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:

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

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

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

* 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.

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

Data analysis

1. Exploring the data with plotting of relevant statistics

2. Justification of the choice of model

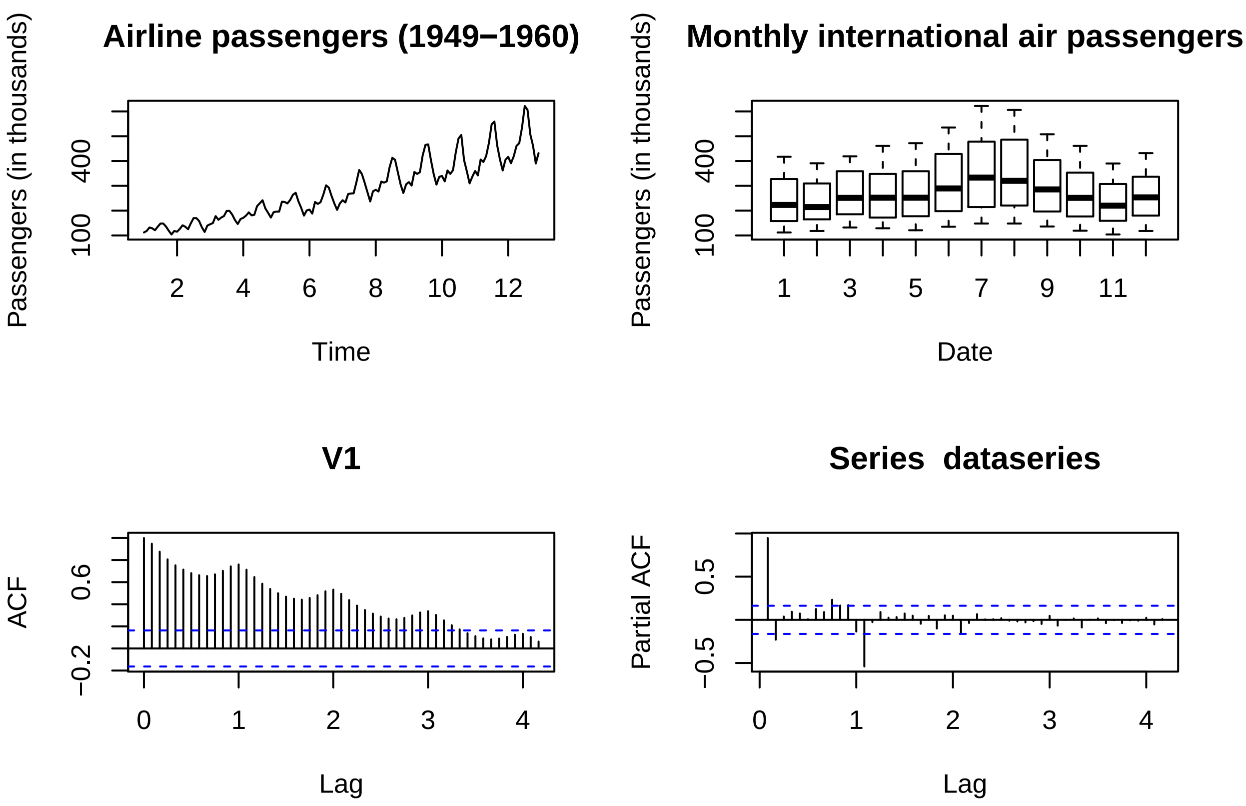
3. Model parameter estimation including uncertainty

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

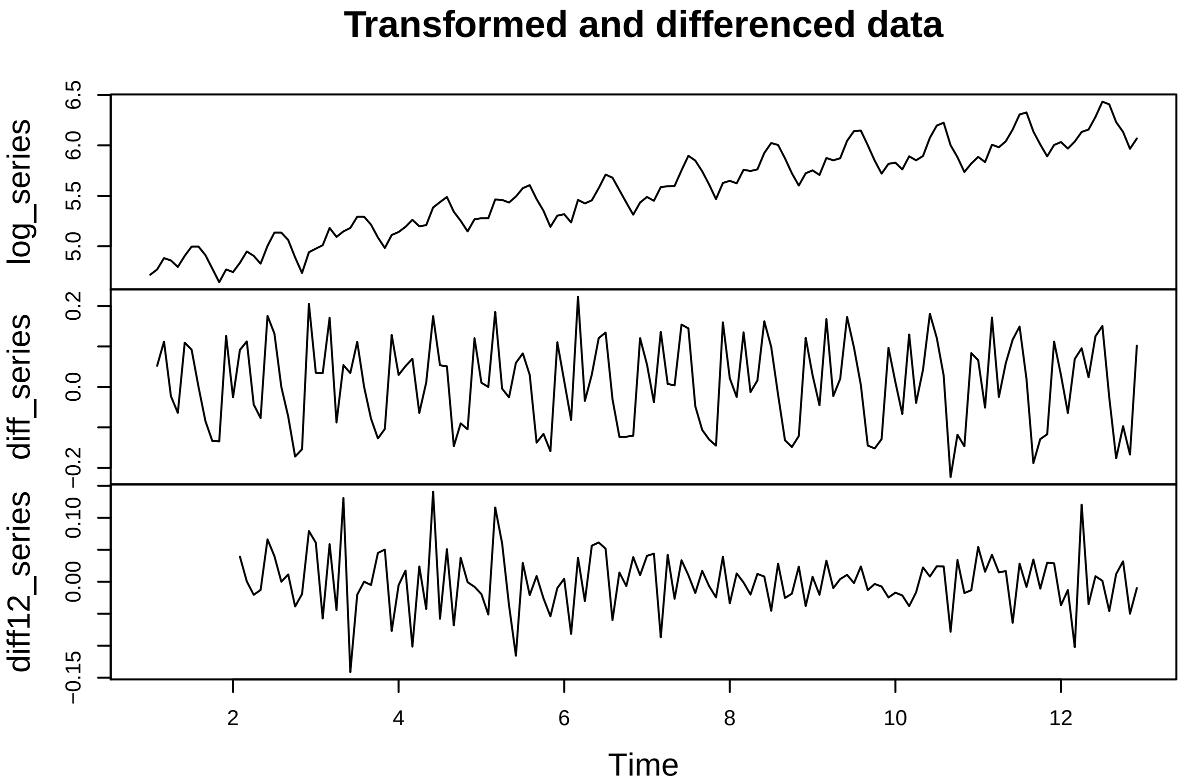
$SARIMA(p,d,q) \times (P,D,Q)_s$ **model**

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.

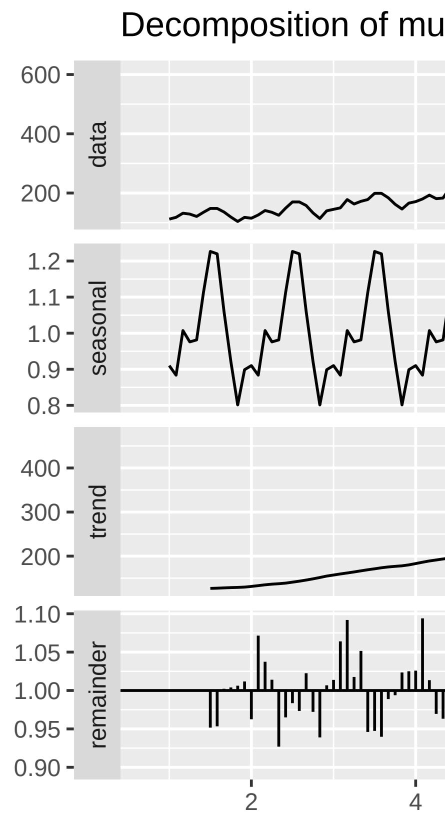


From the time series plot, this is obviously a non-stationary series as both trend and seasonality is visible in the visualization. Trend is there because the total number of airline passengers on average increases for each month, and seasonality because of the wave-like behaviour of the series. Non-stationarity 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. Furthermore, the box plot shows both higher mean number of passengers and also higher variance for months 6 to 9 in the year, i.e. June til September.

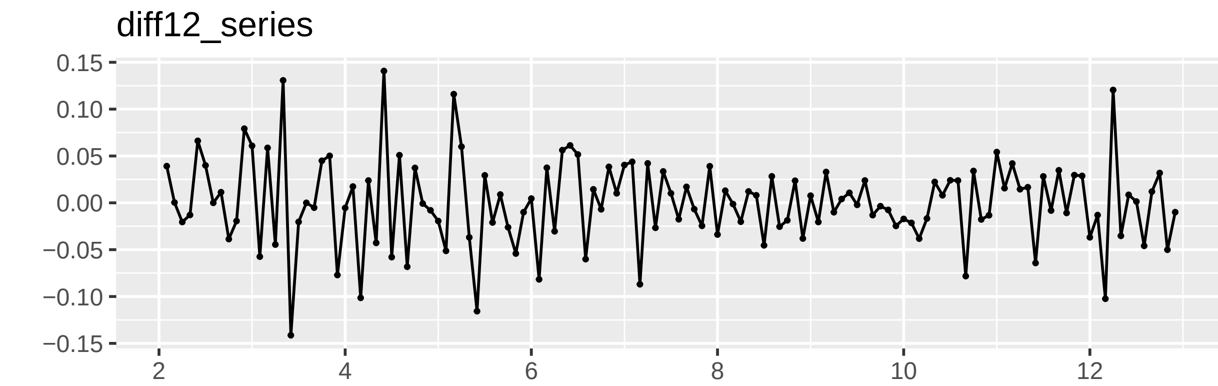
To investigate the data, we first find a transformation to the time series. A normal transformation is the log()-transformation, which seems to stabilize the multiplicative behaviour of the variance so that the variance is not increasing with time.

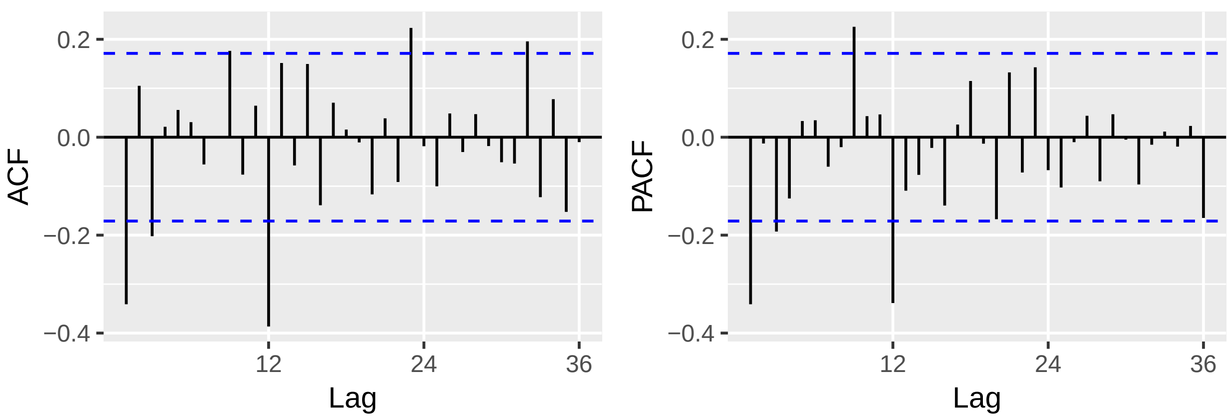


?? Check if this is correct, especially $d = 1 or 2$. Secondly, the data is differenced to remove trend. The parameter d is related to trend within a season, and since this trend seems to be removed by differencing once, we try $d = 1$. See the plot in the middle in the figure for transformed and differenced data. As there after differencing once still seems to be a wavy pattern for a seasonol trend equal to the length of a year, i.e. $s = 12$ in a seasonal ARIMA model, a twelfth-order difference is applied (bottom plot in figure transformed and differenced data). The parameter D is related to this trend between seasons, i.e. the trend one can se 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.



Our model parameter choices can be checked by comparing to the decomposed time series:
which clearly shows linear trend and a season equal to a year. 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 $P = 0$. ?? Ask about what is seen in the plots. Difference between tailing off and cutting off??

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.

Both parameters are significant in the model, and the residual analysis plot shows that our model is a good fit. In addition, the uncertainty of the parameters are given in the table.

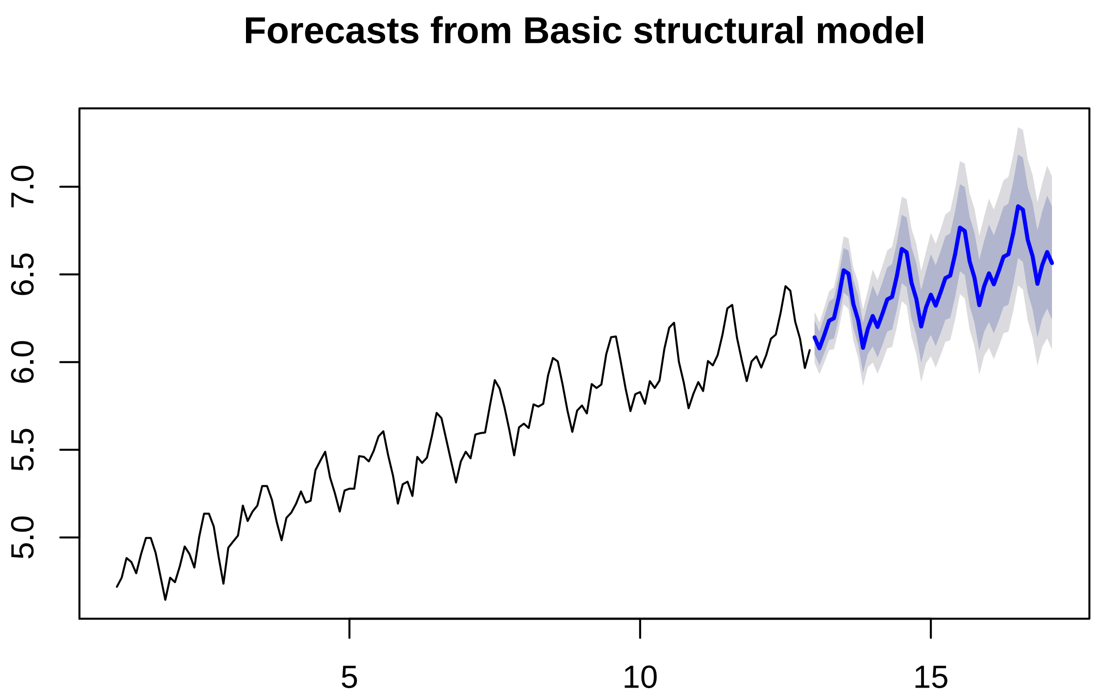
When we have found our model, we can do forecasting for the next twelve months.

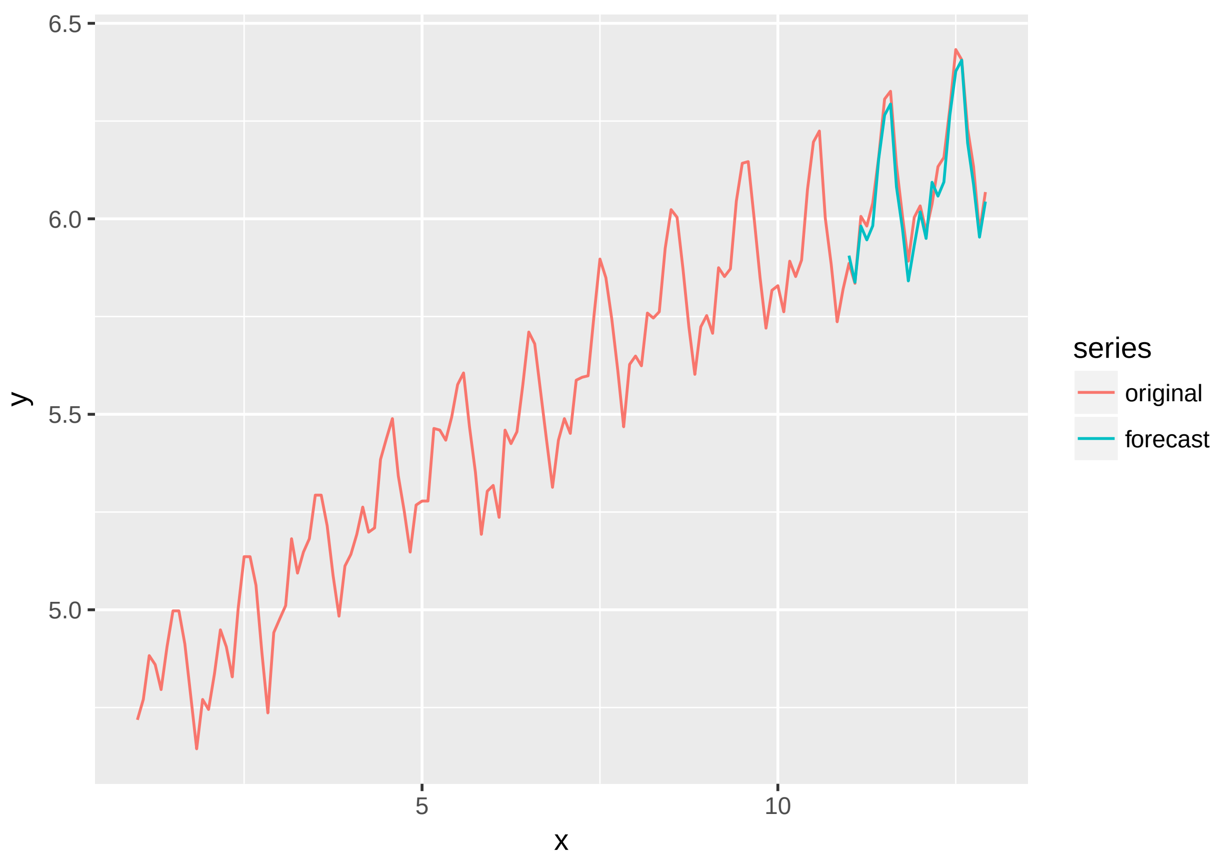
?? Need to forecast for given sample points. Only forecasting for the next year is done now ??

The forecasting seems to continue the season and trend of the data well, and the standard deviation of each predicted value is lower 0.09 for all months, which is low considered the relatively large y-values in the dataseries (monthtly passengers in thousands).

State-space model

We build a state-space model by . . .





?? Also need to forecast for given sample points here and find uncertainty. ??

Bootstrap

?? Missing: * Uncertainty in parameter estimates, at least for state-space. They are given from $ttable for sarima. But still if we manage, bootstrap / simulation is preferred, I think. \ * Model diagnostics for state-space? I don't know. \ * Model comparison + discussion. \ * Theory \ ??

Discussion

Conclusion

Appendix

References