

# TMA4285 Time series models

## Exercise 8: Comparison of a state space model and a SARIMA model

*Sivert Selnes, Kristine L. Mathisen, Gina Magnussen*

*20th of October 2018*

### Remember:

- Note how uncertainty should be written: 0.056(7) (See sources on web page)

## Structure

### Title

### Abstract

### Introduction

### Theory

Theory is written here (in theory.Rmd)

### \* State-space models

General:

$$\begin{aligned} Y_t &= G_t \mathbf{X}_t + W_t & \{\mathbf{W}_t\} &\sim \text{WN}(\mathbf{0}, \{R_t\}) \\ \mathbf{X}_{t+1} &= F_t \mathbf{X}_t + \mathbf{V}_t & \{\mathbf{V}_t\} &\sim \text{WN}(\mathbf{0}, \{Q_t\}) \end{aligned}$$

When F, G, R and Q are time independent:

$$\begin{aligned} Y_t &= G \mathbf{X}_t + W_t & \{W_t\} &\sim \text{WN}(\mathbf{0}, R) \\ \mathbf{X}_{t+1} &= F \mathbf{X}_t + \mathbf{V}_t & \{\mathbf{V}_t\} &\sim \text{WN}(\mathbf{0}, Q) \end{aligned}$$

### \* SARIMA

(p,d,q) is ARIMA model *within* each season. (P,D,Q) is the ARIMA model between each season and is the seasonal component of the SARIMA model.

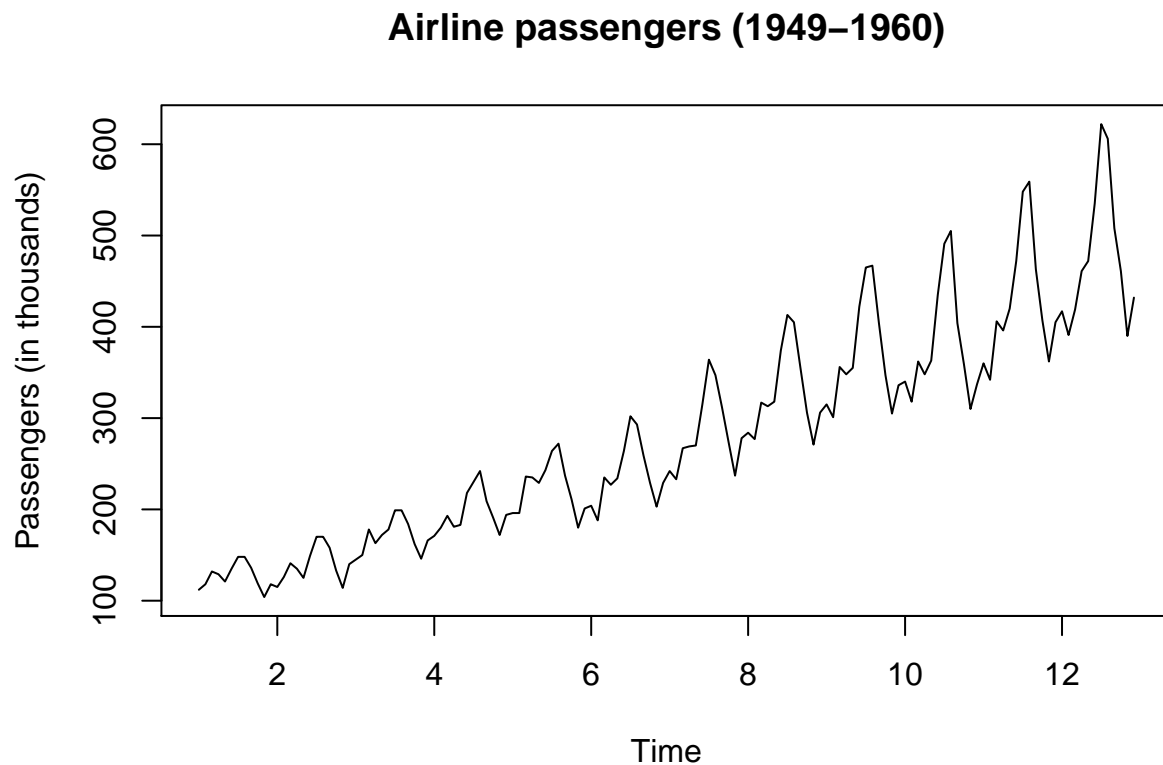
$$\text{SARIMA}(p, d, q) \times (P, D, Q)_s$$

## Data analysis

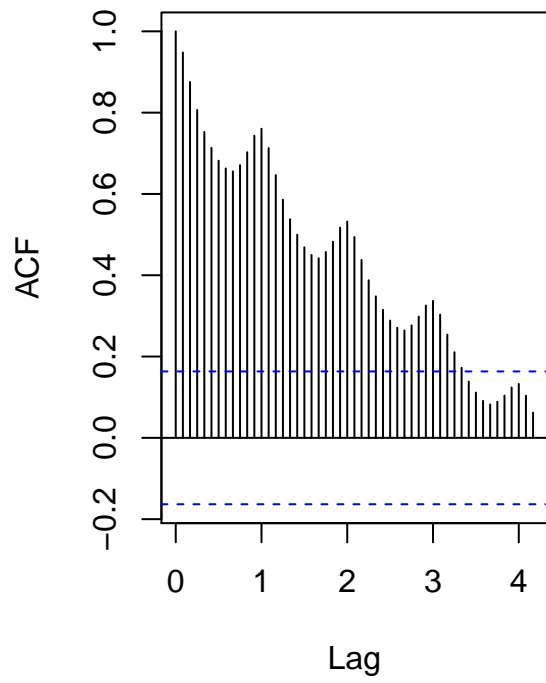
1. Exploring the data with plotting of relevant statistics
2. Justification of the choice of model
3. Model parameter estimation including uncertainty
4. Model prediction at the given sample points and for the next year including uncertainty of the best linear predictions
5. Diagnostics, and model choice discussion including comparison of the two models

$SARIMA(p, d, q) \times (P, D, Q)_s$  model

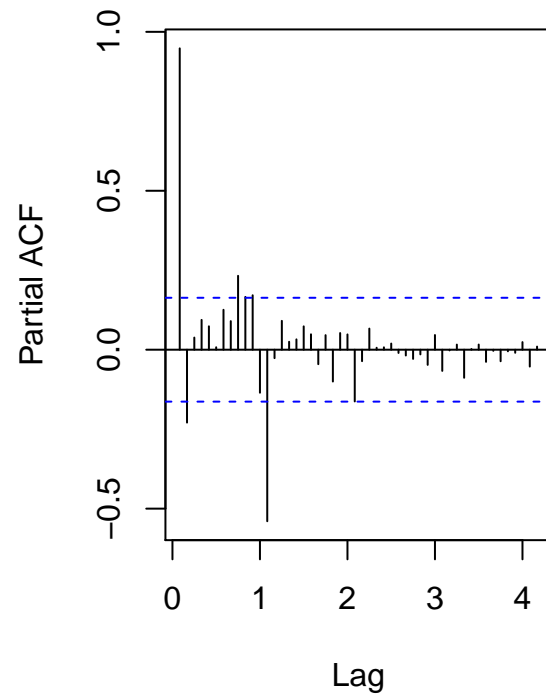
The data set consists of total number of international airline passengers, in thousands, for each month from January 1949 to December 1960(reference to data set), giving a total number of  $N = 144$  observations.



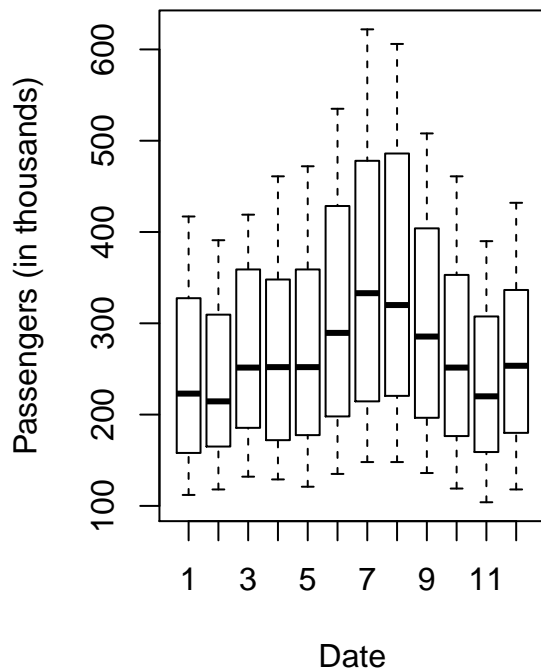
**V1**



**Series dataseries**



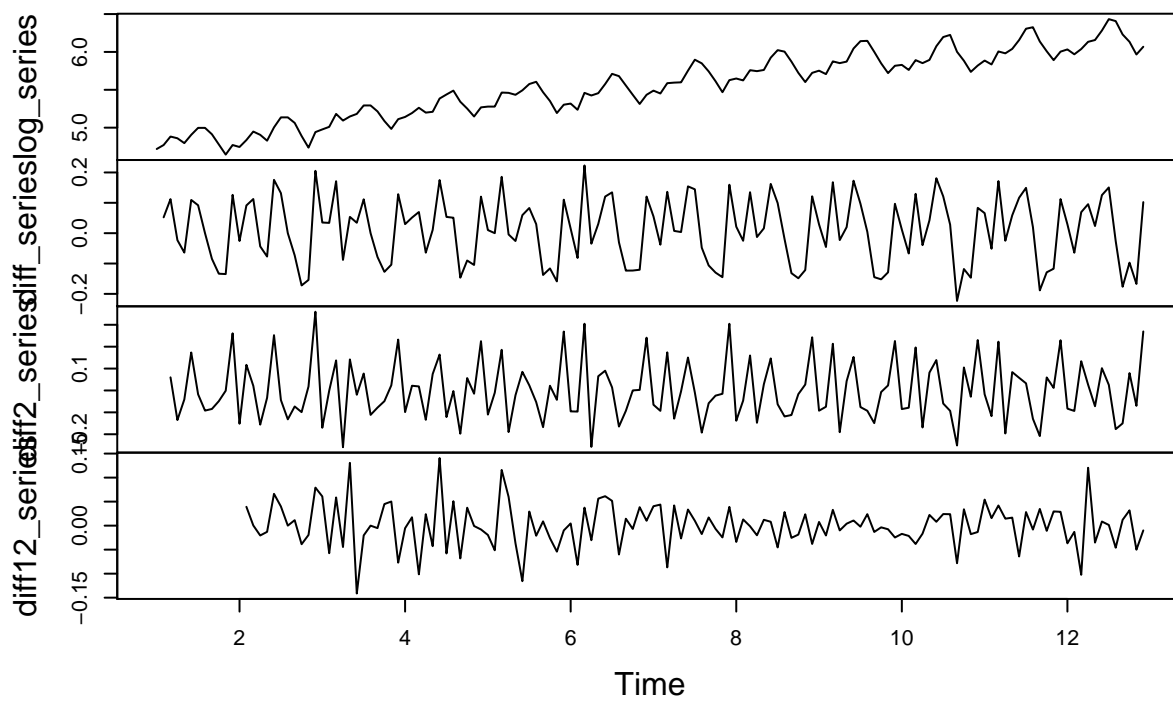
## Monthly international air passengers

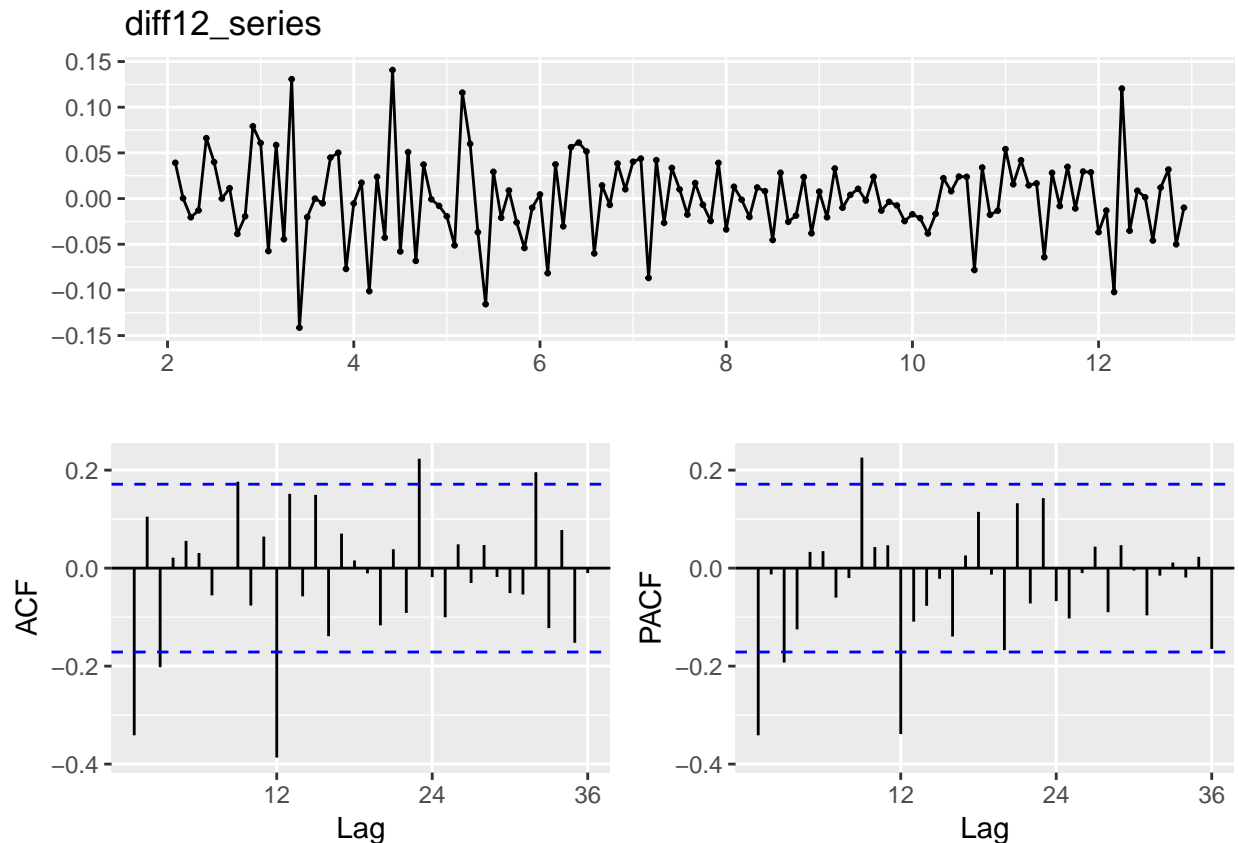


From the time series plot, this is obviously a non-stationary series as both trend and seasonality is visible in the visualization. Trend is there because the total number of airline passengers on average increases for each month, and seasonality because of the wave-like behaviour of the series. Non-stationarity is further supported by looking at the autocorrelation function, which shows considerable correlation between observations even up to lag  $h$  of size 30, and again a periodic wave-like pattern.

To investigate the data, we first find a transformation to the time series. A normal transformation is the  $\log()$ -transformation. This transformation seems to stabilize the multiplicative behaviour of the variance so that the variance is not increasing with time.

## Transformed and differenced data





?? Check if this is correct, especially  $d = 1$  or  $2$ . Secondly, the data is differenced to remove trend. The parameter  $d$  is related to trend within a season, and since this trend seems to be removed by differencing once, we try  $d = 1$ . As there after differencing once still seems to be a pattern for a seasonal trend equal to the length of a year, i.e.  $s = 12$  in a seasonal ARIMA model, a twelfth-order difference is applied. The parameter  $D$  is related to this trend between seasons, i.e. the trend one can see from one January month to the next, the second and so on. We therefore set  $D = 1$ . Differencing once indicates linear trend, both within and between the seasons, which seems reasonable by looking at the time series plot itself.

Consider now the ACF and the PACF of the last differenced series.

The obtained series after applying the twelfth-order difference seems to be stationary without any trend or seasonality. (Show stationarity?) We can then estimate the parameters.

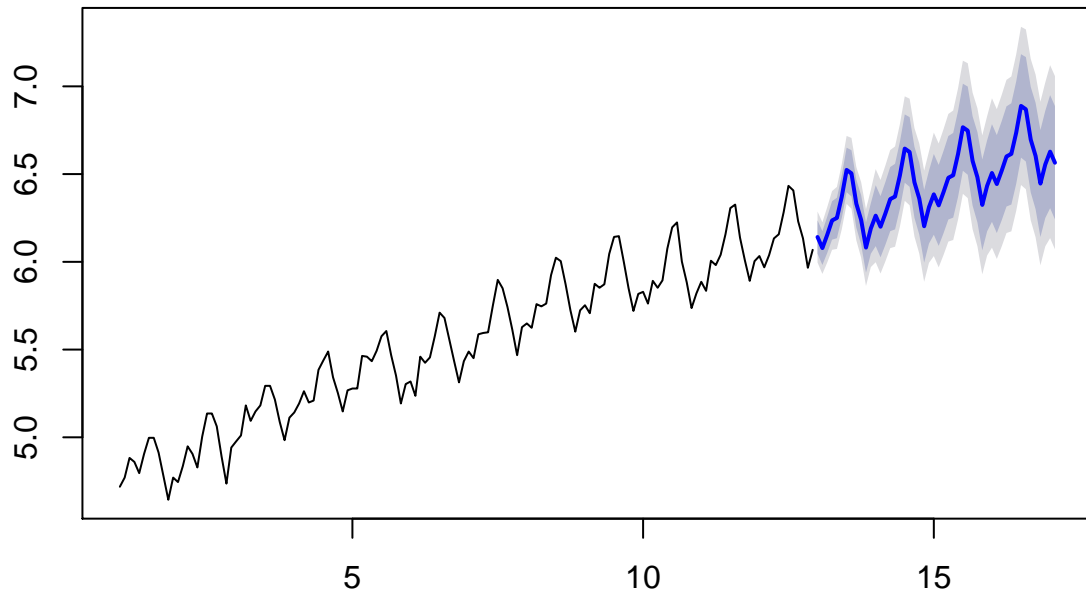
For the seasonal component, it looks like the ACF is cutting off at lag equal to one season, where  $s = 12$ , indicating  $Q = 1$ . The PACF seems to be tailing off after lag equal the length of one season, indicating  $P = 0$ . ?? Is the reasoning correct??

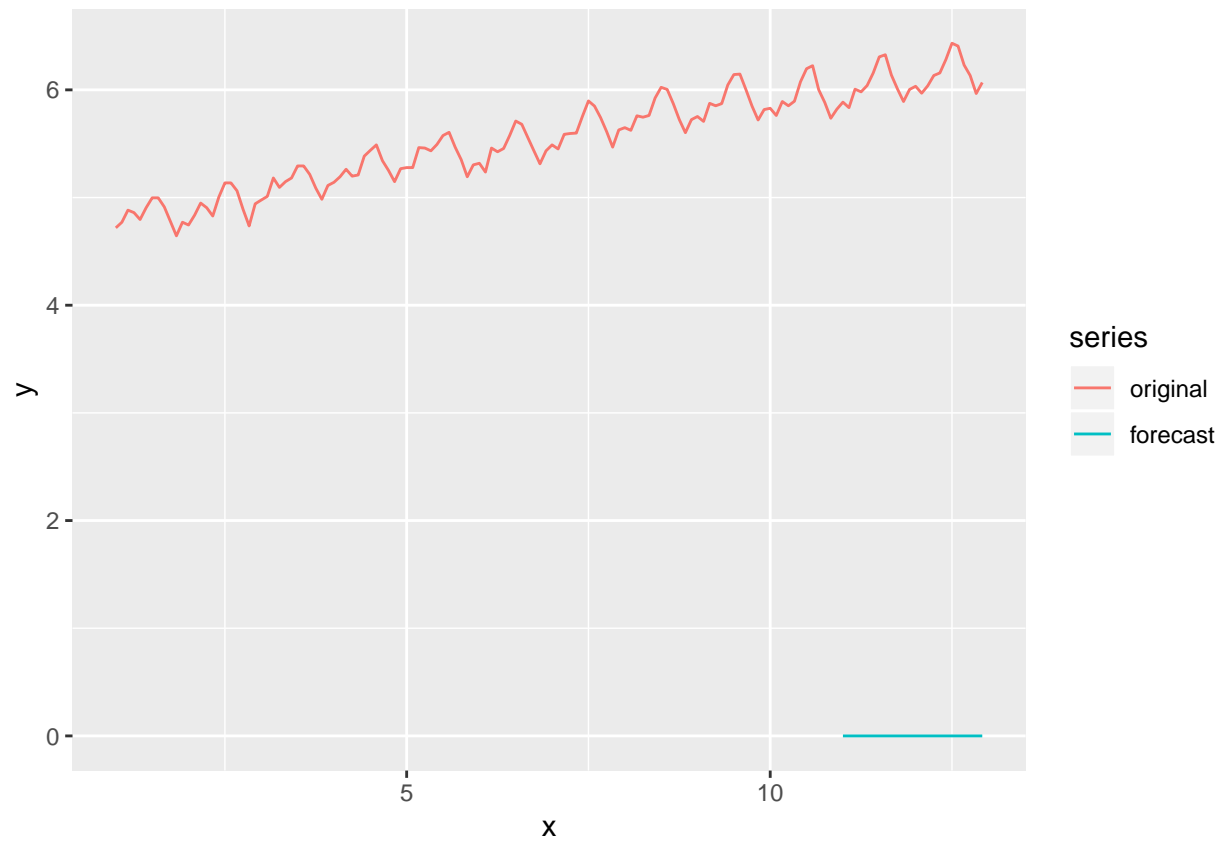
In the non-seasonal component, both the ACF and the PACF seem to be tailing off at lower lags. FILL IN HERE / EXPLANATION for choosing  $p, q$ .  $p = 0, q = 1$  are good parameters, I think.

?? Bootstrapping for estimation of uncertainty is preferred, but SE is also given here ?? \ ?? How to do bootstrapping ?? \ ?? Transform back ?? \

Forecasting SARIMA

### Forecasts from Basic structural model





**Bootstrap**

---

**Discussion**

**Conclusion**

**Appendix**

**References**