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)

Abstract

In this exercise, we analyse a time series which contain the number of international airline passengers every month for 12 years. The series is first explored with SARIMA, and then explored using state-space models. Finally, the two models are compared.

Introduction

The analysis on time series can be used for multiple purposes. In this exercise it will be used to present the data in a way that is easy to interpret. In addition, we use it to do forecasting. The modeling includes a seasonal component that repeats every s observation (every 12th in our data set). This makes the model more precise for time series which has seasonal trends and periodic variations.

Theory

* State-space models

A state space model consists of two equations. The first equation is the observation equation, which expresses Y_t as a linear function of a v-dimensional state variable X_t and a noise term. The second one is the state equation, which shows X_{t+1} at time $t+1$ as a function of the previous state X_t plus noise.

These two equations can be written as:

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

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

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

* SARIMA

A SARIMA model is an extended version of the ARIMA model which can also consider a seasonal component in a time series. Its general form

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

is a seasonal ARIMA process with period s and is written as

$$\phi(B)\Phi\left(B^{S}\right)Y_t=\theta(B)\Theta\left(B^{S}\right)Z_t,\qquad \{Z_t\}\sim \mathrm{WN}(0,\sigma^2)$$

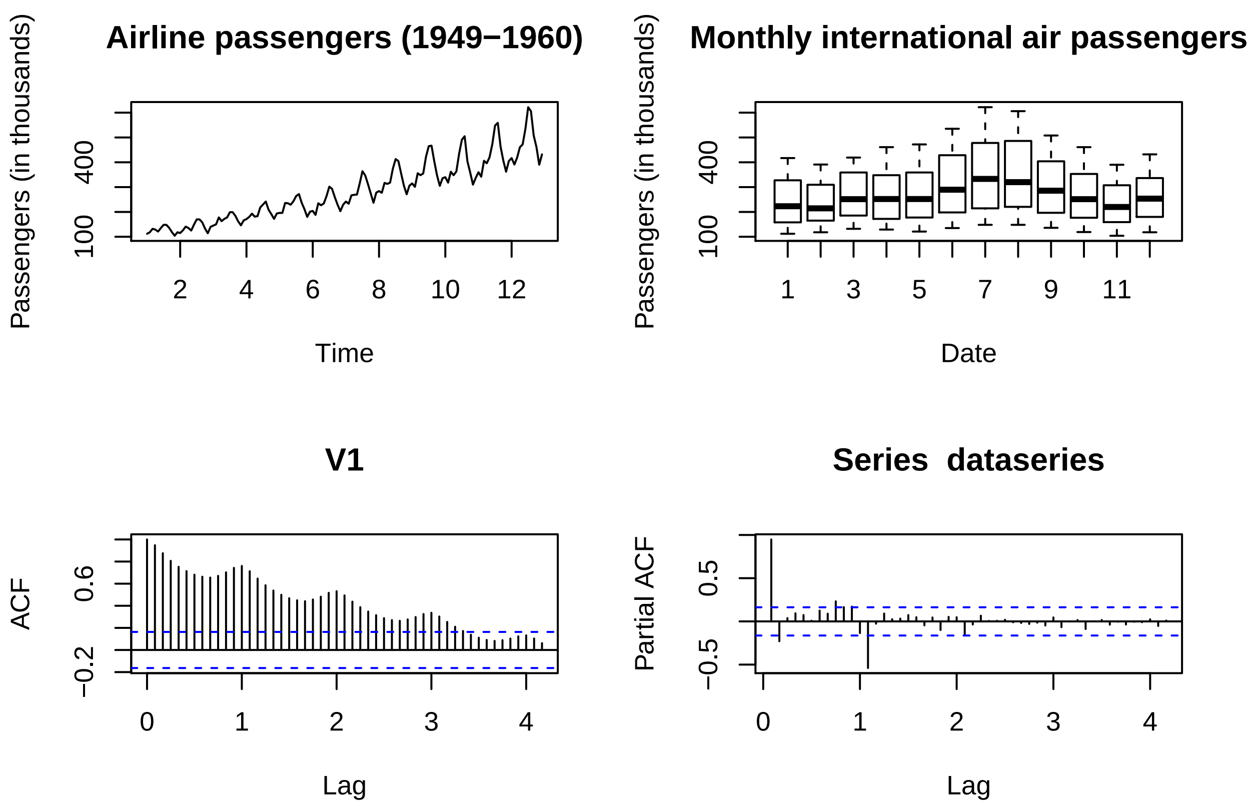
if the differenced series $Y_t=(1-B)^d(1-B^s)^D X_t$ is a causal ARMA process defined as above. Here, $\phi(z)=1-\phi_1 z-\ldots-\phi_p z^p$, $\Phi(z)=1-\Phi_1 z-\ldots-\Phi_P z^P$, $\theta(z)=1+\theta_1 z+\ldots+\theta_q z^q$ and $\Theta(z)=1+\Theta_1 z+\ldots+\Theta_Q z^Q$.

The SARIMA model has two parts: A seasonal component and a non-seasonal component. (P,D,Q) is the ARIMA model between each season and is the seasonal component of the SARIMA model. This part models behaviour for for instance the January months in a time series with monthly observations and where the season is a year long. (p,d,q) is the non-seasonal part of the model and is an ARIMA model *within* each season. The interpretation of p,d,q,P,d and Q are known from previous knowledge on the ARIMA models, and B is the backshift operator as before.

Data analysis

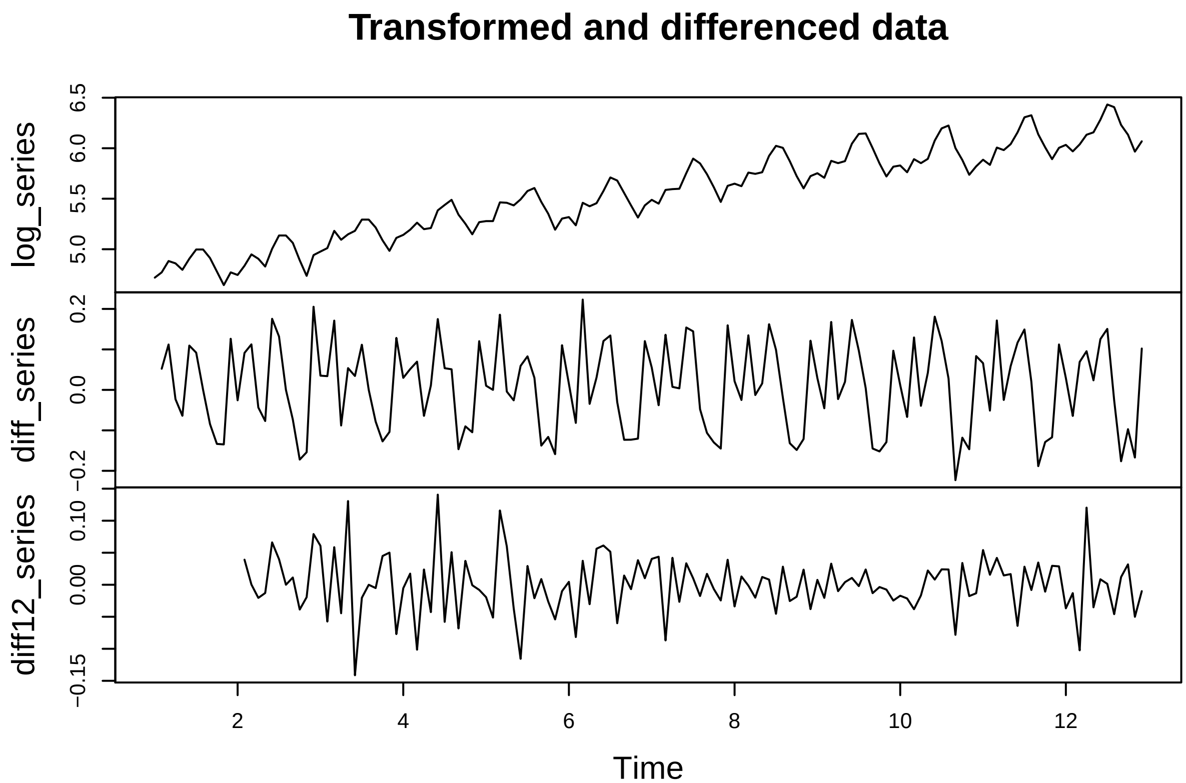
SARIMA$(p,d,q)\times(P,D,Q)_s$ model

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.

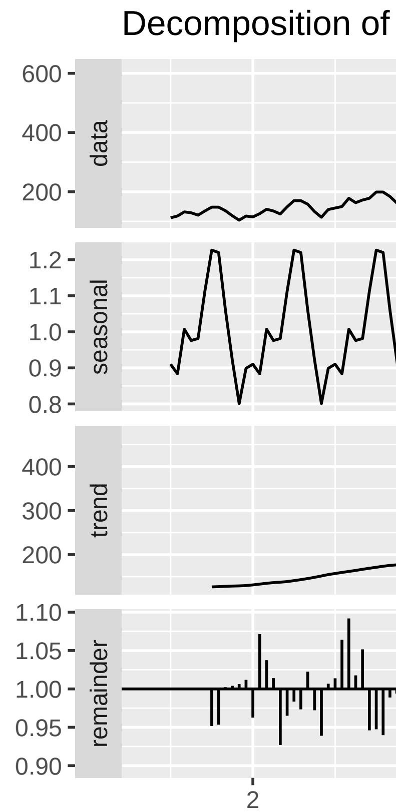


From the time series plot, this is obviously a non-stationary series as both trend and seasonality is visible in the visualization. Trend is there because the total number of airline passengers on average increases for each month, and seasonality because of the wave-like behaviour of the series. Non-stationarity 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.

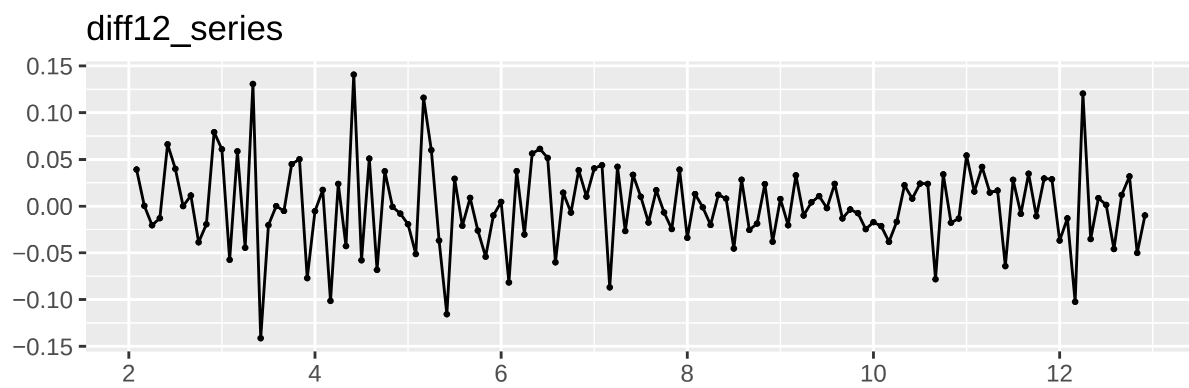
To investigate the data, we first find a transformation to the time series. A typical transformation for making the series more stationary is the log()-transformation. This seems to stabilize the multiplicative behaviour of the variance so that the variance is not increasing with time.

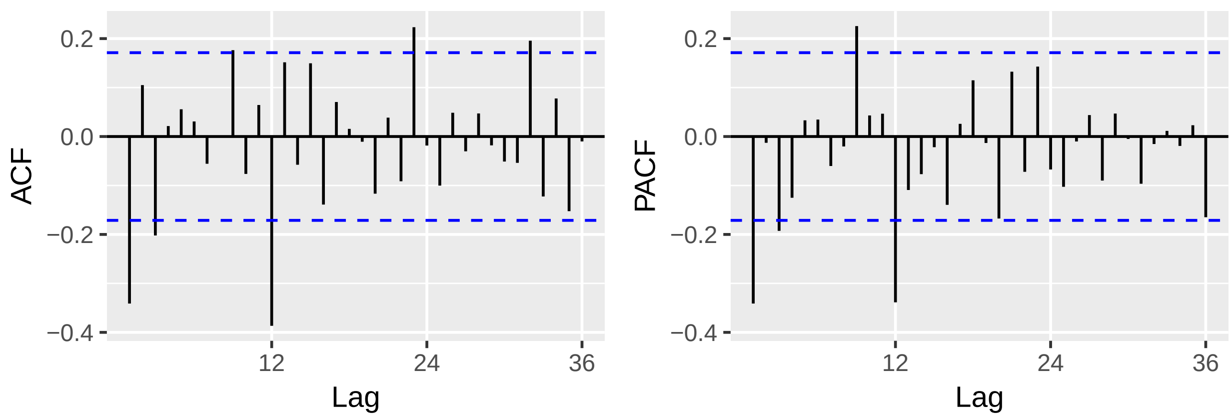


Secondly, the data is differenced to remove trend, i.e. stabilizing the mean. The parameter d is related to trend within a season, and since this trend seems to be removed by differencing once, we try $d = 1$. See the plot in the middle of the figure for transformed and differenced data. After differencing there still is a wave pattern present, one for each year. This indicates a seasonal trend equal to the length of a year, i.e. we suspect a $s = 12$ in a seasonal ARIMA model. Then $(1 - B^{12})$ is applied to the series. The parameter D is related to the trend between seasons, i.e. the trend one can se from one time of the year to the next, and the next, 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.





The obtained series after differencing the second time seems stationary without any trend or seasonality. We can then estimate the remaining parameters by considering the ACF and the PACF. This could be done by looking at lags equal to $1s, 2s, ...,$ $s = 12$ to determine P and Q in the seasonal component and by looking at smaller lags in each season to determine p and q in the non-seasonal component. In this case, however, we have used the AICC criterion and tested models with $d = 1$, $D = 1$, $s = 12$, which is determined before, and $p, q, P, Q < 5$ which is based by looking at the ACF and PACF. In addition we prefer models with smaller parameters for simplicity if they are a good fit.

The `auto.arima()`-function tests a variety of models and chooses the best one given our chosen criteria:

```
##
##  ARIMA(2,1,2)(1,1,1)[12]                    : Inf
##  ARIMA(0,1,0)(0,1,0)[12]                    : -434.799
##  ARIMA(1,1,0)(1,1,0)[12]                    : -474.6299
##  ARIMA(0,1,1)(0,1,1)[12]                    : -483.2101
##  ARIMA(0,1,1)(1,1,1)[12]                    : -481.5957
##  ARIMA(0,1,1)(0,1,0)[12]                    : -449.8857
##  ARIMA(0,1,1)(0,1,2)[12]                    : -481.6451
##  ARIMA(0,1,1)(1,1,2)[12]                    : Inf
##  ARIMA(1,1,1)(0,1,1)[12]                    : -481.582
##  ARIMA(0,1,0)(0,1,1)[12]                    : -467.4644
##  ARIMA(0,1,2)(0,1,1)[12]                    : -481.2991
##  ARIMA(1,1,2)(0,1,1)[12]                    : -481.5633
##
##  Best model: ARIMA(0,1,1)(0,1,1)[12]
```

The chosen model with the lowest AICC sets $p = 0, q = 1, P = 0$ and $Q = 1$. Fitting this model gives the following result:

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 from `fit$ttable`. However, when trying to bootstrap the to estimate the uncertainty of the parameters, we don't get the same results. The first line are supposed to be the parameter estimates, and the second line the standard errors.

```
## [1] -0.99155 -0.92885
```

```
## [1] 0.02668613 0.10747544
```

We nevertheless choose our model to be

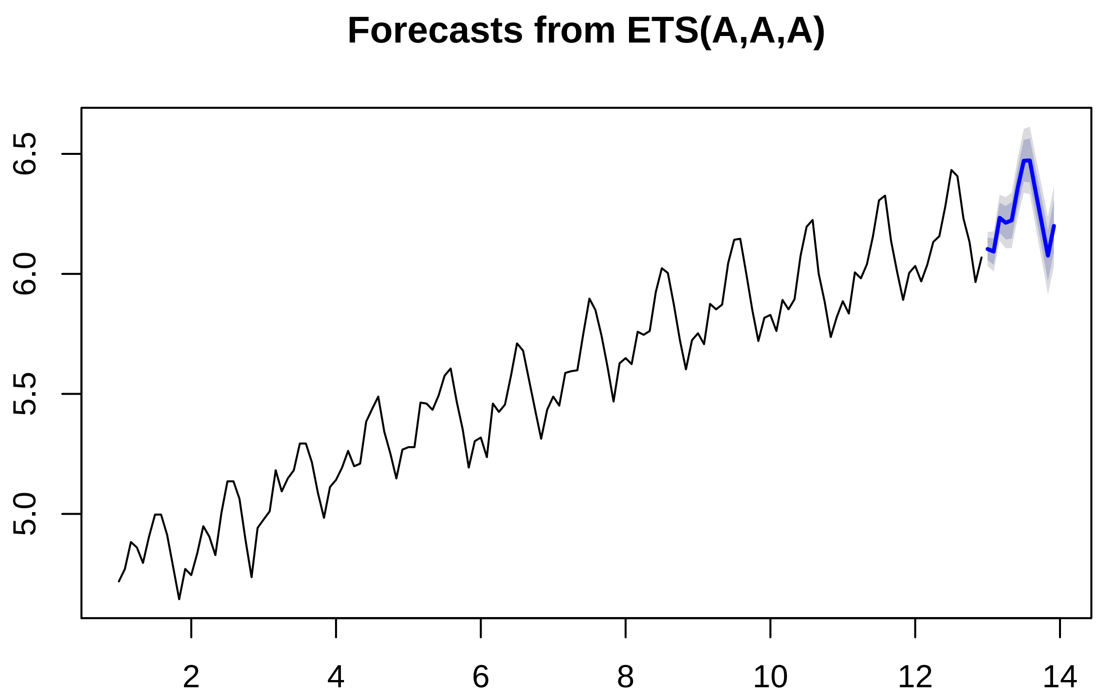
$$\mathrm{SARIMA}(0,1,1) \times (0,1,1)_{12}$$

which is equal to

$$\phi(B)\Phi(B^{12})(1-B)(1-B^{12})X_t = \theta(B)\Theta(B^{12})Z_t$$

When we have found our model, we can do forecasting with the SARIMA model for the next twelve months. From the fitting functions in R, we can also extract credible intervals of the forecasts and the fitted values at our given sample points.

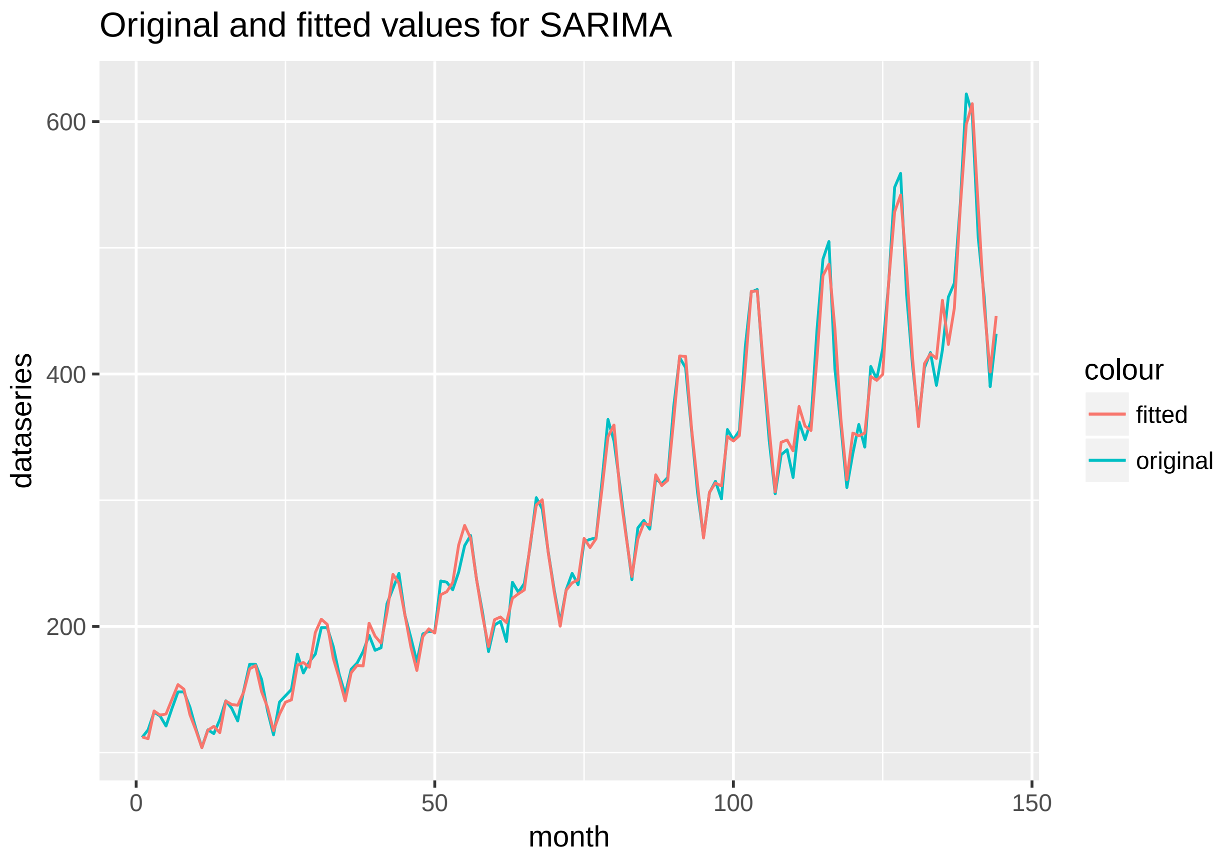
The following plot shows the time series and the forecased values for the next twelve months:



The credible intervals for the forecasts are

```
##    X95percent_lower X95percent_upper
## 1          417.0741         480.1365
## 2          407.0756         481.6434
```

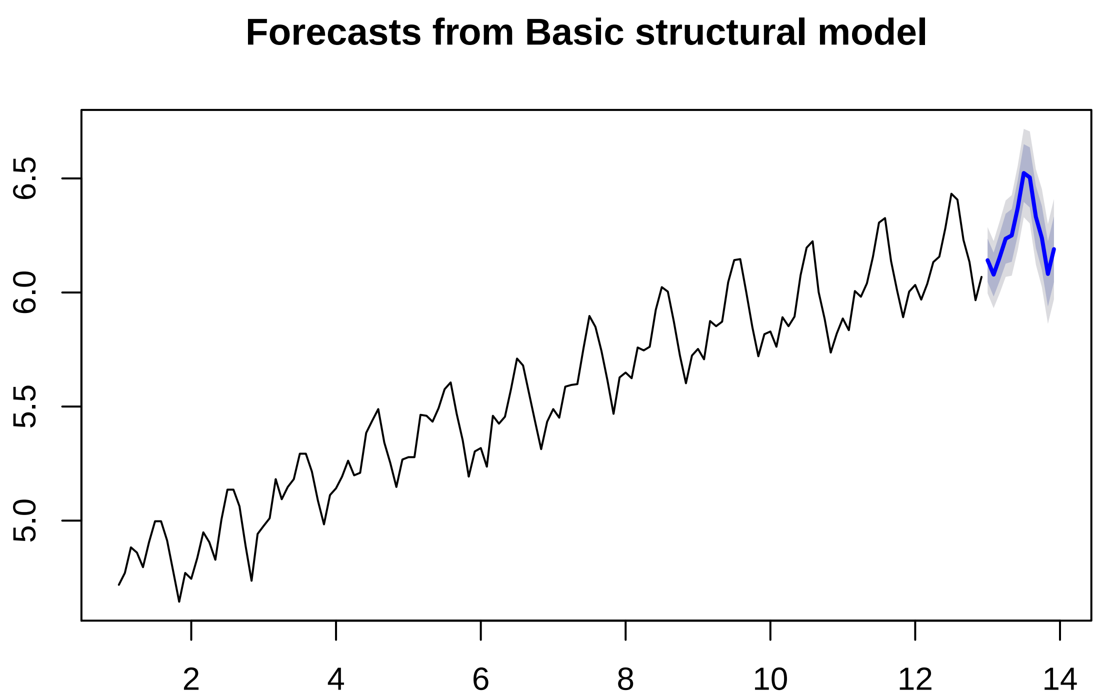
```
## 3          463.1019          560.9776
## 4          448.8949          555.2788
## 5          449.1688          566.3135
## 6          509.8210          654.2095
## 7          565.8695          738.1816
## 8          562.2268          744.8878
## 9          487.8114          655.8636
## 10         425.5134          580.1720
## 11         370.2184          511.5899
## 12         416.3545          582.7973
```



The fitted data looks good compared to the original data and the forecasting seems to continue the season and trend of the data well.

State-space model

We build a state-space model, which gives the following matrices



$F =$

```
##        [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
##  [1,]     1    1    0    0    0    0    0    0    0     0     0     0     0
##  [2,]     0    1    0    0    0    0    0    0    0     0     0     0     0
##  [3,]     0    0   -1   -1   -1   -1   -1   -1   -1    -1    -1    -1    -1
##  [4,]     0    0    1    0    0    0    0    0    0     0     0     0     0
##  [5,]     0    0    0    1    0    0    0    0    0     0     0     0     0
##  [6,]     0    0    0    0    1    0    0    0    0     0     0     0     0
##  [7,]     0    0    0    0    0    1    0    0    0     0     0     0     0
##  [8,]     0    0    0    0    0    0    1    0    0     0     0     0     0
##  [9,]     0    0    0    0    0    0    0    1    0     0     0     0     0
## [10,]     0    0    0    0    0    0    0    0    1     0     0     0     0
## [11,]     0    0    0    0    0    0    0    0    0     1     0     0     0
## [12,]     0    0    0    0    0    0    0    0    0     0     1     0     0
## [13,]     0    0    0    0    0    0    0    0    0     0     0     1     0
```

$G =$

```
##  [1] 1 0 1 0 0 0 0 0 0 0 0 0 0
```

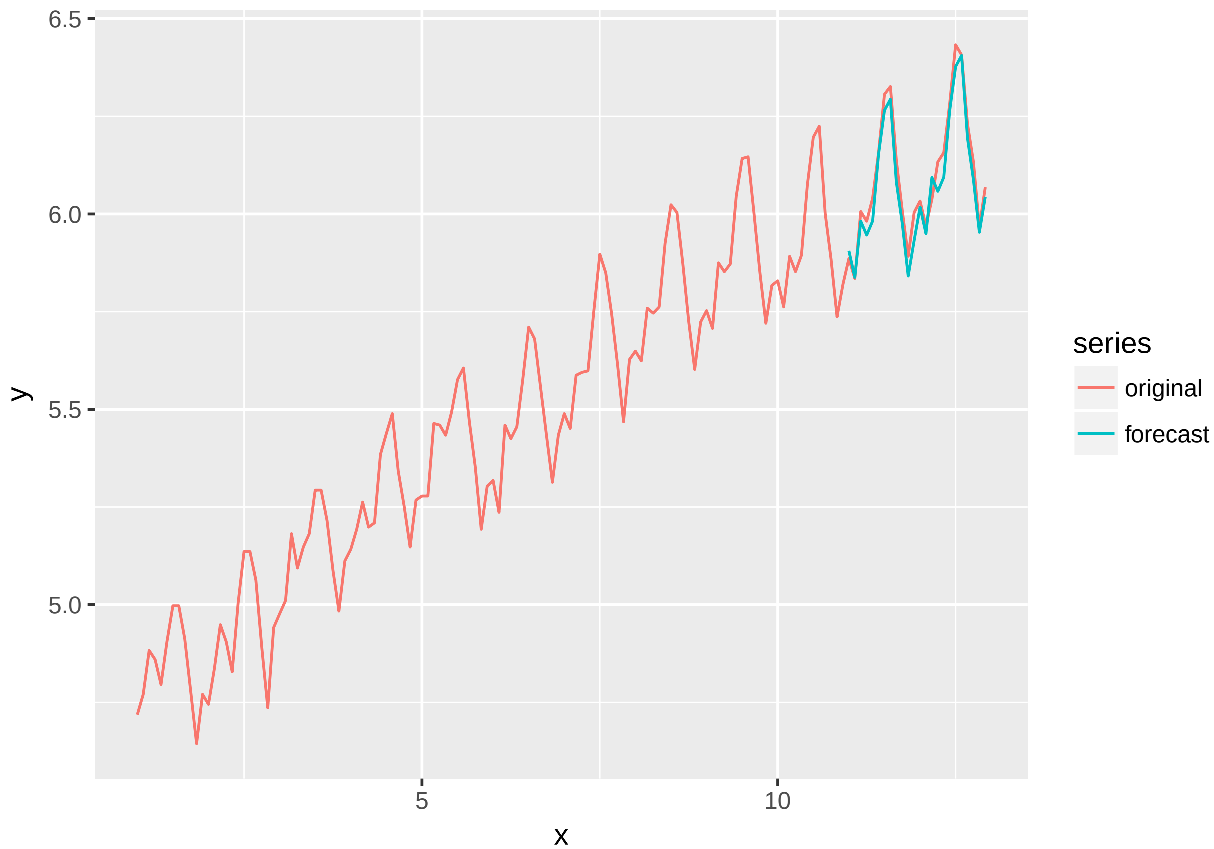
$Q =$

```
##        [,1]     [,2]     [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11]
##  [1,]     0   0.0000  0.00000    0    0    0    0    0    0     0     0
##  [2,]     0 160.9755  0.00000    0    0    0    0    0    0     0     0
##  [3,]     0   0.0000 29.84652    0    0    0    0    0    0     0     0
##  [4,]     0   0.0000  0.00000    0    0    0    0    0    0     0     0
##  [5,]     0   0.0000  0.00000    0    0    0    0    0    0     0     0
```

```
##  [6,]    0   0.0000  0.00000    0    0    0    0    0    0     0     0
##  [7,]    0   0.0000  0.00000    0    0    0    0    0    0     0     0
##  [8,]    0   0.0000  0.00000    0    0    0    0    0    0     0     0
##  [9,]    0   0.0000  0.00000    0    0    0    0    0    0     0     0
## [10,]    0   0.0000  0.00000    0    0    0    0    0    0     0     0
## [11,]    0   0.0000  0.00000    0    0    0    0    0    0     0     0
## [12,]    0   0.0000  0.00000    0    0    0    0    0    0     0     0
## [13,]    0   0.0000  0.00000    0    0    0    0    0    0     0     0
##       [,12] [,13]
##  [1,]     0     0
##  [2,]     0     0
##  [3,]     0     0
##  [4,]     0     0
##  [5,]     0     0
##  [6,]     0     0
##  [7,]     0     0
##  [8,]     0     0
##  [9,]     0     0
## [10,]     0     0
## [11,]     0     0
## [12,]     0     0
## [13,]     0     0
```

This gives $(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_2^3) = (0.0000, 160.9755, 29.84652)$.

```
##  [1] 417.8507614 -29.6143515  14.1492386 -57.4651129 -23.6949346
##  [6] -12.2523959  56.6479592  60.1013022  -9.8206840 -37.0447315
## [11]  -4.5661018  -0.8086603  -7.7654562
```



Bootstrap

We bootstrap the estimated parameters for the state space model.

The estimated parameters and their uncertainties are respectively

```
## [1] -2.213260e-02  7.242622e-02 -2.267926e-05 -2.213260e-02
```

```
## [1] 2.813720e-03 1.852655e-02 2.021231e-09 2.813720e-03
```

- Model diagnostics for state-space \
- Model comparison + discussion. \

Discussion

-It is hard to esimate the encertainty. -We have tried to do bootstrapping for SARIMA, but we did not get the results we wanted. - Does bootstrapping for -We prefer SARIMA? Yes

Conclusion

Appendix

Reference

Brockwell, Peter J., Davis, Richard A. 2002. *Introduction to time series and forecasting.* 2nd ed. New York: Springer Science