Computer Exercise 2

Christian Glissov, s146996

November 6, 2018

DTU Compute - Institute for Mathematics and Computer science
Advanced Time Series Analysis 02427 November 6, 2018
Henrik Madsen & Erik Lindstrøm

Technical University of Denmark

Made in collaboration with Frederik Boe Hüttel, s151689, and Jonas Gottlieb Svendsen, s152992.

Contents

Question 2.1

In part 1 of exercise 1 the following SETAR(2,1,1) model was given:

$$X_t = \begin{cases} 4 + 0.5 \cdot X_{t-1} + \epsilon_{t,1} & X_{t-1} < 0 \\ -4 - 0.5 \cdot X_{t-1} + \epsilon_{t,2} & X_{t-1} \geq 0 \end{cases} \tag{1}$$

Now assume the parameters are unknown, this means

$$X_t = \begin{cases} a_0^{(1)} + a_1^{(1)} \cdot X_{t-1} + \epsilon_{t,1} & X_{t-1} < r \\ a_0^{(2)} + a_1^{(2)} \cdot X_{t-1} + \epsilon_{t,2} & X_{t-1} \geq r \end{cases} \tag{2}$$

To estimate the parameters, a few demands must be made. The first demand is that the regime models are distinguishable and the second is that the upper limit for the delay d is known a priori [2]. Both of these are known in the simulated system. One could estimate the regime shift, using the prediction error method, given enough observations are present and a few good initial guesses. However the results seem too inconsistent and therefore for this case the regime shift is assumed known $r = 0$. Now the prediction error method, (chapter 5.5 [2]), can be used, which is given in eq. (5.37) in [2]. To do this the `Optim` function in `R` is used. It uses the Nelder-Mead algorithm by default. Setting the initial guess for all the parameters to 1, with 1000 sample points gives the following result:

Parameter	$a_1^{(1)}$	$a_1^{(2)}$	$a_0^{(1)}$	$a_0^{(2)}$
Estimated Value	0.5140395	-0.5177555	3.9604616	-3.8778064

All the values are fairly close to the true values. The more sample points, the better the convergence due to the increase of information, which can also be seen in Figure 1.

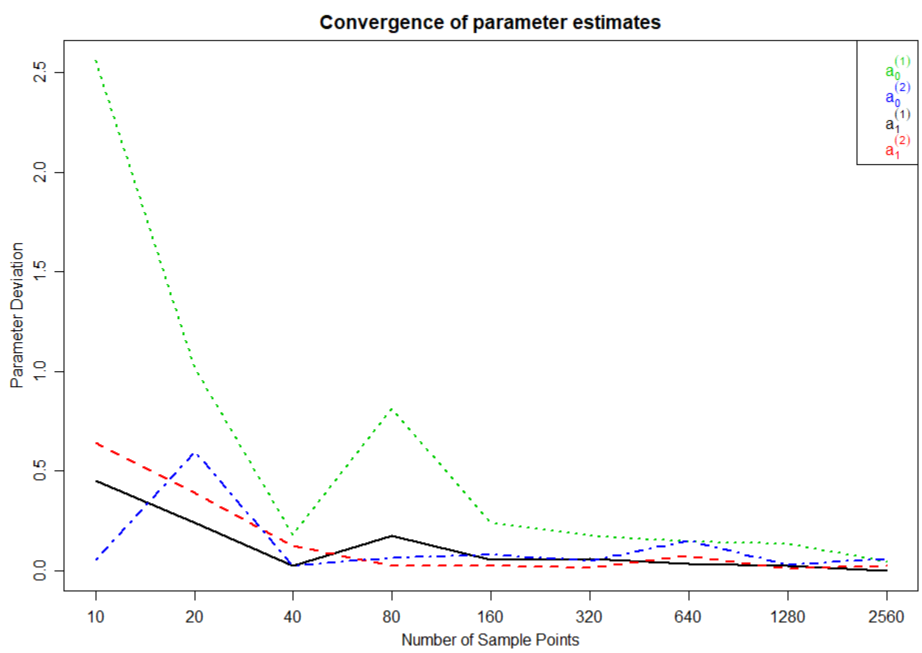


Figure 1: Looking at the deviation of the estimated parameter against the true parameter for different sample points (n).

Clearly the error converges to 0 as $n \to \infty$. When looking at the theoretical mean against the estimated mean, given by

$$E\{X_t|X_{t-1} = x\} = \begin{cases} 3.9604616 + 0.5140395 \cdot X_{t-1} & X_{t-1} < 0 \\ -3.8778064 - 0.5177555 \cdot X_{t-1} & X_{t-1} \geq 0 \end{cases} \tag{3}$$

It is observed in Figure 2 that there's not a large difference between the means as expected:

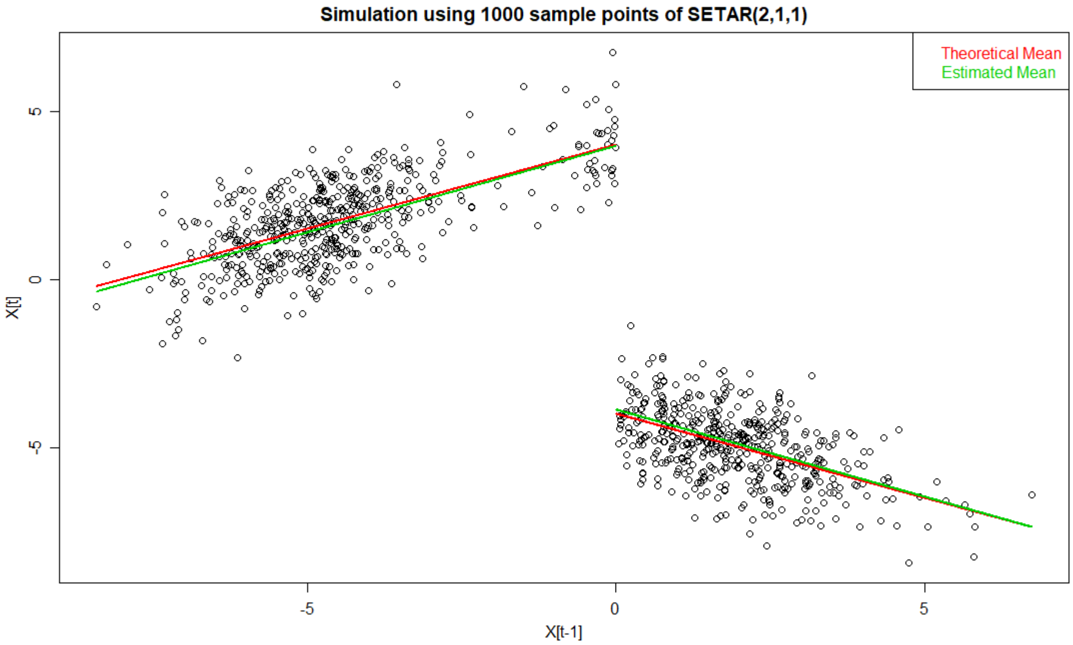


Figure 2: Theoretical against estimated mean of the SETAR(2,1,1) process.

Question 2.2

In this exercise the same model is used as given in Equation 1. However only two parameters will be estimated for illustration purposes. The chosen parameters to be estimated will be $(a_1^{(1)}, a_1^{(2)})$ from Equation 2, with $(a_0^{(1)} = 4, a_0^{(2)} = -4)$. A total of 3000 observations will be used. Looking at the contour plot in Figure 3 it can be seen that indeed the estimated parameters are very close to the true parameters. This is also expected, as was also seen in Figure 1

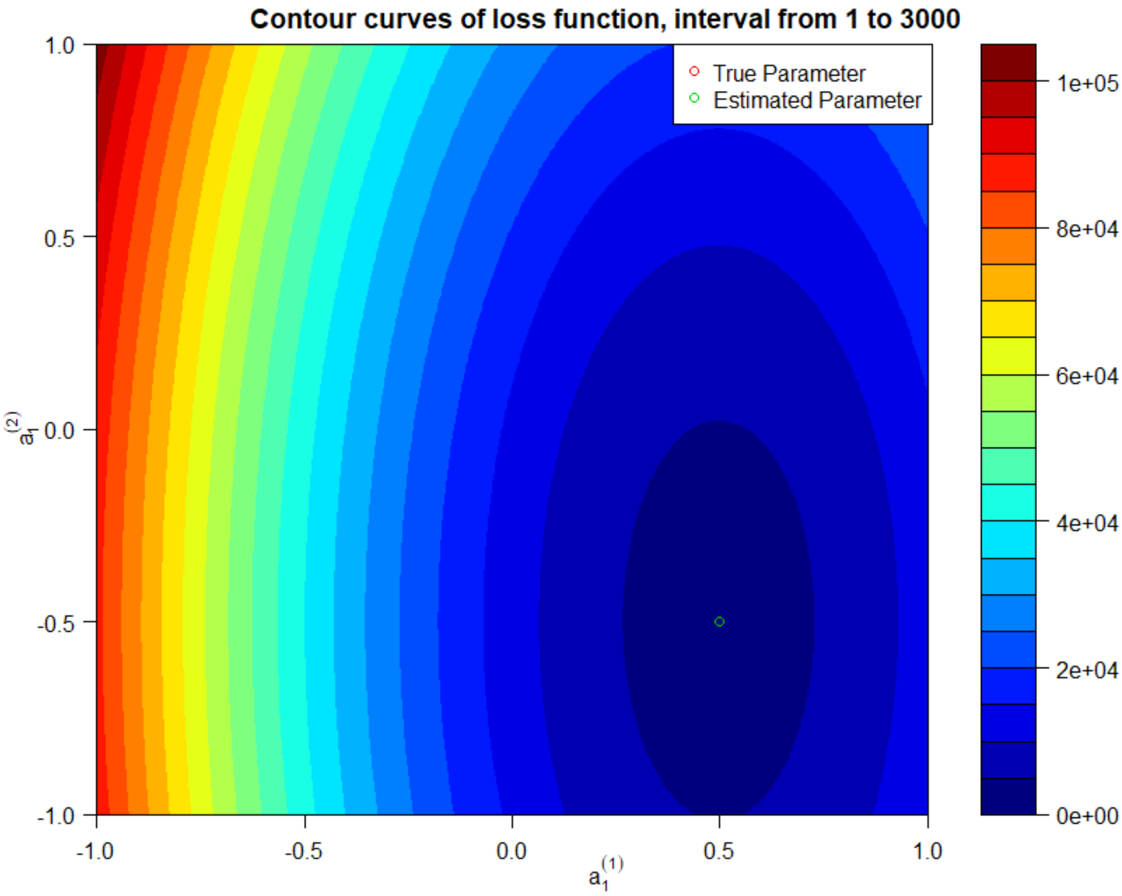


Figure 3: Contour curves of the loss, Q. The green circle is the estimated parameters and the red circle is the true parameters. In this case the red circle can't be seen, due to overlap of the green circle.

Trying with different subsets as given in the exercise the following contour curves for the loss function can be found:

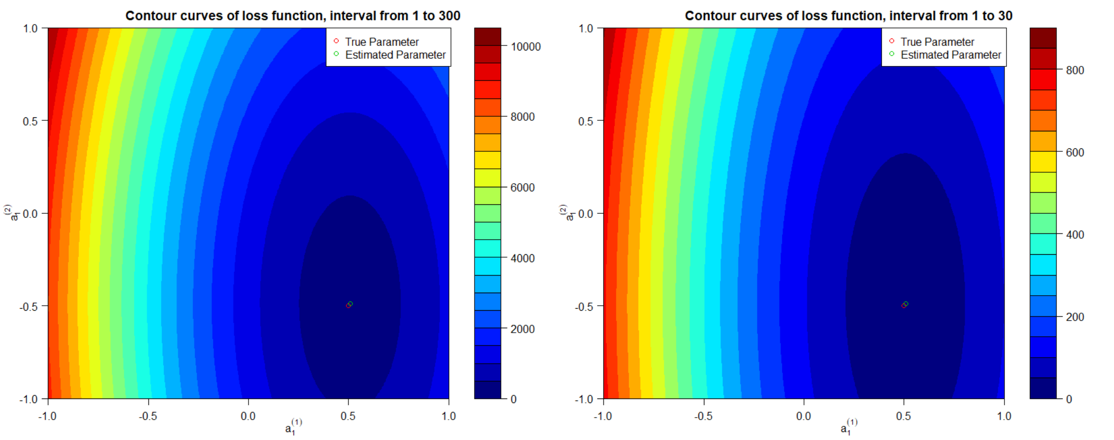


Figure 4: Contour curves of the loss, Q, for respectively 1:300 and 1:30 observations.

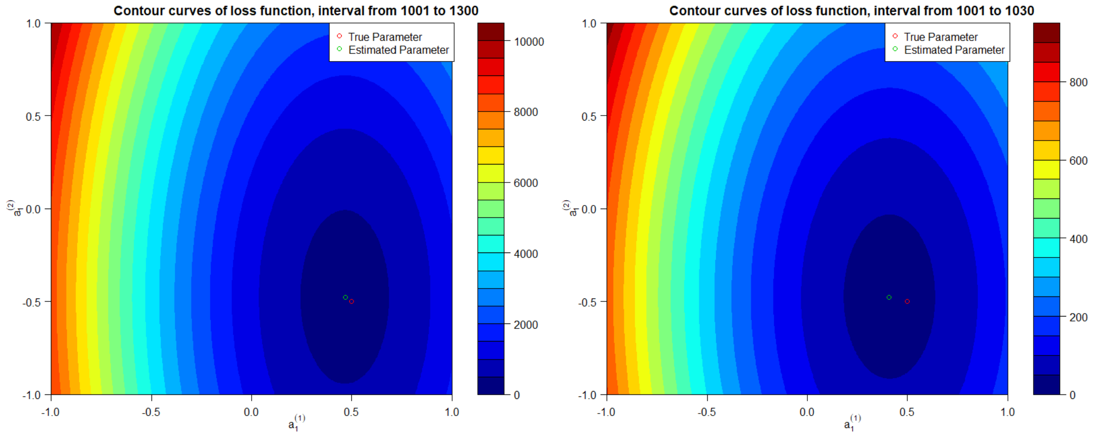


Figure 5: Contour curves of the loss, Q, for respectively 1001:1300 and 1001:1030 observations.

It is seen that for Figure 4 not a huge change can be seen, however a slightly worse estimate is given. The same goes for the contours in Figure 5, and especially for the plot with the subset (1001:1030), here the estimate is quite a bit off, this is due to the lower information and higher uncertainty when using a smaller data set.

Question 2.3

Let's consider the non-linear time series process:

$$\begin{aligned} \Phi_t &= \mu(1-\phi_1-\phi_2)+\phi_1\Phi_{t-1}+\phi_2\Phi_{t-2}+\theta_1\zeta_{t-1}+\zeta_t \\ Y_t &= \Phi_t Y_{t-1}+\Phi_{t-1}Y_{t-2}+\epsilon_t \end{aligned} \tag{4}$$

The process in Equation 4 is an ARMA(2,1)-AR(2) doubly stochastic process, defined respectively by Φ_t and Y_t. Here $\{\zeta\}$ and $\{\epsilon\}$ are both mutually uncorrelated Gaussian noise processes. To write this process in state space form δ is first defined as $\delta = \mu(1-\phi_1-\phi_2)$ as defined in [2]. This can then be written as

$$\begin{pmatrix} \Phi_t \\ \Phi_{t-1} \\ \delta_t \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Phi_{t-1} \\ \Phi_{t-2} \\ \delta_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ \theta_1 \\ 0 \end{pmatrix} \zeta_t$$

$$Y_t = (Y_{t-1}, Y_{t-2}, 0) \begin{pmatrix} \Phi_t \\ \Phi_{t-1} \\ \delta_t \end{pmatrix} + \varepsilon_t$$

Now defining all the necessary values:

μ	ϕ_1	ϕ_2	θ_1	σ_ζ	σ_ϵ	δ
0.5	0.85	-0.6	0.4	0.1	0.4	0.375

From the above, the following doubly stochastic process can be observed:

$$\begin{pmatrix} \Phi_t \\ \Phi_{t-1} \\ \delta_t \end{pmatrix} = \begin{pmatrix} 0.85 & -0.6 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Phi_{t-1} \\ \Phi_{t-2} \\ \delta_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0.4 \\ 0 \end{pmatrix} \zeta_t$$

$$Y_t = (Y_{t-1}, Y_{t-2}, 0) \begin{pmatrix} \Phi_t \\ \Phi_{t-1} \\ \delta_t \end{pmatrix} + \varepsilon_t$$

The above process is now simulated and can be seen in Figure 6 and Figure 7.

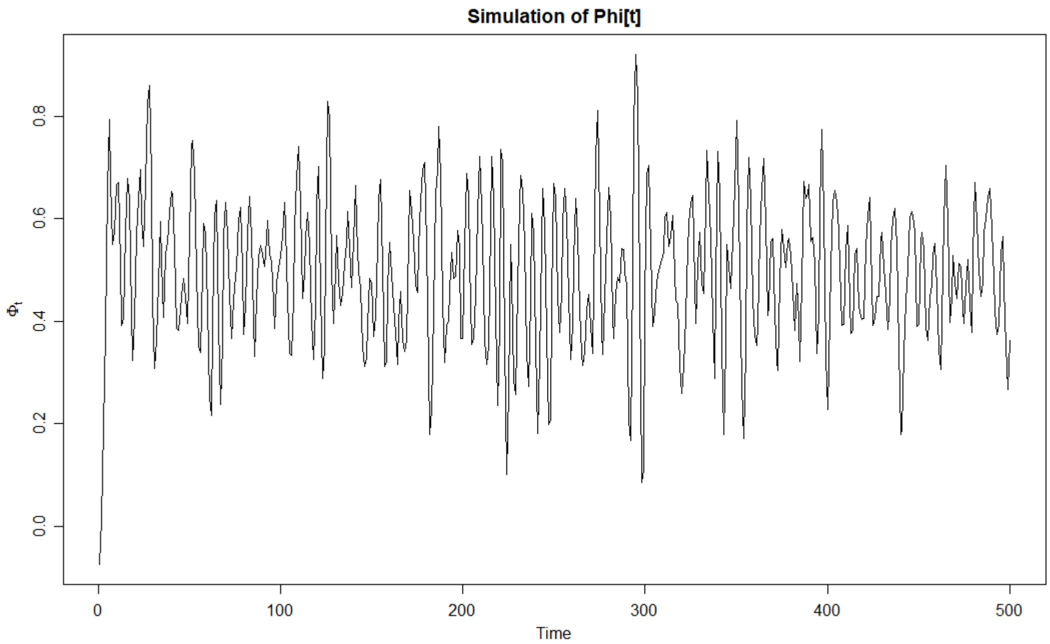


Figure 6: Simulation of Φ_t

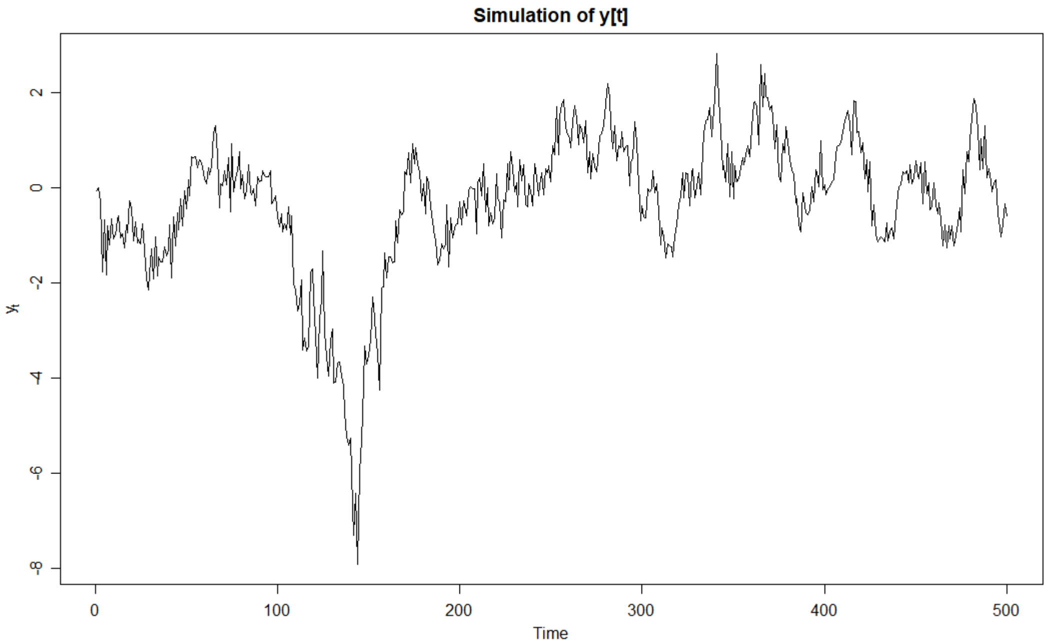


Figure 7: Simulation of Y_t

It can be seen that in Figure 7 that there are some large spikes, this might be due to complex poles in the transfer function from the noise to the value of the parameter [2]. To verify that the doubly stochastic process is indeed an ARMA(2,1)-AR(2) process the ACF and PACF for each process is shown in Figure 8 and Figure 9

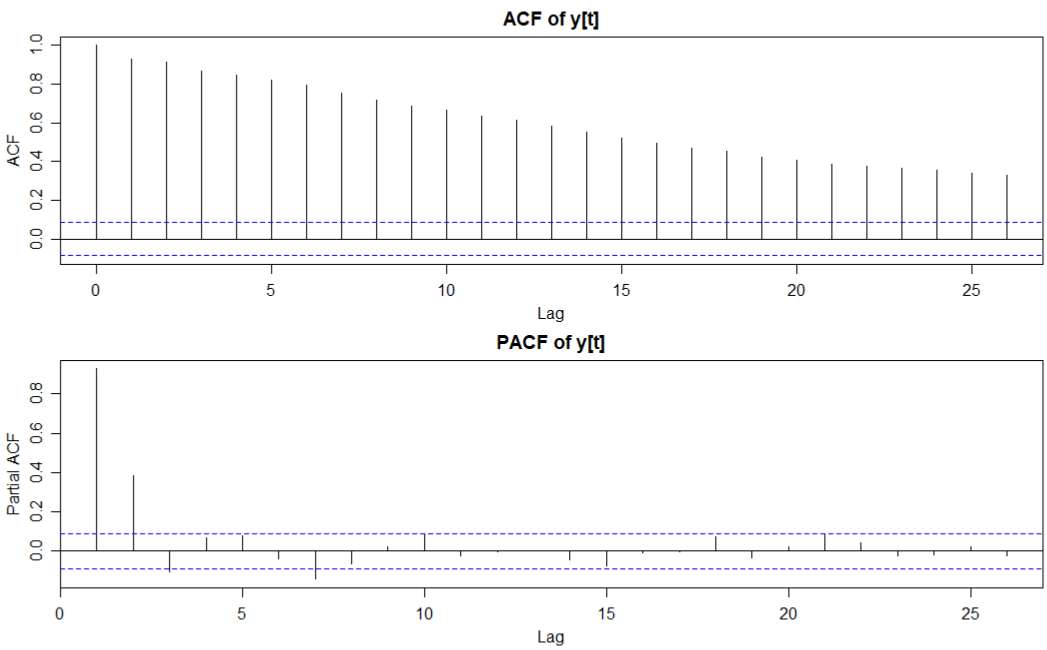


Figure 8: ACF and PACF of Y_t

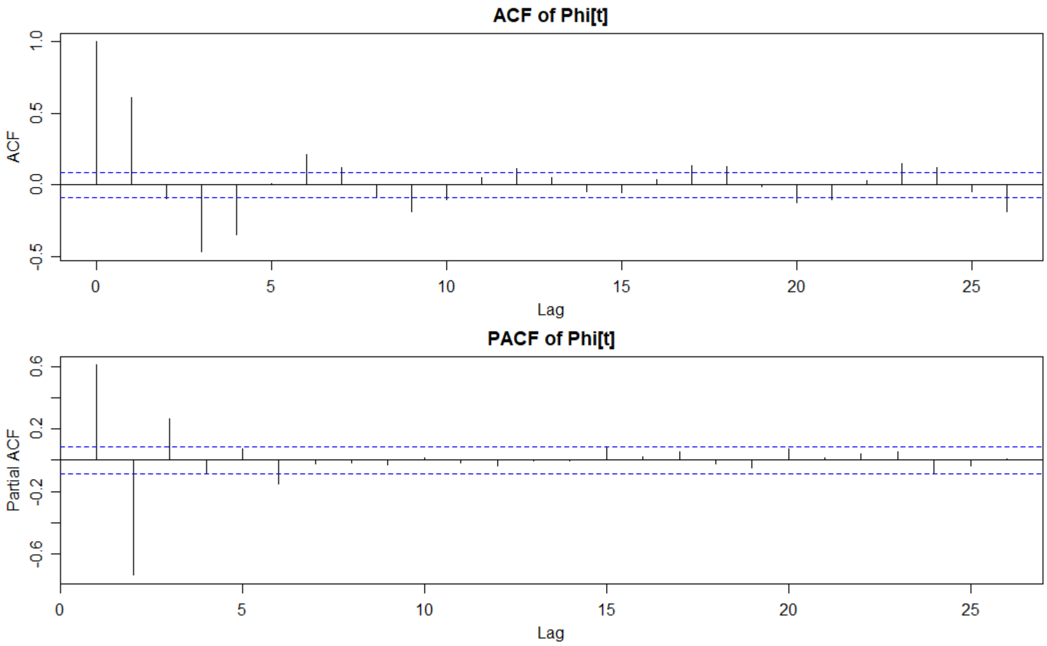


Figure 9: ACF and PACF of Φ_t

Indeed Figure 8 shows significance of the lags which resembles an AR(2) process [1]. The same applies to Figure 9, however the significant lags resembles an ARMA(2,1) process [1].

Question 2.4

Part a

$$\begin{aligned} x_{t+1} &= a x_t + v_t \\ y_t &= x_t + \epsilon_t \end{aligned} \tag{5}$$

First a number of observations will be generated from eq. (1) in the assignment. In this case it will be 3000 observations. Where the system noise is $v_t \sim \mathbb{N}(\mu_v = 0, \sigma_v = 1)$ and the observation noise is $e_t \sim \mathbb{N}(\mu_e = 0, \sigma_e = 1)$, furthermore $a = 0.4$. Now the model needs to be rewritten by including the parameter in the state vector, here the following is given in the script `EKF-example.R`. The state vector is therefore:

$$z_t = \begin{pmatrix} x_t \\ a_t \end{pmatrix} \tag{6}$$

Which is initialised by the following:

$$z_t = \begin{pmatrix} 0 \\ -0.5 \end{pmatrix} \tag{7}$$

The rest of the extended Kalman filter (EKF), can be seen in the mentioned script `EKF-example.R`.

Part b

By doing several simulations (in this case 20 simulations) using the above settings and with an initial value of the variance of a set to 1 and the same for the measurement and observation noise the following results can be shown:

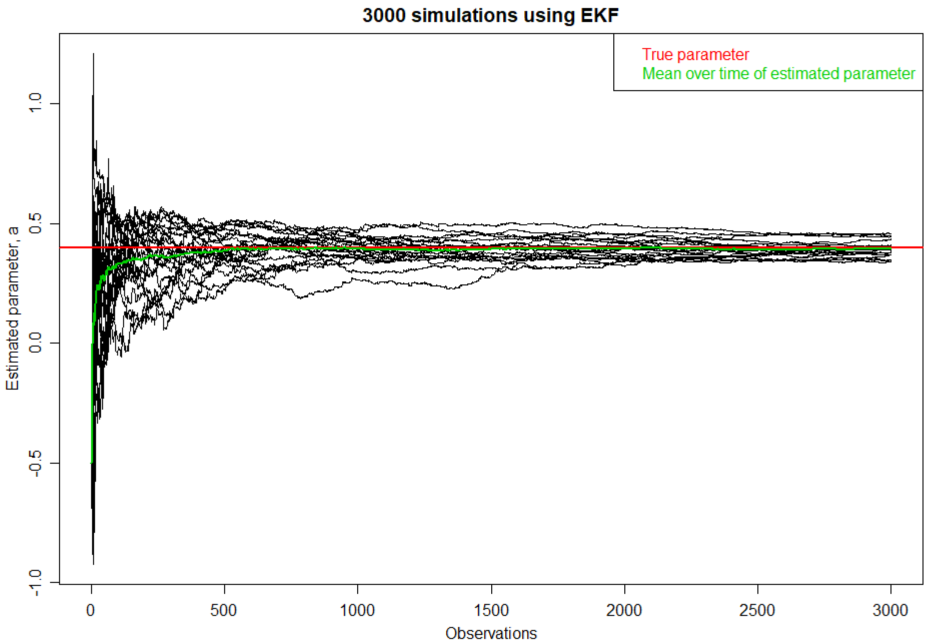
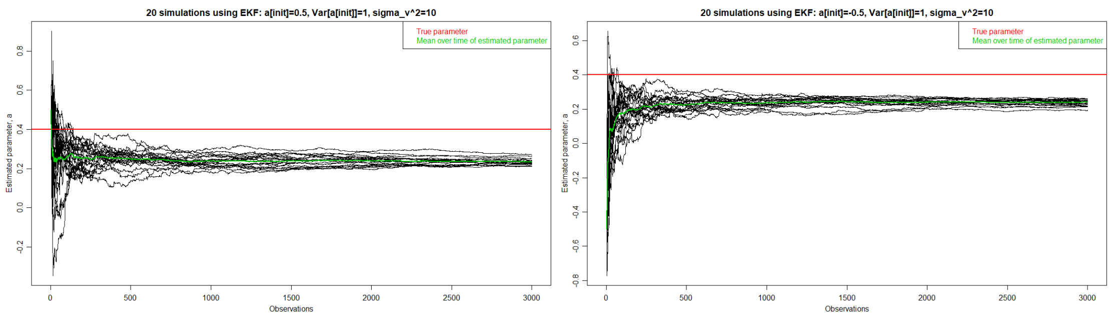


Figure 10: Estimated parameter, a, $\sigma_e = \sigma_v = \sigma_{a_{init}} = 1$

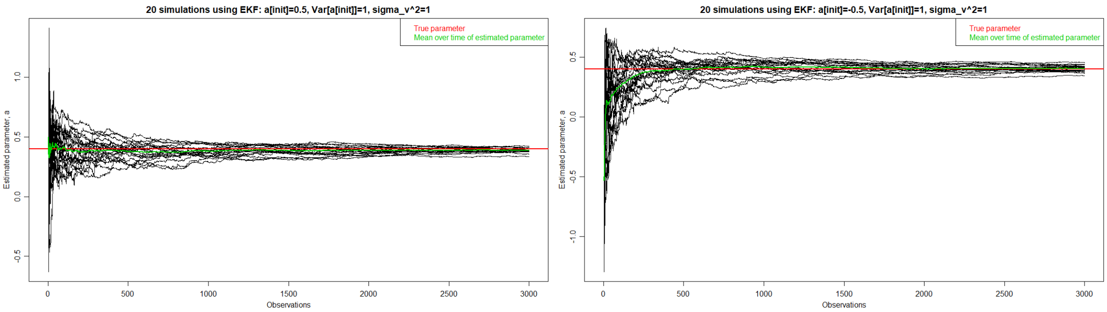
Clearly it can be seen that the EKF seems to asymptotically converge towards the true value of the parameter, a. Now below a different number of combinations of simulations using EKF is shown, the changes of the initial values can be seen in the caption to each figure.



(a) $a_{init} = 0.5$.

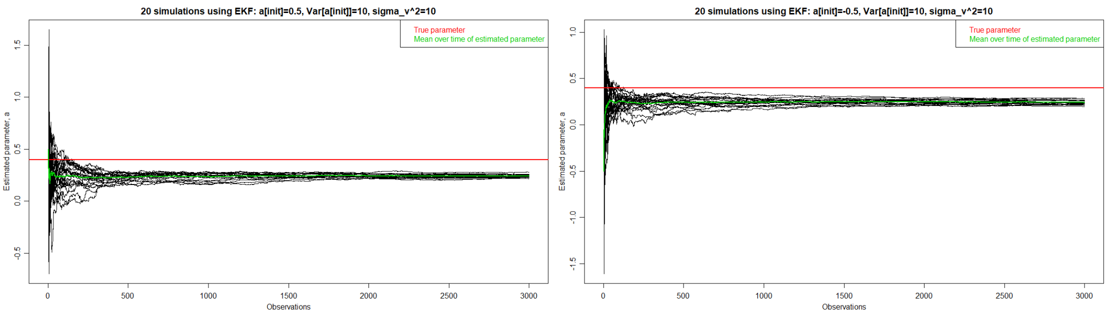
(b) $a_{init} = -0.5$.

Figure 11: 20 simulations, $V_{init}[a] = 1$, $\sigma_v^2 = 10$.



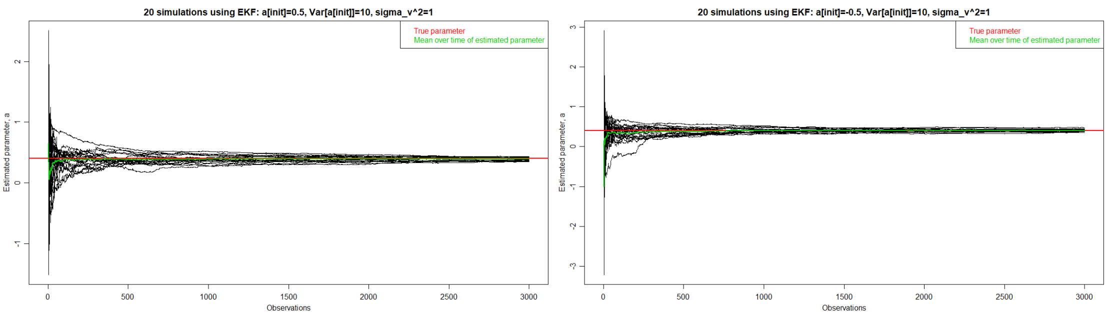
(a) $a_{init} = 0.5$. (b) $a_{init} = -0.5$.

Figure 12: 20 simulations, $V_{init}[a] = 1$, $\sigma_v^2 = 1$.



(a) $a_{init} = 0.5$. (b) $a_{init} = -0.5$.

Figure 13: 20 simulations, $V_{init}[a] = 10$, $\sigma_v^2 = 10$.



(a) $a_{init} = 0.5$. (b) $a_{init} = -0.5$.

Figure 14: 20 simulations, $V_{init}[a] = 10$, $\sigma_v^2 = 1$.

From the different figures it is clear that if the system noise, σ_v, is set too high, it will affect the estimated value of the parameter and the EKF will be influenced by a bias, which underestimates the parameter. However setting the initial variance of the parameter a doesn't seem to affect the convergence at all. The variance for the prediction step P_t will therefore start high if the initial variance of a is set high, but still converge, however the system noise σ_v will keep getting added to the prediction step, making the innovation covariance S_t high. This will in turn make the Kalman gain too low or high based on $sign(a)$. This will make EKF converge to a wrong parameter value. As seen in the figures above. One should therefore be careful initialising the system noise, to get appropriate estimates of the parameters, in this case. The scenario where the system noise is over estimated will create bias towards lower values, than the actual true value. This is also what was found in [3], here it is mentioned that the EKF method may give biased results, when initial estimate are not good enough. In [3] it concludes that the bias is not in the method, but due to wrong noise estimations. This is also observed in figure Figure 13 and Figure 11 where the noise estimates are too high.

References

[1] Henrik Madsen, *Time Series Analysis*, Chapman & Hall/CRC, 2008.

[2] Henrik Madsen & Jan Holst, *Modelling Non-Linear and Non-Stationary Time Series*, IMM, 2006, December.

[3] L. Ljung, *"Asymptotic behavior of the extended kalman filter as a parameter* estimator for linear systems," Automatic Control, IEEE Transactions on, vol. 24, pp. 36 – 50, 03 1979.