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

Contents

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:

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

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

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

*** 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, $\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$, $\Phi(z) = 1 - \Phi_1 z - ... - \Phi_P z^P$, $\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q$ and $\Theta(z) = 1 + \Theta_1 z + ... + \Theta_Q z^Q$.

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

1. Exploring the data with plotting of relevant statistics

2. Justification of the choice of model

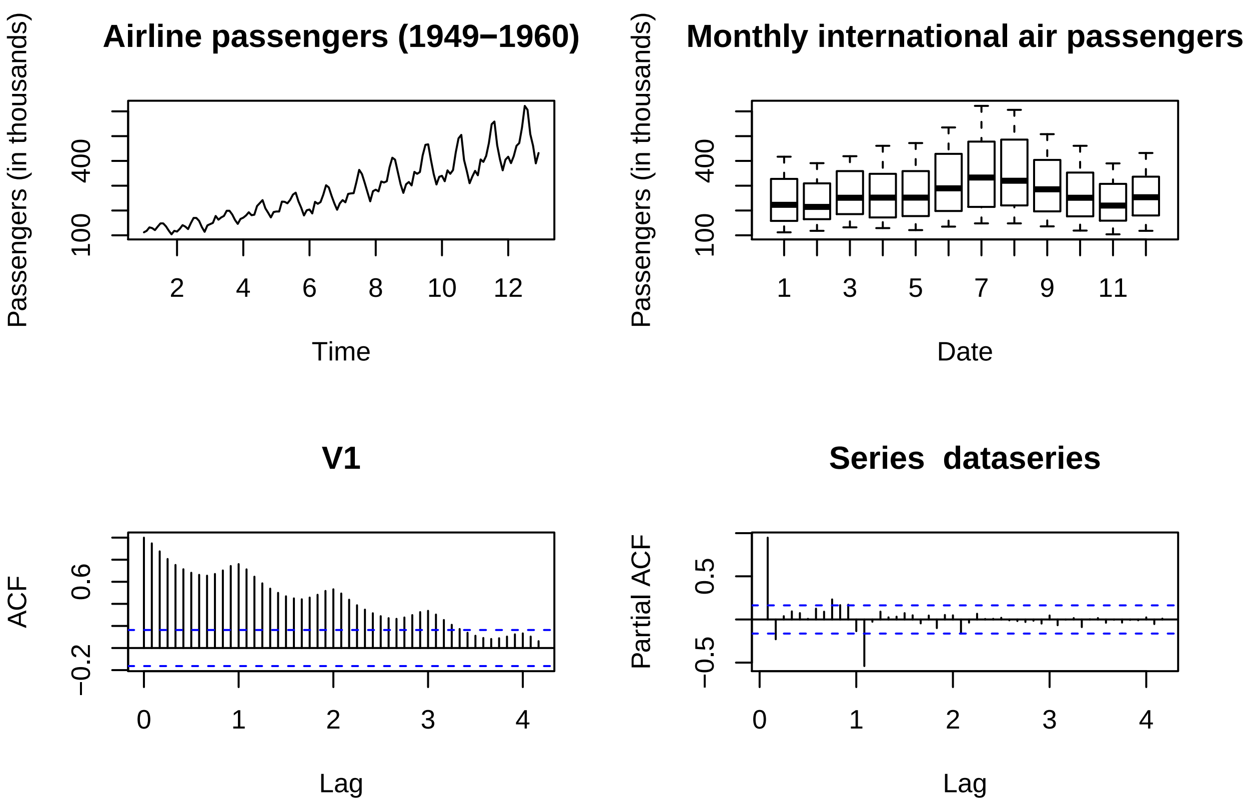
3. Model parameter estimation including uncertainty

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

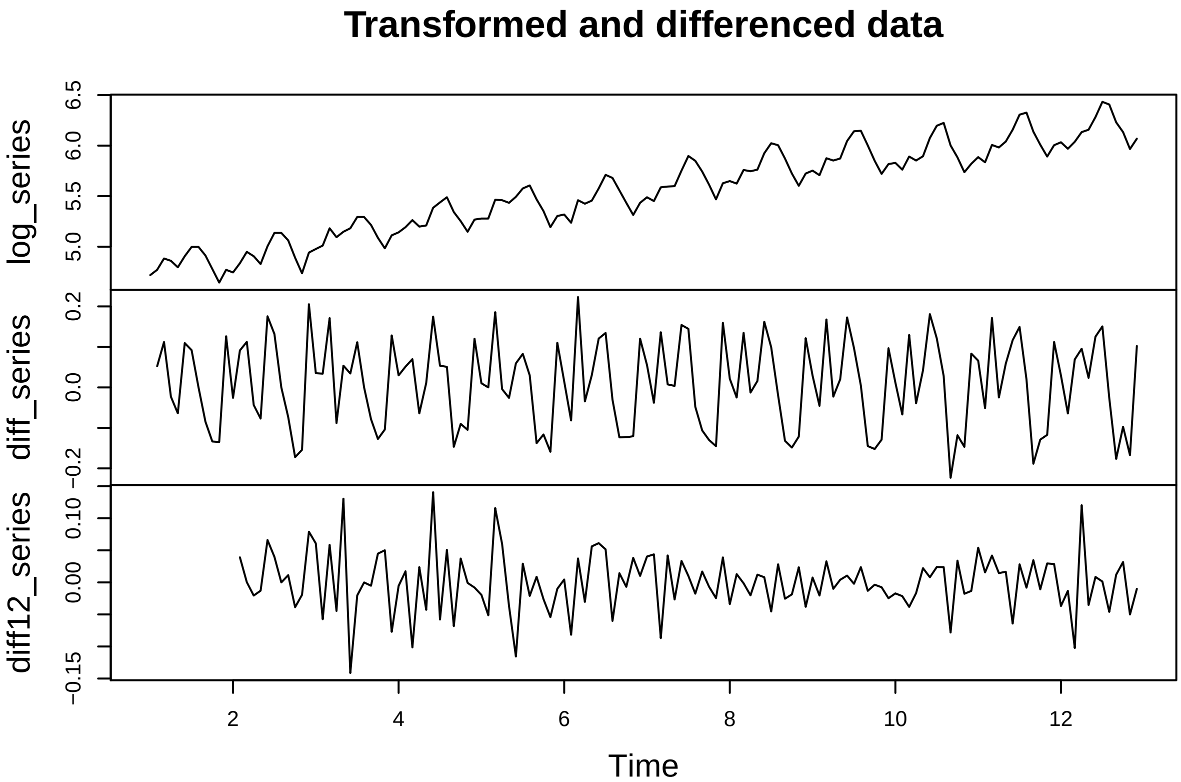
$SARIMA(p,d,q) \times (P,D,Q)_s$ model

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.

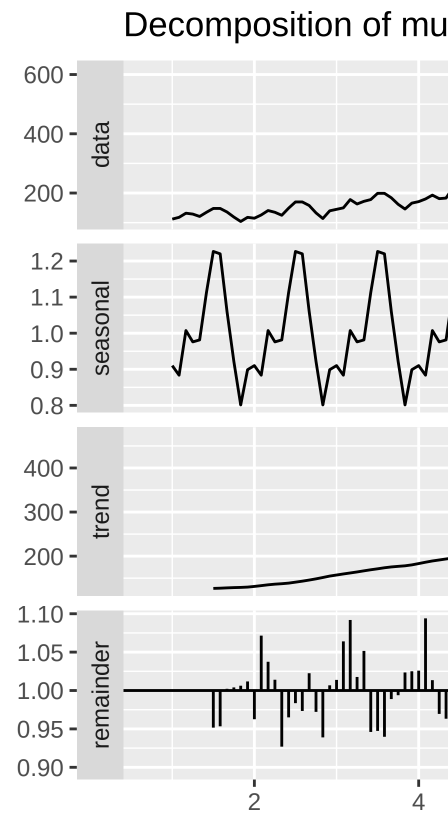


From the time series plot, this is obviously a non-stationary series as both trend and seasonality is visible in the visualization. Trend is there because the total number of airline passengers on average increases for each month, and seasonality because of the wave-like behaviour of the series. Non-stationarity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

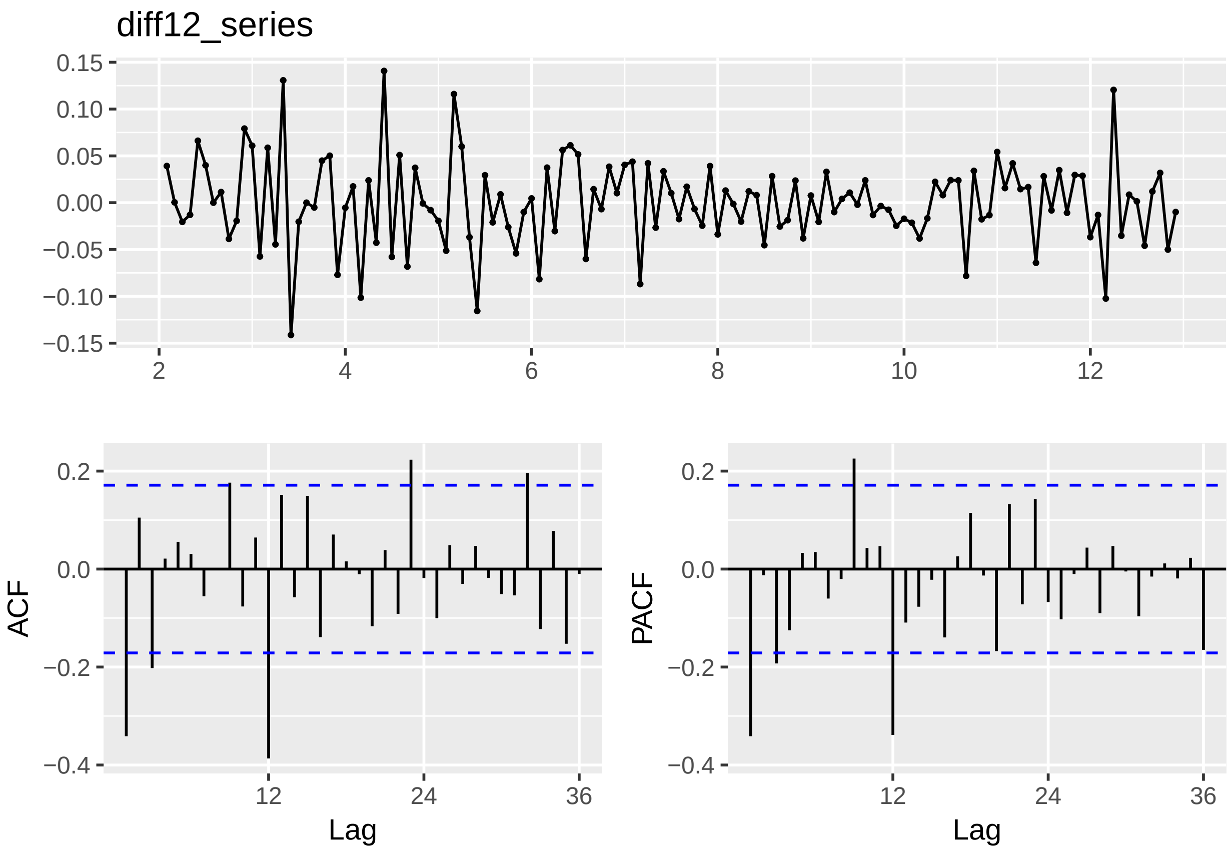
To investigate the data, we first find a transformation to the time series. A normal transformation is the log()-transformation, which seems to stabilize the multiplicative behaviour of the variance so that the variance is not increasing with time.



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 second and the third 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 applying the twelfth-order difference seems to be stationary without any trend or seasonality. We can then try to estimate the remaining parameters by considering the ACF and the PACF. This would 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. However, we chose to look 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 on looking at the ACF and PACF. In addition we prefer models with smaller parameters for simplicity if they are a good fit. The following models were tested:

```
##
##  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]
```

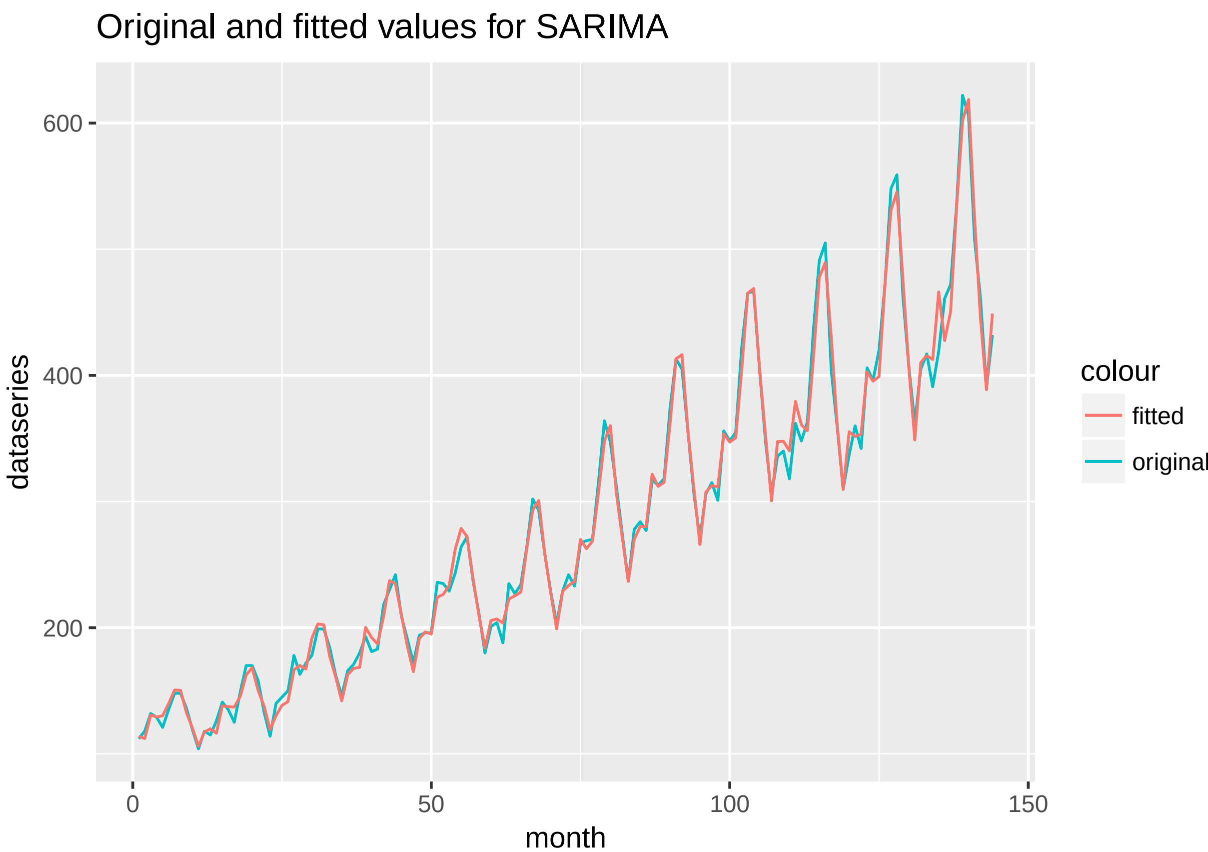
Both parameters are significant in the model, and the residual analysis plot shows that our model is a good fit. In addition, the uncertainty of the parameters are given from `fit$ttable`. Our chosen model is thus

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

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 for the next twelve months.

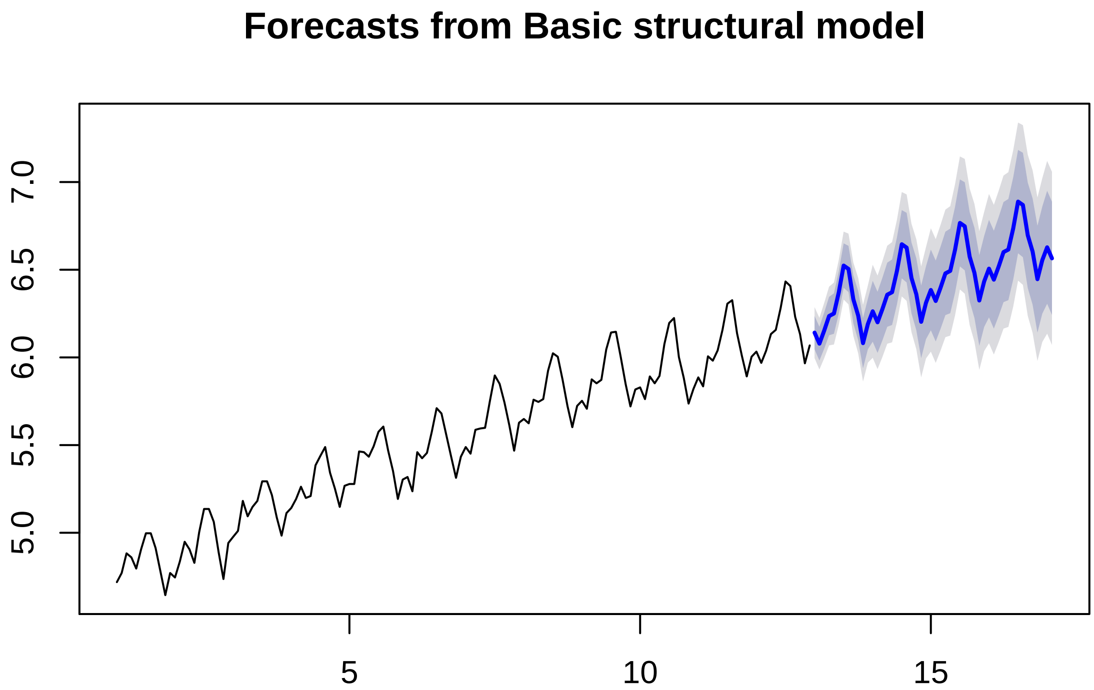


The fitted data looks good compared to the original data and the forecasting seems to continue the season and trend of the data well. Upper and lower bounds for both 80% and 95% confidence intervals are given from

```
##        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 13       6.105184 6.051211 6.159157 6.022639 6.187729
## Feb 13       6.095690 6.033218 6.158161 6.000148 6.191232
## Mar 13       6.251904 6.180118 6.323690 6.142116 6.361691
## Apr 13       6.227339 6.148917 6.305762 6.107403 6.347276
## May 13       6.235570 6.150695 6.320446 6.105764 6.365376
## Jun 13       6.379206 6.286346 6.472065 6.237189 6.521222
## Jul 13       6.500228 6.399853 6.600604 6.346717 6.653740
## Aug 13       6.508293 6.402357 6.614229 6.346277 6.670309
## Sep 13       6.354237 6.245771 6.462702 6.188353 6.520120
## Oct 13       6.205944 6.095320 6.316568 6.036760 6.375129
## Nov 13       6.047271 5.935102 6.159441 5.875723 6.218820
## Dec 13       6.191406 6.072263 6.310549 6.009192 6.373619
```

State-space model

We build a state-space model by ??eqref to theory or just use Rfunctions??, 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
```

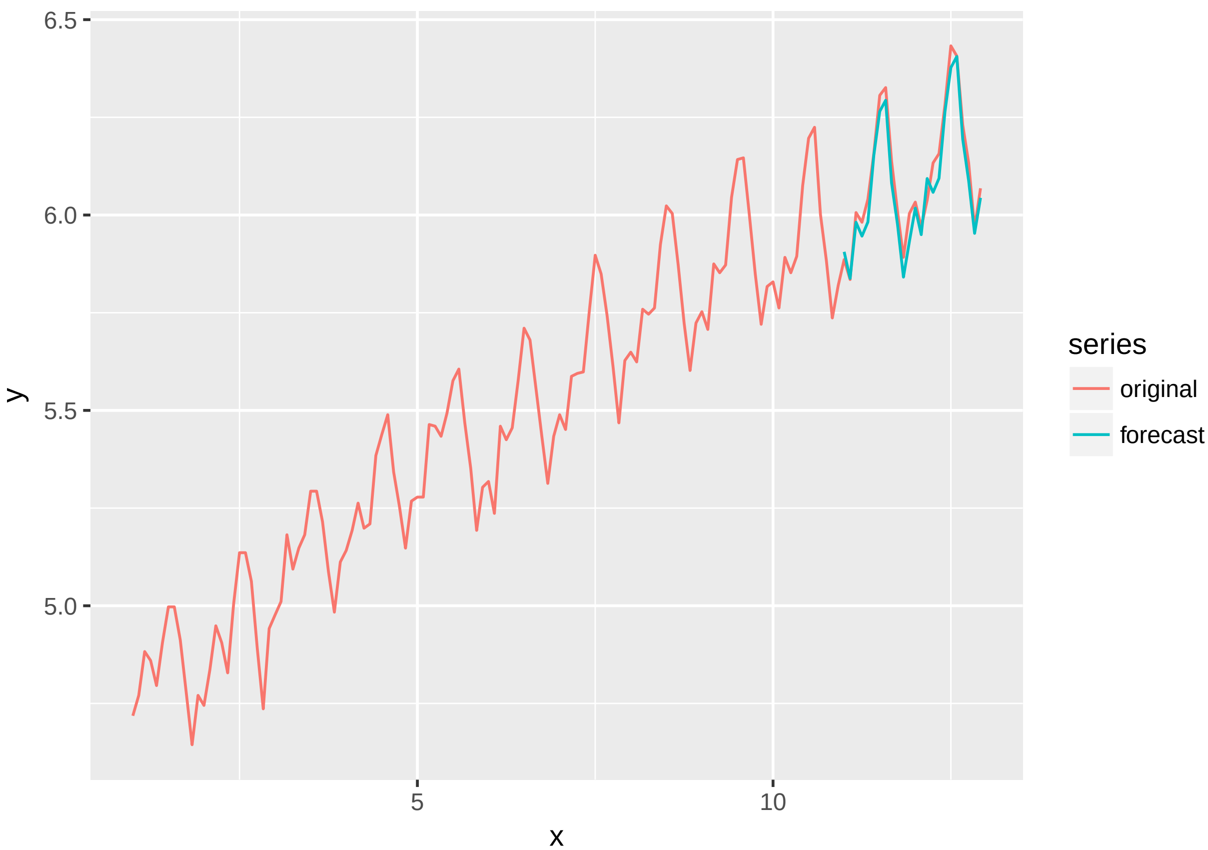
$Q =$

```
##       [,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
```



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

Bootstrap

We bootstrap the estimated parameters for the state space model.

- Model diagnostics for state-space? I don't know. \
- Model comparison + discussion. \
- Theory \ ??

Discussion

Conclusion

Appendix

References