

Time Series Analysis Lab 2

Andrea Bruzzone, Thomas Zhang

Sunday, September 27, 2015

Problem 1.a.

AR(2) process is represented by $Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \epsilon_t$. The Conditional sum of squares is:

$$S_c(\phi_1, \phi_2) = \sum_{t=3}^n [Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2}]^2$$

Problem 1.b and 1.d

Using our function to perform linear regression estimation of the parameters of a simulated AR(2) process, we get some values for ϕ_1 and ϕ_2 close to the true values. For example with a time series of length 30, with

$$\phi_1 = 1.6, \phi_2 = -0.9, \sigma = 5$$

the AR(2) parameter estimations and their std. dev. estimation errors are:

```
Y <- AR2sim(30,1.6,-0.9,5)
```

```
AR2LS(Y)
```

```
##                                     (Intercept)
##                                     0.8352452
##                                     phi1_est
##                                     1.5177365
##                                     phi2_est
##                                     -0.7727427
## Estimate of the std. dev. of the white noise term in AR(2) process
##                                     4.9266147
```

In order to get a sense of the standard errors of the estimated parameter values we run an aggregator function which simulates 100 AR(2) processes with the same parameters.

```
AR2simaggregator(30,1.6,-0.9,5)
```

```
##                               Std. Dev. of phi_1 estimates
##                               0.1247448
##                               Std. Dev. of phi_2 estimates
##                               0.1290378
## mean of estimates of std. dev. of white noise term
##                               5.1629422
```

Let us now compare this to the `arima()` function with `method = "ML"`

```
arima(Y,order=c(2,0,0),include.mean=FALSE,method="ML")
```

```
##
## Call:
## arima(x = Y, order = c(2, 0, 0), include.mean = FALSE, method = "ML")
##
## Coefficients:
##          ar1      ar2
##       1.4841 -0.7271
## s.e.  0.1103  0.1080
##
## sigma^2 estimated as 21.12:  log likelihood = -89.75,  aic = 183.49
```

We can see that the standard errors are smaller for the `arima()` function. We now try to increase the length of the simulated time series from 30 to 100 and see if our standard errors become smaller.

```
Y <- AR2sim(100,1.6,-0.9,5)
```

```
AR2LS(Y)
```

```
##                                     (Intercept)
##                                     0.004962419
##                                     phi1_est
##                                     1.543072534
##                                     phi2_est
##                                     -0.811610517
## Estimate of the std. dev. of the white noise term in AR(2) process
##                                     4.922300105
```

```
AR2simaggregator(100,1.6,-0.9,5)
```

```
##                               Std. Dev. of phi_1 estimates
##                               0.04396508
##                               Std. Dev. of phi_2 estimates
##                               0.04633330
## mean of estimates of std. dev. of white noise term
##                               5.00179295
```

It is seen that the standard deviation of the parameter estimates becomes smaller and even better than the standard errors of the `arima()` function when length of time series is set to 100.

Problem 1.c.

The stationary variance can be calculated by

$$\begin{aligned}
 \gamma_0 &= \mathbf{E} [\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t]^2 \\
 &= \phi_1^2 \mathbf{E} [Y_{t-1}^2] + \phi_2^2 \mathbf{E} [Y_{t-2}^2] + 2\phi_1^2 \phi_2^2 \mathbf{E} [Y_{t-1} Y_{t-2}] + \sigma_\epsilon^2 \\
 &= (\phi_1^2 \phi_2^2) \gamma_0 + 2\phi_1 \phi_2 \gamma_1 + \sigma_\epsilon^2 \\
 \gamma_0 \left[1 - \phi_1^2 - \phi_2^2 - \frac{(2\phi_1^2 \phi_2)}{(1 - \phi_2)} \right] &= \sigma_\epsilon^2 \\
 \gamma_0 &= \frac{1 - \phi_2}{(1 - \phi_2)(1 - \phi_1^2 - \phi_2^2) - (2\phi_1^2 \phi_2)} \sigma_\epsilon^2
 \end{aligned}$$

We used the fact that, for AR(2) processes:

$$\begin{aligned}\gamma_1 &= \frac{\phi_1 \gamma_0}{1 - \phi_2} \\ &= \frac{\phi_1}{1 - \phi_2} \sigma_\epsilon^2 \\ &= \rho_1 \sigma_\epsilon^2\end{aligned}$$

We see that by comparison:

$$\kappa = \frac{1 - \phi_2}{(1 - \phi_2)(1 - \phi_1^2 - \phi_2^2) - (2\phi_1^2\phi_2)}$$

The Likelihood function can be calculated using the distribution of (y_3, y_4, \dots) conditional on (y_1, y_2)

$$f(y_3, y_4, \dots | y_1, y_2) = (2\pi\sigma_\epsilon^2)^{-\frac{(n-2)}{2}} \exp \left[-\frac{1}{2\sigma_\epsilon^2} \sum_{t=3}^n \epsilon_t^2 \right]$$

The Likelihood function is, then given by

$$\begin{aligned}L(\phi_1, \phi_2, \sigma_\epsilon^2) &= f(y_3, y_4, \dots | y_1, y_2) f(y_1, y_2) \\ &= (2\pi\sigma_\epsilon^2)^{-\frac{n}{2}} \frac{1}{\kappa \sqrt{1 - \rho_1^2}} \exp \left[-\frac{1}{2\sigma_\epsilon^2} S(\phi_1, \phi_2) \right]\end{aligned}$$

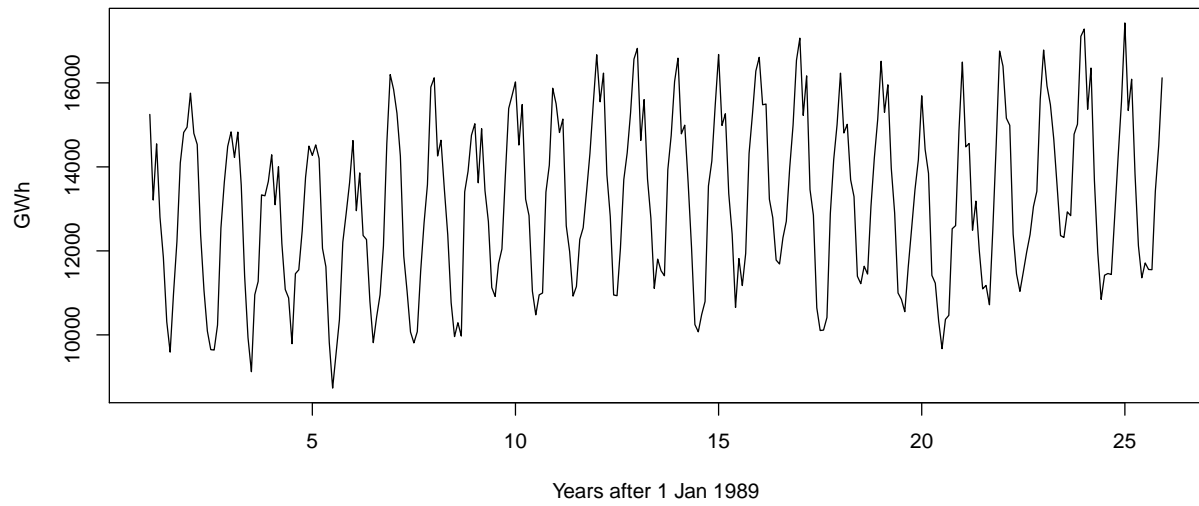
where

$$S(\phi_1, \phi_2) = \sum_{t=3}^n (Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2})^2 + \frac{1}{\kappa(1 - \rho_1^2)} (Y_1^2 + Y_2^2 - 2\rho_1 Y_2 Y_1)$$

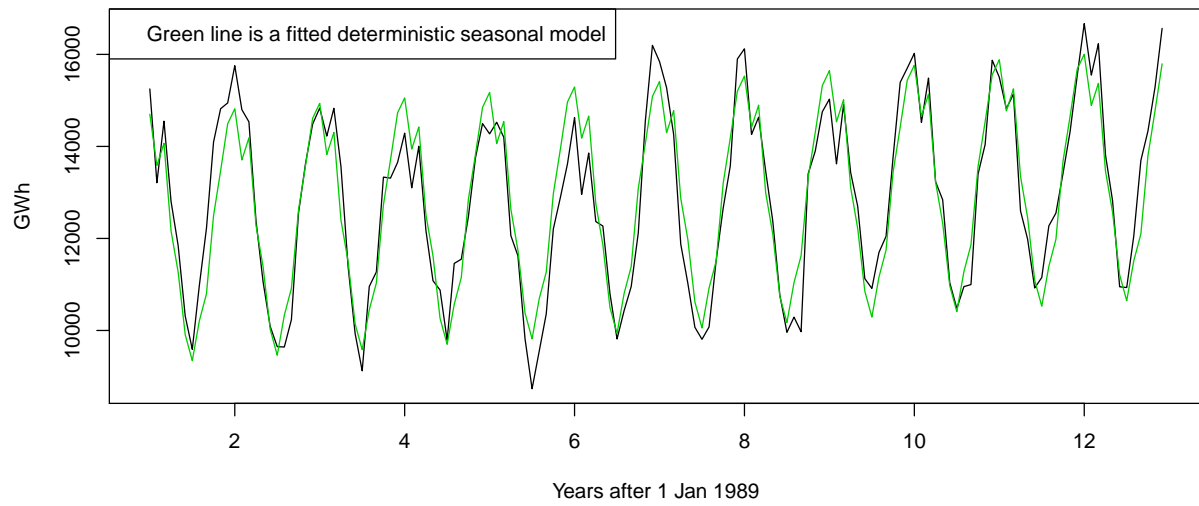
Problem 2

We divided the time series in two parts, the first graph shows the entire time series and the second graph shows the modeling part in which a seasonal trend is evident. The series does not show outliers and it seems to be stationary in mean and variance. Thus, we tried to fit the modeling part with a deterministic seasonal trend model but still we can see some autocorrelations between different lags. The EACF suggests that an AR(1) model can be good to fit to the residuals of the deterministic seasonal model.

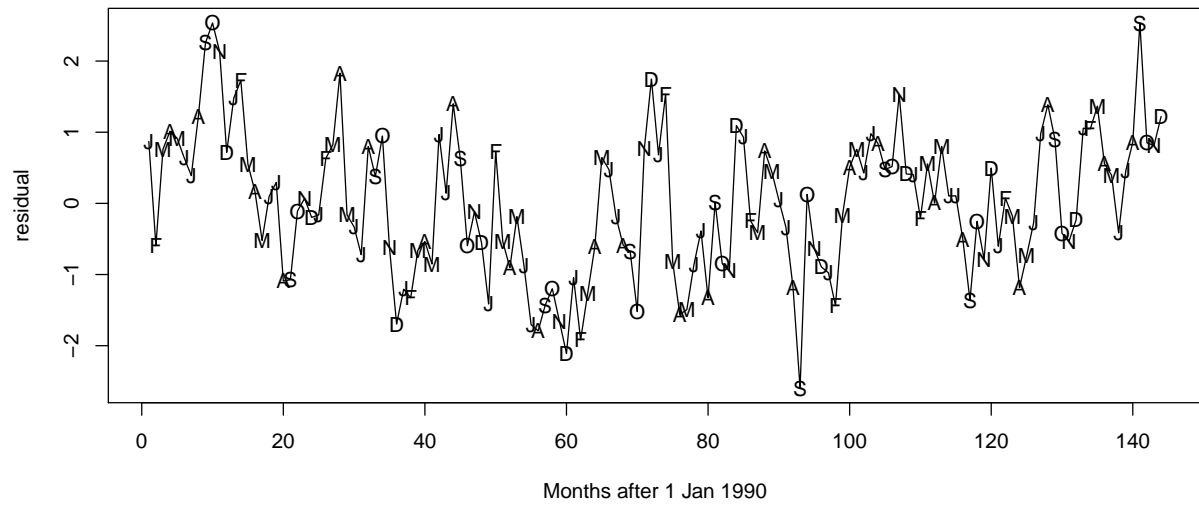
total Swedish electricity consumption in GWh 1990–2014



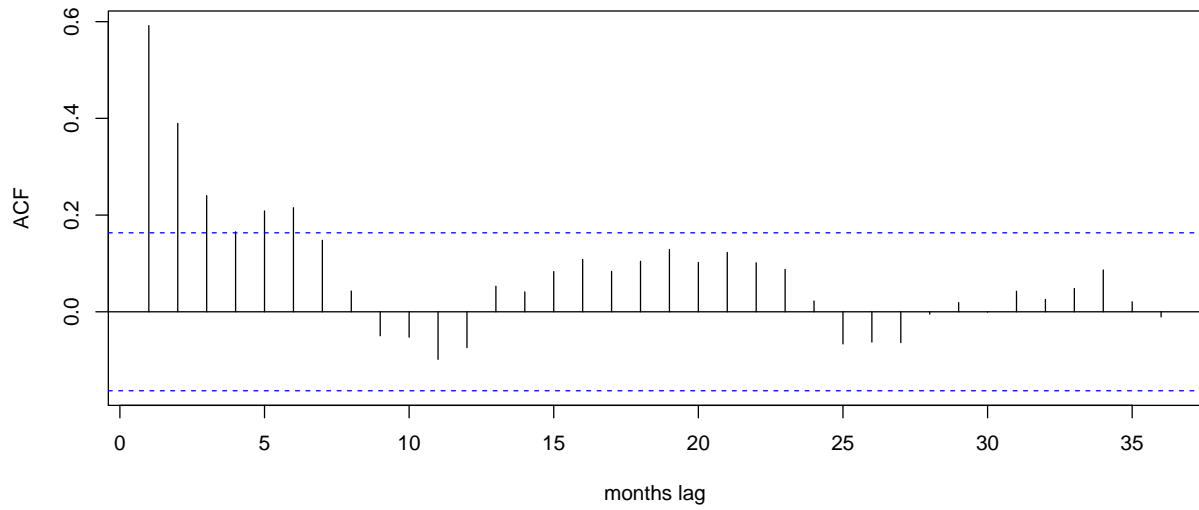
total Swedish electricity consumption in GWh 1990–2014



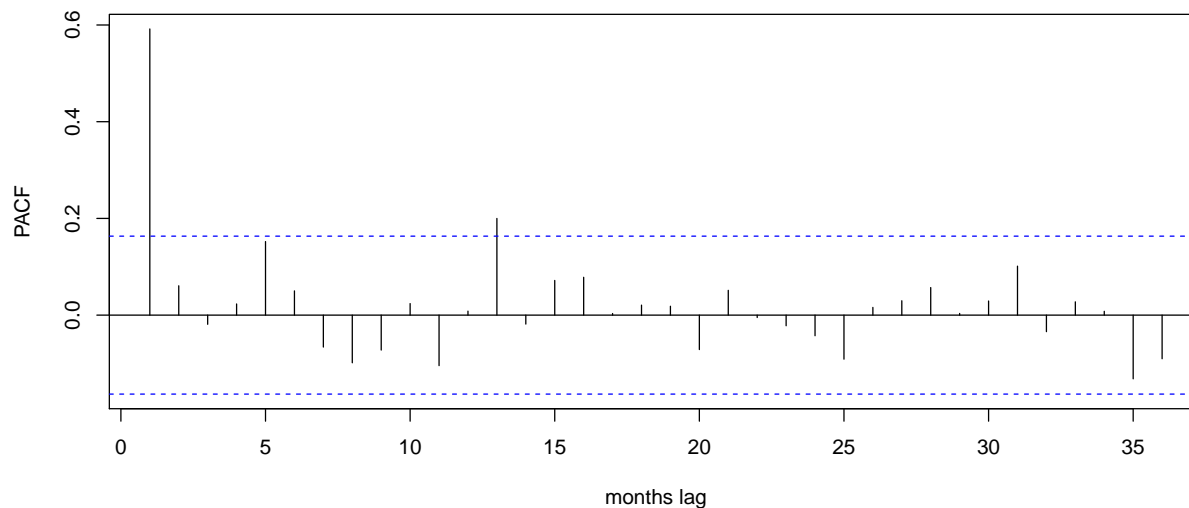
Residuals of the seasonal model



Autocorrelation function of the residuals of the deterministic seasonal model

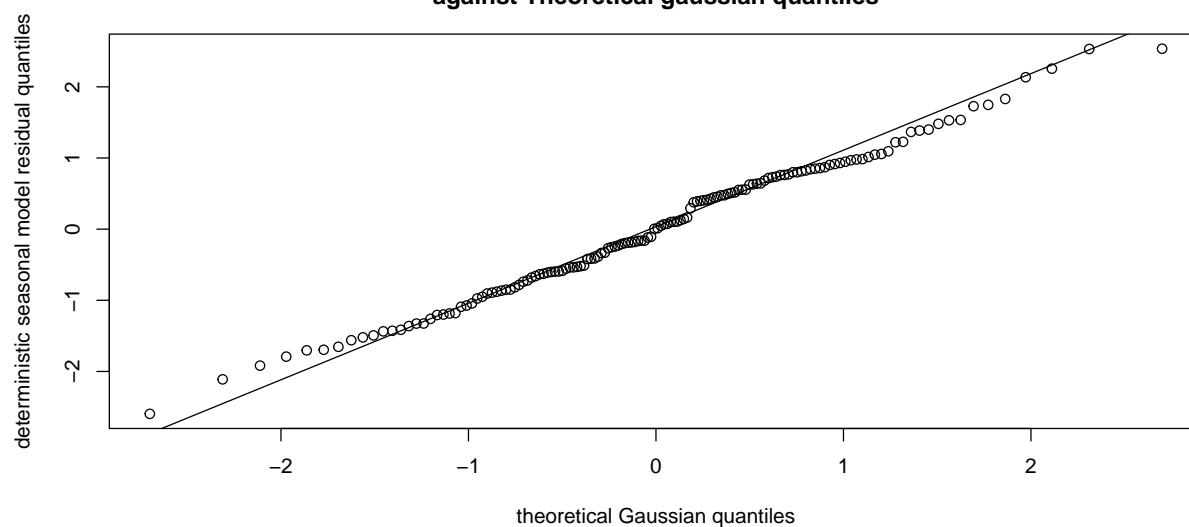


Partial Autocorrelation function of the residuals of the deterministic seasonal model



```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x x o x x o o o o o o o
## 1 o o o o o o o o o o o o o
## 2 x o o x o o o o o o o o o
## 3 x o o o o o o o o o o o o
## 4 x x x o o o o o o o o o o
## 5 x x o o o o o o o o o o o
## 6 x x o x o o o o o o o o o
## 7 x x x o o x o o o o o o o
```

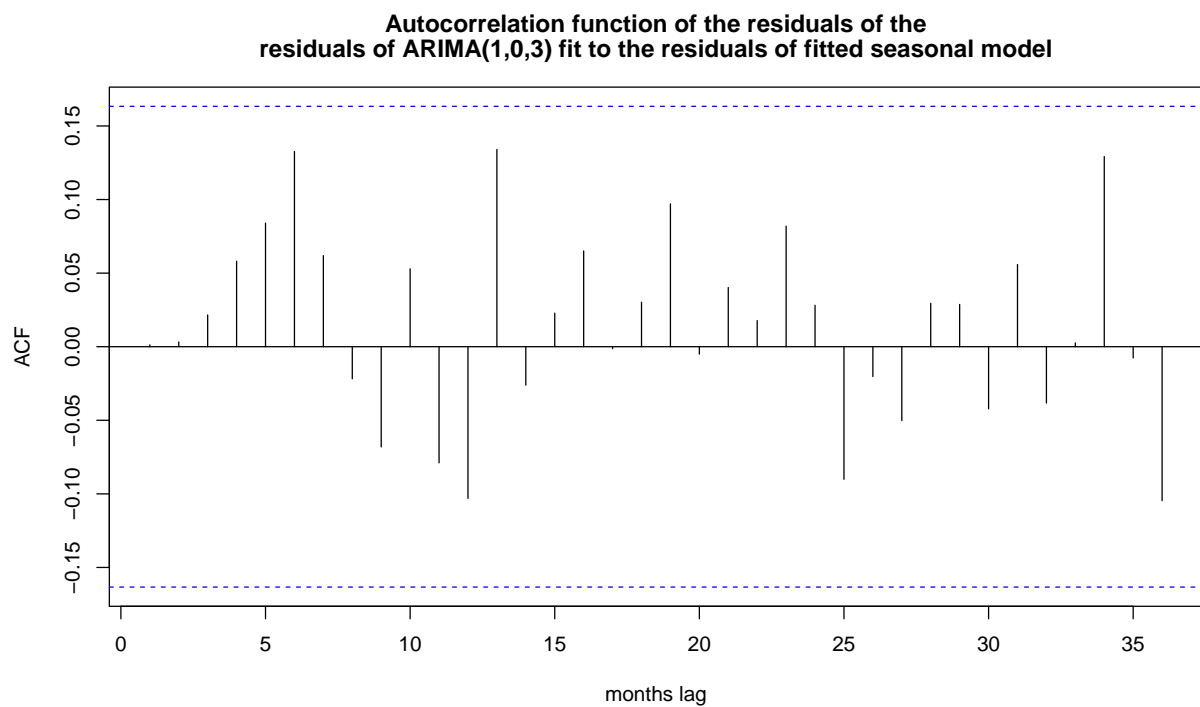
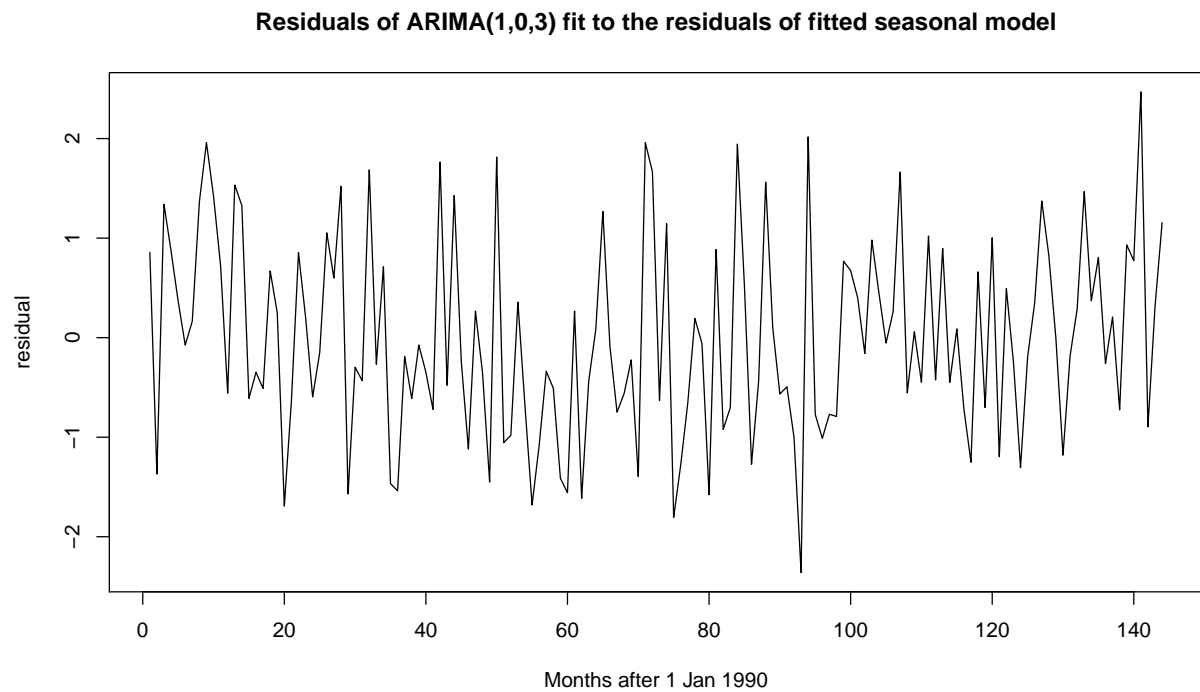
Quantiles of the residuals of the deterministic seasonal model against Theoretical gaussian quantiles



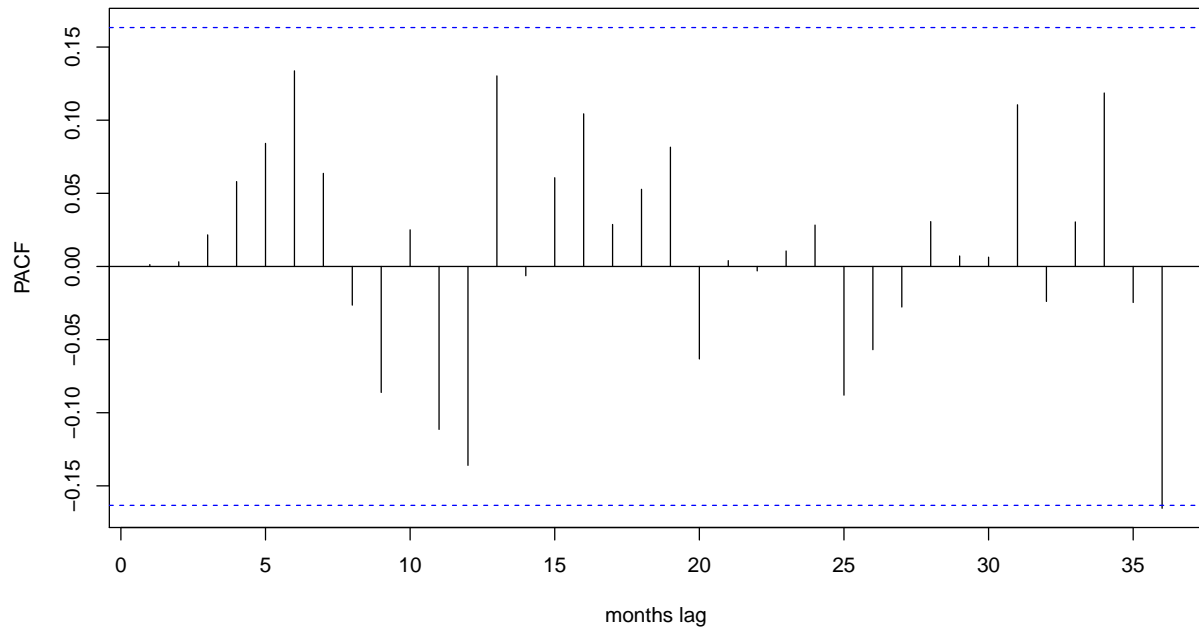
We tried to fit the residuals of the seasonal model using an ARIMA(1,0,3) process. The sample ACF and sample PACF look good but it can be seen in the summary that some parameters do not look statistical

significant.

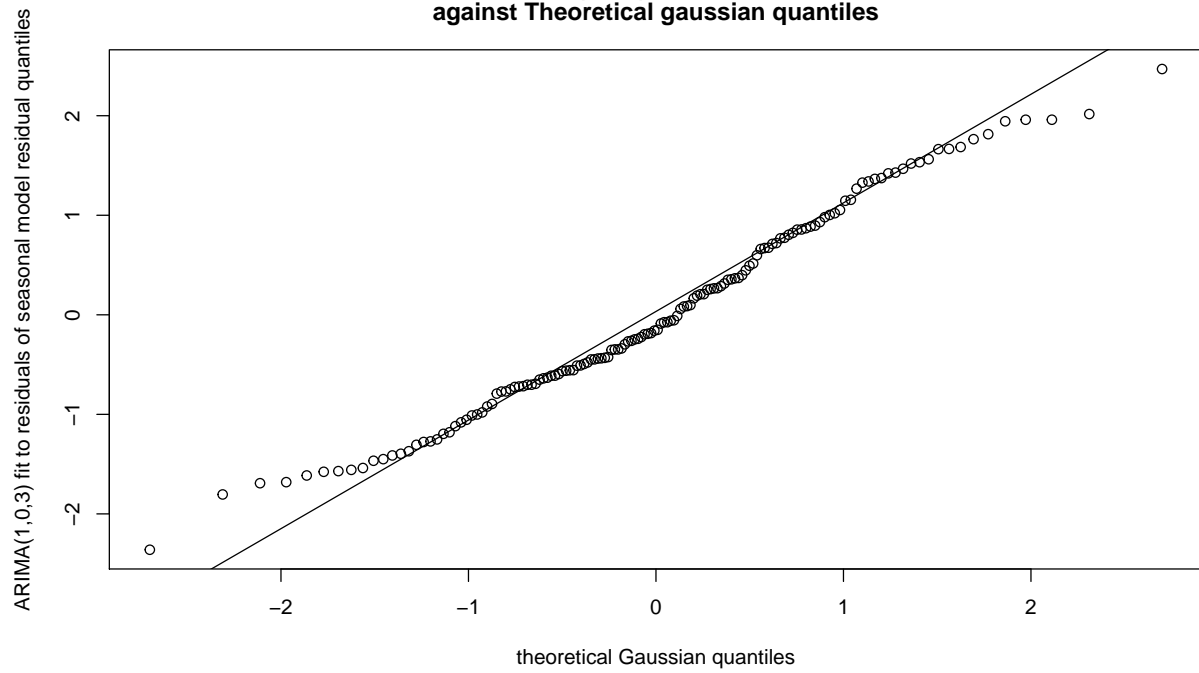
```
ma4 <- arima(resi,order=c(1,0,3), seasonal = list(order = c(0,0,0)))  
fit_ar1_ma3_res <- rstandard(ma4)
```

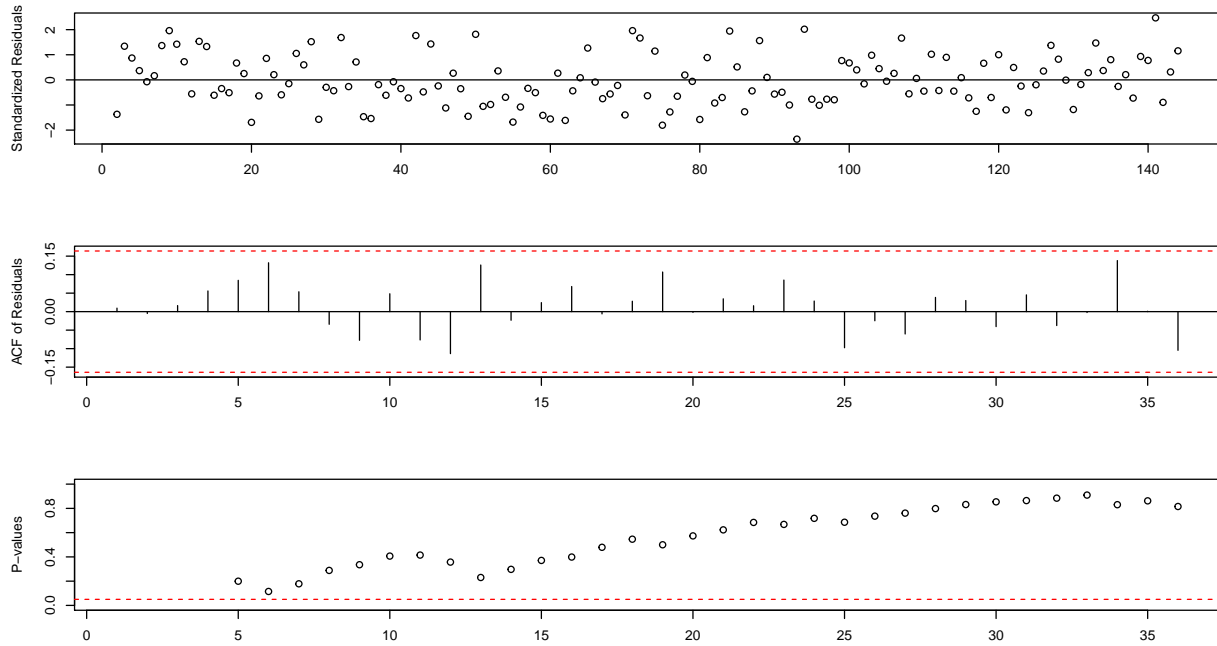


Partial Autocorrelation function of the residuals of ARIMA(1,0,3) fit to the residuals of fitted seasonal model



Quantiles of the residuals of ARIMA(1,0,3) fit to the residuals of fitted seasonal model against Theoretical gaussian quantiles





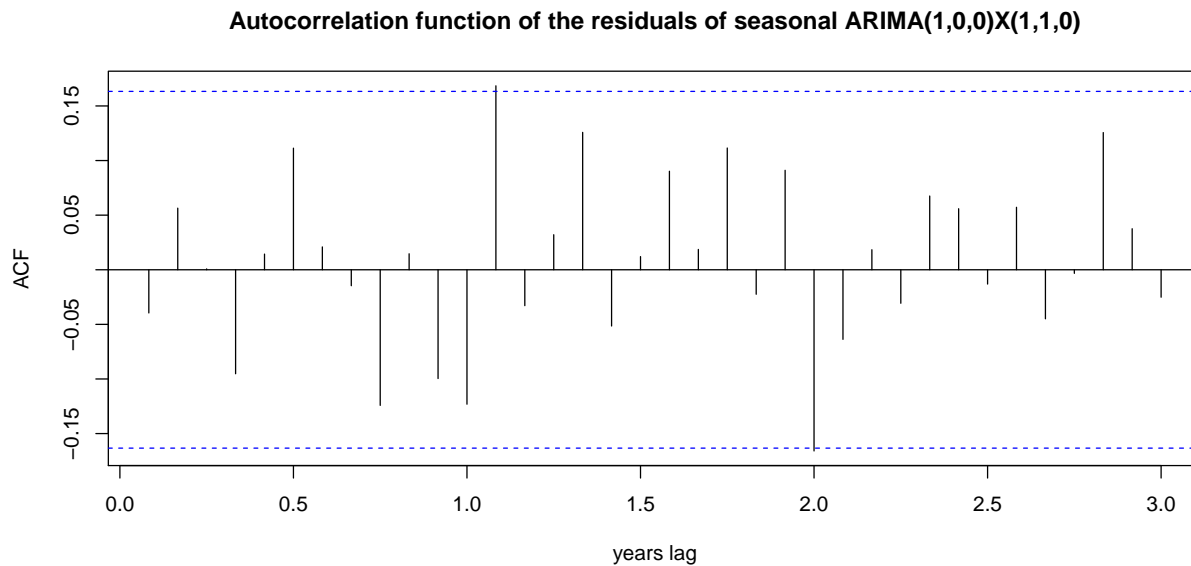
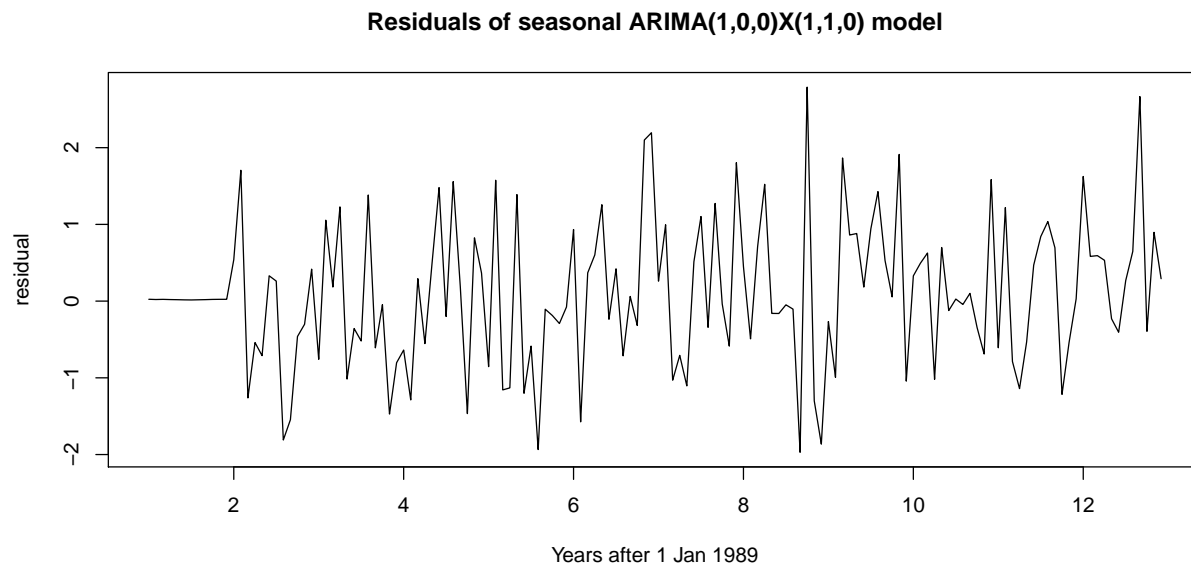
```
##
## Call:
## arima(x = resi, order = c(1, 0, 3), seasonal = list(order = c(0, 0, 0)))
##
## Coefficients:
##          ar1      ma1      ma2      ma3  intercept
##          0.0146  0.5589  0.4062  0.2115      0.0160
## s.e.      0.3655  0.3553  0.2151  0.1540      0.1465
##
## sigma^2 estimated as 0.6412:  log likelihood = -172.58,  aic = 355.16

##
## Box-Ljung test
##
## data:  fit_ar1_ma3_res
## X-squared = 24.7636, df = 36, p-value = 0.9212

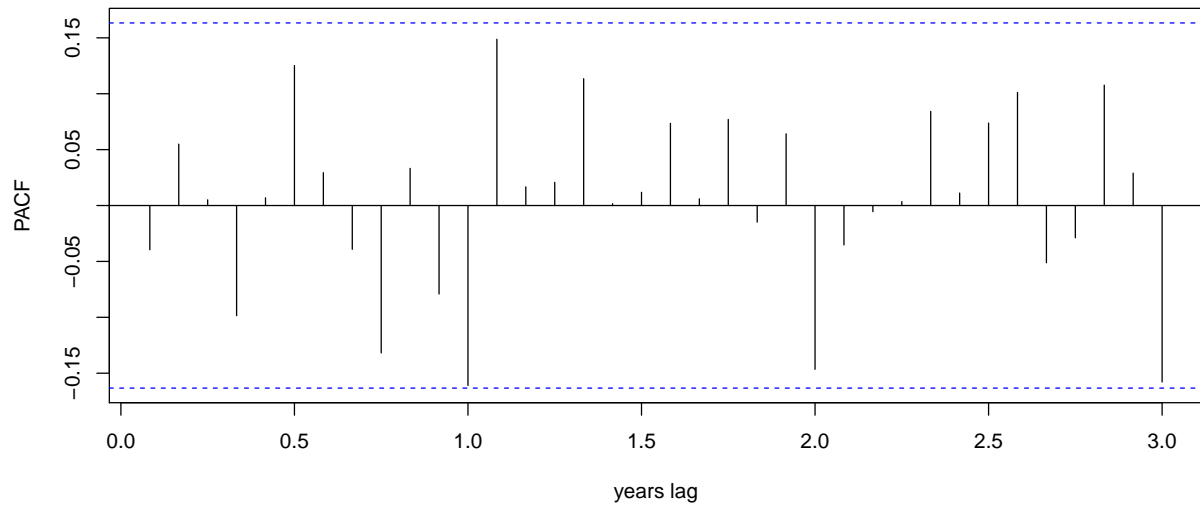
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
## 0 o o o o o o o o o o o o o o o o
## 1 x o o o o o o o o o o o o o o o
## 2 o x o o o o o o o o o o o o o o
## 3 x x x o o o o o o o o o o o o
## 4 x o o o o o o o o o o o o o o
## 5 x o o o o o o o o o o o o o o
## 6 x x o x x o o o o o o o o o o
## 7 x x o o x o o o o o o o o o o
## 8 x x x o x x o o o o o o o o o
## 9 x x x o x x o o o o o o o o o
```

Since the EACF of the residuals of the modeling part suggested an AR(1) process to fit the series and since there is a seasonal trend, we try a new model that is an ARIMA(1,0,0)x(1,1,0).

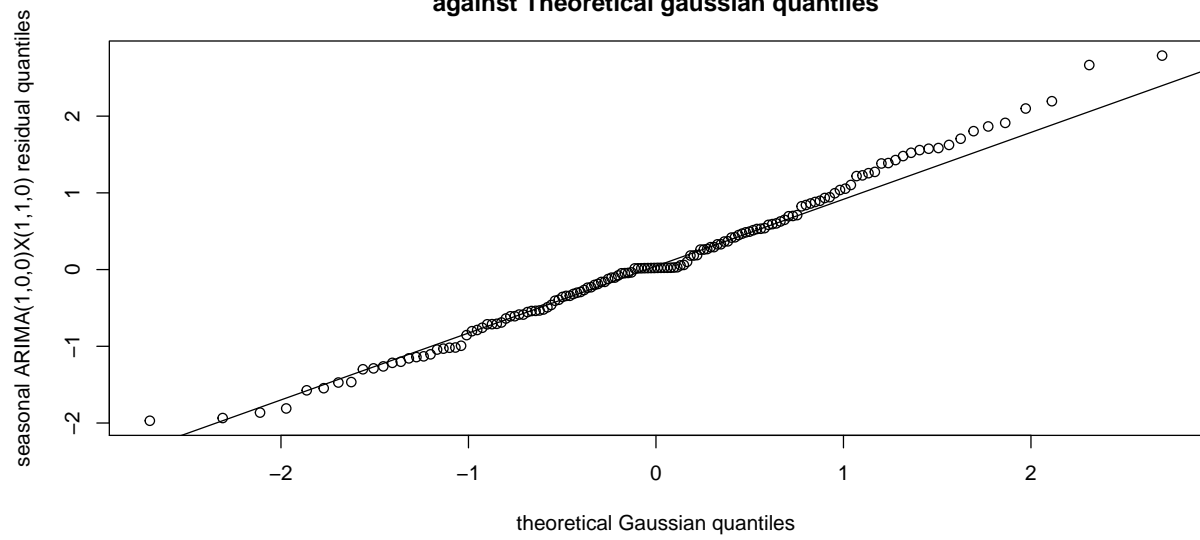
```
ar1 <- Arima(ele_model, order = c(1,0,0), seasonal = list(order = c(1,1,0)))  
res_ar1 <- rstandard(ar1)
```

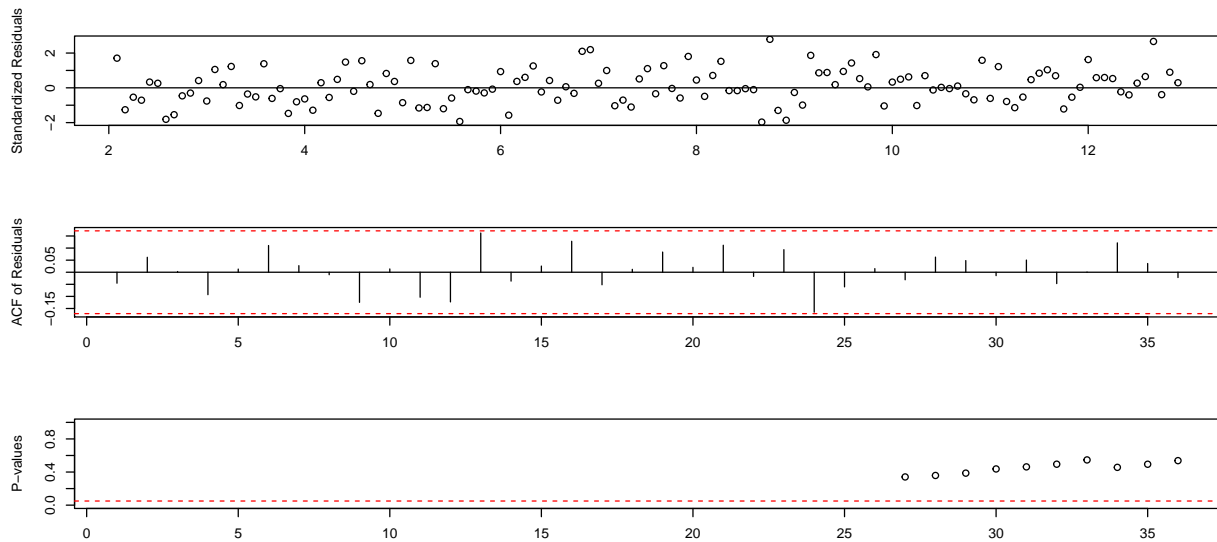


Partial Autocorrelation function of the residuals of seasonal ARIMA(1,0,0)X(1,1,0)



**Quantiles of the residuals of the seasonal ARIMA(1,0,0)X(1,1,0)
against Theoretical gaussian quantiles**





```
## Series: ele_model
## ARIMA(1,0,0)(1,1,0)[12]
##
## Coefficients:
##          ar1      sar1
##      0.5924  -0.5426
## s.e.  0.0698   0.0727
##
## sigma^2 estimated as 402978:  log likelihood=-1041.45
## AIC=2088.89   AICc=2089.08   BIC=2097.54
```

```
##
## Box-Ljung test
##
## data:  res_ar1
## X-squared = 35.1456, df = 36, p-value = 0.509
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
## 0 o o o o o o o o o o o o o o o o
## 1 x o o o o o o o o o o o o o o
## 2 o x o o o o o o o o o o x o o o
## 3 o x x o o o o o o o o o o o o
## 4 o x x o o o o o o o o o o o o
## 5 o x x o o o o o o o o o o o o
## 6 x x x o o o o o o o o o o o o
## 7 x o x o o o o o o o o o o o o
## 8 x x x o o x o o o o o o o o o
## 9 x x x o x x x o o o o o o o o
## 10 x o x x o x x o o x o o o o o
```

The analysis of the residuals shows that an Arima(1,0,0)x(1,1,0) process is a good model and it is better than the ARIMA(1,0,3) process because it is a model with less parameters. We also see that the AIC is larger in this model.

Problem 3.a.

The model equation is

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta \Theta e_{t-13}$$

And the one and two step-ahead predictors are given in terms of the past values and errors as

$$\begin{aligned}\hat{Y}_t(1) &= Y_t + Y_{t-11} - Y_{t-12} - \theta e_t - \Theta e_{t-11} + \theta \Theta e_{t-12} \\ \hat{Y}_t(2) &= \hat{Y}_t(1) + Y_{t-10} - Y_{t-11} + e_t - \Theta e_{t-10} + \theta \Theta e_{t-11}\end{aligned}$$

We know that the forecast error is defined as: $e_t(\ell) = Y_{t+\ell} - \hat{Y}_t(\ell)$

Thus, for our model we have: $e_t(1) = Y_{t+1} - (Y_t + Y_{t-11} - Y_{t-12} - \theta e_t - \Theta e_{t-11} + \theta \Theta e_{t-12})$ but we know that the following holds:

$$Y_{t+1} - Y_t = Y_{t-11} - Y_{t-12} + e_{t+1} - \theta e_t - \Theta e_{t-11} + \theta \Theta e_{t-12}$$

so we can simplify into

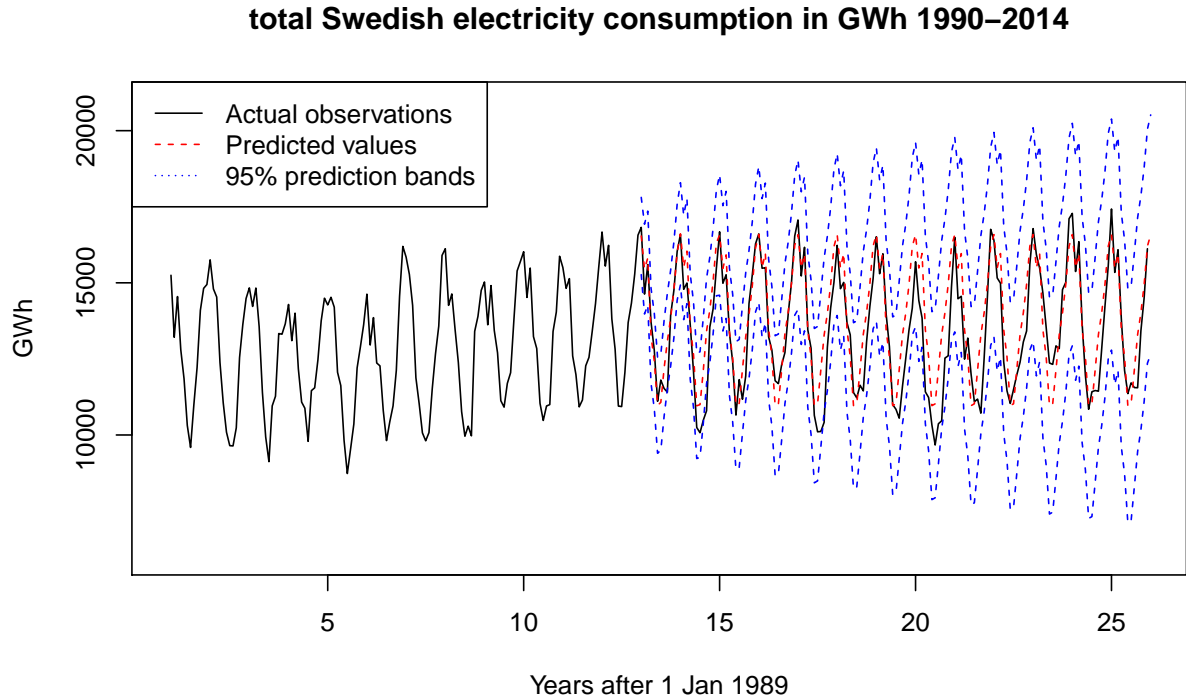
$$e_t(1) = e_{t+1}$$

And thus $Var(e_t(1)) = Var(e_{t+1}) = \sigma_e^2$

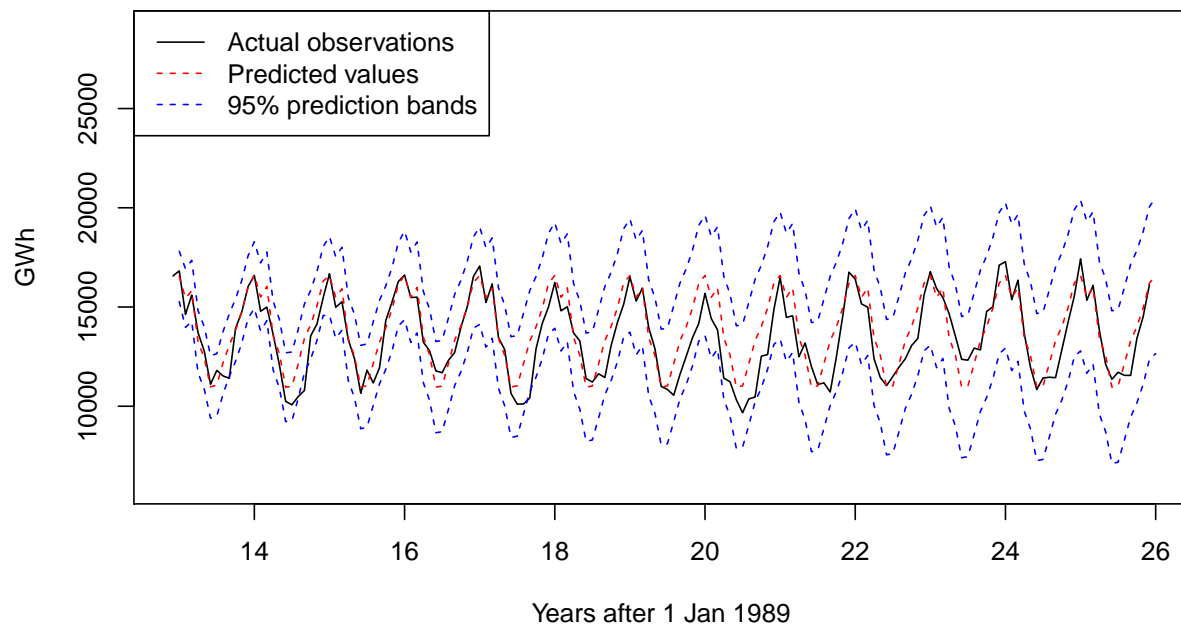
For our model the one and two-step forecast error variances are the same: $Var(e_t(1)) = Var(e_t(2)) = \sigma_e^2$

Problem 3.b.

For predicting the monthly electrical consumption we use the Arima(1,0,0)x(1,1,0) process model.



total Swedish electricity consumption in GWh 1990–2014



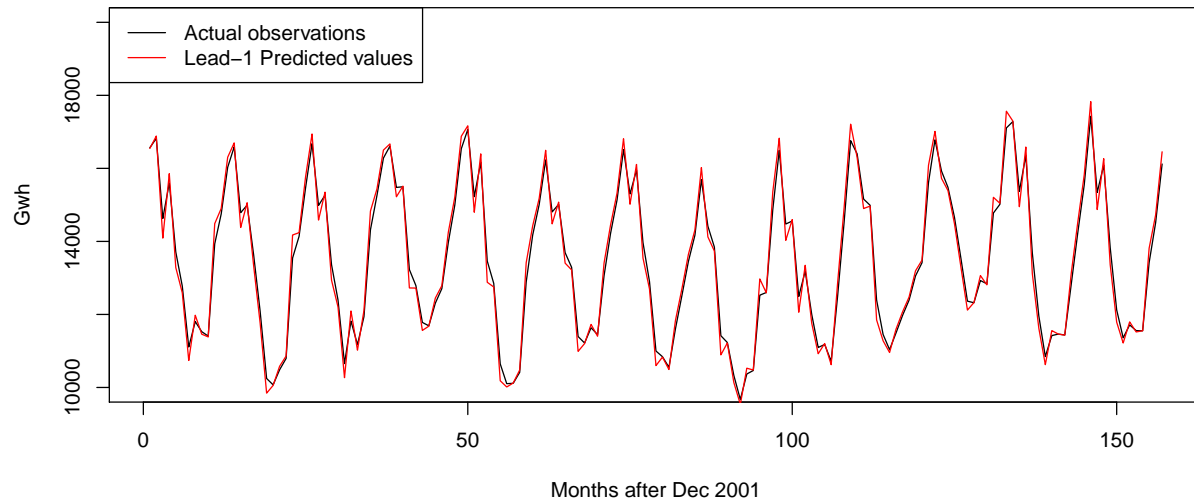
```
## [1] "We have 13 observations outside the 95% prediction bands out of 157 observations."
```

We see that the number of observations out of the 95% prediction bands are a little more than what we would expect. We believe this is due to the observations, out of pure coincidence, in the beginning of the time series exceeding the prediction bands where the prediction bands are the most narrow.

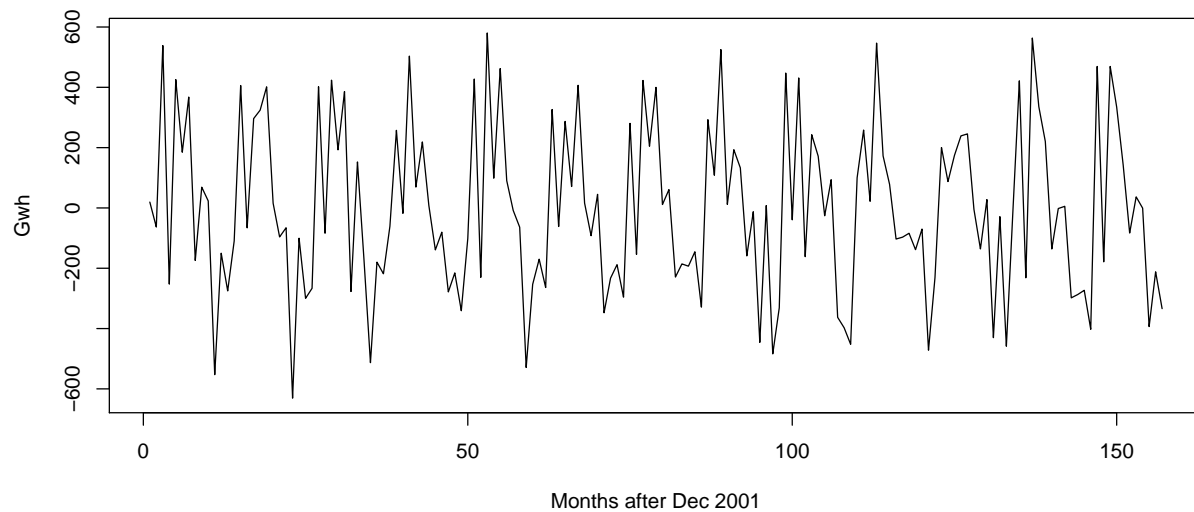
Problem 3.c.

Here we construct the lead-1 predictions for the validation part of the monthly electrical consumption data.

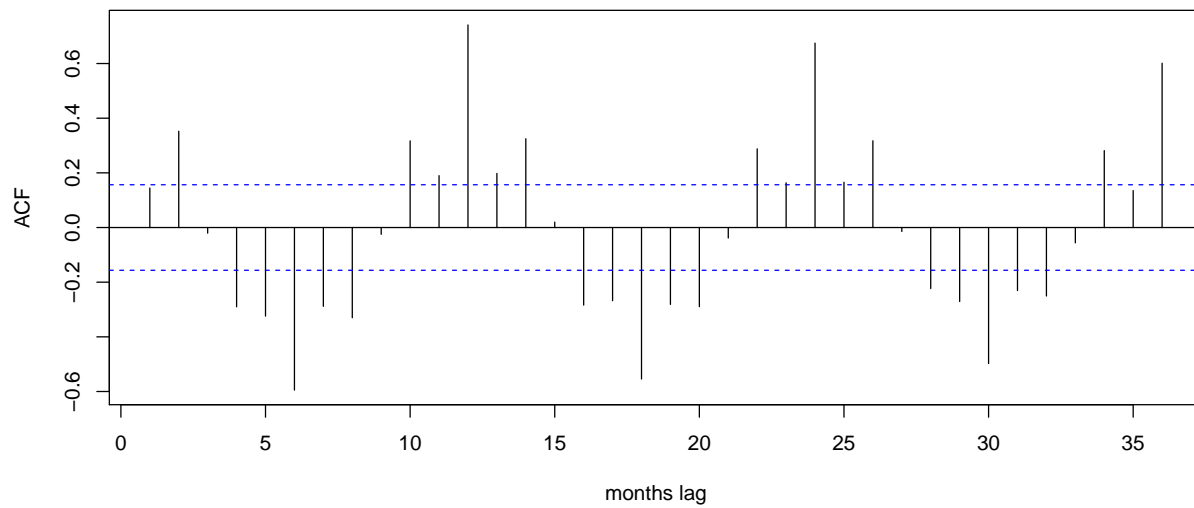
validation part of Electrical consumption data and Lead-1 predictions



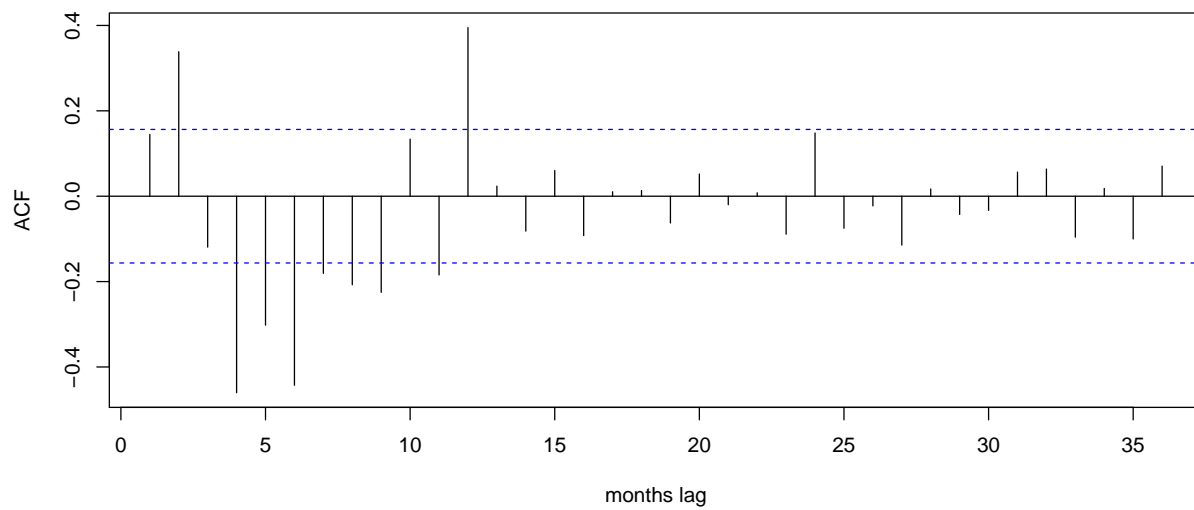
Residuals for the lead-1 predictions of the validation part of Electrical consumption data using ARIMA(1,0,0)X(1,1,0) model



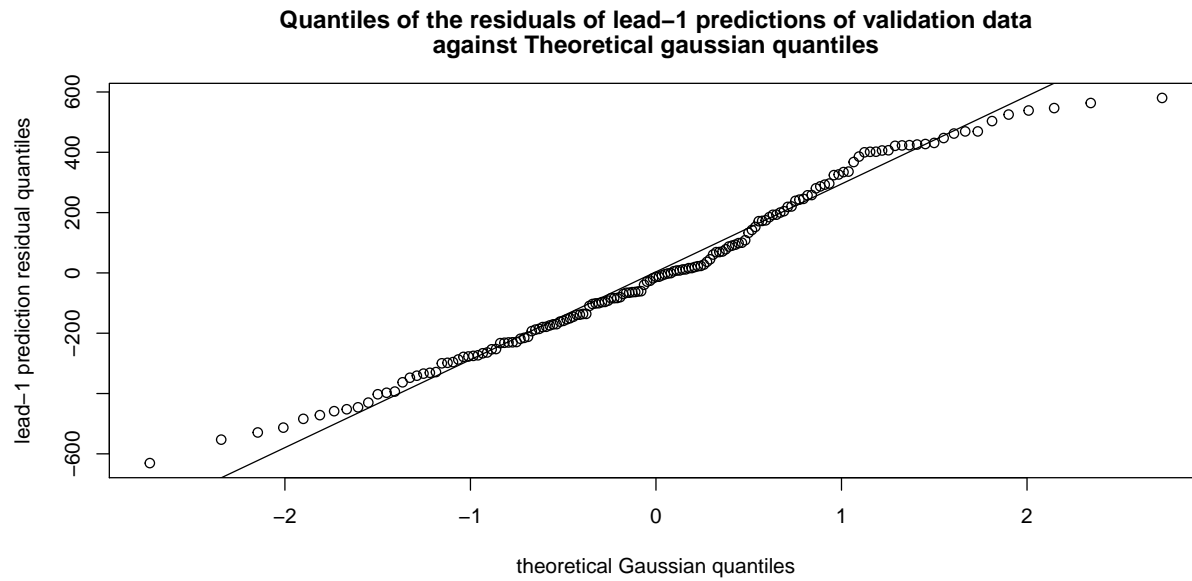
**Autocorrelation function of the residuals of lead-1 predictions of validation data
using ARIMA(1,0,0)X(1,1,0) model**



**Partial Autocorrelation function of the residuals of lead-1 predictions of validation data
using ARIMA(1,0,0)X(1,1,0) model**



```
## AR/MA
##  0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 o x o x x x x x o x x x x
## 1 x x o o o x x x o x o x x x
## 2 x x o x o x x x o x o x x x
## 3 x x o o o x o x o x o x o x
## 4 x x x o x o o o o o o x x o
## 5 x x x o o o o o o o o x x o
## 6 x x x x o o o o o o o x o o
## 7 x o x x o o o o o o o x x o
```

```
##  
## Box-Ljung test  
##  
## data: errortermsfor3c  
## X-squared = 954.2786, df = 50, p-value < 2.2e-16
```

Even though the lead-1 predictions closely follow the validation data, we also see that the Arima(1,0,0)x(1,1,0) process model constantly overshoots the validation observations and thus we have a very seasonal plot of residuals, sample ACF and sample PACF. We are not sure how to fix the overshooting.