model and an additive model, and similar

to the classical decomposition method, the relative performance of the multiplicative and the additive Holt-Winters depends on the characteristics of the seasonal fluctuations (Kalekar, 2004).

Another, widely used, method to handle trend and seasonality is the autoregressive-integrated-moving average (ARIMA) method. ARIMA can be modeled by adjusting to specific characteristics of a time series, thus make it applicable to different types of data (Box et al, 2011). The method is considered objective because its forecasts are based only on historical data which is extrapolated into the future. However, the method can be subjected to model selection problems, especially when the characteristics of the data are difficult to interpret (Hyndman, 2001). Gooijer and Hyndman (2006) suggest that ARIMA is a robust method to handle trend and seasonality and Wong et al (2011) state that the method is useful when data follows a discernible pattern. However, Clarke and Wheelwright (1976) argue that the method is very complex and difficult to understand. Armstrong (2006) states that ARIMA has been immensely studied but that there is little evidence that the method improves forecast accuracy. Pollack-Johnson (1995) concludes that ARIMA, despite its complexity, often gives poorly forecasts compared to less complex methods. Also Hyndman (2001) argues that ARIMA many times perform moderate forecasts because of the difficulty to identify the most appropriate ARIMA-model.

The accuracy of forecasting methods is evaluated by accuracy measures. An accuracy measure is the difference between forecasted value and actual value and a low accuracy measure indicates that a suitable forecasting model is used (Mahmoud, 1984). According to Gooijer and Hyndman (2006), a confusing array of accuracy measure techniques are available. Also Mahmoud (1984) states that the main problem with accuracy measures is the absence of a universally accepted measure for evaluating forecast errors, which in turn, makes it hard for the user to select a suitable accuracy measure. However, Gooijer and Hyndman (2006) argue that the variety of accuracy measures is due to different characteristics of the forecast to be evaluated; different accuracy measures are suitable for different types of forecasts. Also, due to the different characteristics of the accuracy measures, Makridakis, Wheelwright and Hyndman (1998) propose that a fair comparison between different forecasting methods should involve more than one accuracy measures. A literature review conducted by Mahmoud (1984), revealed that up to ten different accuracy measures are commonly employed among researchers. Three of the more frequently used measures are the mean error, the absolute measure, mean squared error (MSE), and relative measure, mean absolute percentage error (MAPE). Mean error is easy to calculate and gives a good indication of whether the forecasting method evaluated under-forecast or over-forecast. MSE, which squares the errors, highlight large errors. Consequently, MSE is a good accuracy measures when the user prefer many small errors rather than a few large. However, due to large numbers, the measures can be difficult to interpret. MAPE give equal weight to all errors, and because the measure is easy to interpret, it is the most applied accuracy measure (Makridakis et al, 1998).

3. Method and data

3.1 Research process

The first step of this thesis involved a discussion with managers of Reinertsen. The discussion included a brief presentation of the firm and its current forecasting method as well as a general discussion about the problem. Also empirical data was obtained at this time. The second step involved an extensive literature study. Some meta-studies were reviewed to get an overview of previous research in the area of forecasting before focus was directed towards time series forecasting methods and their underlying theories. The literature study, together with the discussion, formed the research question for this study. Next, the data was analyzed to enable selection of suitable forecasting methods. To detect the characteristics of the data, the data analysis included describing the data and graphing the data against time. Also the choice of STATA⁵ and Excel as analysis software was made in this step. The fourth step involved selection of forecasting methods. The problem discussion and the characteristics of the data available were evaluated against previous literature. Once selected, the forecasting methods' mathematical equations were reviewed, the analysis software were chosen, and eventually, the forecasts were generated. The last step involved identification of reliable accuracy measures. Again, previous literature was consulted to support the choice. As soon as the measures were decided, the accuracy of each time series forecast as well as Reinertsen's current forecast were computed and compared.

3.2 Research design

This thesis examines whether any of the three mentioned time series forecasting methods can provide more accurate forecasts of chargeable hours at Reinertsen than the qualitative forecasting method currently used. The time series forecasts were extrapolated from historical data obtained from Reinertsen. The data was monthly numbers of chargeable hours which covered the period between 2007 and 2011. The first 48 monthly observations were used to generate the forecasts while the remaining 12 monthly observations were used to evaluate the forecasts. Next, the error of each forecast was assessed by three commonly used accuracy measures. The accuracy measures of each forecast were appraised and then compared to the accuracy measures of Reinertsen's current forecast, thus enabling an analysis of the performance of each forecasting method.

The research design is similar to that used by Koehler (1985), Greer and Liao (1986), and Wong et al (2011). These studies compare the relative performance of different quantitative forecasting methods in the food processing industry, the aerospace industry, and the construction industry, respectively. This thesis focuses, however, on the applicability of time series forecasting in the consulting engineering industry. Further, the focus is on whether the selected times series forecasting methods can provide more

⁵ STATA is a statistics software for data analysis, data management, and graphics. STATA commands are presented in Appendix 2.

accurate forecasts at a single firm in the consulting engineering industry rather than for the industry as a whole.

3.3 Reliability and validity of research

To ensure the reliability of the research only literature by reputed researchers has been used. Most literature originates from business journals. Also a few books and working papers have been consulted. The selected forecasting methods are widely employed (e.g Koehler, 1985; Greer and Liao 1986; Heuts and Brockners, 1998; Chatfield and Yar, 1998; Kalekar, 2004; Theodosiou, 2011; Wong et al, 2011) and evaluated (e.g. Pollack-Johnson 1995; Armstrong 2006; Gooijer and Hyndman, 2006). Also, appropriateness of the chosen accuracy measures have been appraised in several studies (e.g. Mahmoud, 1984; Gooijer and Hyndman, 2006). The forecasts are modeled from actual values covering the period January 2007 to December 2010 and validated against actual values covering the period January 2011 to December 2011. Similar research design is used by Koehler (1985), Greer and Liao (1986), and Wong et al (2011). The data used in this thesis is provided by Reinertsen. It originates from operational documents so there is no reason to believe that the provided data is incorrect or skewed.

A limitation of this thesis is that the selected methods are only a few of all available methods. Several other methods, including explanatory methods and qualitative methods, could be used to forecast chargeable hours at Reinertsen. The choice of other methods would probably lead to different results, and consequently different conclusions. A further limitation is that the result of the forecasts is only reliable for this particular moment; the accuracy of each time series forecast deteriorates over time and need continuous revisions and evaluations. Yet another limitation is the absences of a completely reliable procedure to estimate the ARIMA parameters; although a common estimation procedure is used there is no guarantee that the chosen ARIMA model is optimal. Similarly, there is no standardized approach to generate starting values for the Holt-Winters method. As a result, different approaches can produce substantially different forecasts, although the same Holt-Winters model is used.

3.4 Data

The data used in this thesis was provided by Reinertsen. An Excel document, with financial statements, key indicators and current forecast, was obtained, and after a discussion with managers of Reinertsen, where the problem of this thesis was identified, the data observations associated with chargeable hours were sorted out. The sorted data included monthly observations of chargeable hours, dating back to January 2007. However, a re-organization of Reinertsen in January 2008, where two business units became three, skewed the data. Thus, data before January 2008 was inappropriate to use, resulting in only 48 observations for each business unit. Yet, the total number of chargeable hours was correct. Consequently, the data for each business area was merged, resulting in 60 monthly observations of total chargeable hours for the period January 2007 to December 2011. The first 48 monthly observations were used to generate the forecasts, while the remaining 12 monthly observations were used to test the

forecasted values to the actual values. Consequently, the total sample, the model sample, and the validation sample contain 60 observations, 48 observations, and 12 observations, respectively.

3.4.1 Data analysis

To obtain an overview of the data, descriptive statistics, including mean, median, standard deviation, minimum value and maximum value, was computed for the total sample, the model sample and the validation sample, respectively. Also, two line diagrams were plotted; one to identify the relationship between chargeable hours and time and one to identify the seasonal fluctuations. To control the data for seasonality, a seasonal index (S_t) for each month was computed using equation (1):

$$S_t = \frac{X_t}{X_m} \quad (1)$$

where X_t is the actual value in month t ; and X_m is the average value of all months. If the seasonal index is below 1 it indicates a value below average, and if the seasonal index is above 1, it indicates a value above average (Kalekar, 2004).

To test the data for randomness the Runs test for binary randomness was used. First, the mean value (μ) of the sample was calculated. Second, all observations (n) were modified to binary numbers. If the trend and seasonally adjusted observation is less than the mean value (μ) it is assigned with the binary number 0. On the contrary, if the trend and seasonally adjusted observation is larger than the mean value (μ) it is assigned with the binary number 1. The sum of binary number 0 (n_0) and the sum of binary number 1 (n_1) were then calculated, respectively. Third, the number of runs in the sample was computed. A run is defined as a series of increasing, or decreasing, values. The number of changes, from increasing values to decreasing values, or vice versa, is then equal to the number of runs (r). Fourth, the test statistics (Z) was calculated, using the following series of equations:

$$r_E = \frac{2n_0n_1}{n_0 + n_1} + 1 \quad (2)$$

$$s_R = \sqrt{\frac{2n_0n_1(2n_0n_1 - n_0 - n_1)}{(n_0 + n_1)^2(n_0 + n_1 - 1)}} \quad (3)$$

$$Z = \frac{r - r_E}{s_R} \quad (4)$$

where r_E is the expected value of r ; and s_R is the standard deviation. The null hypothesis, that the data contains randomness, could then be rejected if the absolute test statistics exceeded the critical value, that is $|Z| > Z_{1-\alpha/2}$ (Babbage, 2004).

3.5 Forecasting methods

3.5.1 Classical decomposition

Classical decomposition is a simple forecasting method designed to handle both trend and seasonality. It is a three step procedure including (1) calculation of seasonal indices; (2) developing of trend line equation; and (3) generating of forecast (Makridakis and Wheelwright, 1980). Both a multiplicative decomposition forecast and an additive decomposition forecast were computed.

First, to solve for monthly fluctuations, seasonal indices were computed. In the multiplicative decomposition model, a seasonal index, for each month, is computed by averaging all ratios for that month, where the ratio is computed by dividing that month's actual value by its centered moving average value⁶. In the additive decomposition model, a seasonal index, for each month, is calculated by averaging all differences for that month, where the difference is computed by subtracting that month's centered moving average value from its actual value. A centered moving average value is calculated according to equation (5):

$$x_{t,c} = \frac{1}{n_s} \left[0.5x_{t-n_s/2} + x_{t-5} + \dots + x_{t-1} + x + x_{t+1} + \dots + x_{t+5} + 0.5x_{t+n_s/2} \right] \quad (5)$$

where $x_{t,c}$ is the centered moving average in month t ; and n_s is the number of months (Makridakis and Wheelwright, 1980).

Second, in order to develop a trend line, monthly unseasonalized values were computed. In the multiplicative decomposition model unseasonalized value, for each month, is calculated by dividing its actual value by its seasonal index. In the additive decomposition model unseasonalized value, for each month, is computed by subtracting actual value with monthly index. Next, simple regression was used to derive a trend line equation for each model. The simple regression involves minimizing the least squares differences between the line and the unseasonalized values. That process can be modeled as:

$$SSE = \sum_{i=1}^n (y_i - \gamma_i)^2 = \sum_{i=1}^n [y - (\alpha + \beta * x_i)]^2 \quad (6)$$

where SSE is the sum of squared errors; n is the number of observations (unseasonalized values); y_i the observed y -value; and γ_i is the forecasted average value of the dependent variable; α is the constant; β is the slope; and x is the value of the independent variable (time); y_a is the average value of the dependent variable. The constant (α) and the trend (β) are calculated using equation (7) and equation (8), respectively:

⁶ Centered moving average is preferable when the number of seasons is even. On the contrary, if the number of seasons is odd a so called running moving average is preferable (Makridakis et al, 1998).

$$\alpha = y_a - \beta x_a \quad (7)$$

$$\beta = \frac{\sum xy - nx_a y_a}{\sum x^2 - nx_a^2} \quad (8)$$

Third, the trend line equations were used to compute an unseasonalized forecast. For the multiplicative method the unseasonalized forecast was then multiplied with each month's seasonal index to generate a forecast. For the additive model, each month's seasonal index was then added to the unseasonalized forecast to generate a forecast (Makridakis and Wheelwright, 1980).

3.5.2 Holt-Winters

Holt-Winters is a triple exponential smoothing forecast method constructed to handle trend and seasonality. It is a six step procedure including (1) calculation of seasonal indices; (2) overall smoothing of trend level; (3) smoothing of trend factor; (4) smoothing of seasonal indices; (5) generating of forecast; and (6) optimization of smoothing constants. Both multiplicative Holt-Winters and additive Holt-Winters were computed.

First, equation (1) was used to compute seasonal indices for 2007. Second, an overall smoothing of level (L_t) was performed to deseasonalize the data. The overall smoothing value for multiplicative Holt-Winters is computed in accordance with equation (9):

$$L_{t,m} = \alpha * \frac{D_t}{S_{t-1}} + (1 - \alpha) * (L_{t-1} + T_{t-1}) \quad (9)$$

where $L_{t,m}$ is the trend level at time t in the multiplicative Holt-Winters model ; S_t is the seasonal factor at time t ; D_t is the actual value at time t ; T_t is the trend factor at time t ; L_t is the level of trend at time t ; and α ($0 < \alpha < 1$) is the first smoothing constant. Same equation for the additive Holt-Winters is:

$$L_{t,a} = \alpha * (D_t - S_{t-1}) + (1 - \alpha) * (L_{t-1} + T_{t-1}) \quad (10)$$

where $L_{t,a}$ is the trend level at time t in the additive Holt-Winters model; S_t is the seasonal factor at time t ; D_t is the actual value at time t ; T_t is the trend factor at time t ; L_t is the level of trend at time t ; and α ($0 < \alpha < 1$) is the first smoothing constant. At this stage, α can be arbitrarily set to any number between 0 and 1. The average value (D_t) is the average value of actual values for 2007.

Third, the trend factor was smoothed. Same equation for both the models was used. The trend factor is computed using equation (11):

$$T_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * T_{t-1} \quad (11)$$

where β ($0 < \beta < 1$) is the second smoothing constant. At this stage, β can be arbitrarily set to any number between 0 and 1. The derivative of the trend line equation for 2007 is used as the initial value of the trend factor (T_t).

Fourth, each seasonal index was smoothed. In the multiplicative model equation (12) is used:

$$S_{t,m} = \gamma * \left(\frac{D_t}{L_t} \right) + (1 - \gamma) * S_{t-1} \quad (12)$$

where $S_{t,m}$ is the smoothed seasonal index at time t ; and γ ($0 < \gamma < 1$) is the third smoothing constant. In the additive model equation (13) is used:

$$S_{t,a} = \gamma * (D_t - L_t) + (1 - \gamma) * S_{t-1} \quad (13)$$

where $S_{t,a}$ is the smoothed seasonal index at time t ; and γ ($0 < \gamma < 1$) is the third smoothing constant. At this stage, γ can be arbitrarily set to any number between 0 and 1.

Next, a forecast for each model was generated. The forecast for the multiplicative model is computed using equation (14):

$$F_{t,m} = (L_{t-1} + T_{t-1}) * S_{t-s} \quad (14)$$

where $F_{t,m}$ is the forecast at time t . The forecast for the additive model is computed using equation (15):

$$F_{t,a} = L_{t-1} + T_{t-1} + S_{t-s} \quad (15)$$

where $F_{t,a}$ is the forecast at time t (Kalekar, 2004).

Last, each smoothing constant were optimized using Excel Solver. The optimization procedure includes iteration of each smoothing constant until a particular variable is minimized (Hyndman, Kohler, Snyder and Grose, 2002). The variable chosen to minimize was MAPE⁷. Consequently, both Holt-Winters models' forecast for 2011 were generated with the aim to minimize MAPE.

3.5.3 ARIMA

Auto Regressive Integrated Moving Average (ARIMA) is a complex, but widely used, approach to analyze time series data. The method is a four-step iterative process including the following activities: (1)

⁷ The choice was founded on indications that the managers of Reinertsen preferred to minimize their percentage forecasting error.

identification of the model; (2) estimation of the model; (3) checking the appropriateness of the model; and (4) forecasting of future values (Wong et al, 2011).

ARIMA(p,d,q) involves three parameters, and the first step, the model identification step, involves estimation of these parameters. The ARIMA model is constructed only for stationary data. Thus, the differencing parameter (d) is related to the trend of the time series (Box et al, 2011). If the data is stationary in nature d is set to zero; that is d(0). However, if the data is non-stationary in nature it has to be corrected before implemented in the model. This is done by differencing the data in accordance with equation (16):

$$y_{D1} = y_t - y_{t-1} \quad (16)$$

where y_t is the present value and y_{t-1} is its previous value one period ago. If the data is stationary after the first differencing d is set to one; that is d(1) (Kalekar, 2004).

To check the data for stationarity an Augmented Dickey-Fuller (ADF) test was performed using STATA. The ADF test is modeled in accordance with equation (17):

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{i=1}^{\rho} \beta_i \Delta y_{t-1} + \epsilon_t \quad (17)$$

where Δy_t is $y_t - y_{t-1}$; α is the constant; β is the trend; and γ is the parameter to be estimated. The null hypothesis ($\gamma=0$), that the data is trending over time, is tested with a one tail test of significance (τ)⁸. Thus, if test statics exceeds the critical value ($\tau > \tau_c$), the null hypothesis is not rejected; the data is non-stationary in nature. On the contrary, if the critical value exceeds test statistics ($\tau < \tau_c$), the null hypothesis is rejected; the data is stationary in nature (Wong et al, 2011).

The autoregressive term (AR) is related to autocorrelation. That is, the variable to be estimated depends on past values (lags). Thus, if the variable only depends on one lag it is an AR(1) process. The autoregressive process, AR(p), is calculated using equation (18):

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (18)$$

where y_t is the present value of the variable; α is the constant; ϕ is the AR coefficient; y_{t-p} is the past value p periods ago; and ϵ_t is the present random error term (Box et al, 2011). The first order differencing autoregressive model, ARIMA(1,1,0) can then be expressed in accordance with equation (19):

⁸ Tau statistic (τ) is used instead of t -statistic because a non-stationary time series has a variance that increase as the sample size increase.

$$y_t - y_{t-1} = \alpha + \theta_1(y_{t-1} - y_{t-2}) \quad (19)$$

The moving average term (MA) refers to the error term of the variable to be estimated. If the variable to be estimated depends on past error terms there is a moving average process. Thus, MA(1) refers to a variable that only depends on an error term one period ago. The moving average process, MA(q), is calculated using equation (20):

$$y_t = \alpha + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_3 \epsilon_{t-3} + \dots + \theta_q \epsilon_{t-q} \quad (20)$$

where y_t is the present value of the variable; α is the constant; ϵ_t is the present random error term; θ is the MA coefficient; and ϵ_{t-q} the past random error term q periods ago (Box et al, 2011). The first order differencing moving average model, ARIMA(0,1,1) can then be expressed in accordance with equation (21):

$$y_t - y_{t-1} = \alpha - \theta_1 \epsilon_{t-1} \quad (21)$$

To estimate AR(p) and MA(q) the autocorrelation coefficient (AC) and the partial autocorrelation coefficient (PAC) were derived in STATA. AC shows the strength of the correlation between present and previous values, and PAC shows the strength of the correlation between present value and a previous value, without considering the values between them. The behavior of AC and PAC indicates the values of p and q respectively⁹ (Wong et al, 2011). The statistical significance of each value is tested with the Box-Pierce Q statistic test for autocorrelation, which is calculated using equation (22):

$$Q = n \sum_{m=1}^h r_m^2 \quad (22)$$

where Q is the test statistics value; n is the number of observations; h is the maximum number of lags allowed; and r_m is the number of lags. The null hypothesis is that the data contain no autocorrelations. Thus, if the null hypothesis is not rejected there are no correlations between present value and previous values. On the contrary, if the null hypothesis is rejected there are correlations between present value and previous values (Box and Pierce, 1970).

The second step, after the parameters were assessed, involved estimation of the model. The tentative ARIMA was fitted to the historical data and a regression was run. The third step involved controlling the appropriateness of the model. Thus, the residuals were collected and tested for autocorrelation. Again, the

⁹ This identification procedure involves much trial and error. Other, more information-based procedures are Akaike's information criterion (AIC), Akaike's final prediction error (FPE), and the Bayes information criterion (BIC) (Gooijer and Hyndman, 2006).

Box-Pierce Q statistic test for autocorrelation was conducted. Eventually, when a proper model was identified, a forecast for 2011 was generated (Wong et al, 2011).

3.6 Accuracy measures

Three common accuracy measures were chosen to evaluate the performance of each forecast. The mean error (ME) shows whether the forecasting errors are positive or negative compared to the actual values. It is defined as follows:

$$ME = \frac{1}{n} \sum_{t=1}^n (A_t - P_t) \quad (23)$$

where A_t is the actual value at time t ; P_t is the predicted value at time t ; and n is the number of observations (Makridakis et al, 1998). The mean square error, MSE, which square the error terms to highlight large deviation, yields an absolute error term. The measure is calculated as follows:

$$MSE = \frac{1}{n} \sum_{t=1}^n (A_t - P_t)^2 \quad (24)$$

The mean absolute percentage error, MAPE, yields the percentage error of the forecast. The measure is defined as follows:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left(\frac{|A_t - P_t|}{A_t} \right) \quad (25)$$

where $|A_t - P_t|$ is the absolute error (Kaastra and Boyd, 1995).

4. Empirical results

4.1 Characteristics of data

Table 1 presents the descriptive statistics, including number of observations, mean, median, standard deviation, minimum value and maximum value for for total sample, model sample and validation sample, respectively.

Table 1: Descriptive Statistics

The total sample contains 60 observations and covers the period January 2007 to December 2011. The model sample contains 48 observations and covers the period January 2007 to December 2010. The validation sample contains 12 observations and covers the period January 2011 to December 2011. Each year contains 12 observations. The observations are monthly numbers of chargeable hours (CH) provided by Reinertsen.

Variable	N	Mean	Median	Std. Dev.	Min	Max
Total sample (CH)	60	20505	19758	6480	9243	37320
Model sample (CH)	48	18784	18652	5582	9243	34977
Validation sample (CH)	12	27386	27601	5246	19878	37320
2007	12	14433	13863	3375	9783	20786
2008	12	17333	17827	3585	9243	24865
2009	12	19783	19348	5002	9666	28623
2010	12	23587	22659	5260	13749	34977
2011	12	27386	27601	5246	19878	37320

The total sample contains 60 observations and covers the period January 2007 to December 2011. The model sample contains 48 observations and covers the period January 2007 to December 2011. The validation sample contains 12 observations and covers the period January 2011 to December 2011. The minimum value in the time series is 9243 hours and the maximum value in the time series is 37320 hours. Mean value, median value and standard deviation for the total sample is 20505 hours, 19750 hours and 6480 hours, respectively. Corresponding numbers for the model sample is 18784, 18652 and 5582, and for the validation sample 27386, 27601 and 5246. Both mean and median increase each year for the period 2007 to 2011, thus indicating an upward trend.

The relationship between chargeable hours and time is plotted in figure 1a. The monthly fluctuation of chargeable hours per year is plotted in figure 1b.

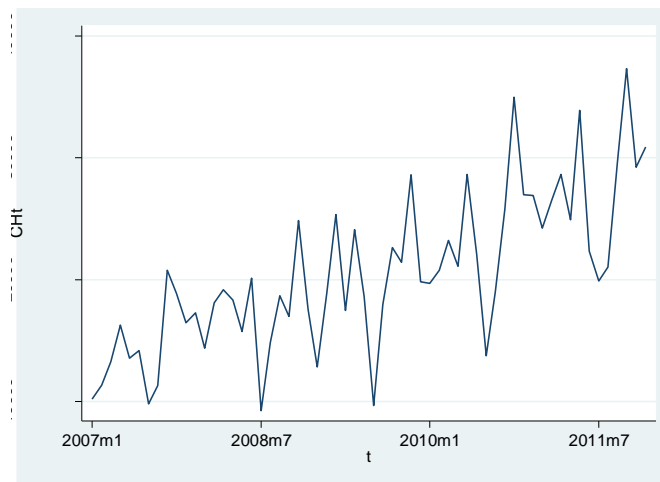


Figure 1a: Chargeable hours over time

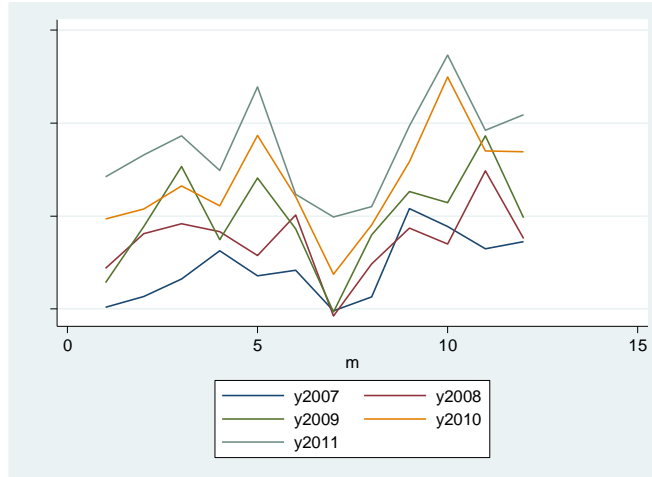


Figure 1b: Monthly fluctuation of chargeable hours

Figure 1a shows an increase in chargeable hours between January 2007 and December 2011, indicating an upward trend. Figure 1b reveals that chargeable hours fluctuate similarly each year, suggesting seasonality.

Table 2 presents seasonal index for each month between 2007 and 2011 as well as the average index value for each month for these years. Table 2 also shows the standard deviation of the seasonal fluctuations each month.

Table 2: Seasonal indices

The table shows the seasonal index value for each month for the total sample (2007-2011). Also, the average index value for each month for the total sample is displayed. If the seasonal index is below 1 it indicates a value below average, and if the seasonal index is above 1, it indicates a value above average. The standard deviation of the monthly fluctuations for each month, which also is presented in table 2, is for the model sample (2007-2010).

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2007	0.71	0.79	0.92	1.13	0.94	0.98	0.68	0.78	1.44	1.31	1.14	1.20
2008	0.83	1.04	1.11	1.06	0.91	1.16	0.53	0.86	1.08	0.98	1.43	1.01
2009	0.65	0.95	1.28	0.88	1.22	0.94	0.49	0.91	1.14	1.08	1.45	1.00
2010	0.84	0.88	0.98	0.89	1.21	0.94	0.58	0.81	1.10	1.48	1.14	1.14
2011	0.88	0.97	1.05	0.91	1.24	0.82	0.73	0.77	1.08	1.36	1.07	1.13
Average	0.78	0.93	1.07	0.97	1.10	0.97	0.60	0.82	1.17	1.24	1.25	1.10
Std. dev. (%)	8.00	9.50	13.80	10.40	14.70	9.10	7.00	4.80	14.70	19.50	14.90	8.30

Consistent with earlier observation, Table 2 displays that the data contain seasonality. The index values fluctuate below and above average several times a year. Table 2 shows that many fluctuations occur on monthly basis, thus indicating that the seasonality, more or less, is monthly in nature. Chargeable hours in January, February, April, June, June, and August are, on average, below average. On the contrary, chargeable hours in March, May, September, October, November and December are, on average, above average. Table 2 also shows that July is the lower outlier and that October is the upper outlier. The

standard deviation of the monthly fluctuations is particularly high in May, October and November. The high standard deviation in monthly fluctuations for these months might lead to forecasting difficulties. Table 2 does not give a clear indication whether the seasonality is multiplicative or additive in nature.

Table 3 present the result of the Runs test for binary randomness. The test includes all 60 monthly observation obtained.

Table 3: Runs test for binary randomness

CHt denotes chargeable hours for the total sample. The test statistics is the absolute value of the calculated test value. The critical values at 10%, 5% and 1% significance level are obtained in a standard normal distribution table. The null hypothesis is tested with a two tail test of significance. Thus, each critical value, obtained in the standard normal distribution table, has been divided by 2 before presented in table 3. The null hypothesis is true if $|Z| < Z_{1-\alpha/2}$ and rejected if $|Z| > Z_{1-\alpha/2}$

Variable	Test statistics $ Z $	Critical values (1%)	Critical values (5%)	Critical values (10%)
CHt	5.446	2.57	1.96	1.65

Table 3 shows that test statistics exceeds critical values at 10%, 5%, and 1% significance level, thus enable rejection of the null hypothesis. The result shows that the monthly fluctuation is systematic with non-random variation. Thus, it is possible to conclude that chargeable hours have followed a discernible pattern between 2007 and 2011.

The data analysis indicates that the historical observations of chargeable hours contain trend and strong monthly fluctuations. Thus, the chosen forecasting methods are designed to handle trend and recurrent fluctuations. The analysis does not reveal whether the trend is additive or multiplicative in nature. As a result, both an additive and multiplicative models are used for classical decomposition and Holt-Winters. July and November are the lower and the upper outlier respectively. The standard deviation in monthly fluctuations is particular high in May, October, and November. These months together with July might cause the largest forecasting errors. Although, strong monthly fluctuations, the historical data does not contains much randomness.

4.2 ARIMA parameters

Table 4 presents the result of the Augmented Dickey-Fuller test for chargeable hours. The ADF test, designed for constant and trend, is used because of the characteristics of the data obtained. Also, because of monthly observation, the data is assumed to correlate with its previous 12 observations. Thus, 12 lags are used in the ADF test.

Table 4: Augmented Dickey-Fuller test of chargeable hours

CH denotes chargeable hours for the model sample C, T, and 12 represent constant, trend and number of lags, respectively. Test statistics (τ) is the derived value which is compared with the critical values (τ_c) at 10%, 5% and 1% significance level. The null hypothesis is true if $\tau > \tau_c$ and rejected if $\tau < \tau_c$

Variable	Test statistics	Critical values (1%)	Critical values (5%)	Critical values (10%)
CH	-0.866 (C,T,12)	-4.288	-3.560	-3.216

Table 4 shows that test statistics exceeds critical values at 10%, 5% and 1% significance level. Accordingly, the null hypothesis cannot be rejected at any significance level. The data is non-stationary in nature. Consequently, the data has to be corrected for stationarity by using the first order differencing equation. The characteristics of first order differencing of chargeable hours is shown in Figure 2.

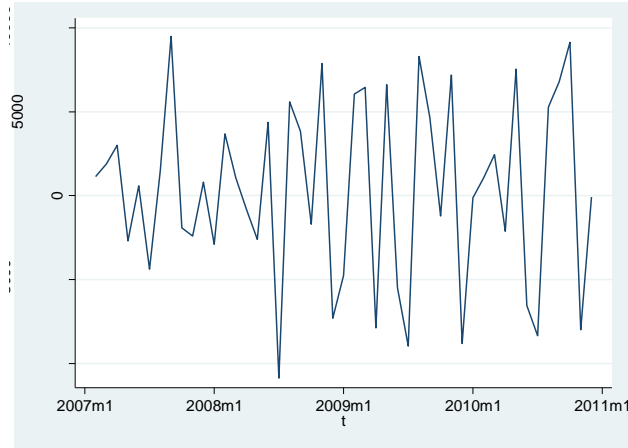


Figure 2: First order differencing of chargeable hours over time

Figure 2 suggests that first order differencing of chargeable hours has a constant trend and variation between January 2007 and December 2010. Figure 2 also shows that the first order differencing of chargeable hours is mean reverting, indicating that the variable is stationary. Table 5 presents the result of the Augmented Dickey-Fuller (ADF) test for first order differencing of chargeable hours. Due to the characteristic of first order differencing of chargeable hours, an ADF test without trend parameter is chosen.

Table 5: Augmented Dickey-Fuller test (ADF) of first order differencing of chargeable hours

CHD1 denotes first order differencing of chargeable hours for model sample. C and 12 represent constant and number of lags, respectively. Test statistics (τ) is the derived value which is compared with the critical values (τ_c) at 10%, 5% and 1% significance level. The null hypothesis is true if $\tau > \tau_c$ and rejected if $\tau < \tau_c$.

Variable	Test statistics	Critical values (1%)	Critical values (5%)	Critical values (10%)
CHD1	-3.933 (C,12)	-3.689	-2.975	-2.619

Table 5 shows that critical values, at 10%, 5% and 1% significance level, exceed test statistics. Consequently, the null hypothesis can be rejected at any significance level. The first order differencing of chargeable hours is stationary, indicating that the d in ARIMA(p,d,q) should be set to one.

Table 6 shows the autocorrelation coefficient (AC) and the partial autocorrelation coefficient (PAC) of first-order differencing of chargeable hours for 12 lags. Also the result of the Box-Pierce statistic test (Prob>Q) is shown in the table.

Table 6: AC, PAC and Q of first order differencing of chargeable hours

AC shows the correlation between the present value of first order differencing of chargeable hours and the previous value of the first order differencing of chargeable hours. PAC shows the correlation between first order differencing of chargeable hours and a previous value of the first order differencing of chargeable hours without the effect of the lags between them. Q refers to Box-Piece statistic tests. Prop>Q refers to the null hypothesis that all correlation up to lag m (m=1,2,3,...,m) are equal to 0. If Prop>Q is less than 0.05 the null hypothesis can be rejected at 5% significance level. On the contrary, if Prop>Q is more than 0.05 the null hypothesis cannot be rejected at 5% significance level. Lag refers to previous month m (m=1,2,3,...,12).

Lag	1	2	3	4	5	6	7	8	9	10	11	12
AC	-0.2768	-0.2484	0.0624	-0.2483	0.0358	0.4146	-0.1264	-0.1504	0.1680	-0.4182	0.1185	0.4126
PAC	-0.2769	-0.3739	-0.1613	-0.5148	-0.6315	-0.0789	0.0941	-0.2211	0.1747	-0.5075	-0.6243	-0.3023
Prob>Q	0.0501	0.0303	0.0658	0.0328	0.0606	0.0025	0.0036	0.0041	0.0040	0.0001	0.0002	0.0000

A positive value for both coefficient indicates an AR(p) process while a negative value for both coefficient indicates and a MA(q) process. Table 6 shows that the coefficients have more negative values, thus indicating a MA(q) process. This is clearly demonstrated in Figure 3, which graphically shows the values of AC and PAC.

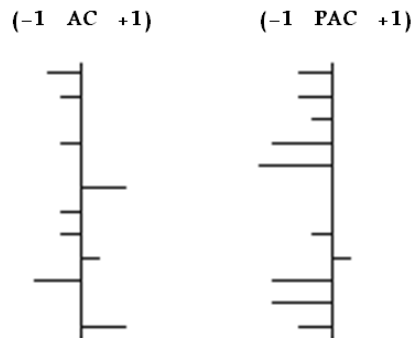


Figure 3: AC and PAC of first order differencing of chargeable hours

Figure 3 shows that most spikes, for both AC and PAC, are negative. This is an indication of a MA(q) process. Table 6 also shows that most of the autocorrelation between present value of first order differencing of chargeable hours and previous values is statistically significant at 5% level (Prop>Q is less than 0,05). The low values of prop>Q for lag 10, 11, and 12, however, indicate that the first order differencing of chargeable hours have stronger correlation to lag 10, 11 and 12. This indicates that the first order differencing of chargeable hours depends much on the previous 10th, 11th, and 12th month, respectively¹⁰. Consequently MA(10), MA(11) and MA(12) are chosen, leading to three tentative ARIMA models; ARIMA(0,1,10), ARIMA(0,1,11), and ARIMA(0,1,12).

¹⁰ Although lag 2,4, 6, 7, 8, and 9 show autocorrelation that are statistically significant at 5% level, the residuals of tentative ARIMA(0,1,2), ARIMA(0,1,4), ARIMA(0,1,6), ARIMA(0,1,7), ARIMA(0,1,8) and ARIMA(0,1,9) give low p-values. This indicates autocorrelation among residuals, which in turn, make these models unsuitable to forecast chargeable hours at Reinertsen.

A regression of each tentative ARIMA model is run and the residuals are collected and tested for autocorrelation. Table 7 shows the AC and PAC for residuals of ARIMA(0,1,10), ARIMA (0,1,11) and ARIMA(0,1,12). The result of Box- Pierce statistic test (Prob>Q) is also shown in Table 7.

Table 7: AC, PAC and Q for residuals of ARIMA(0,1,10), ARIMA(0,1,11) and ARIMA(0,1,12)

AC shows the correlation between the present value of first order differencing of chargeable hours and the previous value of the first order differencing of chargeable hours. PAC shows the correlation between first order differencing of chargeable hours and a previous value of the first order differencing of chargeable hours without the effect of the lags between them. Q refers to Box-Pierce statistic tests. Prop>Q refers to the null hypothesis that all correlation up to lag m (m=1,2,3,...,n) are equal to 0. If Prop>Q is less than 0.05 the null hypothesis can be rejected at 5% significance level. On the contrary, if Prop>Q is higher than 0.05 the null hypothesis cannot be rejected at 5% significance level. Lag refers to previous month m where m is equal to 10,11 and 12.

Lag	ARIMA(0,1,10)			ARIMA(0,1,11)			ARIMA(0,1,12)		
	AC	PAC	Prop>Q	AC	PAC	Prop>Q	AC	PAC	Prop>Q
10	-0.1858	-0.2473	0.8536	-0.1454	-0.2231	0.9055	-0.1544	-0.2315	0.8299
11	0.1810	.03080	0.7469	0.0431	0.1052	0.9361	0.0532	0.0509	0.8731
12	0.2891	0.5675	0.3605	0.2365	0.2690	0.7386	0.2210	0.2251	0.6843

Table 7 indicates that there is no autocorrelation among residuals in any of the tentative models (Prop>Q is higher than 0.05). However, of these tentative ARIMA models, ARIMA(0,1,11) has the highest Prop>Q for each lag tested for autocorrelation (lag 10 (0.9055), lag 11 (0.9361) and lag 12 (0.7386)), indicating that this model is the most suitable selection for this time series. Thus, an ARIMA(0,1,11) model is chosen for forecasting chargeable hours for 2011. Figure 4 graphically shows AC and PAC for ARIMA(0,1,11). The few and small spikes prove that there is no autocorrelation among residuals in ARIMA(0,1,11).

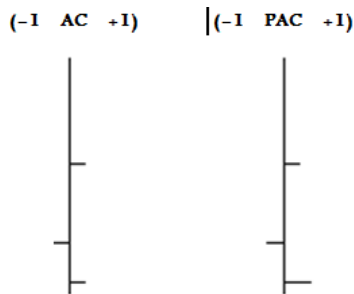


Figure 4: AC and PAC for residuals of ARIMA(0,1,11)

4.3 Forecasting performance

Figure 5a, 5b, 5c, 5d, 5e and 5f demonstrate the relative performance of the current forecasting method, multiplicative decomposition, additive decomposition, additive Holt-Winters, multiplicative Holt-Winters and ARIMA to the actual values of 2011, respectively. The diagrams are based on the forecasted values shown in appendix 1.

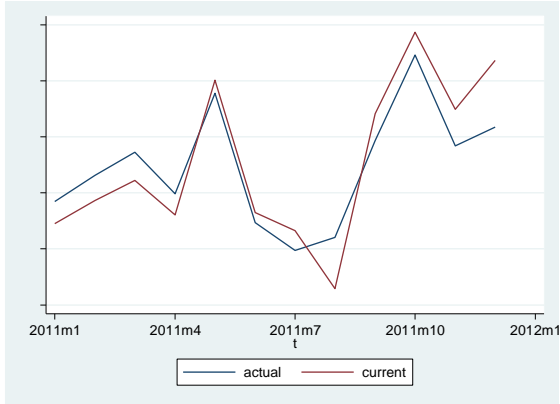


Figure 5a: Current forecasting method

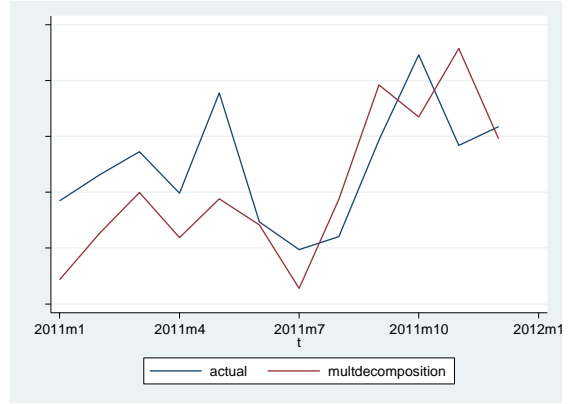


Figure 5b: Multiplicative decomposition

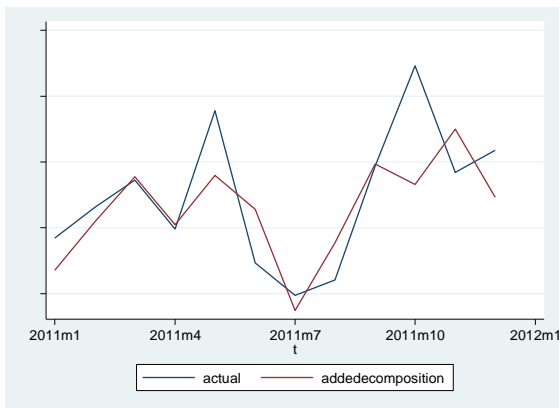


Figure 5c: Additive decomposition

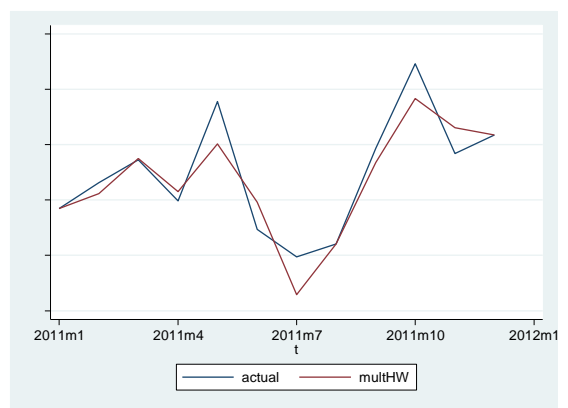


Figure 5d: Multiplicative Holt-Winters

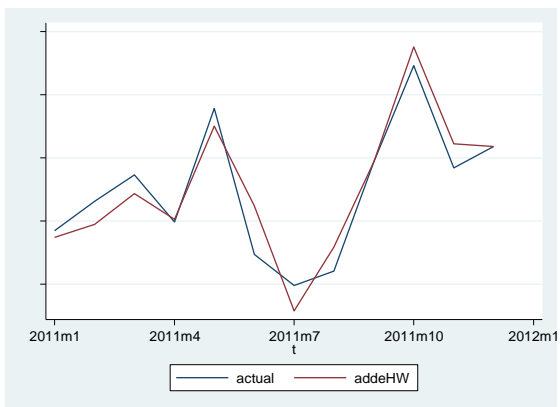


Figure 5e: Additive Holt-Winters

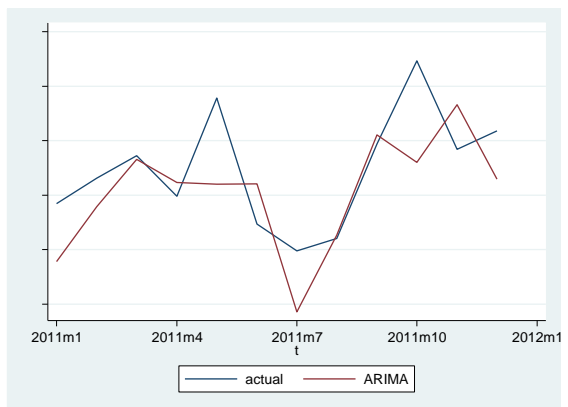


Figure 5f: ARIMA

Figure 5a shows that the current method used at Reinertsen captures the fluctuations in the data fairly well. However, figure 5a also shows a large underestimation in the beginning of the second half of 2011 as well as in the end of the year. Figure 5b indicates that multiplicative decomposition has problem to estimate the actual values in 2011, particularly during the first half of the year. The figure also indicates that the method tends to be one period ahead in the end of the year. Figure 5c shows that additive

decomposition forecast the lower outlier fairly well but that the method has problem to capture upper outliers. Figure 5d indicates that multiplicative Holt-Winters forecast most of the actual values well. However, the method clearly under-forecast the lower outlier. Figure 5e shows that additive Holt-Winters smooth the actual values well. Also this method has a tendency to under-forecast the lower outlier. Figure 5f shows that ARIMA has large problems to capture all outliers in the data.

Table 8 presents each forecasting methods MAPE. The average MAPE as well as the MAPE for each month 2011 is displayed.

Table 8: MAPE

The mean absolute percentage error, MAPE, shows the forecast error in percent. Table 8 shows MAPE for each method, each month, for 2011. Also the average MAPE for each method is displayed. All error is in relation to actual values for 2011. All figures represent the percentage error.

%	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Average
Qualitative	8.10	8.38	8.82	7.50	3.47	4.07	8.81	21.76	8.01	5.47	11.09	19.29	9.57
Add. decompos.	10.15	4.12	0.86	1.28	14.53	18.19	5.77	13.43	0.50	24.20	11.20	11.55	9.65
Mul. decompos.	29.17	19.68	12.70	15.96	27.97	1.21	17.60	15.92	16.62	14.95	29.62	3.56	17.08
Mul. Holt-Winter	0.00	3.75	0.38	3.40	11.35	11.05	17.23	0.00	4.38	8.42	7.97	0.00	5,66
Add. Holt-Winter	2.23	6.87	5.14	8.2	4.13	17.24	10.05	9.07	0	3.97	6.47	0	5.50
ARIMA	6.64	10.86	5.08	8.16	32.33	26.55	17.50	2.30	6.29	21.32	25.64	9.57	14.36

Table 8 shows that the currently used qualitative forecasting method at Reinertsen has a somewhat low MAPE (9.57%) in comparison to most of the quantitative forecasting methods. Only additive Holt-Winters and multiplicative Holt-Winters yield a lower MAPE (5.50%, 5.66%). Additive decomposition has a comparable MAPE (9.65%). The remaining methods, ARIMA and multiplicative decomposition, yield a MAPE (14.36%, 17.08%) that is significantly larger than of the qualitative method.

Table 9 shows MSE of each forecast generated by the different forecasting methods. The average MSE as well as the MSE for each month for 2011 is displayed.

Table 9: MSE

The mean square error, MSE, shows a squared absolute error term. Table 9 shows MSE for each method, each month, for 2011. Also the average MSE for each month is shown. All errors are in relation to the actual values for 2011. All figures represent a value that should be multiplied by 10^6 .

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Average
Qualitative	3.85	4.96	6.38	3.49	1.39	0.82	3.07	20.9	5.66	4.17	10.5	35.5	8.40
Add. decompos.	6.05	1.12	0.06	0.10	24.27	16.51	1.32	7.98	0.02	81.54	10.70	12.72	13.54
Mul. decompos.	49.97	27.32	13.23	15.82	89.96	0.07	12.24	11.21	24.37	31.13	74.91	1.21	29.29
Mul. Holt-Winter	2.55	1.19	0.29	0.26	31.47	16.71	11.70	8.38	0.05	80.28	5.42	12.23	4.30
Add. Holt-Winter	0.29	3.33	2.16	0.04	1.96	14.83	3.99	3.64	0	2.20	3.58	0	3.00
ARIMA	2.59	8.33	2.11	4.14	120	35.19	12.10	0.23	3.49	63.30	56.12	8.74	26.38

Table 9 shows that the currently used qualitative forecasting method at Reinertsen has a fairly low MSE (8.40) compared to most of the quantitative forecasting methods. Only, additive Holt-Winters and multiplicative Holt-Winters yield a lower MSE (3.00, 4.30). Multiplicative decomposition and ARIMA have significantly larger MSE (29.29, 26.38). These results are consistent with the results shown in table. Table 9 also shows that additive decomposition has a MSE (13.54) that is much larger than that of the currently used forecasting method. This result is a deviation from the result shown in table 8.

Table 10 demonstrates the actual error of each forecast generated by the different forecasting methods. The error for each month as well the mean error for 2011 is displayed.

Table 10: Mean error

Table 10 shows the mean error (ME) for each method, 2011. Also all actual error for each method, each month are displayed. All errors are in relation to actual values. All values are presented in hours (h).

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	ME
Qualitative	1963	2227	2526	1868	-1177	-909	-1752	4576	-2380	-2041	-3241	-5959	-358
Add. decompos.	2460	1094	-246	-319	4926	-4063	1147	-2825	-149	9030	-3271	3567	946
Mul. decompos.	7069	5227	3638	3977	9485	270	3498	-3348	-4936	5580	-8655	1101	1909
Mul. Holt-Winter	1	996	-108	-848	3848	-2468	3425	0	13	3144	-2327	-1	473
Add. Holt-Winter	539	1825	1471	-204	1400	-3852	1997	-1908	0	-1482	-1892	0	-176
ARIMA	-1609	-2886	-1453	-2035	10963	-5932	-3479	-485	-1869	7956	-7492	2956	-447

Table 10 shows that ME for the currently used forecasting method at Reinertsen is -358 hours. This indicates a tendency of underestimation. Also additive Holt-Winters and ARIMA has a negative ME (-176h, -447h). Consequently, these methods tend to under-forecast actual values. On the contrary, remaining methods, additive decomposition, multiplicative decomposition, and multiplicative Holt-Winters have a positive ME (946h, 1909h, 473h). These findings indicate that these methods have a tendency to over-forecast.

5. Analysis

Mahmoud (1984) suggests that plus or minus 10% is a generally acceptable accuracy limit. The evaluation of the qualitative method currently used at Reinertsen gives forecasts within this range. This is an indication of an elaborate qualitative forecasting method. However, Reinertsen is not satisfied with current forecasting and wants it more accurate.

The result obtained clearly shows that both Holt-Winters models yield the most accurate forecasts amongst the evaluated forecasting methods. This indicates that there is no correlation between forecasting accuracy and the complexity of the method used, which is also suggested of Pollack-Johnson (1995). The result also shows that both Holt-Winters models can provide more accurate forecasts of chargeable hours than the qualitative method currently used at Reinertsen. This suggests that time series forecasting has the

potential to improve the forecasting of chargeable hours at Reinertsen. That time series forecasting methods can provide better forecasts than these provided by qualitative forecasting methods is a result similar to that obtained by others, including Makridakis (1981) and Mahmud (1984).

Basset (1973) suggests that improved forecasting has the potential to increase the profitability of a firm. In this case, the improved forecasting accuracy of chargeable hours will enable Reinertsen to make more rational manpower planning decisions. For instance, with improved forecasting of chargeable hours, Reinertsen can schedule holidays better and time hiring more exactly. This will lead to a more efficient use of available consultant hours, which in turn improves the utilization rate. As a result, the better forecasting of chargeable hours has the potential to make Reinertsen to a more efficient and profitable firm.

Table 8 and 9 shows that both Holt-Winters methods can provide more accurate forecasts than the qualitative method currently used at Reinertsen. The relative success of Holt-Winters to forecast chargeable hours at Reinertsen might be attributable to the methods flexibility. Kalekar (2004) suggests that Holt-Winters is a flexible method because it adds more weights to recent observation, and according to Gelper et al (2008), the flexibility of Holt-Winters makes it a robust method to handle seasonality with outliers. As shown in table 2, the historical data of chargeable hours contain much monthly fluctuations with some outliers. Table 8 and 9 and Figure 5d and 5e clearly show that both Holt Winters models capture these outliers better than the other methods.

Additive decomposition provides a forecast that is comparable to that of the qualitative method currently used at Reinertsen, only when MAPE is taken into account. When also MSE is taken into account, these methods are no longer comparable. The higher MSE for additive decomposition is most likely due to the large forecasting error in October. Consequently, Reinertsen should not consider this model if they prefer many small errors rather than a few large errors. One reason that the additive decomposition did not perform better might be attributable to the methods underlying inertia; all historical values in the time series are equally weighted, regardless of occurrence in time. This makes additive decomposition less applicable when standard deviation in monthly fluctuations is high. Table 2, suggests that May (14.70%), October (19.50%), and November (14.90%) have the highest standard deviation amongst all months in the time series. Table 8 and figure 5c show that these are months when additive decomposition has large forecasting errors, especially October where its forecasting error is as high as 24.20%. That the classical decomposition methods have difficulties to forecast when the seasonal fluctuations are not stable is also suggested by Makridakis et al (1998). They argue, instead, that classical decomposition is a more suitable forecasting method when standard deviations in seasonal fluctuations are low.

The poor performance of ARIMA indicates that the method is not suitable to forecast chargeable hours at Reinertsen. Hyndman (2001) suggest that the major problem with ARIMA is to estimate its parameters, especially when data is difficult to interpret. The estimation procedure used in this thesis shows that this is the case; although the estimation procedure is commonly used it did not give a clear indication of which model to choose. Another problem with ARIMA, suggested by Wong et al (2011) is

the methods inability to forecast strongly fluctuating time series. Table 8 shows that ARIMA has problem to yield accurate forecasts for May, October and November; that are the months with the largest standard deviation in monthly fluctuations. Also, figure 5e indicates problem to capture the outliers in the data. Consequently, the poor performance of ARIMA can be attributable to the methods inability to capture outliers and to handle deviation in the monthly fluctuations.

The reason to the poor performance of multiplicative decomposition is probably attributable to the same underlying inertia that additive decomposition exhibit. A further reason to the poor performance of multiplicative decomposition could be that the historical data is additive in nature. According to Kalekar (2004), the additive model should be used when the magnitudes of the seasonal fluctuations are constant to the trend. On the contrary, the multiplicative model should be used when the magnitudes of the seasonal fluctuations are proportional to the trend. The relative performance of additive and multiplicative decomposition, where the additive model outperform the multiplicative model, suggests that the data of chargeable hours exhibit additive characteristics rather than multiplicative.

It is clear that both Holt-Winter models have the potential to improve the forecasting accuracy of chargeable hours at Reinertsen. However, whether to choose the additive or multiplicative model is not equally clear and a deeper analysis of the relative performance between the Holt-Winters models should be made.

According to Makridakis et al (1998), the result of classical decomposition could be used as a tool to identify whether the seasonal fluctuations is additive or multiplicative in nature. As discussed above, the relative performance between the decomposition models indicates that this time series is additive in nature. Accordingly, from this point of view, it is suggested that Reinertsen should choose additive Holt-Winters in front of multiplicative Holt-Winters.

Another important issue to consider when selecting between the two Holt-Winters models is their relative tendency to forecast either above or below the actual value of chargeable hours. Table 10 shows that additive Holt-Winters has many negative errors. This is an indication that the model tend to under-forecast. On the contrary, multiplicative Holt-Winters has many positive errors, thus suggesting that the model has a tendency to over-forecast. Under-forecast of chargeable hours can lead to improved utilization rate. On the other hand, under-forecast of chargeable hours can also lead to manpower shortage.

Many studies, including Bunn (1989), Pollack-Johnson (1995) and Armstrong (2006), suggest a combination of qualitative and quantitative forecasting. Thus, a further issue that is interesting to examine is how the Holt-Winters models compliment the quantitative method currently used at Reinertsen. The current forecast yields large errors in August and December. The multiplicative model forecasts these months perfectly while the additive model yields a somewhat high error in August and no error in December. The multiplicative model has a large error in July and the additive model has large error in June. The currently used forecasting method used at Reinertsen yields fairly low error for both of these months. Seemingly there is a good match both between additive Holt-Winters and the current forecasting

method and multiplicative Holt-Winters and the current forecast. Yet, the latter seems to be a better compliment.

Regardless of which Holt-Winters methods to select, Reinertsen should evaluate their preferences considering a few large errors in opposition to many small errors. As mentioned earlier, the smoothing constants have been optimized with respect to MAPE, which do not value large errors to the same extent as MSE. If Reinertsen value many small errors in front of a few large errors it would be preferable to optimize the smoothing constants with respect to MSE. It will lead to a higher MAPE but a more even forecast.

6. Conclusion

This thesis targets the forecasting of chargeable hour at Reinertsen. The research question refers to the performance of classical decomposition, Holt-Winters and ARIMA in relation to the qualitative forecasting method currently used at Reinertsen. The performance of each method is evaluated and compared to see whether an improvement of current forecast of chargeable hours is possible.

The results obtained clearly indicate that both multiplicative and additive Holt-Winters have the potential to provide more accurate forecasts of chargeable hours at Reinertsen than the qualitative method currently used at Reinertsen. The findings from this thesis also show that additive decomposition can provide a comparable forecast to the method currently used when only MAPE is taken into consideration. However, larger MSE for additive decomposition suggests that the method should be avoided. Moreover, the results also indicate that ARIMA and multiplicative decomposition are inappropriate forecasting methods to forecast chargeable hours at Reinertsen.

This thesis also attempts to explain why or why not each forecasting method could improve the forecasting accuracy at Reinertsen. The success of the Holt-Winters method is believed to be attributable to the method's flexibility. The forecasted time series fluctuates much and the Holt-Winters method, which focuses on recent observations, might be better suited to capture these fluctuations. On the contrary, the relatively poor performance of classical decomposition and ARIMA is believed to be attributable to these methods inability to forecast varying fluctuations.

This thesis forecasts chargeable hours at Reinertsen because a more accurate forecast of chargeable hours is assumed to enable better manpower planning. As a result, Reinertsen can use its available consultant hours more efficiently, which in turn, can lead to an improvement of the utilization rate. Eventually, improved forecasting accuracy of chargeable hours is assumed to enhance the profitability of the firm.

This thesis contributes to the literature on forecasting by further assessing the applicability of the selected forecasting methods in a real situation. Although only one time series is used to generate the forecasts, the results give an indication of the applicability of the evaluated forecasting methods when data contains trend and strong monthly fluctuations. This thesis also provides possible reasons to the

methods performance. A finding of this thesis is that simple methods can provide forecasts that are more accurate than those derived from more complex methods. Another finding is that time series forecasting method has the potential to provide better forecast than those provided by qualitative forecasting methods, even when the data contain strong monthly fluctuations.

A major limitation of this thesis is that the evaluated methods are only a few of all available methods. Several other methods, including explanatory methods and qualitative methods, could be used to forecast chargeable hours at Reinertsen. The choice of other methods would probably lead to different results, and consequently different conclusions. For instance, the results obtained gave an indication that Holt-Winters and the currently used forecasting method could complement each other. Thus, an interesting direction of further research is to investigate whether Holt-Winters can be efficiently combined with qualitative forecasting.

Bibliography

- Armstrong, J. S. (2006) Findings from evidence-based forecasting: Methods for reducing forecast error, *International Journal of Forecasting* **22**: 583-598
- Babbage, S. H. (2004). Runs test for binary randomness, *Dissertation, University of London, England, Great Britain*
- Bassett, G. A. (1973). Elements of Manpower Forecasting and Scheduling, *Human Resource Management* **12**: 35-40
- Bechet, T. P. and Maki, W. R. (2002). Modeling and Forecasting Focusing on People as a Strategic Resource, *Human Resource Planning* **10**: 209-217
- Berridge, K. (2003). Irrational pursuit: Hyper-incentives from a visceral brain, *The psychology of economic decisions* **1**: 17-40
- Box, G., Jenkins, G. and Reinsel, G. (2011). Time Series Analysis: Forecasting and Control, Fourth edition, *New York, John Wiley & Sons*
- Box G. and Pierce D. (1970). Distribution of Residual Autocorrelation in Autoregressive-Integrated Moving Average Times Series Models, *Journal of American Statistical Association* **65**: 1509-1526
- Bunn, D. (1989). Forecasting with more than one model, *Journal of Forecasting* **8**: 161-166
- Chen, C. (1997). Robustness properties of some forecasting methods for seasonal time series: a Monte Carlo study, *International Journal of Forecasting* **13**: 269-280
- Chatfield, C. and Yar, M. (1998). Holt-Winters Forecasting: Some practical Issues, *Journal of the Royal Statistical Society* **37**: 129-140
- Clarke, D. G. and Wheelwright, S. C. (1976). Corporate forecasting: promise and reality, *Harvard Business Review* **6**: 40-64
- Gelper, S., Fried, R. and Croux, C. (2008)., Robust Forecasting with Exponential and Holt-Winters Smoothing, *Working Paper, Katholieke universiteit, Leuven, Belgium*
- Gooijer, J. G. and Hyndman, R. J. (2006). 25 years of time series forecasting, *International Journal of Forecasting* **22**: 443-473
- Greer, W. and Liao, S. (1986). Forecasting capacity and capacity utilization in the U.S. aerospace industry, *Journal of Forecasting* **5**: 57-67

- Hawkins, J. (2005). Economic Forecasting: history and procedures, *Economic Roundup* 2: 1-10
- Heuts, R. and Brockners, J. (1998). Forecasting the Dutch Heavy Truck Market: A Multivariate Approach, *International Journal of Forecasting* 4: 57-79
- Hyndman, R. J. (2001). It's time to move from what to why, *International Journal of Forecasting* 17: 5-13
- Hyndman, R.J., Koehler A.B., Snyder, R.D. and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting* 18:439-454
- Hyndman, R.J. and Kostenko, A.V. (2007). Minimum sample size requirements for seasonal forecasting models, *Foresight* 6: 12-15
- Ito, J., Pynadath, D. and Marsella S. (2010). Modeling Self-Deception within a Decision-Theoretic Framework, *Autonomous Agents and Multi-Agent Systems* 20: 3-13
- Kaastr, I. and Boyd, M. S. (1995). Forecasting futures trading volume using neural networks, *The Journal of Futures Markets* 15: 953-970
- Kalekar, P. S. (2004). Time series Forecasting using Holt-Winters Exponential Smoothing, *Working paper, Kanwal Rekhi School of Information Technology, Bombay, India*
- Koehler, A. B. (1985). Simple VS. Complex Extrapolation models, *International Journal of Forecasting* 1: 63-68
- Mahmoud, E. (1984). Accuracy in Forecasting: a Survey, *Journal of Forecasting* 3: 139-159
- Makridakis, S. (1981). If we Cannot Forecast How Can we Plan, *Long Range Planning* 14: 10-20
- Makridakis, S., Wheelwright, S. C. and Hyndman, R. J. (1998). Forecasting: Methods and Applications, Third edition, *New York, John Wiley & Sons*
- Makridakis, S. and Wheelwright, S. C. (1980). Forecasting Methods for Management, Third edition, *New York, John Wiley & Sons*
- Meade, N. (2000). Evidence for the Selection of Forecasting Methods, *Journal of Forecasting* 19: 515-535
- Pollack-Johnson, B. (1995). Hybrid structures and improving forecasting and scheduling in project management, *Journal of Operation Management* 12: 101-117
- Simon, H. (1979). Rational Decision Making in Business Organizations, *The American Economic Review* 69: 493-513

Theodosiou, M. (2011). Forecasting monthly and quarterly time series using STL decomposition, *International Journal of Forecasting* **27**: 1178-1195

Wong, J., Chan, A. and Chiang, Y. (2011). Construction manpower demand forecasting – A comparative study of univariate time series, multiple regression and econometric modeling techniques, *Engineering, Construction and Architectural Management* **18**: 7-29

Appendix 1: Forecasted values of chargeable hours

Table 11: Forecasted values

Table 11 shows the forecasting values generated by each forecasting model for each month, 2011. Also, the total value for each model is displayed.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Actual	24230	26567	28635	24918	33910	22340	19878	21032	29697	37320	29217	30890	328634
Qualitative	22267	24340	26109	23050	35087	23249	21630	16456	32077	39361	32458	36849	332933
Add.decompos.	21770	25473	28881	25237	28984	26403	18731	23857	29846	28290	32488	27323	317283
Mul.decompos.	17161	21340	24997	20941	24425	22070	16380	24380	34633	31740	37872	29789	305728
Mul. Holt-Winter	24229	25571	28743	25766	30062	24808	16453	21032	28397	34176	31544	30891	321672
Add. Holt-Winter	23691	24742	27164	25122	32510	26192	17881	22940	29697	38802	31109	30890	330738
ARIMA	25839	29453	30089	26953	22947	28272	23357	21517	31566	29364	36709	27934	334000

Appendix 2: STATA commands

Table 12: STATA commands

STATA commands used to generate the ARIMA forecast.

Description	STATA command
Augmented Dickey-Fuller test, trend and constant	dfuller CH, trend lags(12)
First-order differencing of chargeable hours	generate CHD1=D1.CH
Line diagram	tsline CHD1
Augmented Dickey-Fuller, constant	dfuller CHD1, lags(12)
AC, PAC and Q of first order differencing of chargeable hours	corrgram CHD1, lags(12)
Collect residuals	predict ema11, resid
AC, PAC and Q of residuals	corrgram ema11, lags(12)
ARIMA forecasting	arima CH, arima(0,1,11)