NTNU

Norwegian University of
Science and Technology

Stochastic modelling and prediction of call data from TrønderTaxi.

Tormod Østhus

Master of Science in Physics and Mathematics
Submission date: June 2016
Supervisor: Håkon Tjelmeland, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Preface

This report concludes my Master of Science degree in Industrial Mathematics at the Department of Mathematical Sciences, Norwegian University of Science and Technology (NTNU). The work on this report has been carried out in the months from January to June 2016. I would like to sincerely thank my supervisor at NTNU, Håkon Tjelmeland for his excellent feedback and guidance during these months. I would also like to thank TrønderTaxi for allowing me to use their call data. Finally, I would like to thank Pål Clausen, my primary contact at TrønderTaxi, for bringing this interesting task to the attention of my supervisor.

Tormod Østhus
Trondheim
June 2016

Abstract

In this report we examine a data set from TrønderTaxi, which contains information about all phone calls their call center received from 1 March 2014 to 31 January 2016. We model the number of phone calls received per 30 minute time interval using two different models. The first model is a seasonal autoregressive integrated moving average model, or SARIMA model. The second model is a naive model, which treats all of the weeks of the data set as independent and identically distributed. We use both the SARIMA and naive models to make predictions of both one and two weeks into the future. The results show that the SARIMA model appears to make slightly better predictions for one week into the future, while the naive model is better for more distant predictions.

Sammendrag

I denne rapporten undersøker vi et datasett fra TrønderTaxi som inneholder informasjon om alle telefonsamtaler deres telefonsentral mottok fra 1. mars 2014 til og med 31. januar 2016. Vi modellerer antall mottatte samtaler per 30-minutters tidsintervall ved hjelp av to forskjellige modeller. Den første modellen er en SARIMA-modell. Den andre modellen er en naiv modell, som behandler alle ukene i datasettet som uavhengige og identisk fordelt. Både SARIMA-modellen og den naive modellen ble brukt til å predikere verdier både én og to uker fremover i tid. Resultatene viser at SARIMA-modellen ser ut til å være litt bedre til å predikere én uke fremover i tid, mens den naive modellen er bedre til å predikere lengre fremover i tid.

Contents

1 Introduction

Every day TrønderTaxi receives hundreds of phone calls to their call center from customers wanting to order taxis. The motivation behind this project is to create a model which can predict the number of phone calls per time interval TrønderTaxi will receive during a certain period of time in the future. It is very important for TrønderTaxi to be able to predict the number of calls they will receive during a given time interval, since this information can be used to set up a timetable for the employees at their call center. Thus, TrønderTaxi will be able to have the appropriate number of people at work at any given time. This reduces the amount of overstaffing and understaffing, both of which can be very costly for TrønderTaxi. Exactly how many people this turns out to be depends on how many calls each employee is able to process per hour, and how large a percentage of dropped calls TrønderTaxi finds acceptable. TrønderTaxi has other practical considerations as well. For instance, they cannot ask someone to work for a single hour in the middle of the night just because they expect more calls during that hour. Such practical considerations are not taken into account when we are selecting our models.

Several different attempts have been made to successfully staff a call center. In 1995, queuing theory was used to economically optimise the staffing levels of the telephone operators working for L.L. Bean, a large American telemarketer (Andrews and Parsons, 1993). This was done by expected-total-cost minimisation, with expected savings amounting to more than $500 000 per year. An ARIMA model has successfully been used to predict the number of calls L.L. Bean would receive three weeks into the future (Andrews and Cunningham, 1995). This model used intervention analysis to handle the abnormal number of calls received during the holidays and during advertisement campaigns. Furthermore, ARIMA models have been used for predicting calls to AT&T's call center, and were shown to perform better than a simpler Holt-Winters smoothing (Bianchi et al., 1998). This method also used intervention analysis. The number of calls received by a call center has also been modelled as a Poisson process (Soyer and Tarimcilar, 2008), a bivariate mixed-effects model (Ibrahim and

L'Ecuyer, 2013) and as a Poisson equally weighted moving average (PEWMA) model (Brandt et al., 2000).

In this report we present a SARIMA model for the number of calls received per 30 minute time interval by TrønderTaxi, and use this model to make predictions. We start by performing a preliminary investigation of the data set to look for patterns. To aid this process, we plot short intervals of the data set, as well as the autocorrelation and partial autocorrelation functions of the entire data set. Following this, we build a model based on our investigation of the data set. We start by testing the data set for level and trend stationarities. After this, we use the ACF and PACF in order to make a first guess of the model. We improve the model by adding regressors to account for the difference between the weekdays and weekends. This eventually leads to the final SARIMA model. Estimates of this model are plotted together with the observed values. The standardised residuals are also examined in order to determine how well the model fits the data. After deciding on a model, we use this model to make predictions of the last part of the data set, based on all of the data leading up to the time intervals we want to predict. These predictions are compared to the predictions of a naive model, which only looks at the mean values calculated for every 30 minute time interval of the week. Various error checking statistics for both the SARIMA and naive models are calculated and compared. Finally, we calculate the coverage probability of forecasts of one and two weeks into the future using both the SARIMA and naive models.

In Section 2 of this report theory about time series models is presented. Section 3 contains a preliminary investigation of the data set. In Section 4 we arrive at our final models and use them to perform predictions based on the data set. Finally, in Section 5 we summarise the findings from Section 4 and discuss implications of the results.

2 Time series models

A time series is an ordered sequence of equally spaced data points. It is most common to have ordering through time, but it is also possible to have ordering through space. Examples of time series include yearly tobacco production, daily maximum temperature and the number of calls received at a call center per hour. An important aspect of time series is that they are commonly used to predict future values. A more thorough introduction can be found in the references (Box and Jenkins, 1990). We define a time series Y as

$$Y = \{y_t, t \in T\}, \tag{2.1}$$

where y_t is the value of the time series at time t in the index set T. Before introducing various time series models, we introduce the backshift operator B in order to ease notation. The backshift operator B applied to the time series Y at time t is defined as (Hyndman and Athanasopoulos, 2013)

$$By_t = y_{t-1}, \tag{2.2}$$

giving the value of the previous step of the time series. In the following, we introduce the autocorrelation function (ACF) and partial autocorrelation function (PACF). Following that, we define the autoregressive (AR) and moving average (MA) models, as well as various extensions, culminating in the seasonal autoregressive integrated moving average (SARIMA) model.

2.1 Autocorrelation and partial autocorrelation functions

A time series Y is said to be stationary if its joint probability distribution remains the same when shifted in time. This means that we have

$$E(y_t) = \mu \quad \forall t \in T, \tag{2.3}$$

$$\mathrm{Var}(y_t) = E(y_t - \mu)^2 = \gamma_0 \quad \forall t \in T. \tag{2.4}$$

For a stationary time series, the correlation between y_t and y_{t+k} is also independent of time, and is defined as

$$\rho_k = \frac{\mathrm{Cov}(y_t, y_{t+k})}{\sqrt{\mathrm{Var}(y_t)}\sqrt{\mathrm{Var}(y_{t+k})}} = \frac{\gamma_k}{\gamma_0}, \tag{2.5}$$

and the partial autocorrelation between y_t and y_{t+k} is defined as

$$P_k = \frac{\mathrm{Cov}\left[(y_t - \hat{y}_t), (y_{t+k} - \hat{y}_{t+k})\right]}{\sqrt{\mathrm{Var}(y_t - \hat{y}_t)}\sqrt{\mathrm{Var}(y_{t+k} - \hat{y}_{t+k})}}, \tag{2.6}$$

where $\hat{y}_{t+k}$ is the best linear estimate of y_{t+k} (Wei, 2006). These expressions are known as the ACF and PACF, respectively. Notice that these are not dependent on the time t, but on the lag parameter k.

2.2 The AR(p) model

The autoregressive model of order p, or AR(p) model of a time series is defined as

$$\boldsymbol{\phi}_p(B)\dot{y}_t = e_t, \tag{2.7}$$

where

$$\boldsymbol{\phi}_p(B) = 1 - \sum_{i=1}^{p} \phi_i B^i, \tag{2.8}$$

$\dot{y}_t = y_t - \mu$, and e_t is Gaussian $N(0, 1)$ white noise, while μ is the mean of the time series (Wei, 2006). A process is called a white noise process if it contains uncorrelated random variables with a constant mean and a constant variance (Kay, 1993). For ease of notation, we from now on use y_t when referencing $\dot{y}_t$, without any loss of generality. The AR(p) model assumes that the current value of the time series is dependent on the p previous values. This can be shown by writing out the definition of an AR(1) and AR(2) model with respect to y_t, giving

$$y_t = e_t + \phi_1 y_{t-1}, \tag{2.9}$$

$$y_t = e_t + \phi_1 y_{t-1} + \phi_2 y_{t-2}, \tag{2.10}$$

respectively. Correspondingly, the general AR(p) model can be written as

$$y_t = e_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p}, \tag{2.11}$$

which is clearly dependent on the p previous values of the time series. Given an arbitrary AR(p) model, it is possible to estimate p by looking at plots of the autocorrelation and partial autocorrelation functions of the time series. For an AR(p) model, the theoretical autocorrelation function (ACF) tails off exponentially or as a damped sine wave. Meanwhile, the theoretical partial autocorrelation function (PACF) cuts off after lag p. This means that all values of the theoretical partial autocorrelation after lag p are zero. An example from an AR(2) model with $(\phi_1, \phi_2) = (0.9, -0.5)$ is shown in Figure 1. In this case, we see that the ACF appears to tail off as a damped sine wave, while the PACF cuts off after lag 2.

2.3 The MA(q) model

The moving average model of order q, or MA(q) model of a times series is defined as

$$y_t = \boldsymbol{\theta}_q(B) e_t, \tag{2.12}$$

where

$$\boldsymbol{\theta}_q(B) = 1 - \sum_{i=1}^{q} \theta_i B^i \tag{2.13}$$

and e_t is as defined above. We rewrite the MA(1) and MA(2) models with respect to y_t, giving

$$y_t = e_t + \theta_1 e_{t-1}, \tag{2.14}$$

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}. \tag{2.15}$$

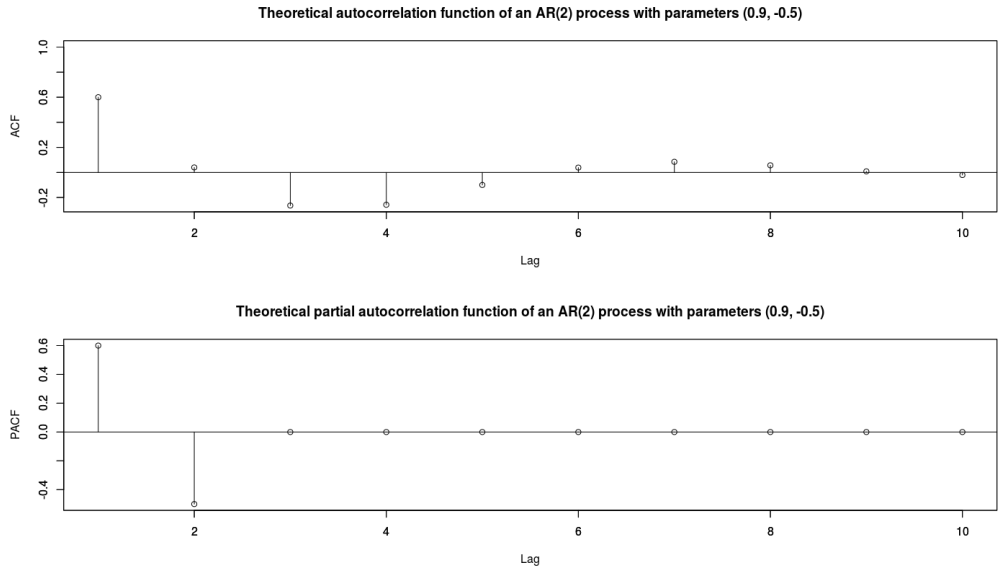

Figure 1: Theoretical ACF and PACF of an AR(2) series with $(\phi_1, \phi_2) = (0.9, -0.5)$.

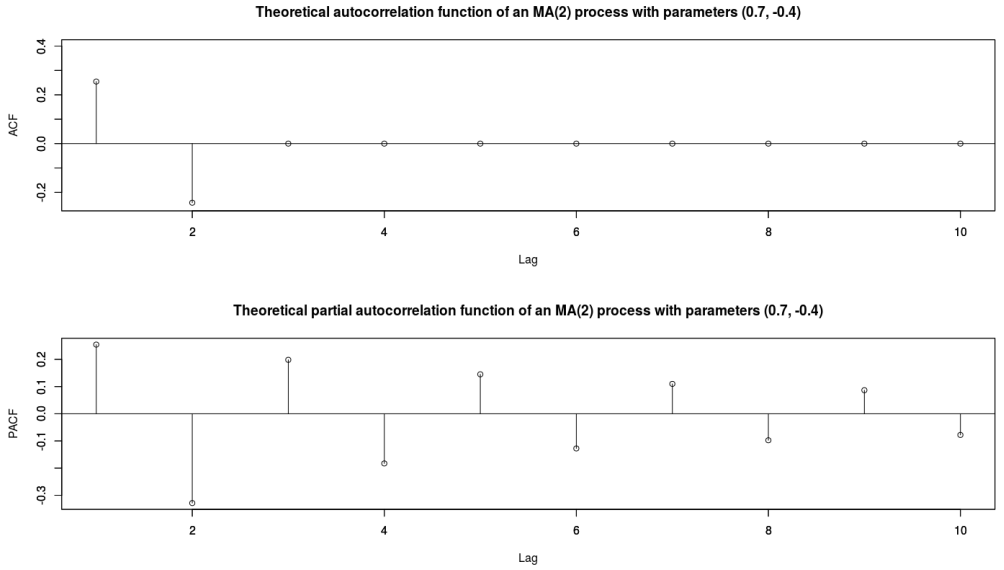

Figure 2: Theoretical ACF and PACF of an MA(2) series with $(\theta_1, \theta_2) = (0.7, -0.4)$.

This leads to the the general MA(q) model with respect to y_t being

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}. \tag{2.16}$$

For MA(q) models, it is also possible to estimate q by looking at the ACF and PACF of the time series. For an MA(q) model, the ACF cuts off after lag q, while the PACF tails off exponentially or as a damped sine wave. An example from an MA(2) model with $(\theta_1, \theta_2) = (0.7, -0.4)$ is shown in Figure 2. In this case, the ACF cuts off after lag 2, while the PACF appears to tail off as a damped sine wave.

Table 1: Characteristics of the ACF and PACF of AR(p), MA(q) and ARMA(p, q) models.

Model	ACF	PACF
AR(p)	Tails off	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off
ARMA(p, q)	Tails off after lag $(q - p)$	Tails off after lag $(p - q)$

2.4 The ARMA(p, q) model

The autoregressive moving average model, or ARMA(p, q) model, is a combination of the AR(p) and MA(q) models described above, and is defined as

$$\boldsymbol{\phi}_p(B)y_t = \boldsymbol{\theta}_q(B)e_t. \tag{2.17}$$

This model is used when the time series contains both an autoregressive and moving average process. We see from the definition that setting $q = 0$ or $p = 0$ gives us the previously defined AR(p) and MA(q) models, respectively. It is more difficult to estimate the values of p and q in an ARMA(p,q) model, since the theoretical ACF tails off after lag $(q - p)$, and the theoretical PACF tails off after lag $(p - q)$. An example of the theoretical ACF and PACF of an ARMA(1,1) model with $(\phi_1 = 0.9, \theta_1 = 0.5)$ is shown in Figure 3. A summary of the characteristics of the ACF and PACF for an ARMA(p, q) model is presented in Table 1 (Rosadi et al., 2012).

2.5 Homogeneous nonstationarity and differencing

Until now, the models we have defined all assume that the time series is stationary. A process is said to be homogeneous nonstationary if its behaviour stays the same, apart from the differences in local means. If this is the case, differencing an appropriate number of times d will make the time series stationary.

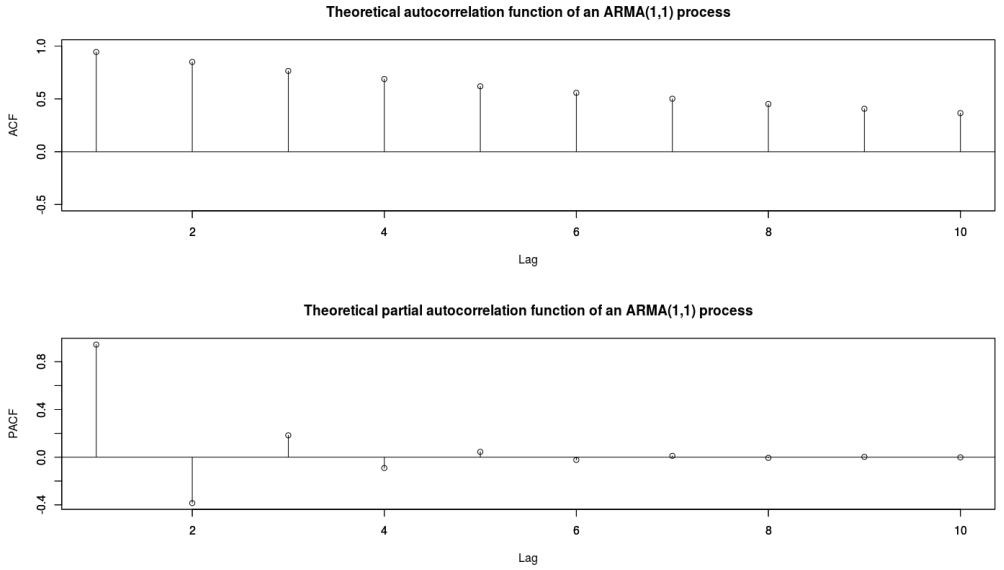

Figure 3: Theoretical ACF and PACF of an ARMA(1,1) series with ($\phi_1 = 0.9, \theta_1 = 0.5$).

Differencing y_t d times is defined as

$$(1 - B)^d y_t, \tag{2.18}$$

where B is the previously defined backshift operator. Rewriting this, with $d = 1$ and $d = 2$, we get

$$y'_t = y_t - y_{t-1}, \tag{2.19}$$

$$y''_t = y_t - 2y_{t-1} + y_{t-2}. \tag{2.20}$$

Intuitively, when we look at y'_t, we look at the difference between every successive value of the original time series. Differencing can be combined with the ARMA(p,q) model to make the autoregressive integrated moving average model, or ARIMA(p,d,q) model. This model is defined as

$$\boldsymbol{\phi}_p(B)(1 - B)^d y_t = \boldsymbol{\theta}_q(B)e_t. \tag{2.21}$$

As previously explained, the p corresponds to the order of the autoregressive part, while the q corresponds to the order of the moving average part. The d corresponds to the degree of first differencing involved.

2.6 The SARIMA$(p, d, q)(P, D, Q)_m$ model

When a time series appears to contain seasonality, a seasonal autoregressive integrated moving average (SARIMA) model may be used. The SARIMA(p,d,q)(P,D,Q)$_m$ model is defined as

$$\boldsymbol{\phi}_p(B)\boldsymbol{\Phi}_P(B^m)(1 - B)^d(1 - B^m)^D y_t = \boldsymbol{\theta}_q(B)\boldsymbol{\Theta}_Q(B^m)e_t, \tag{2.22}$$

with $\boldsymbol{\phi}_p(B)$ and $\boldsymbol{\theta}_q(B)$ as defined for the ARIMA(p,d,q) model. $\boldsymbol{\Phi}_P(B)$ and $\boldsymbol{\Theta}_Q(B)$ are defined similarly, with

$$\boldsymbol{\Phi}_P(B) = 1 - \sum_{i=1}^{P} \Phi_i B^i \tag{2.23}$$

and

$$\boldsymbol{\Theta}_Q(B) = 1 - \sum_{i=1}^{Q} \Theta_i B^i. \tag{2.24}$$

Similarly to the nonseasonal ARIMA model, the P corresponds to the order of the seasonal autoregressive part, and the Q corresponds to the order of the seasonal moving average part. The D corresponds to the degree of first differencing involved in the seasonal part of the model, while m represents the number of periods in each season. For instance, if we were dealing with monthly data, the period would be $m = 12$.

2.7 Adding regressors to the SARIMA model

It is possible to modify the original SARIMA model to include regressors. The model then becomes

$$\boldsymbol{\phi}_p(B)\boldsymbol{\Phi}_P(B^m)(1-B)^d(1-B^m)^D(y_t - \boldsymbol{\beta}^T\boldsymbol{x}_t) = \boldsymbol{\theta}_q(B)\boldsymbol{\Theta}_Q(B^m)e_t, \tag{2.25}$$

where $\boldsymbol{x}$ is a matrix with a number of rows equal to the length of the time series, and a number of columns equal to the number of desired regressors. The vector $\boldsymbol{\beta}$ contains the values of the regressor parameters, indicating in what way the various regressors affect the time series. This is a very important expansion of the model. It allows us to include additional information in the model, such as information about the weather, holidays and so on.

2.8 Variance-stabilising transformations

If a time series is found to be nonstationary in variance, a variance-stabilising transformation is performed. The most common variance-stabilising transformation is the Box-Cox power transformation, which is defined as

$$T(y_t) = \frac{y_t^\lambda - 1}{\lambda}, \tag{2.26}$$

where λ can be any value. However, if $\lambda = 1$ we do not get a transformation. The next step is to test for level stationarity and seasonal stationarity.

2.9 Tests of level stationarity

Tests of level stationarity are used to determine the value of d in the SARIMA$(p,d,q)(P,D,Q)_m$ model. There are several tests of level stationarity, including the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and the augmented Dickey-Fuller (ADF) tests.

2.9.1 The KPSS test of level stationarity

The KPSS test has a null hypothesis that a time series is trend stationary or level stationary. By trend stationary, we mean that the time series is stationary around a deterministic trend. Level stationarity means that the time series behaves like white noise. The KPSS test assumes that the time series can be written as

$$y_t = \xi t + r_t + \varepsilon_t,$$

where ξ is a deterministic trend, r_t is a random walk, and ε_t are independent and identically distributed as $N(0, \sigma_\varepsilon^2)$. The random walk can be written as $r_t = r_{t-1} + u_t$, where u_t are independent and identically distributed as $N(0, \sigma_u^2)$. The stationarity hypothesis is $\sigma_u^2 = 0$. If this is true, y_t is either trend stationary around ξ, or level stationary around r_0 (if $\xi = 0$). The test statistic is defined as (Kwiatkowski et al., 1992)

$$\text{KPSS} = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_e^2}. \tag{2.27}$$

In this case

$$S_t = \sum_{i=1}^{t} e_i, \qquad t = 1, 2, \cdots, T, \tag{2.28}$$

where $e_t = y_t - \bar{y}$, and $\hat{\sigma}_e^2$ is the sum of the squared residuals, divided by T.

2.9.2 The ADF test of level stationarity

The ADF unit root test is an augmented version of the Dickey-Fuller test (Dickey and Fuller, 1979), and assumes the model

$$(1-B)y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1(1-B)y_{t-1} + \cdots + \delta_{p-1}(1-B)y_{t-p+1} + \varepsilon_t, \tag{2.29}$$

where α, β are constants, $\delta_1, \cdots \delta_{p-1}$ are parameters and p is the lag order of the autoregressive process. A common approach to determining the value of p is to start with a large value, and then reduce the number of lags sequentially until the longest lags are significant. The null hypothesis for the augmented Dickey-Fuller test is $\gamma = 0$, while the alternative hypothesis is $\gamma < 0$. The test statistic is

$$\text{ADF} = \frac{\hat{\gamma}}{SE(\hat{\gamma})}. \tag{2.30}$$

If the test statistic is more negative than the critical value, the null hypothesis is rejected and there is no unit root.

2.10 Test of seasonal stationarity

Testing for seasonal stationarity is a way of determining the value of the D in a SARIMA$(p,d,q)(P,D,Q)_m$ model. There are several different tests for seasonal stationarity, but this paper will only briefly explain the Canova and Hansen (CH) test (Canova and Hansen, 1995). This test for seasonal stationarity has a null hypothesis that the seasonal pattern of a time series is deterministic. It assumes that the time series can be written as

$$y_t = \mu + x_t^{\mathrm{T}}\beta + S_t + e_t,$$

where x_t^{T} is a vector of explanatory variables, S_t is a deterministic seasonal component, and e_t is an uncorrelated error with a $N(0, \sigma_e^2)$ distribution. The CH test resembles the KPSS test described earlier, as they are both based on the Lagrange Multiplier statistic.

2.11 Competing models and parameter estimation

After deciding on the values of (p, d, q, P, D, Q), the parameter values are estimated by maximum likelihood estimation. At this point, we may have several competing models which all could be candidates for the best model. This could for instance be the case if it is difficult to discern (p, q) from the ACF and PACF. To decide on which model to use, we can calculate various information criteria. There are several information criteria that can be used to select the best model, including the AIC (Akaike, 1974), AIC_c and BIC.

The Akaike information criterion (AIC) is defined as (Burnham and Anderson, 2004)

$$\text{AIC} = -2\ln(L) + 2k, \tag{2.31}$$

where $\ln(L)$ is the natural logarithm of the maximum likelihood and k is the total number of parameters within the model. For finite sample sizes, it may be more beneficial to use the AIC_c, defined as

$$\text{AIC}_\text{c} = \text{AIC} + \frac{2k(k+1)}{n-k-1}, \tag{2.32}$$

where n is the sample size. This information criterion penalises having a large number of parameters more than the AIC does. However, if n is much larger than k^2, the difference between the two criteria becomes negligible. Extensive simulation work shows that the AIC_c is the preferred criterion of the two (Canova and Hansen, 1995). For large n, the AIC_c converges to the AIC. Finally, the Bayesian information criterion (BIC) is defined as

$$\text{BIC} = -2\ln(L) + k\ln(n). \tag{2.33}$$

2.12 Error checking

Another way to check how good various models are, is by looking at how good they are at forecasting, or predicting future values. This can be measured

by error checking. There exists several error checking criteria, including the mean percentage error (MPE), mean squared error (MSE), mean absolute error (MAE) and the mean absolute percentage error (MAPE). Given a forecast interval of length l, the MPE is defined as (Wei, 2006)

$$\text{MPE} = \frac{1}{l}\sum_{t=1}^{l}\frac{y_t - \hat{y}_t}{y_t}, \tag{2.34}$$

where y_t is the observed value of the time series at time t, and $\hat{y}_t$ is the predicted value of the time series at time t. One advantage of the MPE is that it measures forecast bias. If a forecast is biased, it will generally predict values that are either greater or less than the observed values. The MPE is only defined if none of the observed values are zero. The MSE can be used regardless of whether any of the observed values are zero, and is defined as (Wei, 2006)

$$\text{MSE} = \frac{1}{l}\sum_{t=1}^{l}(y_t - \hat{y}_t)^2. \tag{2.35}$$

A possible disadvantage of the MSE is that outliers are heavily weighted. This happens because all of the errors are squared. The MAE is defined as (Wei, 2006)

$$\text{MAE} = \frac{1}{l}\sum_{t=1}^{l}|y_t - \hat{y}_t|. \tag{2.36}$$

The MAPE is defined as (Wei, 2006)

$$\text{MAPE} = \frac{1}{l}\sum_{t=1}^{l}\left|\frac{y_t - \hat{y}_t}{y_t}\right|. \tag{2.37}$$

As with the MPE, the MAPE is not defined if any of the observed values are zero. Another disadvantage of the MAPE is that it penalises negative errors heavier than positive errors. This can be seen by calculating the MAPE for $(y_t = 150, \hat{y}_t = 100)$ and $(y_t = 100, \hat{y}_t = 150)$. The first case gives a MAPE of $50/150 = 0.33$, while the second case gives a MAPE of $50/100 = 0.50$, even though both of the predicted values are equally far away from the observed values.

3 Introducing the data set

A preliminary investigation of the data set provided by TrønderTaxi is performed before deciding on a specific model. Before plotting the data, it is discretised from near-continuous time to intervals of 30 minutes. Doing this makes the data set computationally easier to work with, without sacrificing any degree of generality needed by TrønderTaxi. A similar investigation has previously been performed on a truncated version of this data set (Østhus, 2015). The number of calls received per 30 minutes by TrønderTaxi for the period 01.03.2014 through 31.01.2016 is plotted in Figure 4. We immediately notice that there are a couple of time periods where there were no recorded calls, namely for intervals in June 2014, May 2015 and January 2016. The reason for these gaps is that TrønderTaxi had technical difficulties during those time intervals. Hence, the information about the calls received during those days was not recorded.

3.1 Data analysis

The local peak values of the data set appear to be evenly spaced, indicating a form of seasonality in the data set. This is what we would expect to find, since we know that more people order taxis at certain times of the day, and certain days of the week. To further verify this, we plot the number of calls received per 30 minutes for two consecutive weeks on top of each other. This is shown in Figure 5. We see that the number of calls received per 30 minute interval for two consecutive weeks appear to be similar to each other. This was further corroborated by looking at several other pairs of consecutive weeks of the data set. Since the weeks resemble each other, it is natural to plot an average week from the entire data set. The average week is found by taking

$$\tilde{\mu}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} y_{j+336(i-1)}, \qquad j = 1, \cdots, 336, \tag{3.1}$$

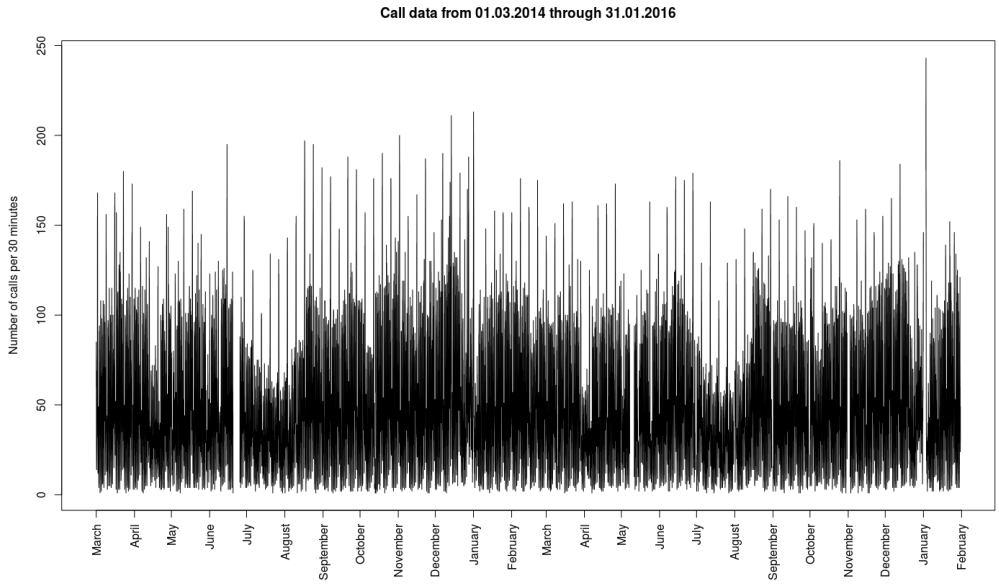

Figure 4: Number of calls received by TrønderTaxi per 30 minute interval from 01.03.2014 through 31.01.2016.

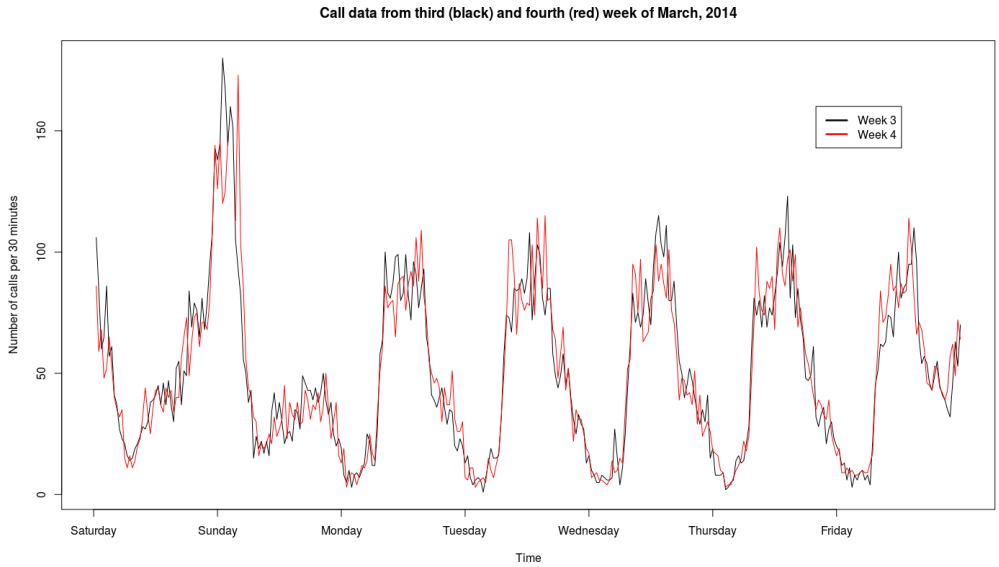

Figure 5: Plot of the number of calls received per 30 minutes for the third and fourth week of the data set, starting in March 2014. The data from the third and fourth weeks are shown in black and red, respectively. It should be noted that there are no holidays during this time period.

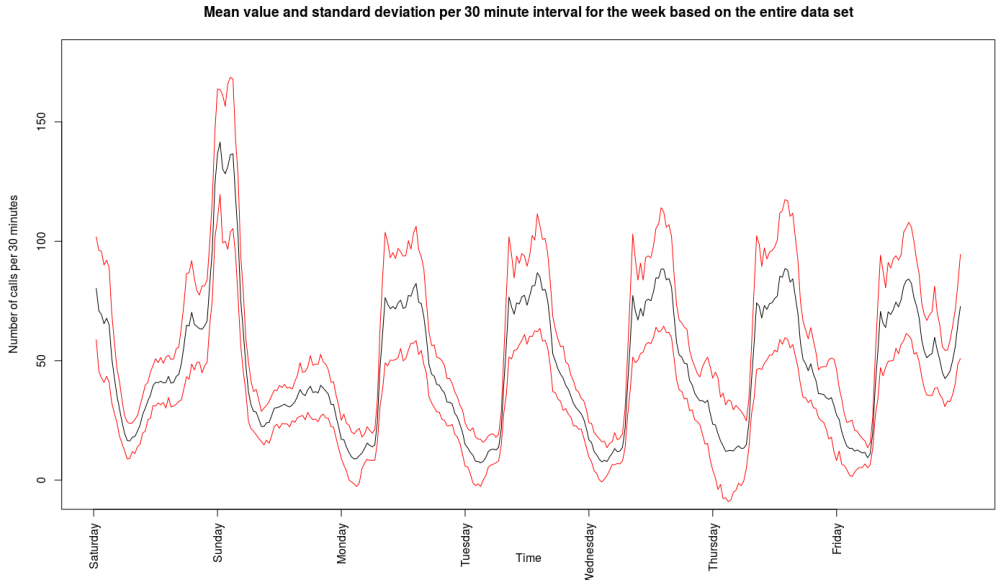

Figure 6: Plot of the mean value with standard deviation of number of calls received per 30 minutes of the week.

where N_j is the number of 30 minute time intervals in the jth time interval of the week. The index j goes from 1 to 336 because there are 336 intervals of 30 minutes in a week. The estimated standard deviation for each 30 minute time interval of the week is defined as

$$\tilde{\sigma}_j = \sqrt{\frac{1}{N_j}\sum_{i=1}^{N_j}(y_{j+336(i-1)} - \tilde{\mu}_j)^2} \quad j = 1, \cdots, 336, . \tag{3.2}$$

The plot of this average week, as well as its standard deviation, based on the entire data set is found in Figure 6. We observe that the weekdays Monday through Thursday seem to follow a very similar pattern, while the weekend days Friday, Saturday and Sunday are very different. For instance, there are more calls during the last hours of Friday than during the last hours of the other weekdays. The rest of the Friday behaves quite similarly to the other weekdays.

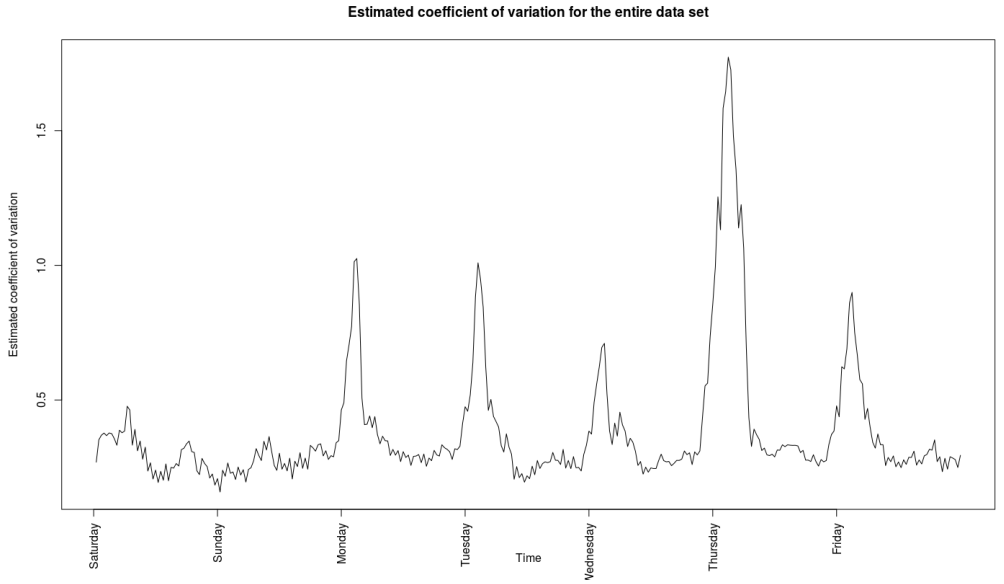

Figure 7: Plot of the estimated coefficient of variation per 30 minutes for the entire data set.

3.2 Coefficient of variation

Another way of looking at how much the number of calls received per 30 minute interval varies, is to plot the coefficient of variation c_v. The coefficient of variation for the jth 30 minute time interval of the week is defined as

$$c_v^j = \frac{\tilde{\sigma}_j}{\tilde{\mu}_j}, \qquad j = 1, \cdots, 336, \tag{3.3}$$

where $\tilde{\mu}_j$ is the estimated mean value of the jth 30 minute time interval of the week, and $\tilde{\sigma}_j$ is the estimated standard deviation of the jth 30 minute time interval of the week. These estimates are defined in equations (3.1) and (3.2), respectively. A plot of the estimated coefficient of variation is found in Figure 7. Here we see that the coefficient of variation is the lowest for Saturday and

Sunday, while the largest values are found in the hours after midnight on the other days of the week. There are several possible reasons for this. We can imagine that a holiday falling on a weekday would have a much larger impact on the number of calls received per 30 minute interval around midnight than if the holiday were to happen during the weekend. From Figure 6, we also see that the mean around midnight on the weekdays is very low. Dividing by these values leads to a large coefficient of variation.

3.3 Comparing the days of the week

From Figure 6, we see that the weekdays resemble each other more than the other days of the week. To further investigate how similar the average days are to each other, we plot the average days of the week on top of each other in Figure 8. This figure confirms that the weekdays are the most similar to each other, especially Monday, Tuesday, Wednesday and Thursday. Friday follows a similar pattern to the other weekdays in the beginning, but we see that there are on average more calls received on Friday evenings than the other weekdays. We also observe that there are fewer calls received during daytime on Saturday and Sunday, but much more calls received during the evenings and nights. Figure 8 shows that the two busiest nights are the nights between Friday and Saturday, and Saturday and Sunday. This makes sense, since many people want to get earlier home on the night to Monday, due to work or school in the morning.

3.4 ACF and PACF

It may also be of interest to plot the autocorrelation and partial autocorrelation functions of the data set to look for underlying patterns. The autocorrelation function is plotted in Figure 9, while the partial autocorrelation function is plotted in Figure 10. The autocorrelation function takes the shape of a damped sine wave. We see that the period of the damped sine wave is 48, corresponding to a period of 24 hours. This means that there is a positive correlation between

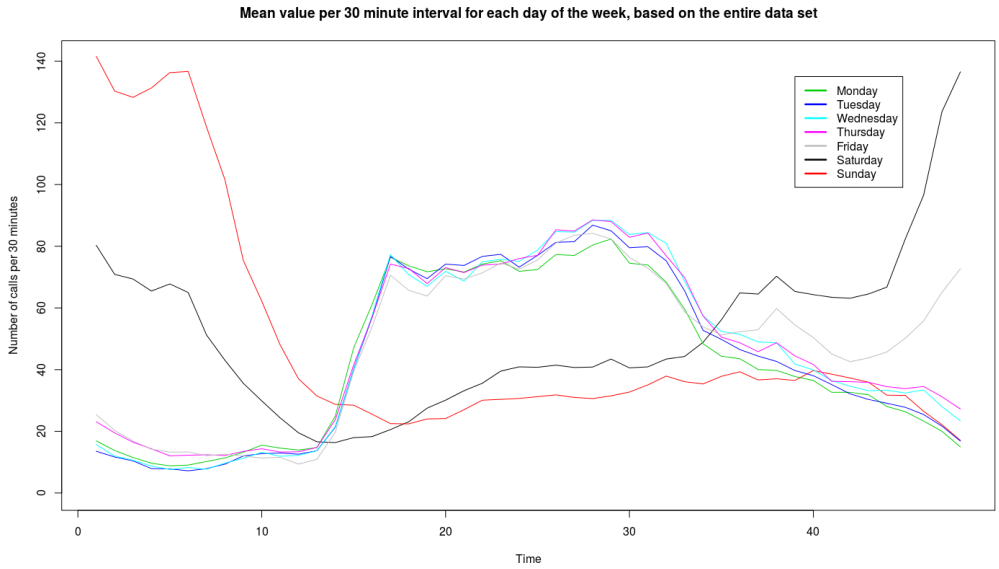

Figure 8: Plot of the mean values per 30 minute interval for each day of the week on top of each other.

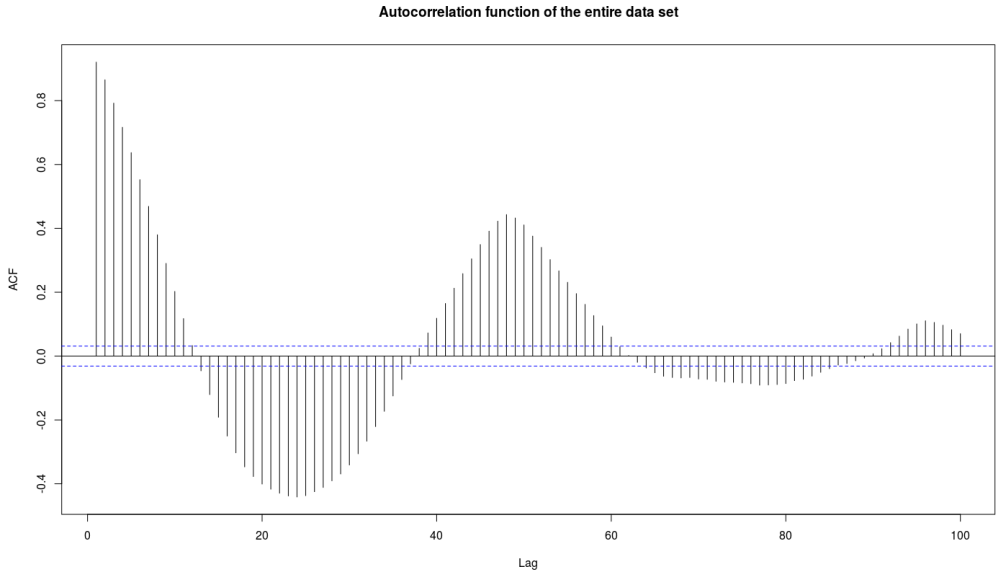

Figure 9: Plot of the autocorrelation function of the entire data set.

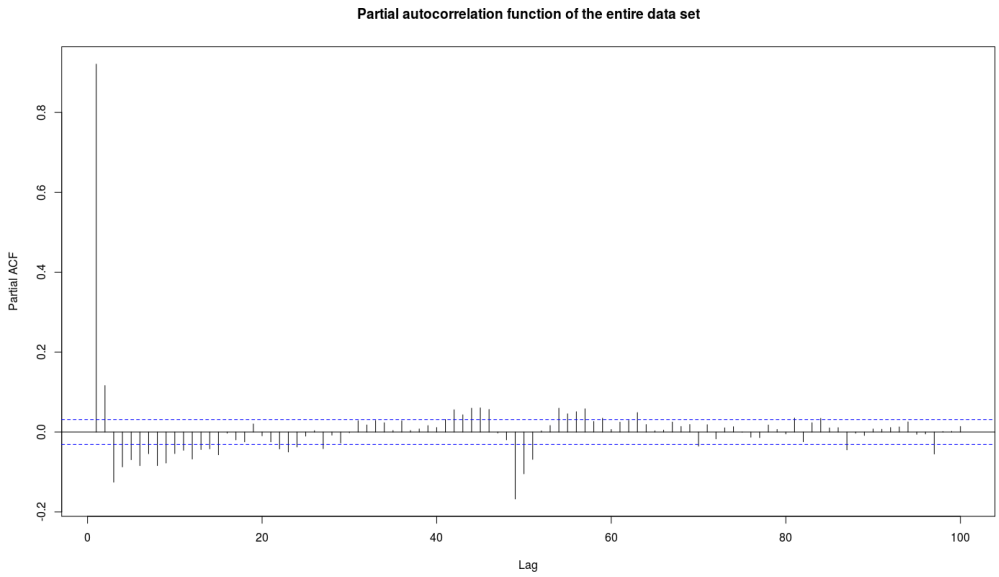

Figure 10: Plot of the partial autocorrelation function of the entire data set.

the number of calls in a 30 minute time interval on one day and the same 30 minute time interval on the following day. We also observe that the largest negative correlation comes after 24 time lags, or 12 hours. From Figure 10, we observe that the partial autocorrelation function appears to cut off quickly. The implications of these plots are discussed in further detail later.

4 Model building

In this section, we present models for the time series data obtained from TrønderTaxi. A series of tests are performed on the data set to ensure fitting models, followed by an examination of the standardised residuals, forecasting and error checking diagnostics.

4.1 Tests of level stationarity and seasonal stationarity

To determine whether the call data provided by TrønderTaxi is level stationary, we perform the KPSS and ADF tests in R, using the functions kpss.test and adf.test found in the tseries package (Trapletti and Hornik, 2016). The KPSS test returned a p-value greater than 0.05, so we cannot reject the null hypothesis of level stationarity. Meanwhile, the ADF test returned a p-value less than 0.05, leading us to reject the null hypothesis of nonstationarity. Together, the results of these two tests imply that the time series is level stationary. This leads us to initially set $d = 0$.

Next, we check for seasonal stationarity. This is done by using the R function nsdiffs in the forecast package (Hyndman and Khandakar, 2008). This function uses the Canova-Hansen test in order to determine the number of seasonal differences required to make the time series stationary in season. To check for seasonal stationarity, we first decide on the period m of the data set. In this case, we choose $m = 48$, since we are dealing with daily data and considering 30 minute time intervals. For this time series, the result of the CH test was 0, leading us to believe that the time series is already stationary in season without the need for seasonal differencing. Similarly to the conclusion of the test for level stationarity, this leads us to initially set $D = 0$.

4.2 Examining ACF and PACF

After deciding on initial values for d and D, we examine the plots of the autocorrelation and partial autocorrelation functions of the data set found in Figures 9 and 10, respectively. This is done in order to obtain an initial guess for (p, q, P, Q). From Figure 9, we see that the autocorrelation function of the data set takes the shape of a damped sine wave. The partial autocorrelation function of the data set plotted in Figure 10 does not give as clear an answer. However, we observe that the partial autocorrelation quickly falls off after lag 1. The combined observations from Figures 9 and 10 lead us to set $(p = 1, q = 0)$ as initial values.

We find initial values for P and Q by looking at every m lags of the autocorrelation function and partial autocorrelation function of the data set. Plots of the ACF and PACF with a larger maximum lag can be seen in Figure 11. The ACF appears to tail off, while the PACF appears to cut off after two periods of m. This leads to an initial guess of $(P = 2, Q = 0)$.

4.3 Adding regressors to the model

Further examination of Figures 6 and 8 leads to the conclusion that all the days of the week cannot be treated equally. However, we see that the number of calls received per 30 minute time interval looks very similar for Mondays through Thursdays. Due to this, we decide to calculate a new average day, based only on the data for Mondays through Thursdays. This is presented in Figure 12. In order to find out how much the various days of the week differ from this average weekday, we examine the difference between the average week plotted in Figure 6 and the average weekday based on Mondays through Thursdays found in Figure 12. This difference is plotted in Figure 13. We see that the number of calls received per 30 minute time interval changes on Fridays, Saturdays and Sundays. This plot is used to create a regressor matrix to be included in the SARIMA model, in order to better fit the data set. The regressor matrix has a number of rows equal to the length of the data set, and a number of columns

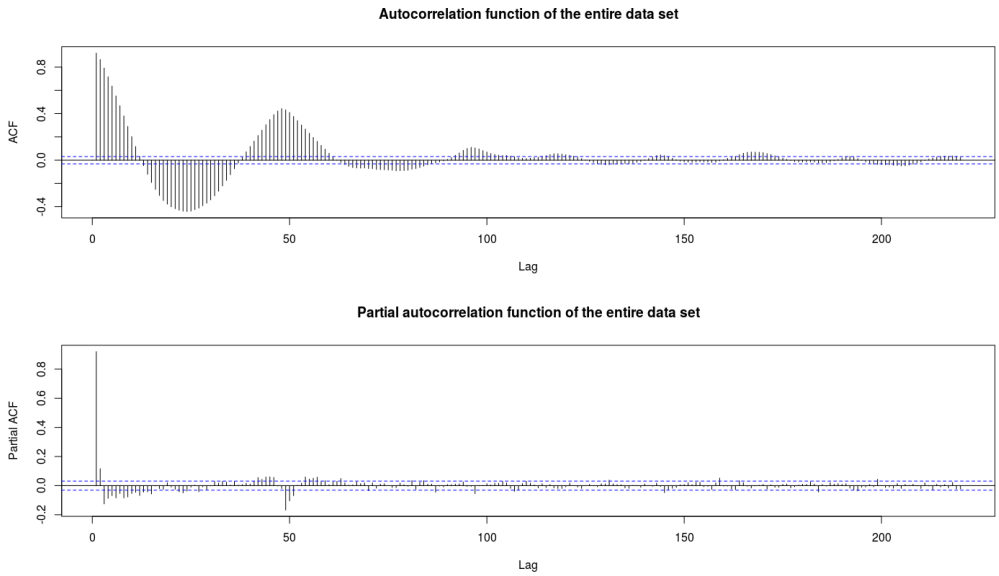

Figure 11: Plot of the autocorrelation function and partial autocorrelation function of the entire data set up to a maximum lag of 220.

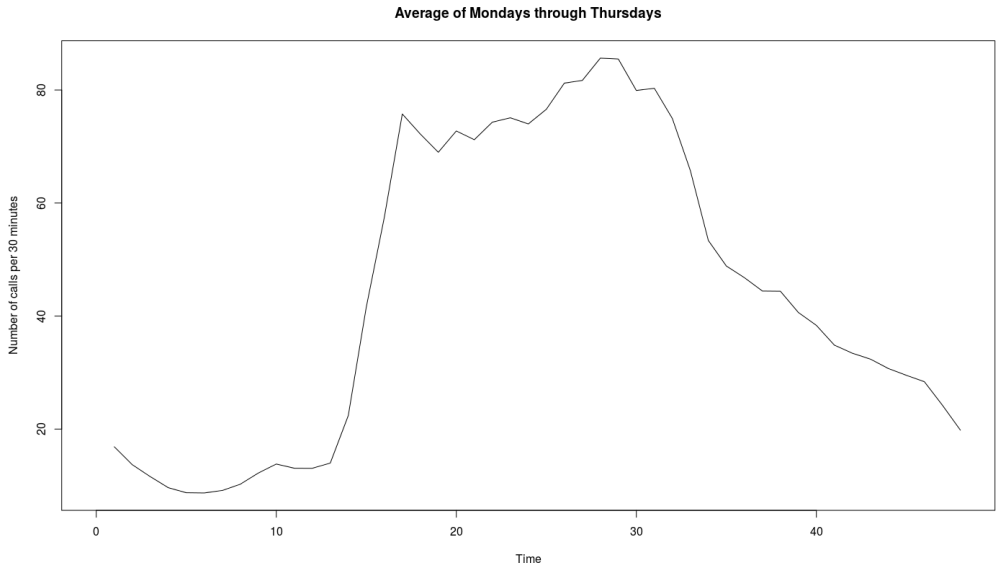

Figure 12: Plot of the average number of calls received per 30 minute time interval, based on all Mondays through Thursdays in the data set.

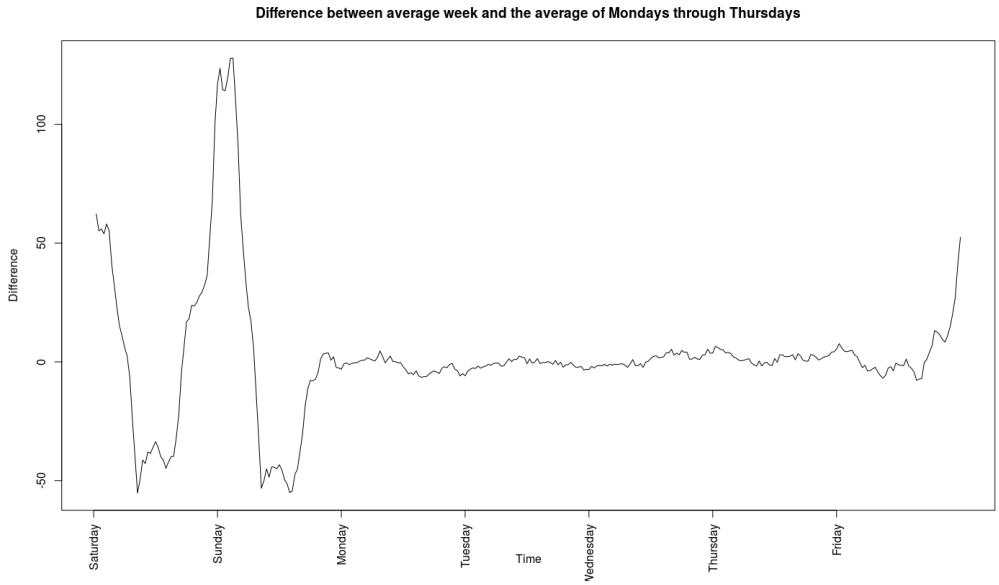

Figure 13: Plot of the difference per 30 minute time interval between the average week and the average of Mondays through Thursdays, based on the entire data set.

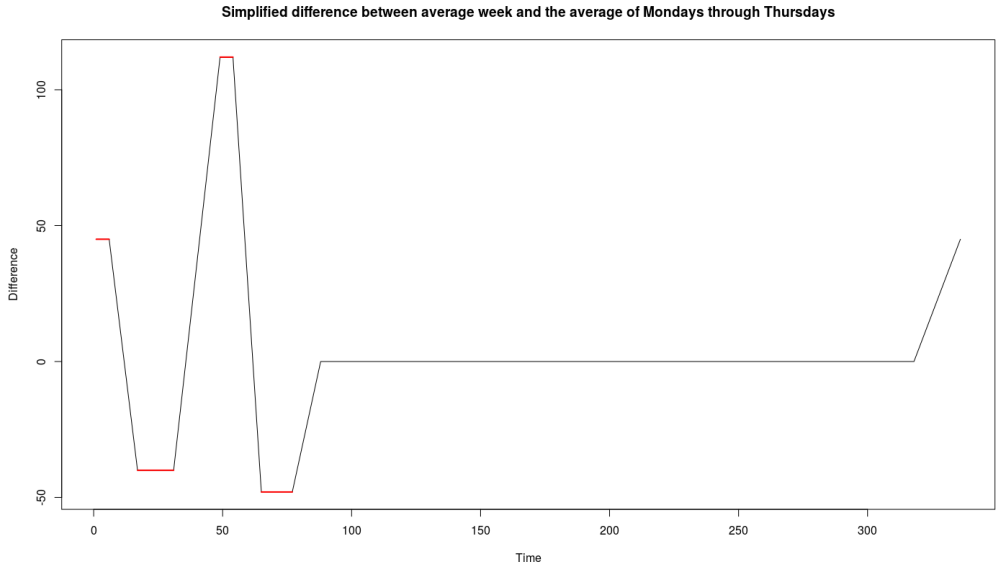

Figure 14: Simplified plot of the difference per 30 minute time interval between the average week and the average of Mondays through Thursdays, based on the entire data set. The four intervals of interest are plotted in red.

equal to the number of regressors. In this case, we choose to model the weekly variation using linear combinations of four regressors. A simplified version of the difference plotted in Figure 13 can be found in Figure 14. The regressors in our regressor matrix correspond to the four constant nonzero time intervals found in Figure 14, corresponding to the time intervals of [1,6], [17,31], [49,54] and [65,77]. These intervals are plotted in red. To make the regressor matrix, we start by making a matrix with 4 columns and 336 rows, one row for each 30 minute time interval of a week. We build the matrix one row at a time. The first 6 rows are only affected by the first regressor, while the rows in the interval [7,16] are affected by a linear combination of the first two regressors. Similarly, the rows in the interval [17,31] are only affected by the second regressor, and the rows in the interval [32,48] are affected by a linear combination of the second and third

regressor. This procedure is repeated until we have covered the entire interval [1,336]. It should be noted that the interval [78,87] only contains a fraction of the fourth regressor, while the interval [319,336] only contains a fraction of the first regressor. After this matrix is built, we repeat it until the number of rows matches the length of the data set. The values of the regressor matrix are normalised to make them easier to interpret later. This means that for the intervals [1,6], [17,31], [49,54] and [65,77], we set the value of the respective regressors to 1. In the parts of the matrix where a row contains two different regressors, we also make sure that their sum is always equal to 1. The sections which only contain fractions of a single regressor are also scaled appropriately.

4.4 Initial model guess

The following procedure is used in order to find the best SARIMA model for the data set. Based on the previous tests of stationarity and seasonal stationarity, and on the ACF and PACF of the original data set, we guess starting values for (p, d, q, P, D, Q). Additionally, a regressor matrix is built as described above. After this, we use a slightly modified Hyndman-Khandakar algorithm (Hyndman and Khandakar, 2008) to find the best SARIMA model for the data set. This is done by testing neighbouring models to our initial guess model. The neighbouring models are defined as models where either one of (p, q, P, Q) varies by ± 1 of the current model, where both (p, q) vary by ± 1 of the current model, or both (P, Q) vary by ± 1 of the current model. The $\mathrm{AIC_c}$ is then calculated for all of these neighbouring models, and the model with the lowest $\mathrm{AIC_c}$ is chosen as our new current model. If any errors occur while estimating neighbouring models, we reject those neighbouring models. The initial guess for our SARIMA model is SARIMA$(1,0,0)(2,0,0)_{48}$. After finding an initial model, we use the R function Arima from the forecast package to calculate the $\mathrm{AIC_c}$ of this model and all neighbouring models. Eventually, we arrive at our final model.

Table 2: Estimated values and standard errors for the parameters of the SARIMA(1,0,0)(2,0,2)$_{48}$ model.

Parameter	Estimate	Standard error
μ	43.9	4.40
ϕ_1	0.787	0.00390
Φ_1	0.175	0.0450
Φ_2	0.824	0.0449
Θ_1	-0.128	0.0403
Θ_2	-0.826	0.0387
β_1	45.5	1.35
β_2	-40.9	1.13
β_3	112	1.40
β_4	-48.6	1.22

4.5 Final model

We end up with a SARIMA(1,0,0)(2,0,2)$_{48}$ model, which can be written as

$$\boldsymbol{\phi}_1(B)\boldsymbol{\Phi}_2(B^{48})(y_t - \boldsymbol{\beta}^T x_t) = \boldsymbol{\Theta}_2(B^{48})e_t, \tag{4.1}$$

or

$$(1-\phi_1 B^1)(1-\Phi_1 B^{49}-\Phi_2 B^{50})(y_t-\mu-\boldsymbol{\beta}^T x_t) = (1-\Theta_1 B^{49}-\Theta_2 B^{50})e_t, \tag{4.2}$$

with the estimated parameter values presented in Table 2. We see from Table 2 that all of the parameters are significantly different from zero. A plot of the estimated number of calls received per 30 minute time interval is shown in Figure 15. Again, we observe the missing values for certain time intervals. This is because the R function Arima from the forecast package does not estimate a value for these time intervals, since the original data was not available. To examine how closely the estimates fit the observed values, we plot the estimated values and observed values on top of each other, for 6 weeks spread fairly evenly

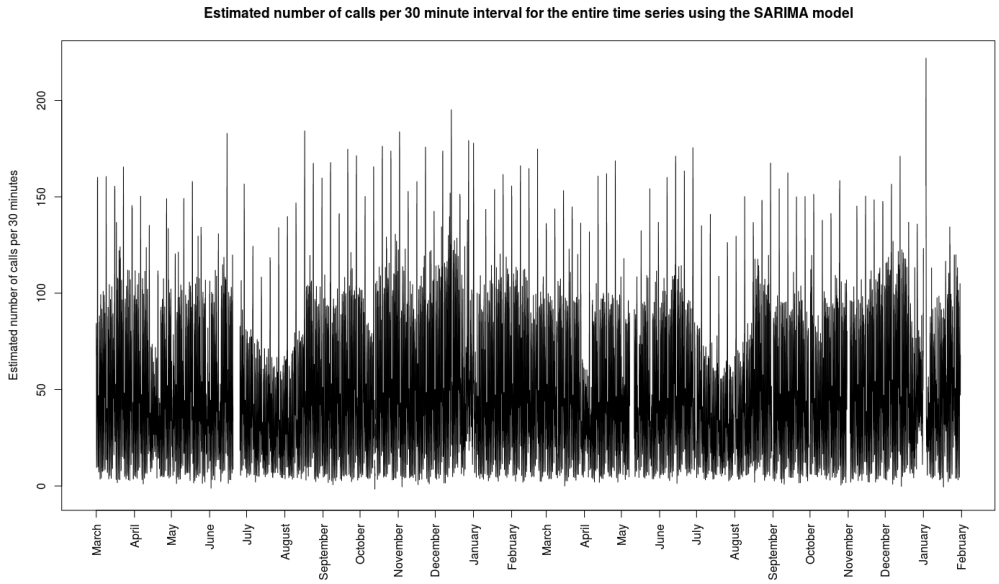

Figure 15: Plot of the estimated number of calls received per 30 minutes from the SARIMA model.

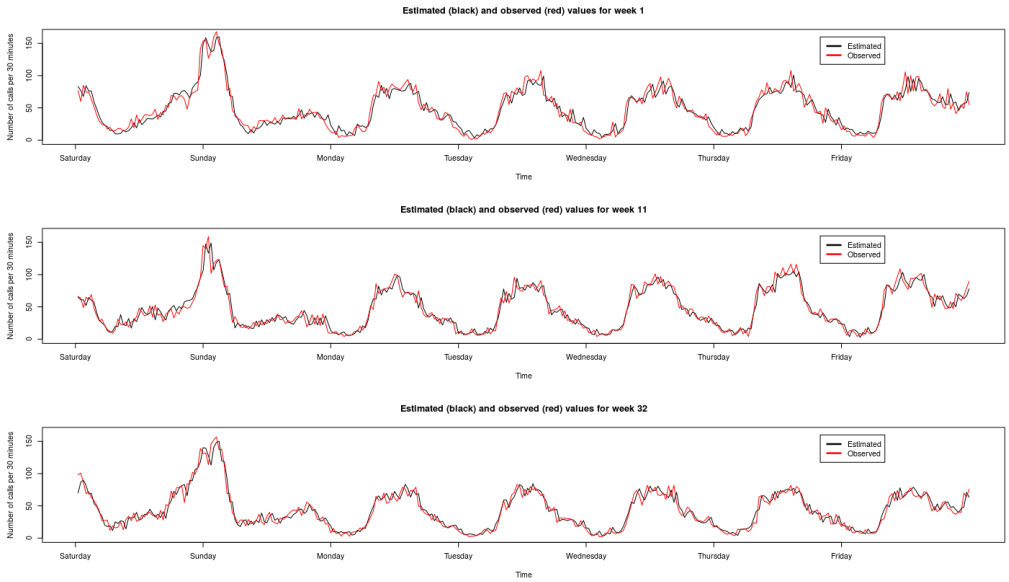

Figure 16: Plot of the estimated number of calls received per 30 minutes for weeks 1, 11 and 32 of the data set, starting in March 2014. The estimated and observed values are shown in black and red, respectively.

through the data set. These weeks are chosen so that they do not contain any holidays. The results are shown in Figures 16 and 17. The plots show that the estimated values fit the observed values fairly well, but the estimated values appear to lag a little behind the observed values. In order to determine how well this SARIMA model fits the data set, we examine the standardised residuals.

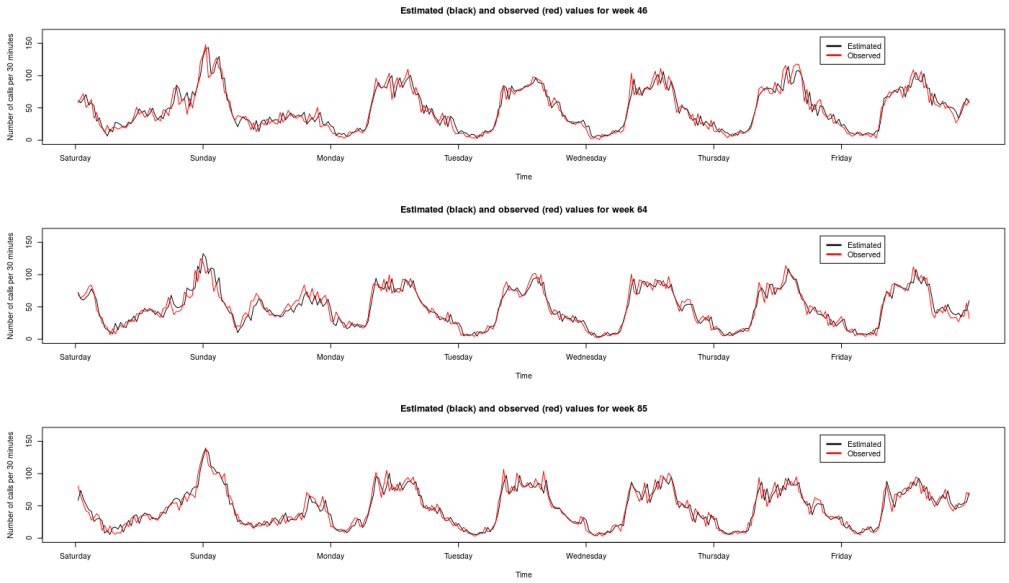

Figure 17: Plot of the estimated number of calls received per 30 minutes for weeks 46, 64 and 85 of the data set, starting in March 2014. The estimated and observed values are shown in black and red, respectively.

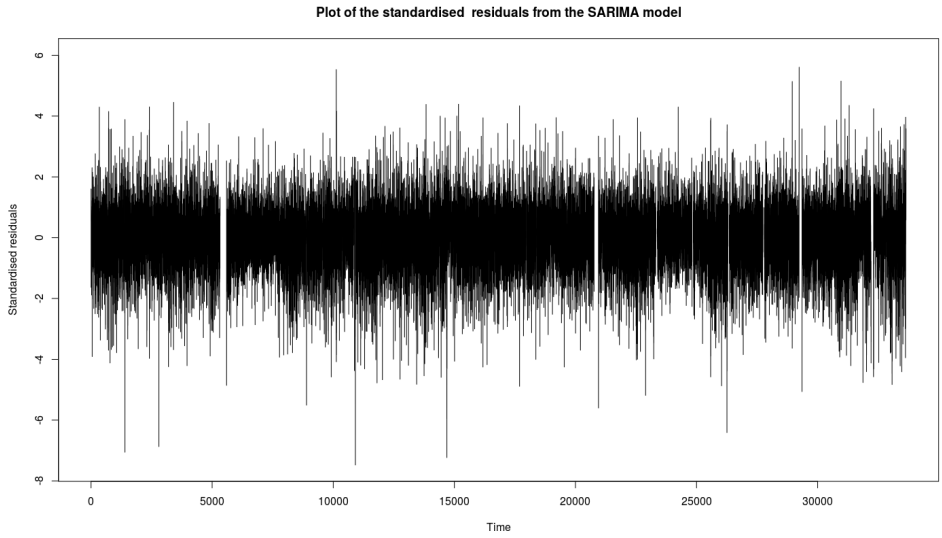

Figure 18: Plot of the standardised residuals of the SARIMA model.

4.6 Examining the standardised residuals

In order to evaluate the final model, we examine the standardised residuals of the model. The standardised residual r_t of a time series at time t is defined as

$$r_t = \frac{y_t - \hat{y}_t}{\hat{\sigma}_t}, \tag{4.3}$$

where $\hat{y}_t$ is the predicted value and $\hat{\sigma}_t$ is the estimated standard deviation of $(y_t - \hat{y}_t)$. A plot of the standardised residuals can be found in Figure 18. The missing standardised residuals in Figure 18 correspond to the missing data in the original data set. Next, we examine the distribution of the standardised residuals. This is done by plotting a histogram of the standardised residuals, as well as a Q-Q plot, shown in Figures 19 and 20, respectively. Figure 19 shows a histogram of the standardised residuals of the SARIMA model, as well as a standard normal distribution. We observe that there are more standardised residuals close to

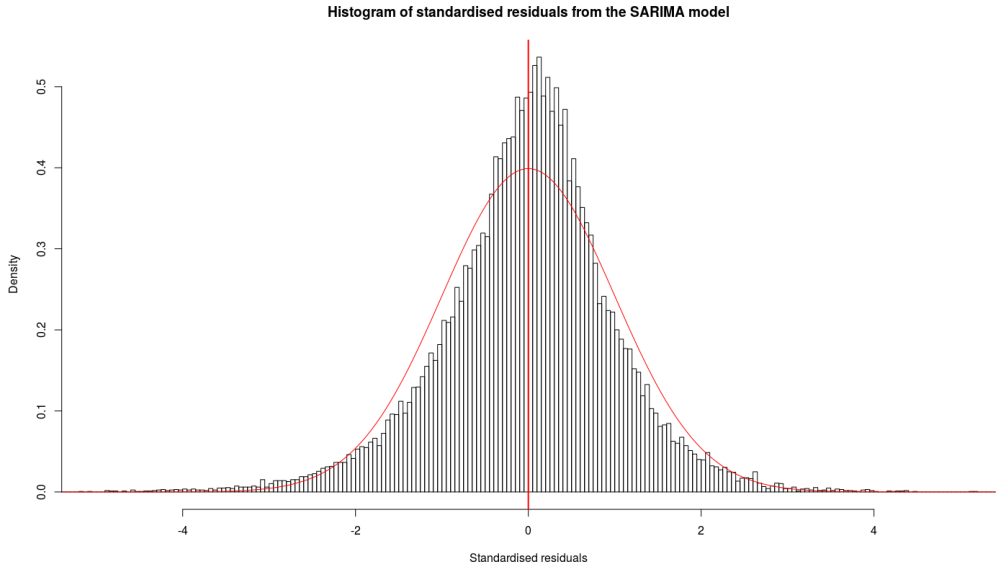

Figure 19: Histogram of the standardised residuals of the SARIMA model. The red curve is a standard normal distribution $N(0, 1)$. A red vertical line is also plotted through zero to make is easier to distinguish the positive standardised residuals from the negative ones.

zero than what would be expected for a standard normal distribution. There also appears to be more slightly positive standardised residuals than slightly negative ones. In addition to this, we observe that the tails of the standardised residuals are slightly heavier than expected. From the Q-Q plot presented in Figure 20, we see that the majority of the standardised residuals follow a straight line. The Q-Q plot also shows that the distribution is heavy-tailed, especially when it comes to the negative standardised residuals. Furthermore, we find that approximately 94 % of the standardised residuals lie within the (-1.96, 1.96) interval. This is slightly less than the 95 % we would have expected to find for a standard normal distribution. It is also of interest to look at the autocorrelation function and partial autocorrelation function of the standardised

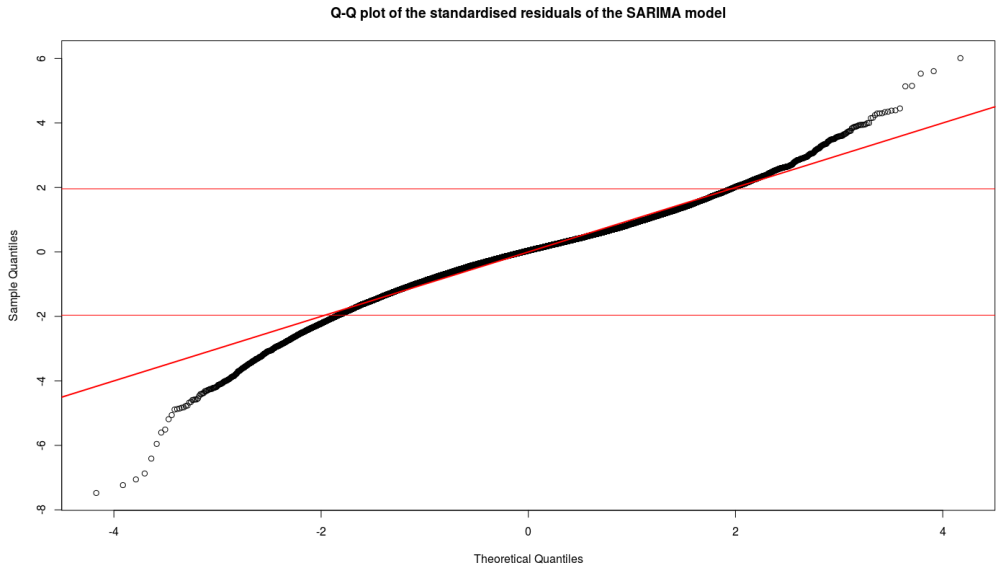

Figure 20: Q-Q plot of the standardised residuals of the SARIMA model. The red horizontal lines go through -1.96 and 1.96.

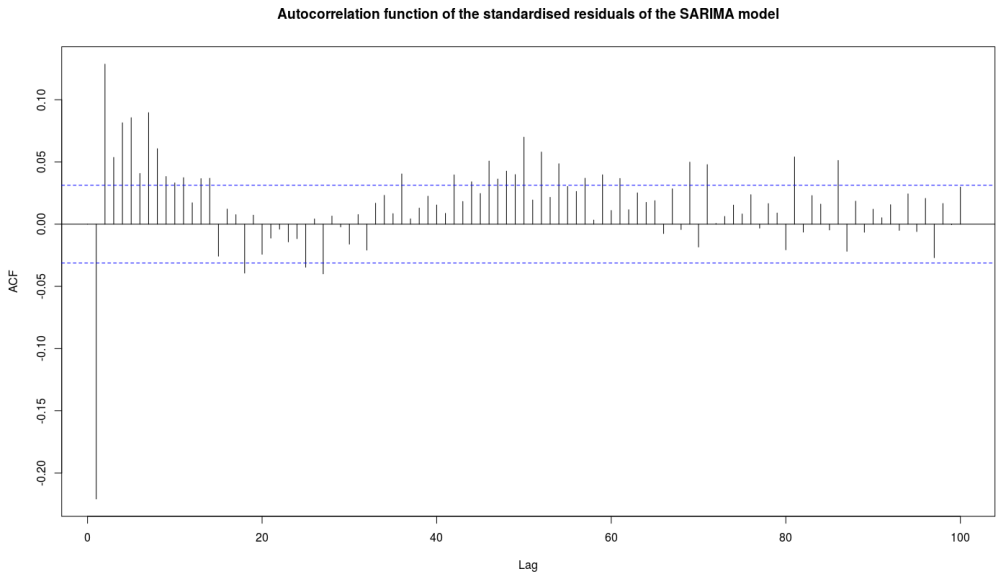

Figure 21: Plot of the autocorrelation function of the standardised residuals of the SARIMA model.

residuals. These plots can be found in Figures 21 and 22, respectively. If the SARIMA model was a perfect fit for the data set, we would expect the standardised residuals to behave like a white noise process. For a white noise process, there is no correlation between the values in the time series. We see from Figures 21 and 22 that this is clearly not the case for our standardised residuals. Both the ACF and PACF seem to have a larger peak at the 48th time lag, in addition to a large peak at lag 1. This could indicate that there is still some seasonality in the data set which is not accounted for by the SARIMA model.

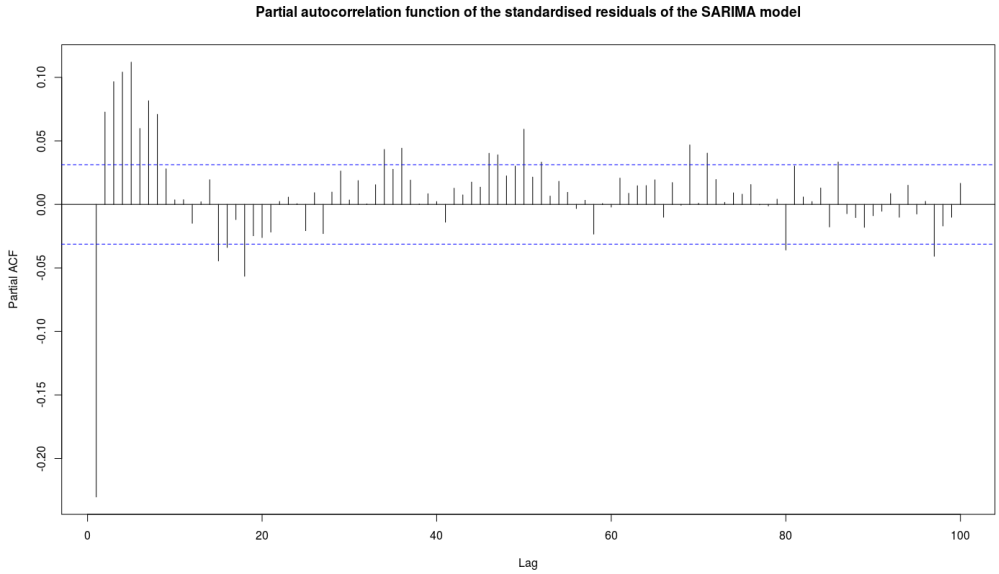

Figure 22: Plot of the partial autocorrelation function of the standardised residuals of the SARIMA model.

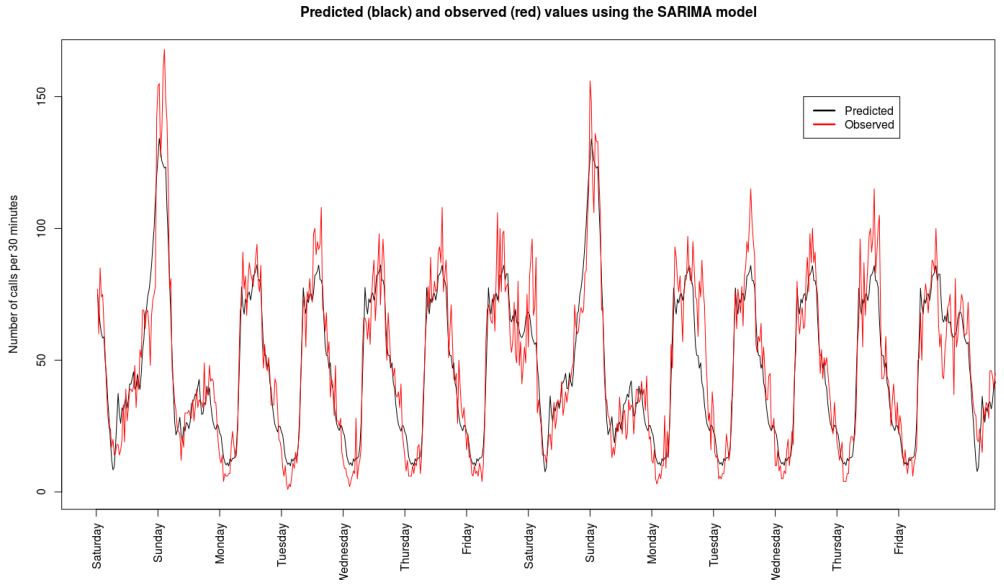

Figure 23: Plot of the predicted number of calls received per 30 minute time interval for two consecutive weeks, starting on 21 November 2015, using the SARIMA model. The predicted and observed values are shown in black and red, respectively.

4.7 Prediction and comparison with naive model

Once we have settled on a model and checked how well it fits the underlying data set, we use the model to make predictions. In order to test how good the predictions of the model are, we estimate the parameters of the SARIMA$(1,0,0)(2,0,2)_{48}$ models based on only a part of the data set. This allows us to perform predictions that we can compare to observed values. In this case, we decide to include the first 90 weeks of the data set. This means that the first week we predict starts on Saturday 21 November 2015. A plot of the first two weeks of prediction can be found in Figure 23. We see that the predicted values follow the observed values quite well, but that they tend to un-

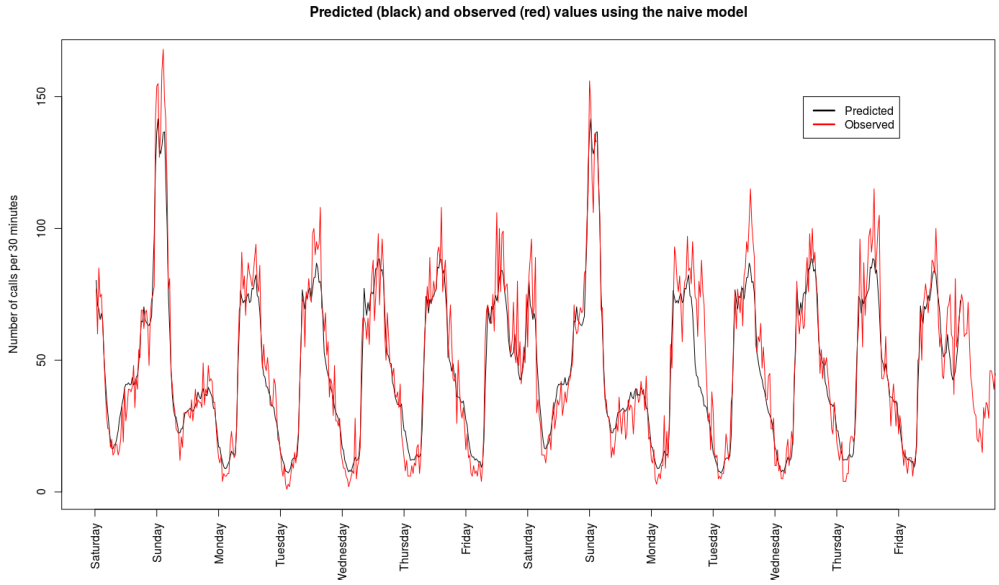

Figure 24: Plot of the predicted number of calls received per 30 minute time interval for two consecutive weeks, starting on 21 November 2015, using the naive model. The predicted and observed values are shown in black and red, respectively.

derestimate the number of calls received during the busiest hours. Additionally, the predicted values for the least busy hours tend to be a bit larger than the observed values for these time periods.

To find out how well our SARIMA model performs, we compare its predictions to predictions made using a naive model. This naive model consists of the previously estimated mean value of a week, and can be found in Figure 6. As with the SARIMA model, we plot the predicted values of the naive model together with the observed values of the same two weeks. The results can be found in Figure 24. The predictions follow the same pattern as the SARIMA model. We observe that the predicted values for the busiest hours are generally

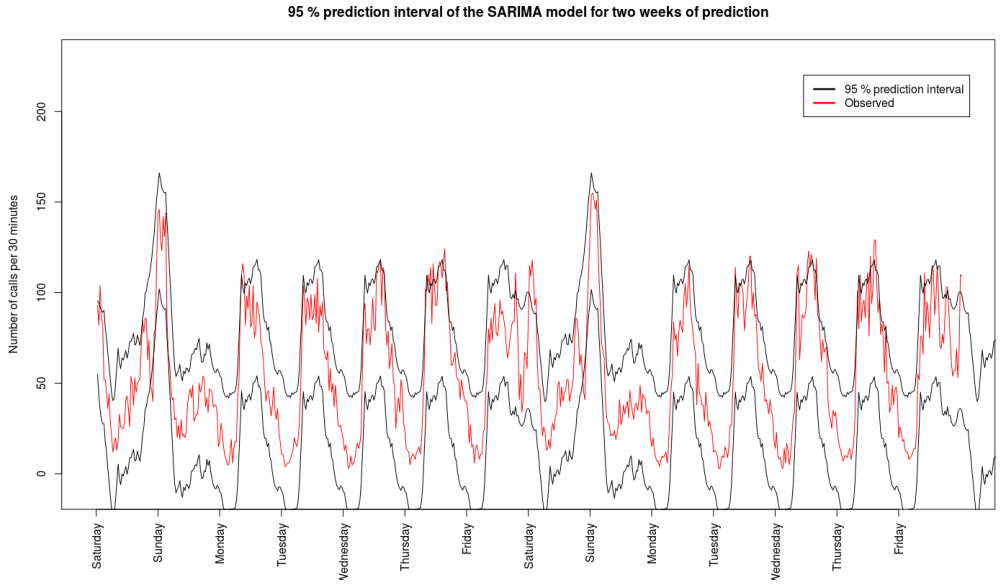

Figure 25: 95 % prediction interval of the number of calls received per 30 minute time interval for two consecutive weeks, starting on 21 November 2015, using the SARIMA model. The predicted and observed values are shown in black and red, respectively.

lower than the observed values, while the predicted values for the least busy hours are often slightly larger than the observed values. We plot the 95 % prediction intervals for the same two weeks, using both the SARIMA and naive models. The results can be seen in Figures 25 and 26, respectively. This is done so we can examine the certainty of our predictions. Figure 25 shows that the prediction intervals for the SARIMA model fit the shape of the observed values quite well, though the peak values for some of the days still lie outside of the interval. The 95 % prediction interval for the naive model in Figure 26 also seems to follow the observed values well. However, we observe that the prediction intervals for Sunday nights, which have the largest observed values, are much larger than for the SARIMA model. We also observe that the prediction interval

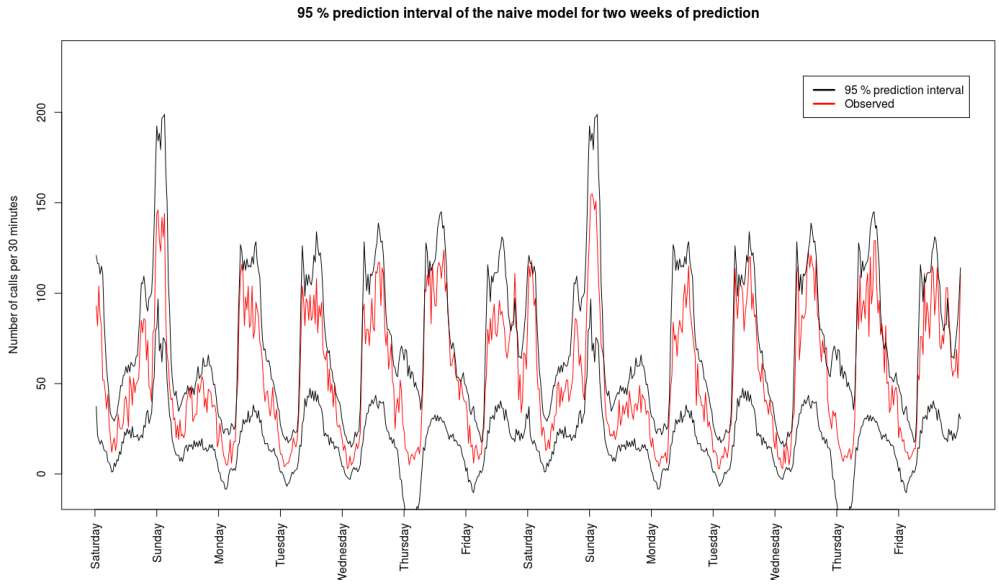

Figure 26: 95 % prediction interval of the number of calls received per 30 minute time interval for two consecutive weeks, starting on 21 November 2015, using the naive model. The predicted and observed values are shown in black and red, respectively.

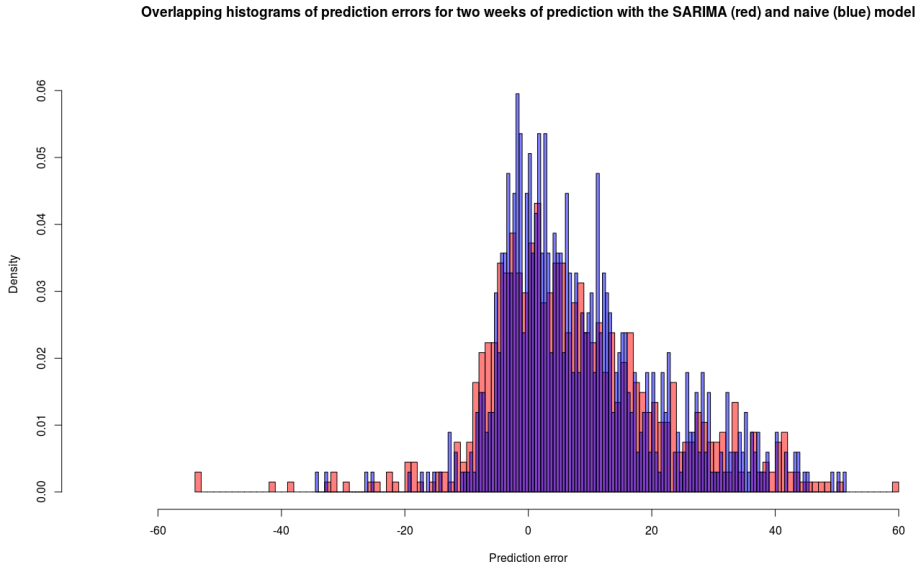

Figure 27: Overlapping histograms of the prediction error for the number of calls received per 30 minute time interval for two weeks, starting on 21 November 2015. The prediction errors for the SARIMA and naive models are shown in red and blue, respectively.

for the SARIMA model is more smooth than the prediction interval for the naive model. To further compare the two models, we plot overlapping histograms of the prediction errors of the models. In this case, the prediction error is defined as the difference between the observed values and the predicted values, or $(y_t - \hat{y}_t)$. The overlapping histograms are found in Figure 27. We see that there is a large amount of overlap between the histograms of prediction errors of the two models, though the SARIMA model appears to have more negative prediction errors.

Table 3: Calculated error statistics for the SARIMA and naive model for two weeks of prediction, starting on 21 November 2015.

Statistic	SARIMA	Naive
MPE	-0.0200	0.0282
MSE	277	243
MAE	12.1	11.2
MAPE	0.271	0.240

4.7.1 Error checking statistics

To further compare the forecasts made by the SARIMA and naive models, we calculate the error statistics defined in equations (2.34) through (2.37). The results of these calculations can be found in Table 3. The results are inconclusive for this specific forecast. The forecast from the SARIMA model has the lowest mean percentage error, while the forecast from the naive model has the lowest mean squared error, mean absolute error and mean absolute percentage error. However, Figure 27 shows that the SARIMA forecast has some very heavy negative errors, which impact the MSE and MAPE in particular. This could also just be the case for this specific forecast. In order to get a better indication of which model provides the best forecast, we decide to produce several one-week and two-week forecasts, each time removing a different number of weeks from the data set. By two-week forecast, we mean a forecast of the second week after the end of the data set. For the SARIMA model, we expect the one-week forecasts to generally be better than the two-week forecasts. Once again, we calculate the MPE, MSE, MAE and MAPE for each one-week forecast, and find their average values. In this case, we examine forecasts where the data set ends on weeks 89, 90, 91 and 92. We decide to use these particular weeks close to the end of the data set so that the SARIMA model can be based on a large data set. These weeks also do not contain any holidays. The results can be found in Table 4. Similarly, we find the average error statistics of the two-week forecasts from the SARIMA and naive model, when the data set ends on weeks 88, 89, 90

Table 4: Calculated average error statistics for the SARIMA and naive models for one-week forecasts of weeks 90, 91, 92 and 93.

Statistic	SARIMA	Naive
$\overline{\text{MPE}}_{1\text{w}}$	-0.0170	0.0300
$\overline{\text{MSE}}_{1\text{w}}$	248	278
$\overline{\text{MAE}}_{1\text{w}}$	11.6	12.0
$\overline{\text{MAPE}}_{1\text{w}}$	0.277	0.250

Table 5: Calculated average error statistics for the SARIMA and naive models for two-week forecasts of weeks 90, 91, 92 and 93.

Statistic	SARIMA	Naive
$\overline{\text{MPE}}_{2\text{w}}$	-0.143	0.0300
$\overline{\text{MSE}}_{2\text{w}}$	390	278
$\overline{\text{MAE}}_{2\text{w}}$	13.7	12.0
$\overline{\text{MAPE}}_{2\text{w}}$	0.393	0.250

and 91. This leaves us with one-week and two-week forecasts of the same four weeks. The results can be found in Table 5. We see from Tables 4 and 5 that the values of all of the statistics increase for the two-week predictions. This is as we expected, since it is more difficult to accurately predict two weeks into the future than one.

4.7.2 Coverage probability

In addition to the various error checking statistics examined above, we can use the coverage probabilities of the forecasts from the SARIMA and naive models to evaluate them. The coverage probability for a prediction interval is the percentage of observed values that are found within the specified prediction interval (Agresti and Coull, 1998). As with the error statistics, we calculate the average coverage probabilities for one-week and two-week forecasts made using

Table 6: Calculated coverage probabilities based on a 95 % prediction interval for the SARIMA and naive model for one-week forecasts of weeks 90, 91, 92 and 93.

Week	SARIMA	Naive
90	93.8 %	97.6 %
91	92.3 %	96.7 %
92	94.1 %	94.6 %
93	93.2 %	92.9 %
Average	93.3 %	95.5 %

Table 7: Calculated coverage probabilities based on a 95 % prediction interval for the SARIMA and naive model for two-week forecasts of weeks 90, 91, 92 and 93.

Week	SARIMA	Naive
90	90.5 %	97.6 %
91	92.1 %	96.7 %
92	91.7 %	94.6 %
93	83.0 %	92.9 %
Average	89.3 %	95.5 %

the SARIMA and naive models on weeks 90, 91, 92 and 93. In these cases, we look at the percentage of observed values found within the 95 % prediction intervals of the models. The results can be found in Tables 6 and 7. The average coverage probability of the naive model seems to fit the wanted 95 % coverage probability well. Furthermore, the individual coverage probabilities of the SARIMA model indicate that the one-week forecasts perform better than the corresponding two-week forecasts. We see this from the fact that the coverage probabilities are larger for the one-week forecasts, and thus closer to the wanted 95 % coverage probability that we would hope to get from a 95 % prediction interval.

5 Closing remarks

In this report, we examined a data set from TrønderTaxi containing the number of calls received per 30 minute time interval from 1 March 2014 through 31 January 2016. The time series was modelled as a seasonal autoregressive integrated moving average model that was fitted to the data set. One-week and two-week forecasts were made using this SARIMA model and a naive model. Various error checking statistics were used to compare the two models with each other. When it comes to performing forecasts for a single week into the future, the SARIMA model appears to outperform the naive model. The SARIMA model has a smaller MPE, MSE and MAE, but its MAPE is slightly larger. This means that if TrønderTaxi is able to set up their staffing schedule as late as only a single week ahead of time, it is advisable to use the SARIMA model instead of the naive model. All of the error checking statistics of the two-week forecasts from the SARIMA model were worse than the error checking statistics for the naive model. Due to this, there is little reason to use the more complicated and computationally demanding SARIMA model when performing forecasts for two weeks into the future.

The models in this report contain several limitations. Neither the SARIMA nor the naive models account for any seasonality other than the daily and weekly seasonality previously explained. However, TrønderTaxi has reported that there seems to be an underlying yearly trend in the number of calls received. For instance, TrønderTaxi receives fewer calls during the summer months. This could be related to the fact that many people from Trondheim decide to spend at least parts of their summer vacation somewhere else than in Trondheim. Information about when in the year the data points are from could be included in the regressor matrix. Furthermore, it would be of interest to examine the effects of various holidays on the number of calls received. Intuitively, we would expect fewer calls to be received during the Easter holiday, as this is also a period of the year when many people from Trondheim temporarily leave the city. This coincides with the experiences of TrønderTaxi. On the other hand, the data set provided by TrønderTaxi shows that some holidays lead to more calls being

received. This becomes apparent when looking at the two New Year's Eves in the data set, which have the busiest 30 minute time intervals of the entire data set. If we were to use our current SARIMA or naive models to predict the number of calls received on these days, we would surely underestimate the values. To obtain a good understanding of how the various holidays affect the number of calls received, it would be beneficial to look at a larger data set, spanning further back in time. Unfortunately, TrønderTaxi does not have call data going further back in time than 1 March 2014. However, TrønderTaxi has detailed logs over all trips made by their taxis spanning further back in time. We would imagine that this data is highly correlated with the number of calls made. If this proves to be the case, the data of the number of trips made could be used to estimate the effects of various holidays. A natural extension from this would be to look at the effects of large annual events, such as music festivals.

In addition to examining the effects of various holidays on the number of calls received, it would be interesting to take a look at how the weather impacts the number of calls. More specifically, the regressor matrix of the SARIMA model could be extended to include a parameter for the recorded precipitation per 30 minute time interval, and another parameter for the recorded air temperature for the same time intervals.

The histogram of the standardised residuals of the SARIMA model has heavier tails than a standard normal distribution. A natural continuation of this project would be to examine where in the data set these standardised residuals come from. If we find out where these extreme standardised residuals come from, we may be able to account for them in our model.

References

Agresti, A. and Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723.

Andrews, B. and Parsons, H. (1993). Establishing telephone-agent staffing levels through economic optimization. *Interfaces*, 23(2):14–20.

Andrews, B. H. and Cunningham, S. M. (1995). L. L. Bean improves call-center forecasting. *Interfaces*, 25(6):1–13.

Bianchi, L., Jarrett, J., and Hanumara, R. C. (1998). Improving forecasting for telemarketing centers by ARIMA modeling with intervention. *International Journal of Forecasting*, 14(4):497–504.

Box, G. E. P. and Jenkins, G. (1990). *Time Series Analysis, Forecasting and Control*. Holden-Day, Incorporated, San Francisco.

Brandt, P. T., Williams, J. T., Fordham, B. O., and Pollins, B. (2000). Dynamic modeling for persistent event-count time series. *American Journal of Political Science*, 44(4):823–843.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference. Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33(2):261–304.

Canova, F. and Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business and Economic Statistics*, 13(3):237–252.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive times series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431.

Hyndman, R. J. and Athanasopoulos, G. (2013). Forecasting: principles and practice. "`http://otexts.org/fpp/`". [Online; accessed 6-June-2016].

Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22.

Ibrahim, R. and L'Ecuyer, P. (2013). Forecasting call center arrivals: Fixed-effects, mixed-effects, and bivariate models. *Manufacturing & Service Operations Management*, 15(1):72–85.

Kay, S. M. (1993). *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, second edition.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:159–178.

Rosadi, D. et al. (2012). New procedure for determining order of subset autoregressive integrated moving average (ARIMA) based on over-fitting concept. In *International Conference on Statistics in Science, Business, and Engineering (ICSSBE), 2012*, pages 1–5.

Soyer, R. and Tarimcilar, M. M. (2008). Modeling and analysis of call center arrival data: a Bayesian approach. *Manage. Sci.*, 54(2):266–278.

Trapletti, A. and Hornik, K. (2016). tseries: Time series analysis and computational finance. `http://CRAN.R-project.org/package=tseries`. R package version 0.10-35. [Online; accessed 6-June-2016].

Wei, W. W. (2006). *Time Series Analysis - Univariate and Multivariate Methods*. Pearson Addison Wesley, Boston, second edition.

Østhus, T. (2015). An analysis of call data from TrønderTaxi. Project report, Department of Mathematical Sciences, Norwegian University of Science and Technology. Trondheim, Norway.