Hand-in Assignment 1

1MS014 Analysis of Time Series

Uppsala University

Aditya Khadkikar, Aviral Jain

July 1, 2025

Task 1:

Handin Assignment 1

1) Since $\{W_t\}$ is a white noise Process, all distinct W_t are independent with zero mean.

$$\mathbb{E}[x_t] = \mathbb{E}[1 + W_t W_{t-1} + 0.6\, W_{t-1} W_{t-2}]$$
$$= 1 + \mathbb{E}[W_t W_{t-1}] + 0.6\, \mathbb{E}[W_{t-1} W_{t-2}]$$

$$\Rightarrow \mathbb{E}[x_t] = 1 + 0 + 0 = 1$$

Autocovariance $\gamma(h) = Cov(x_t, x_{t+h})$
$$= \mathbb{E}[(x_t - \mathbb{E}[x_t])(x_{t+h} - \mathbb{E}[x_{t+h}])]$$

let $A_t = W_t W_{t-1}$, $B_t = 0.6\, W_{t-1} W_{t-2}$

Case $h=0$: $\gamma(0) = \mathbb{E}[(A_t + B_t)^2] = \mathbb{E}[A_t^2] + 2\mathbb{E}[A_t B_t] + \mathbb{E}[B_t^2]$

$\mathbb{E}[A_t^2] = \mathbb{E}[W_t^2 W_{t-1}^2] = \sigma_w^2 \cdot \sigma_w^2 = 100$

$\mathbb{E}[B_t^2] = 0.6^2\, \mathbb{E}[W_{t-1}^2 W_{t-2}^2] = 0.36 \cdot \sigma_w^2 \cdot \sigma_w^2 = 36$

$\mathbb{E}[A_t B_t] = 0.6\, \mathbb{E}[W_t W_{t-1} \cdot W_{t-1} W_{t-2}] = 0.6 \times 0 \times 10 \times 0 = 0$

$\gamma(0) = 100 + 0 + 36 = 136$

Case $h=1$: We want $\gamma(1) = \mathbb{E}[(A_t + B_t)(A_{t+1} + B_{t+1})]$

- $A_t = W_t W_{t-1}$, $A_{t+1} = W_{t+1} W_t$

  $\mathbb{E}[A_t A_{t+1}] = \mathbb{E}[W_t W_{t-1} W_{t+1} W_t] = 10 \times 0 \times 0 = 0$

- $A_t B_{t+1} = W_t W_{t-1} \cdot 0.6\, W_t W_{t-1} = 0.6\, W_t^2 W_{t-1}^2$

  $\mathbb{E}[A_t B_{t+1}] = 0.6\, \mathbb{E}[W_t^2 W_{t-1}^2] = 0.6 \times 10 \times 10 = 60$

- $B_t A_{t+1} = 0.6\, W_{t-1} W_{t-2} \cdot W_{t+1} W_t = 0$
- $B_t B_{t+1} = 0.6\, W_{t-1} W_{t-2} \cdot 0.6\, W_t W_{t-1} = 0.36\, W_{t-1}^2 W_{t-2} W_t$

  $= 0.36 \times 10 \times 0 \times 0 = 0$

$\therefore \gamma(1) = 60$

Case $h = 2$: Only Possible Overlap is

$A_t = W_t W_{t-1}$, $B_{t+2} = 0.6\, W_{t+1} W_t$

$A_t B_{t+2} = 0.6\, W_t^2 W_{t-1} W_{t+1} = 0$

$\therefore \gamma(2) = 0$

Autovariance function:

$\gamma(0) = 136$, $\gamma(1) = 60$, $\gamma(h) = 0$ for $|h| \geq 2$

\\ $\{x_t\}$ is stationary as ~~its~~ mean is constant over time : $\mathbb{E}[x_t] = 1$
and autocovariance $\gamma(h)$ depends only on lag h and not on time t

Task 2:

2) Determine if the Process is stationary →
This is an ARMA(2,1) model

- AR component: $0.5\, x_{t-2}$
- MA component: $W_t + 0.4\, W_{t-1}$

Characteristic eq^n for AR is
$1 - 0.5\, z^2 = 0 \quad \Rightarrow \quad z = \pm\sqrt{2}$

Since the root lie Outside unit circle, the Process is stationary.

- Find autocovariance function $\gamma(h)$
- $\gamma(0) = \mathbb{E}[x_t^2]$

$$= \mathbb{E}[(W_t + 0.4\, W_{t-1})^2] + 0.5^2\, \mathbb{E}[x_{t-2}^2] + 2 \cdot 0.5 \cdot \mathbb{E}[(W_t + 0.4\, W_{t-1})x_{t-2}]$$
$$= 1.16 + 0.25\, \gamma(0) + 0$$
$$= 1.16 + 0.25\, \gamma(0)$$

$\therefore \gamma(0) = 1.16 + 0.25\,\gamma(0) \Rightarrow \gamma(0)(1 - 0.25) = 1.16$
$\Rightarrow \gamma(0) = 1.546$

- $\gamma(1) = \mathbb{E}(x_t\, x_{t-1})$.

Only Taking Terms where noise Overlaps

$\mathbb{E}[0.4\, W_{t-1} \cdot W_{t-1}] = 0.4\, \mathbb{E}[W_{t-1}^2] = 0.4$

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} \approx 0.258$$

- $\gamma(2) = \mathbb{E}[x_t x_{t-2}] = \mathbb{E}[0.5\, x_{t-2}^2]$
  $= 0.5\, \gamma(0) = 0.773$

$\rho(2) = 0.5$

- $\gamma(3) = 0.5 \times \gamma(1)$ because x_{t-3} only shows up in x_{t-1}

$\gamma(3) = 0.2$

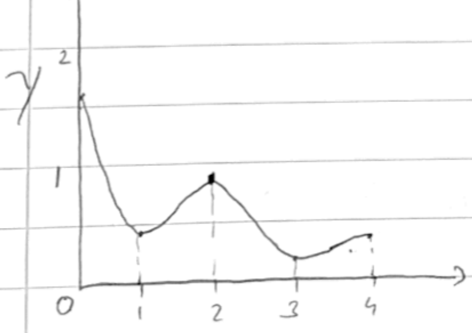
$\rho(3) \approx 0.129$

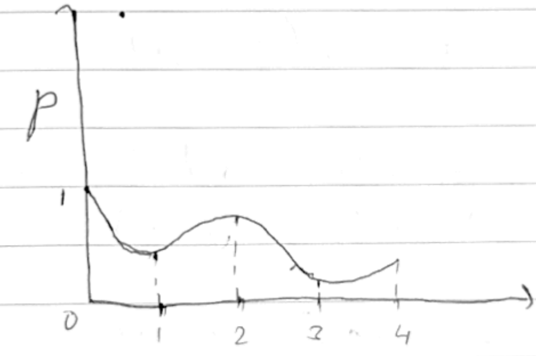
- $\gamma(4) = 0.5 \times \gamma(2) = 0.386$

$\rho(4) \approx 0.25$

→ ~~Lg~~ Lag h

Lag h	$\rho(h)$
1	0.258
2	0.50
3	0.129
4	0.25





Task 3:

3) A) Roots of AR Polynomial

$\phi = 1 - 0.1B - 0.3B^2 = \dfrac{0.1 \pm \sqrt{0.01 + 1.2}}{-0.6} = \dfrac{0.1 \pm 1.1}{-0.6}$

$= -2,\ 1.66$

Both roots are outside unit disk

$\rightarrow$ Casual

Roots of MA Polynomial

$\theta = 1 + 0.4B \Rightarrow B = -2.5$

root is outside unit disk

$\Rightarrow$ Invertible

b) $x_{201} = 0.1\, x_{200} + 0.3 x_{199} + W_{201} + 0.4\, W_{200}$

$= 0.09$

We assume $W \rightarrow 0$ as m increases because mean is 0

$x_{202} = 0.1\, x_{201} + 0.3 x_{200} + W_{202} + 0.4\, W_{201}$

$= 0.099$

c) $W_t \sim N(0, 0.1)$

One step ahead $(x_{201}) \Rightarrow$

$Var(x_{201} \mid info) = Var(W_{201}) = \sigma_w^2 = 0.1$

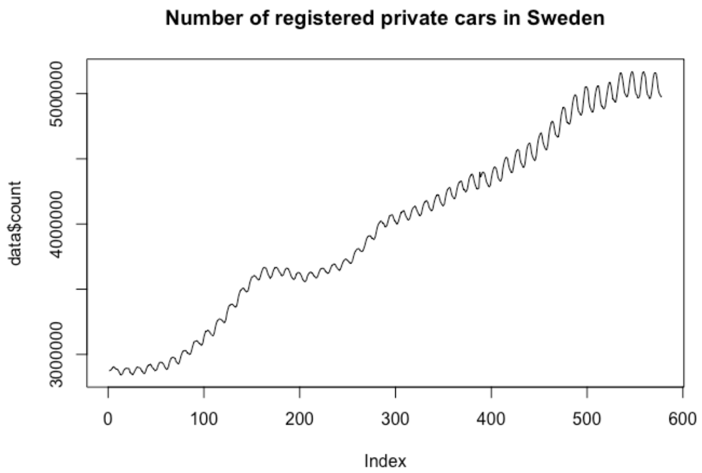
$SE = \sqrt{0.1} \approx 0.316$

Task 4:

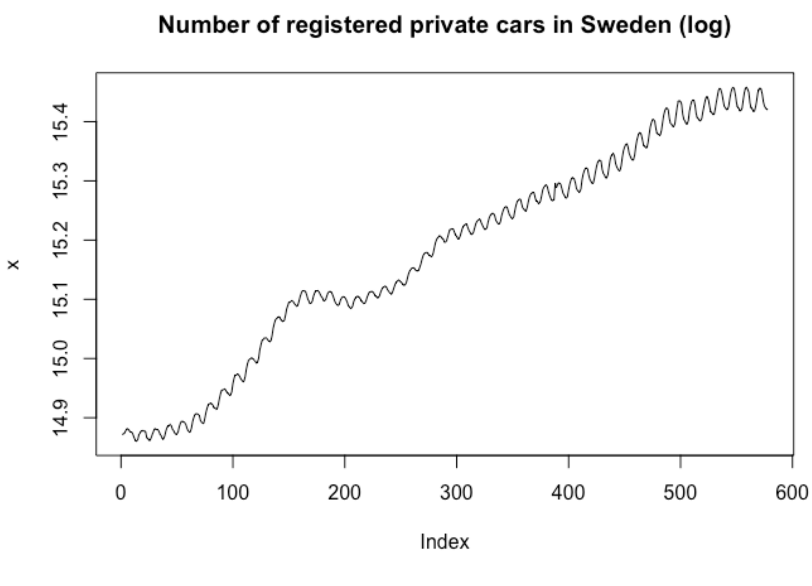
The number of registered private cars in Sweden for the years 1977 until February 2025, monthly data, is given in the second column of the file carsmon.dat at Studium.

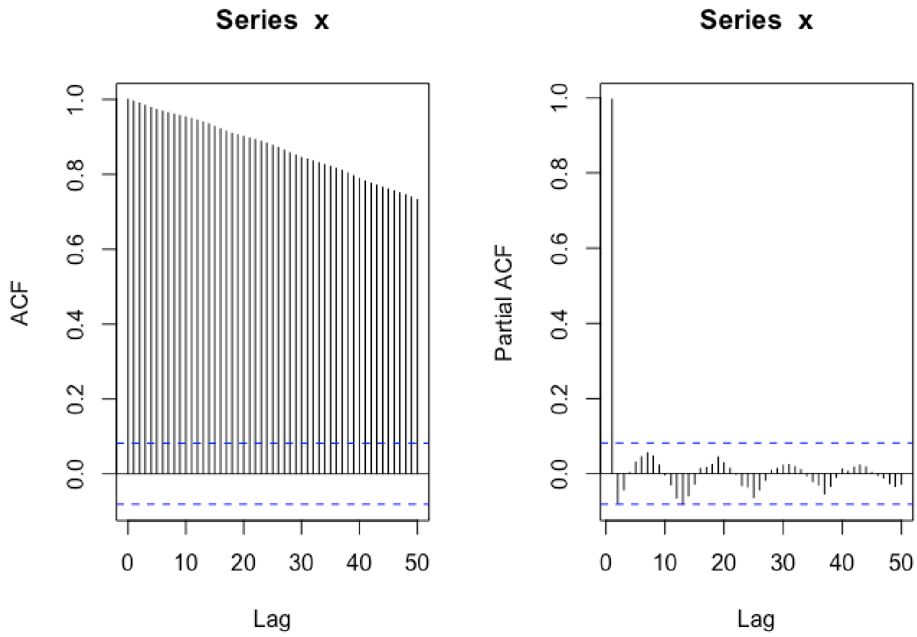
Find a suitable ARIMA (or SARIMA) model for these data, or a transformation thereof. Analyze the model residuals carefully, in order to make sure that the model provides a good description of the data. It might be a good idea to try transformations, like the logarithm.

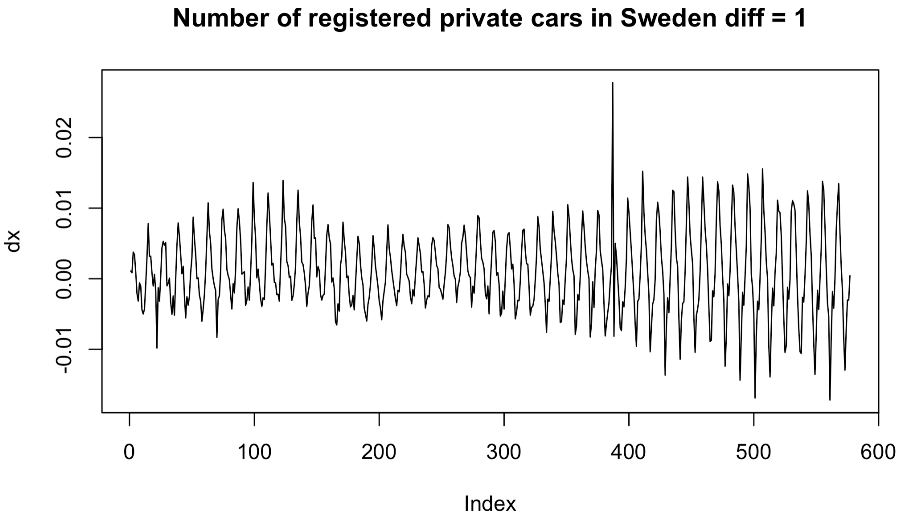
Below is the time-series data plotted for the dataset of registered private cars from 1977-2025:



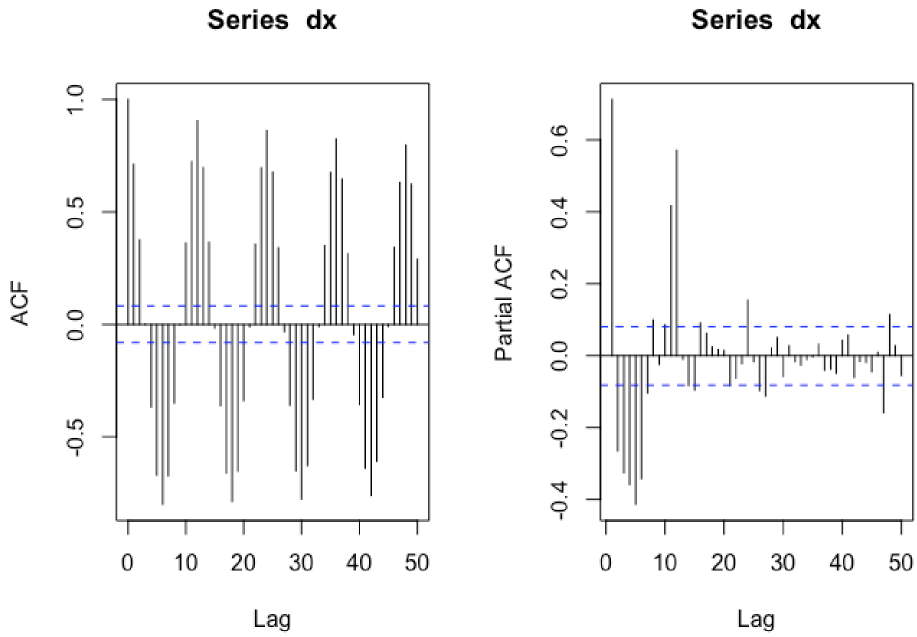
As the mean of the data with respect to time is changing, along with slight differences in the variance noticeable after around index 350, the data appears to be non-stationary initially. Upon applying a log transformation, to reduce the magnitude of the variance, and by applying a difference of lag=1, the graph looks like the following:





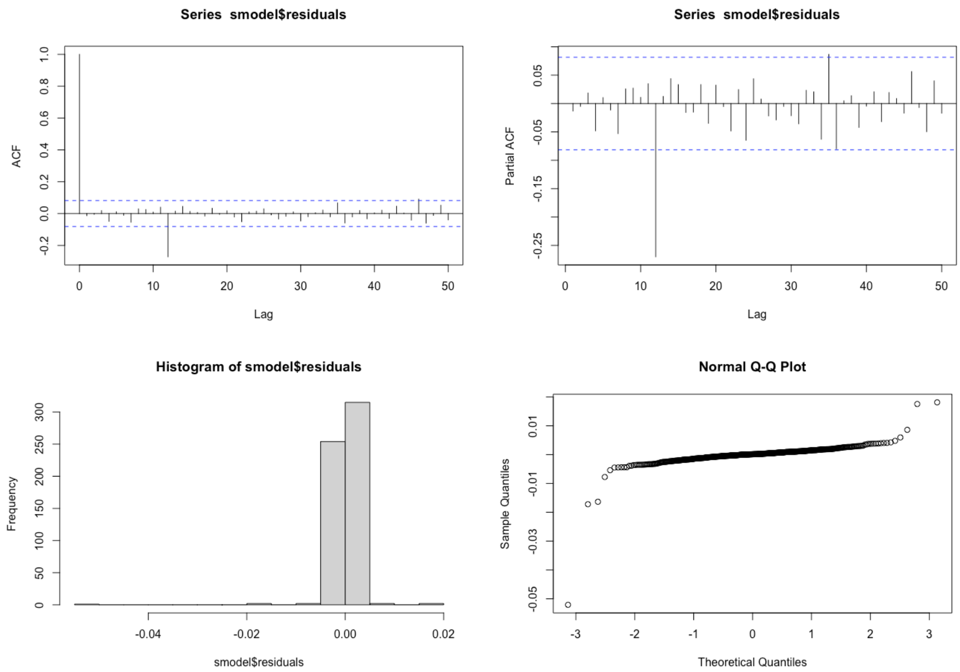


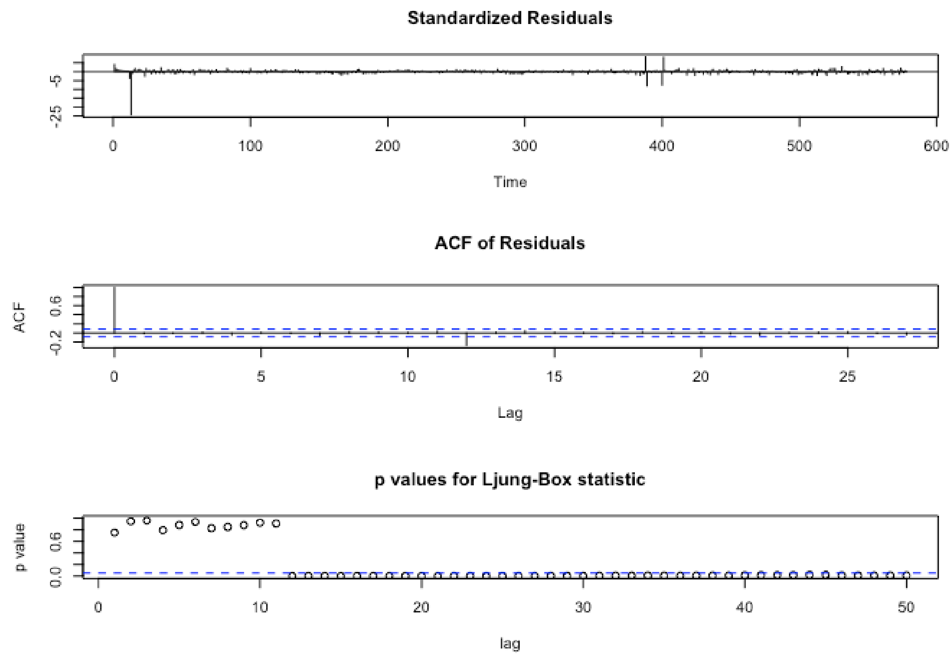
We now compute the autocorrelation (ACF) and partial autocorrelation (PACF) curves for the data, now that it is made to be more stationary. From the graphs, periodic peaks can be seen at lags 12, 24, 36 and so on, with a general slow decay. From the PACF, the noticeable peaks are present in lags up until 12, however, more information is yet to be derived.



As some seasonal pattern is very likely present, we first try fitting an empty seasonal model, where the period is set to 12.

$$\text{ARIMA}(0,1,0).(0,1,0)_{12}$$





Description of the p-values, and relevant statistics: From the empty SARIMA model, log likelihood was calculated as 2670.76 and the Akaike Information Criterion (aic) was -5339.52.

A general SARIMA model looks like the following (with difference kept constant to 1, as the stationarity aspect was achieved, and over-differencing may negatively impact the results):

$$\text{SARIMA}(p, 1, q).(P, 1, Q)_{12}$$

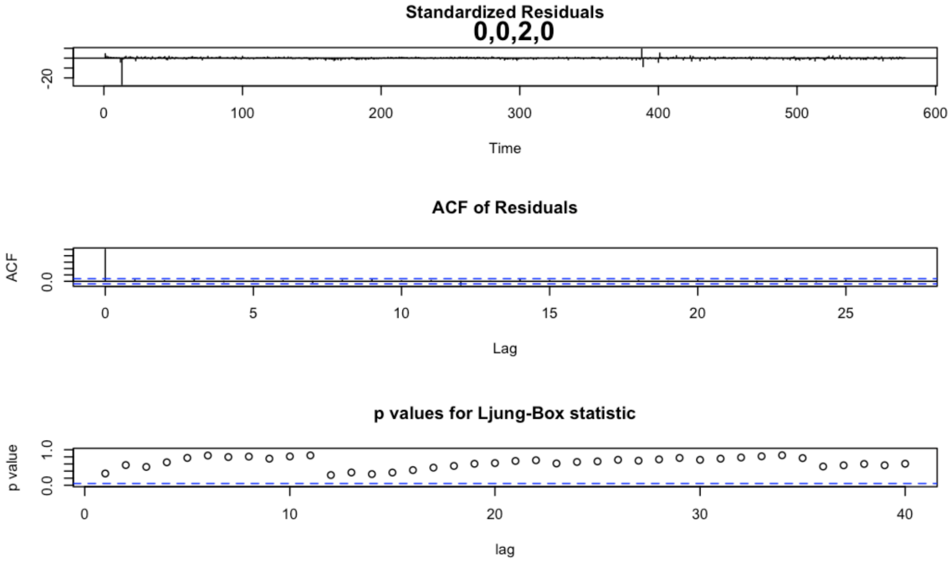
The main signs to find the model best fitting to the time-series data are i) *inspecting the standardized residuals, as well as their ACF/PACF* to see if white-noise like behaviour is shown, ii) *Ljung-Box Statistic p-values* and their behvaiour with increasing lags (not going below the error bound for longer lags), and iii) the AIC.

A grid-search based approach was done, where p, q (for the non-seasonal component), and P, Q (seasonal component) were varied from range of 0 to 4. The first iteration gave the benchmark empty seasonal model described earlier.

In the first few iterations, an example is $\text{SARIMA}(0, 1, 0).(1, 1, 0)_{12}$, which had good p-values for later lags than 12, above the acceptance threshold (seasonal AR(1)). By means of inspection after the received ACF/PACF plots, 15 models with significant p-values in the Ljung-Box plots were:

- $\text{SARIMA}(0, 1, 0).(1, 1, 0)_{12}$, $\text{SARIMA}(0, 1, 0).(2, 1, 0)_{12}$
- $\text{SARIMA}(3, 1, 3).(0, 1, 1)_{12}$, $\text{SARIMA}(3, 1, 3).(0, 1, 2)_{12}$, $\text{SARIMA}(3, 1, 3).(0, 1, 3)_{12}$, $\text{SARIMA}(3, 1, 3).(0, 1, 4)_{12}$
- $\text{SARIMA}(3, 1, 3).(1, 1, 1)_{12}$, $\text{SARIMA}(3, 1, 3).(1, 1, 2)_{12}$, $\text{SARIMA}(3, 1, 3).(1, 1, 3)_{12}$, $\text{SARIMA}(3, 1, 3).(1, 1, 4)_{12}$
- $\text{SARIMA}(3, 1, 3).(2, 1, 1)_{12}$, $\text{SARIMA}(3, 1, 3).(2, 1, 2)_{12}$, $\text{SARIMA}(3, 1, 3).(2, 1, 3)_{12}$
- $\text{SARIMA}(3, 1, 3).(3, 1, 0)_{12}$, $\text{SARIMA}(3, 1, 3).(3, 1, 1)_{12}$

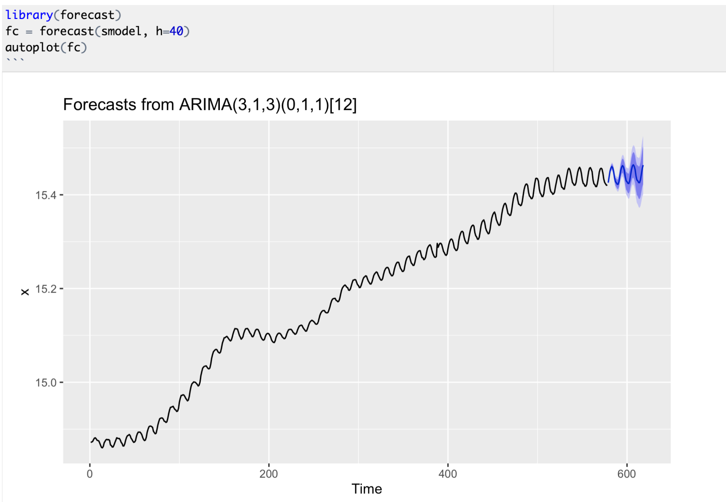
An example of a model with p-values above the confidence threshold for longer lags from h=12 is shown here:



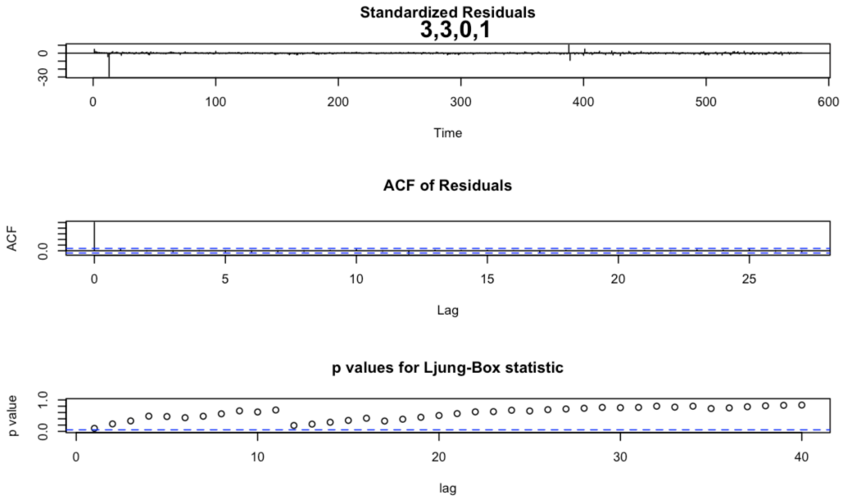
The top-3 smallest AICs from all of the observed good p-value SARIMA models, are marked below:

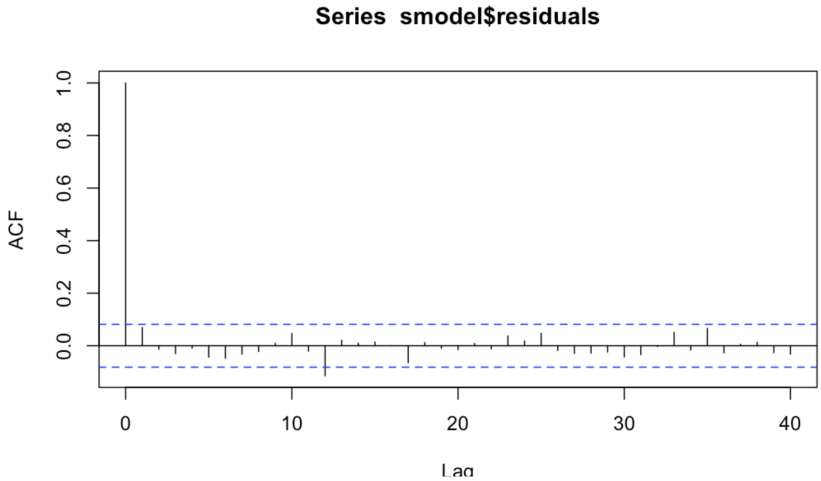
- ... $\text{SARIMA}(0,1,0).(2,1,0)_{12}$ -5441.355
- $\text{SARIMA}(3,1,3).(0,1,1)_{12}$ **-5532.681**, $\text{SARIMA}(3,1,3).(0,1,2)_{12}$ **-5530.727**, $\text{SARIMA}(3,1,3).(0,1,3)_{12}$ -5528.845, $\text{SARIMA}(3,1,3).(0,1,4)_{12}$ -5528.377
- $\text{SARIMA}(3,1,3).(1,1,1)_{12}$ **-5530.729**, $\text{SARIMA}(3,1,3).(1,1,2)_{12}$ -5528.817, $\text{SARIMA}(3,1,3).(1,1,3)_{12}$ -5527.005, $\text{SARIMA}(3,1,3).(1,1,4)_{12}$ -5526.928
- ...

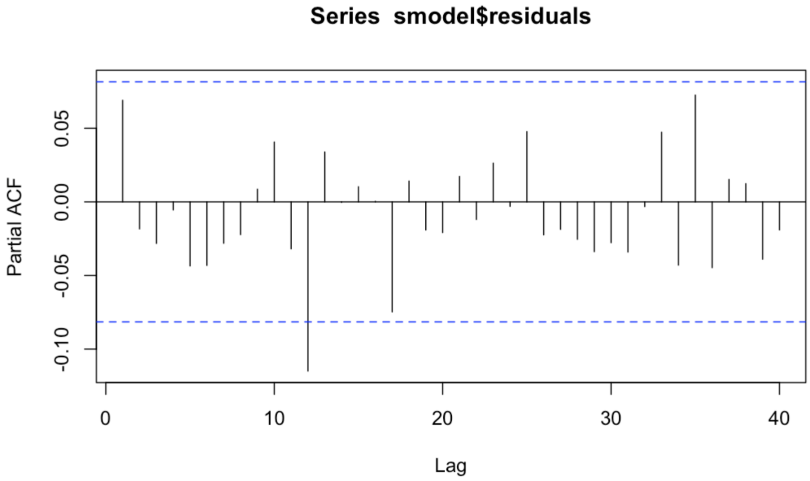
Hence, a SARIMA model to fit the registered cars time-series dataset chosen, based on the achieved AICs, and manual inspection of the Ljung-Box Statistics, and the residuals having a random, white-noise like behaviour, apart from small outliers, is $\text{SARIMA}(3,1,3).(0,1,1)_{12}$ (AIC: **-5532.681**).

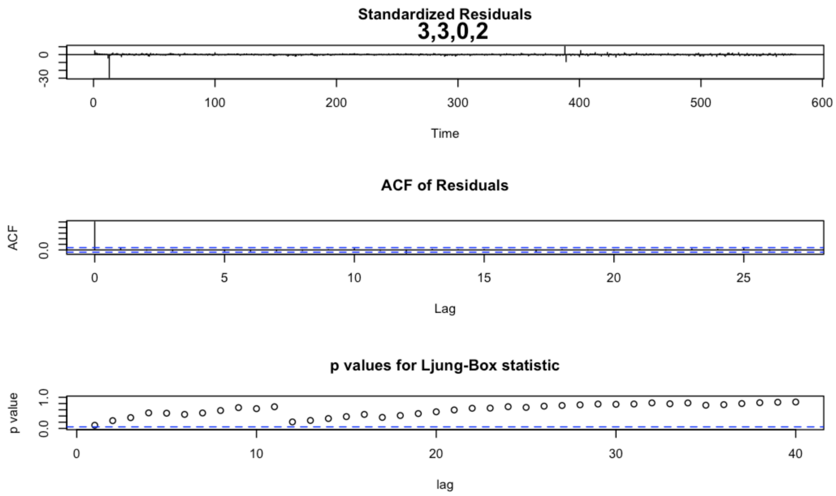


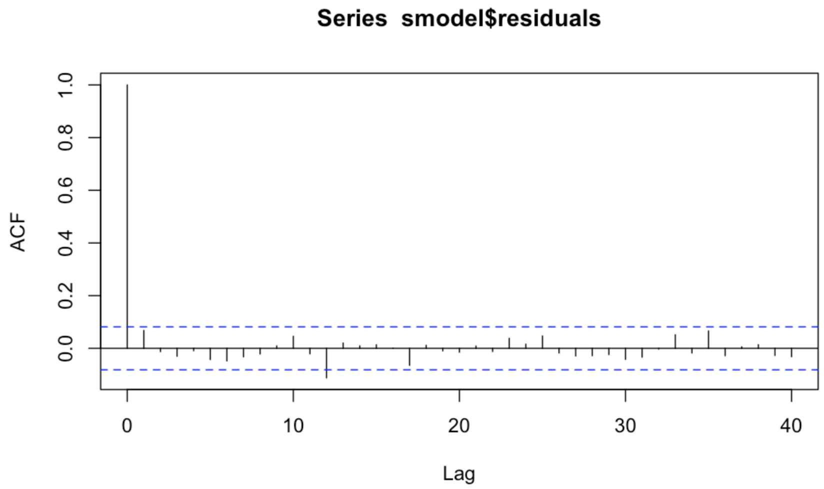
Detailed plots of the top-3 lowest-AIC SARIMA models fitted, AND significant p-values:

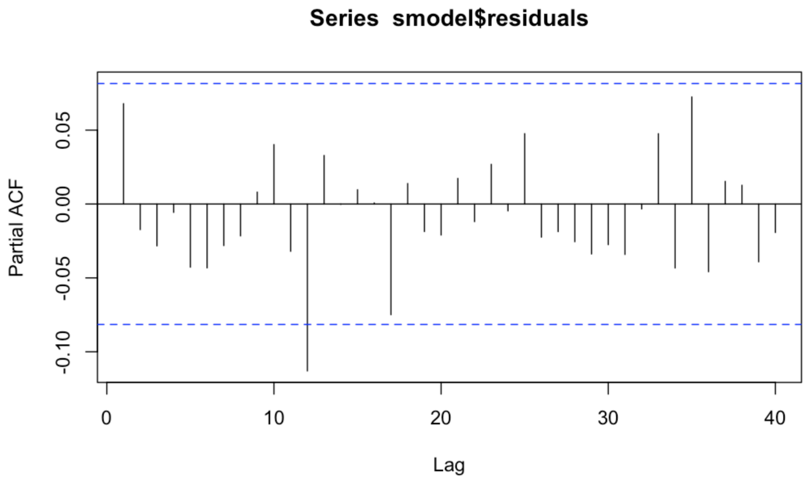


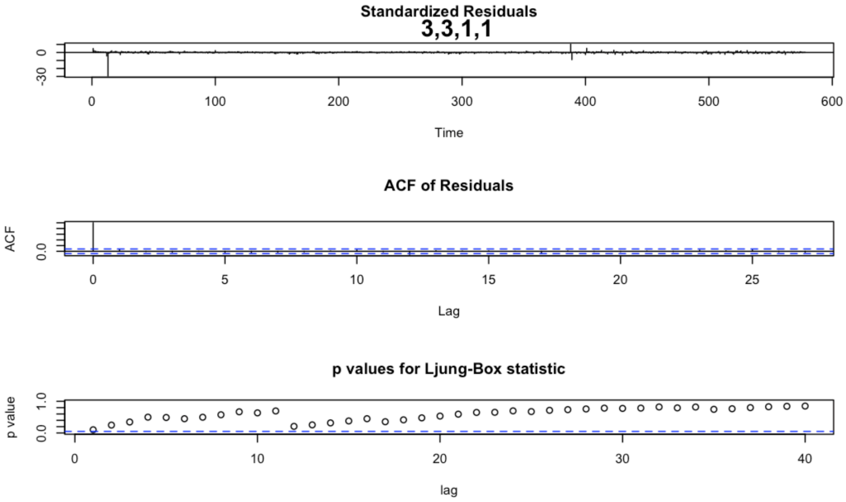


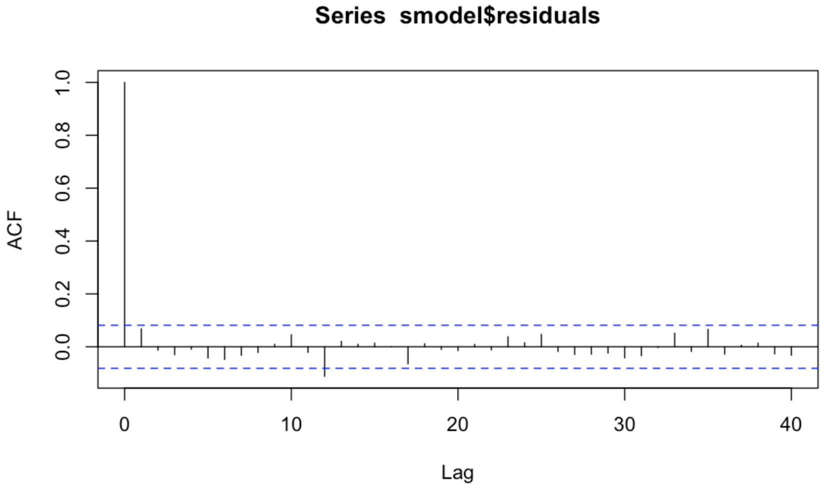


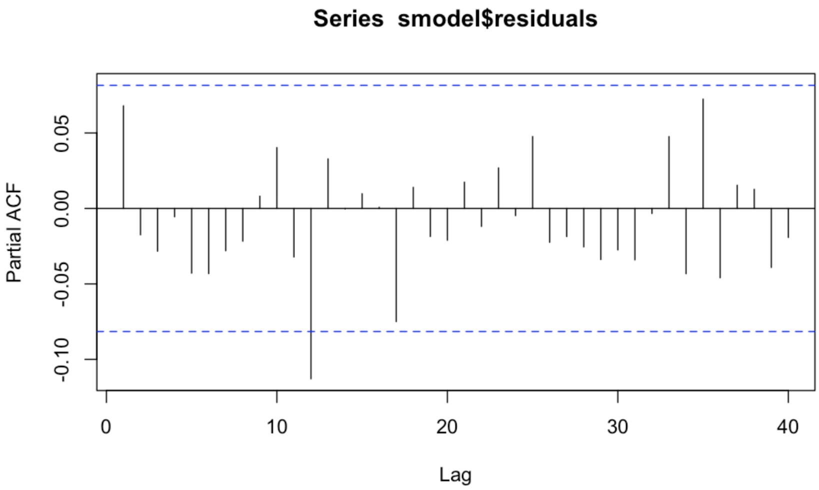

Standardized Residuals
3,3,0,2
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag


Series smodel$residuals
ACF
Lag


Series smodel$residuals
Partial ACF
Lag


Standardized Residuals
3,3,1,1
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag


Series smodel$residuals
ACF
Lag


Series smodel$residuals
Partial ACF
Lag

A Code

```r
#########################
library(stats)
library(here)

data = read.table(here('carsmon.dat'), header=TRUE)
x = log(data$count) # logarithm operation

plot(data$count, type='l')
title(main = "Number of registered private cars in Sweden")
plot(x, type='l')
title(main = "Number of registered private cars in Sweden (log)")

par(mfrow=c(1,2)) # Making the display of 2 plots
acf(x, lag.max = 50) # ACF
pacf(x, lag.max = 50) # PACF

## DIFFERENCING
lag = 1

dx = diff(x, lag=lag)
plot(dx, type='l')
title(main = paste("Number of registered private cars in Sweden diff =", lag, sep=""))

par(mfrow=c(1,2)) # Making the display of 2 plots

acf(dx, lag.max = 20)
pacf(dx, lag.max = 20)

## SARIMA MODEL EXPLORATION

aic_vals <- c()

for(p in 0:4){
  for(q in 0:4){
    for(P in 0:4){
      for(Q in 0:4){
        tryCatch({
          smodel=arima(x,order=c(p,1,q),
           seasonal=list(order=c(P,1,Q),period=12)
           ) # varying p,q,P,Q

          acf(smodel$residuals, lag.max = 40) # ACF
          pacf(smodel$residuals, lag.max = 40) # PACF

          tsdiag(smodel, gof.lag = 40)

          aic<-AIC(arima(x,order=c(p,1,q),seasonal=list(order=c(P,1,Q),period=12)))
          title(paste(p, q, P, Q, sep=","))
          aic_vals[paste(p, q, P, Q, sep=",")]<-aic
        }, error = function(e) {
          # Handle the error
          cat("An error occurred:", conditionMessage(e), "\n")
        }, finally = {
          cat("This block is always executed.\n")
        })
      }
    }
  }
}
```

```
# get the output (key-value output, where key=the configuration of the SARIMA, value
    =AIC)
aic_vals

##########################
smodel=arima(x,order=c(3,1,3),
         seasonal=list(order=c(0,1,1),period=12)
         )

acf(smodel$residuals, lag.max = 40) # ACF
pacf(smodel$residuals, lag.max = 40) # PACF

# Histogram
hist(smodel$residuals)
qqnorm(smodel$residuals)

tsdiag(smodel, gof.lag = 40)

## FORECAST
library(forecast)
fc = forecast(smodel, h=40)
autoplot(fc)
```

Listing 1: R code used for Task 4.