# Advanced Time Series Analysis: Computer Exercise 2

Autumn 21, 01622
Assessment 2
November 17, 2021

Magne Egede Rasmussen, s183963,
Nicolaj Hans Nielsen, s184335
& Anton Ruby Larsen, s174356

**Advanced Time Series Analysis: Computer Exercise 2**

| Part | Anton | Magne | Nicolaj |
|---|---|---|---|
| 1 | 56% | 22% | 22% |
| 2 | 56% | 22% | 22% |
| 3 | 22% | 56% | 22% |
| 4 | 22% | 22% | 56% |

Danmarks
Tekniske Universitet

# Contents

# 1 Part 1

In this exercise we are to estimate the parameters of one of the models simulated in the first computer exercise. We have chosen the SETAR model and to estimate the parameters we will apply the prediction error method described on page 115-118 in [1]. We first define a loss function $Q_N(\theta)$, where $\theta$ are our parameters.

$$Q_N(\theta) = \sum_{t=1}^{N} (x_t - \mathrm{E}(x_t|\mathcal{X}_{t-1}, \theta))^2 \tag{1.1}$$

Here $x_t$ is the observation at time $t$ and $\mathrm{E}(x_t|\mathcal{X}_{t-1}, \theta)$ is the conditional mean given all past observations, $\mathcal{X}_{t-1}$ and a set of parameters. The optimal $\hat{\theta}_N$ is now found by

$$\hat{\theta}_N = \arg\min_{\theta} Q_N(\theta) \tag{1.2}$$

To optimise 1.2 we can use algorithms like Newton-Rhapsody but it is not worth much if we cannot guarantee consistency of our estimator, i.e.

$$\lim_{N\to\infty} \|\theta_0 - \hat{\theta}_N\| = 0 \tag{1.3}$$

where $\theta_0$ is the true underlying set of parameters. For a SETAR model where $d$ and the number of regimes are known, Doric( [2]) proves consistency of the M-estimator for the SETAR model under some mild regularity conditions. The prediction error method is a subset of the M-estimator and hence we can guarantee consistency for 1.1.

The specific SETAR model we will estimate the parameters for is a SETAR(1,1;1,2):

$$X_t = \begin{cases} -3 - 0.5X_{t-1} + \varepsilon_t & if \quad X_{t-1} > 0 \\ 2 + X_{t-1} + \varepsilon_t & if \quad X_{t-1} \leq 0 \end{cases} \tag{1.4}$$

where $\varepsilon_t \sim \mathcal{N}(0, 1)$. We therefore have 5 parameters to estimate,

$$\theta_0 = (a_0^{(1)}, a_0^{(2)}, a_1^{(1)}, a_1^{(2)}, k_1) = (-3, 2, -0.5, 1, 0)$$

We will estimate $\hat{\theta}_N$ for $N \in (1000, 5000, 10000, 50000, 100000)$. To calculate 1.2, we will use the R function 'optim'. It implements several optimisation algorithm including the Newton-Rhapsody algorithm. One downside of the Newton-Rhapsody algorithm is though it can be sensitive to bad stating points. Hence we will find a good starting point using simulated annealing first and then apply the Newton-Rhapsody algorithm.

| | $\theta^{(1)}$ | $\theta^{(2)}$ | $\theta^{(3)}$ | $\theta^{(4)}$ | $\theta^{(5)}$ |
|---|---|---|---|---|---|
| $\theta_0$ | $-3$ | 2 | $-0.5$ | 1 | 0 |
| $\hat{\theta}_{1000}$ | $-2.934$ | 2.081 | $-0.546$ | 1.020 | $-0.039$ |
| $\hat{\theta}_{5000}$ | $-3.009$ | 1.921 | $-0.489$ | 0.969 | 0.001 |
| $\hat{\theta}_{10000}$ | $-3.033$ | 2.003 | $-0.480$ | 0.996 | 0.000 |
| $\hat{\theta}_{50000}$ | $-2.985$ | 2.013 | $-0.512$ | 1.002 | 0.000 |
| $\hat{\theta}_{100000}$ | $-3.001$ | 2.002 | $-0.504$ | 1.002 | 0.000 |

**Table 1** – Parameter estimates for $N \in (1000, 5000, 10000, 50000, 100000)$

We see from table 1 that as we get more data the estimates gets closer to the underlying true parameter $\theta_0$ as we expected due to the proven consistency. This is even clearer illustrated in figure 1 where the $l_2$-distance from the true parameter is plotted.

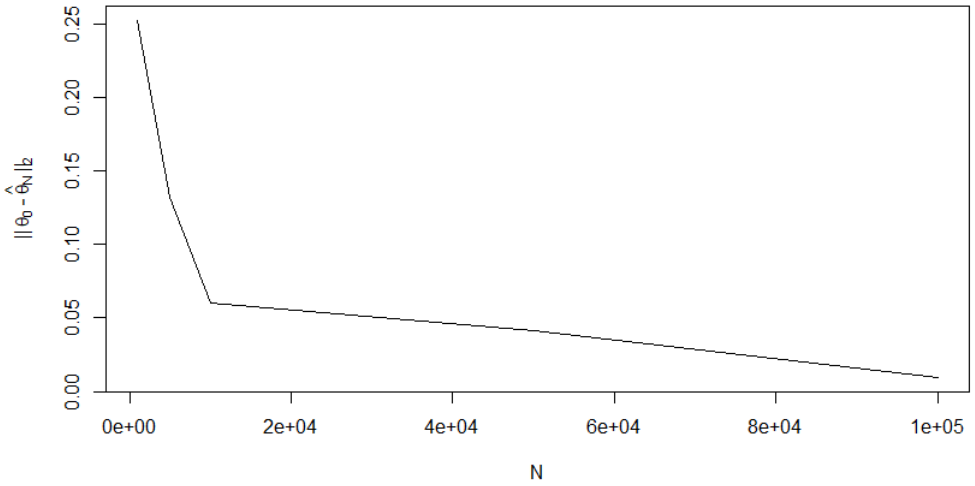



**Figure 1** – $l_2$-distance from the estimated parameter to the true underlying parameter

## 2 Part 2

In this part, we will look into the contour curves of the loss function (1.1), as we vary parameters pairwise. We will use 5000 simulated observations from the same SETAR model as in part 1.
First we compare all the parameters pairwise using all 5000 oberservations. In the following $\theta_1$ and $\theta_3$ are respectively the intercept and slope in regime 1, $\theta_2$ and $\theta_4$ are the intercept and slope in regime 2, and lastly $\theta_5$ is the value for which we shift regime.

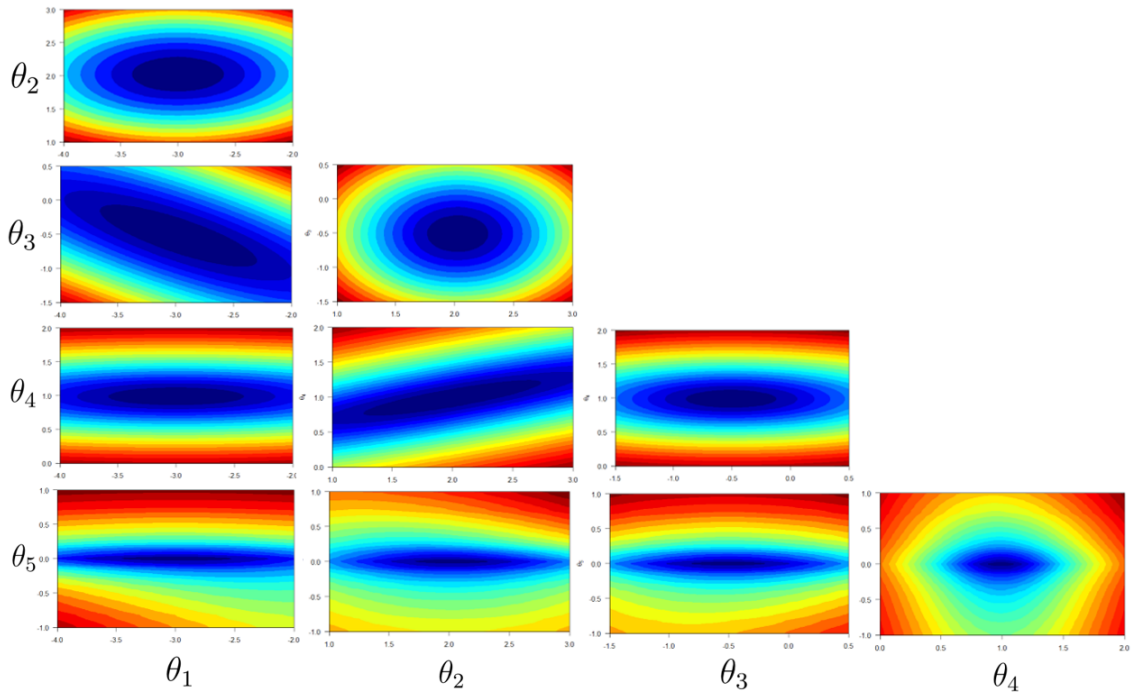



**Figure 2** – Contour curves for all combinations of parameters. The values of the loss function goes from the lowest represented by dark blue to the highest represented by a dark red.

In figure 2 we observe three major things.

1. We see that the intercepts and slops of the AR-models are not correlated between the regimes, i.e. $\theta_1$ and $\theta_3$ are not correlated with $\theta_2$ and $\theta_4$.

2. We see that $\theta_1$ and $\theta_3$ are negatively correlated and $\theta_2$ and $\theta_4$ are positively correlated. This also makes sense due to their underlying true signs.

3. We see all parameters except $\theta_5$ has nice quadratic contour curves as we know them from the Gaussian distribution. This is due to the AR-models being linear. On the contrary, we see that all contour plots where $\theta_5$ is one of the parameters, exhibits very non-Gaussian nature. We see wired curves and at the value where the regime shifts we have a sharp kink in the contour curves. We see this very non-linear behaviour because the regime shift is what makes the model non-linear.

Next we only compare $\theta_1$ and $\theta_3$ but change the number of observations. We use 30, 300 and the full 5000 observation to generate the contour curves of the loss function.

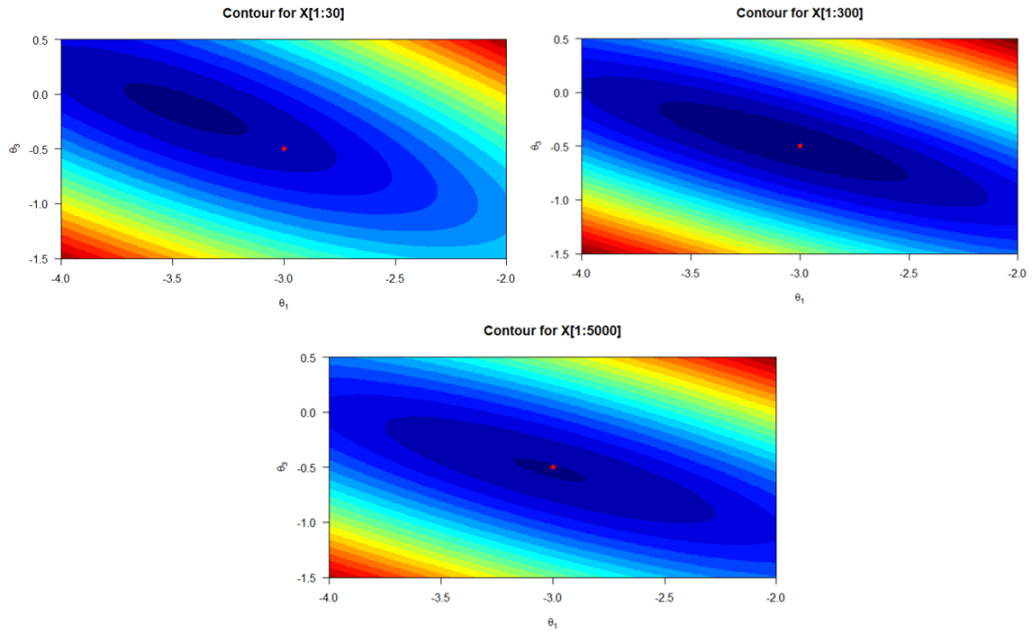



**Figure 3** – The loss function for $\theta_1$ and $\theta_3$ using 30, 300 and 5000 observations. The values of the loss function goes from the lowest represented by dark blue to the highest represented by a dark red.

We see in figure 3 that the minimum of the loss function converges towards the true underlying parameters which are indicated by a red dot. This is in line with the consistency of the estimator which we both argued for theoretically and observed empirically in part 1.

# 3 Part 3

A doubly stochastic model has the ability to describe general non-linear dynamics and can in many cases be used as a substitute for other non-linear models.

Tjøstheim [3] designed the doubly stochastic model as:

$$X_t = A_0(t-1;X) + \sum_{i=1}^{P} A_i(t-1;X)\theta_i X_{t-i} + \sum_{j=0}^{Q} B_j(t-1;X)\varepsilon_{t-j} \tag{3.1}$$

with $\theta_i$, $A_i$ and $B_i$ being random variables. Hence, (3.1) is constructed with time varying coefficients.

We create the following non-linear AR(2) - ARMA(3, 2) model by:

$$X_t = \Phi_t X_{t-1} + \Phi_{t-1} X_{t-2} + \varepsilon_t \tag{3.2}$$

$$\Phi_t - \mu = \phi_1(\Phi_{t-1} - \mu) + \phi_2(\Phi_{t-3} - \mu) + \zeta_t + \theta_1\zeta_{t-1} + \theta_2\zeta_{t-2} \tag{3.3}$$

(3.2) is a doubly stochastic model with $P = 2$, $\{\theta\} = \mathbf{1}$ and $\{A\} = \{0, \Phi_t, \Phi_{t-1}\}$. As $\Phi_t$ is varying through time, (3.2) and (3.3) combined define a doubly stochastic system. Let us now show that this specific system can be written as a dynamic linear state space model.

We start rewriting the dynamic equation (3.3) by isolating $\Phi_t$:

$$\Phi_t = \mu(1 - \phi_1 - \phi_2) + \phi_1\Phi_{t-1} + \phi_2\Phi_{t-3} + \zeta_t + \theta_1\zeta_{t-1} + \theta_2\zeta_{t-2} \tag{3.4}$$

Defining the constant $\mu(1 - \phi_1 - \phi_2)$ by $\delta$, we can write the embedded states and their internal relationships as:

$$\begin{bmatrix} \Phi_t \\ \Phi_{t-1} \\ \Phi_{t-2} \\ \delta_t \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 & \phi_2 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Phi_{t-1} \\ \Phi_{t-2} \\ \Phi_{t-3} \\ \delta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & \theta_1 & \theta_2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \zeta_t \\ \zeta_{t-1} \\ \zeta_{t-2} \\ \zeta_{t-3} \end{bmatrix} \tag{3.5}$$

Finally, we can write $X_t$ using vector notation by:

$$X_t = \begin{bmatrix} X_{t-1} & X_{t-2} & 0 & 0 \end{bmatrix} \begin{bmatrix} \Phi_t \\ \Phi_{t-1} \\ \Phi_{t-2} \\ \delta_t \end{bmatrix} + \varepsilon_t \tag{3.6}$$

Hence, by (3.5) and (3.6) we have written our initial non-linear AR(2) - ARMA(3, 2) model on a state space form.

## 3.1 Simulations

We now turn to the second task in part 3, namely simulations of our selected model. We set the following parameters:

| $\mu$ | $\phi_1$ | $\phi_2$ | $\sigma_\zeta$ | $\theta_1$ | $\theta_2$ | $\sigma_\varepsilon$ |
|---|---|---|---|---|---|---|
| 0.5 | 0.4 | -0.7 | 0.1 | 0.4 | 0.8 | 0.3 |

**Table 2** – Parameter selection for simulations of (3.5) and (3.6).

With the parameters defined in table 2 and the starting conditions $(\Phi_1, \Phi_2, X_1, X_2) = (\sigma_\zeta, \sigma_\zeta, \sigma_\varepsilon, \sigma_\varepsilon)$ we have simulated $t \in [0, 500]$ for 501 discrete values of $t$ separated by the identical event length of 1:

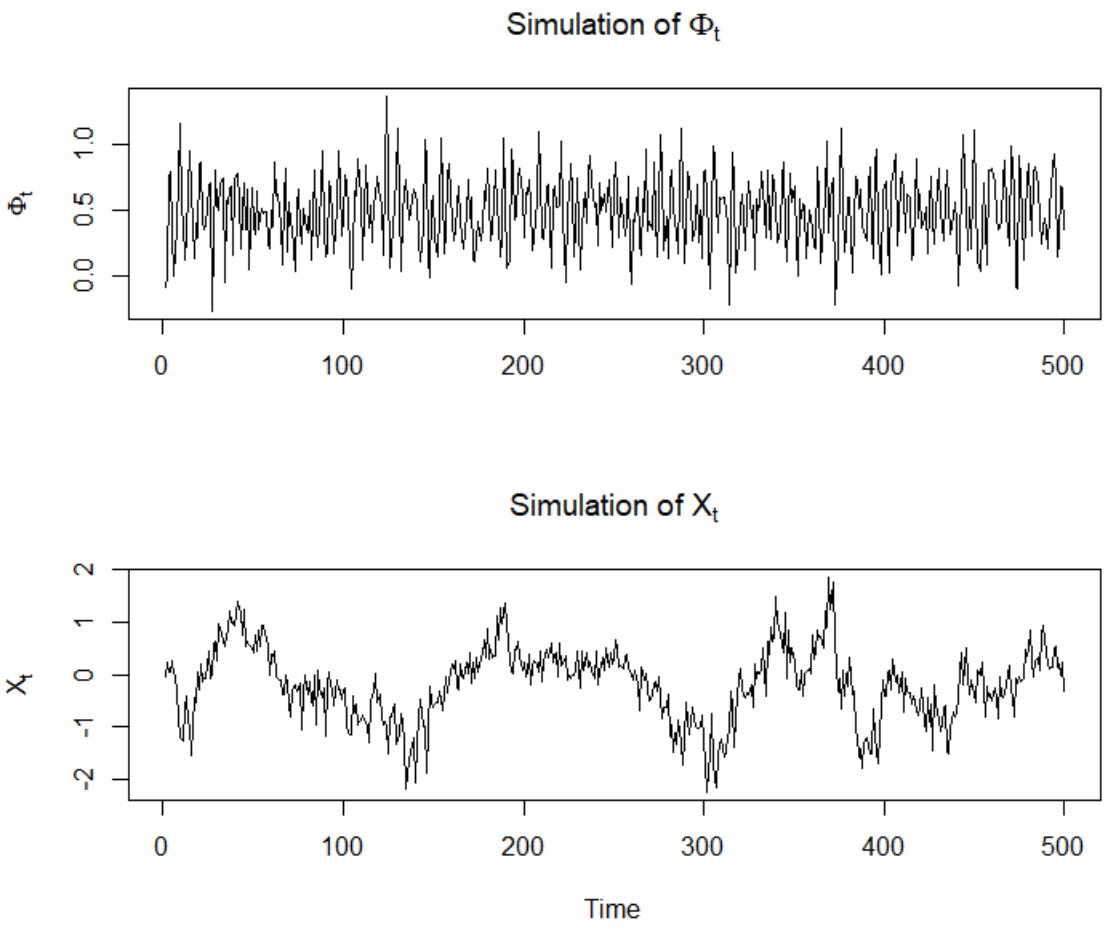



**Figure 4** – Simulation of $X_t$ constructed by a simulation of $\phi_t$ for $t \in [0, 500]$.

From fig. 4, we notice that large absolute values of the process $X_t$ are generally obtained in periods where the system describing the embedded parameter variation is unstable. Providing a deeper analysis of the simulations from fig. 4 we set up ACF and PACF for respectively $\Phi_t$ and $X_t$. The auto correlations are illustrated on the following page (page 8).

In fig. 5 we look at autocorrelation functions with respect to the embedded parameter $\Phi_t$. We see that their distributions align with a standard ARMA$(3, 2)$ model: Both ACF and PACF are dominated by exponential / sinusoidal damping already starting from early lags with a few outliers. The start at lag 0 for ACF is indeed expected, since $\max\{0, 2-3\} = 0$, while for PACF the decay starts at lag 1 because $\max\{0, 3-2\} = 1$. [4].

Fig. 6 instead looks at the series $X_t$. Here it is evident that ACF are damped exponentially. It is a bit more tricky to evaluate PACF in fig. 6, though it agrees with an AR(2) process that lag 1 and 2 are the most significant. Hence, the lags in the simulation of $X_t$ resembles the lags of the underlying process AR(2).

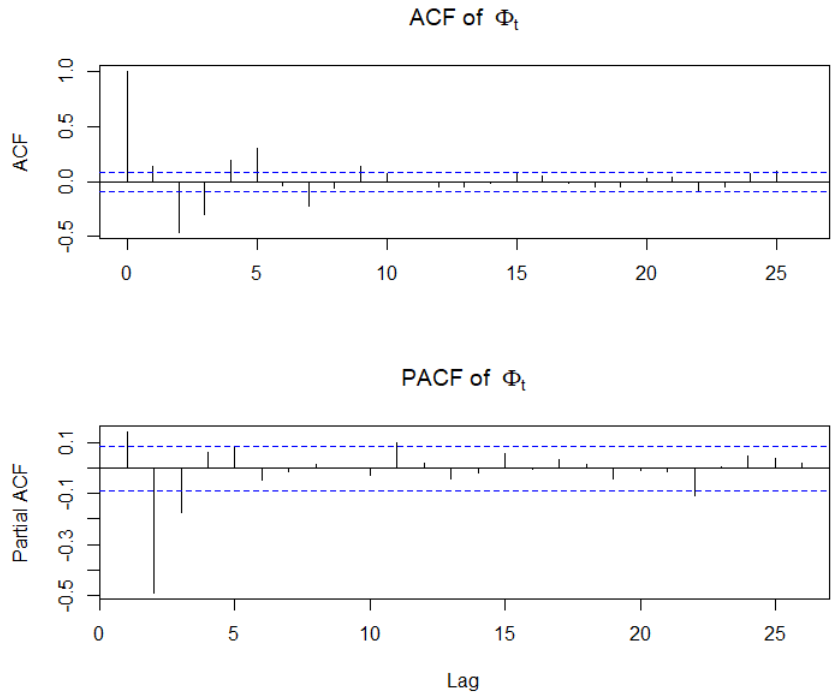



**Figure 5** – Complete and partial auto correlation function for the embedded parameter $\Phi_t$.

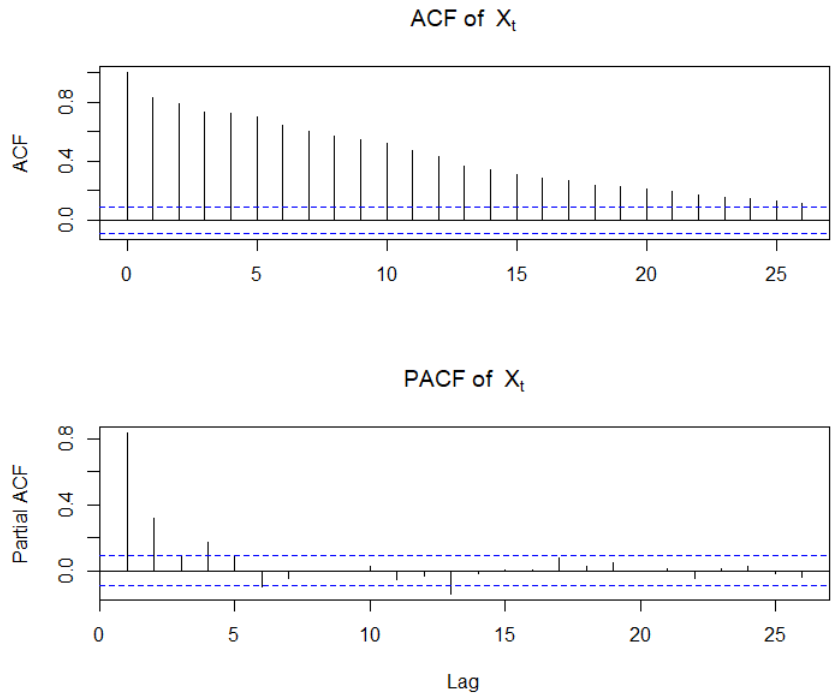



**Figure 6** – Complete and partial auto correlation function for the process $X_t$.

# 4 Part 4

## 4.1 Part 4 prelude

*The following sections are not to be seen as a direct part of the assignment. These are mostly sections created for our reference to quickly grasp the intuition of the Kalman filter.*

### 4.1.1 Part 4 prelude - The linear Kalman filter

*This is a brief recap of state-space models and Kalman filtering through an intuitive example.*
Consider a simple train on a straight track with motors, we can control, and sensors at each end that can measure the distance to nearby objects.
We can model this with the generic linear setup

$$\begin{aligned} \boldsymbol{X}_t &= \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{B}\boldsymbol{u}_{t-1} + v_t \\ \boldsymbol{Y}_t &= \boldsymbol{C}\boldsymbol{X}_t + e_t \end{aligned} \tag{4.1}$$

where

1. $\boldsymbol{A}$ would be the matrix that describes how the train would evolve from $t-1$ to $t$ without control or noise. We could derive $\boldsymbol{A}$ from standard equations of motions and the appropriate states $\mathbf{X}_t$.

2. $\boldsymbol{B}$ encapsulates what would happen if we applied some control $\boldsymbol{u}_{t-1}$. In our case, $\boldsymbol{u}_{t-1}$ could be that we turned on the motors, and $\boldsymbol{B}$ would describe how that would change the dynamics.

3. $\boldsymbol{C}$ will map where we think we will go $\mathbf{X}_t$ to specific measurements we will then expect $\boldsymbol{Y}_t$. In our case, the expected distance the sensors would measure.

4. $v_t$. The deterministic modeled system will not perfectly match reality and likewise, the applied control might not perfectly match the actual outcome. This uncertainty is expressed as a system noise term $v_t$.

5. $e_t$. The sensors are not perfect hence we encapsulate the uncertainty of the measurement in $e_t$.

We assume that $v_t$ and are independent and normal distributed with $\mathbb{E}\left[vv^\intercal\right] = \Sigma_1$ and $\mathbb{E}\left[ee^\intercal\right] = \Sigma_2$.
The derivation of the Kalman filter will not be covered but can be found on p 289-294 [4]. However, the essential update equation is found under the assumptions of linearity and Gaussianity:

$$\widehat{X}_{t|t} = \widehat{X}_{t|t-1} + K_t\left(Y_t - C\widehat{X}_{t|t-1}\right) \tag{4.2}$$

The most important objective for the filtering is to update our internal state, given new measurements. To do that, we use our estimate of where we think the state of the train is $\widehat{X}_{t|t-1}$ and the prediction error from the measurements $\left(Y_t - C\widehat{X}_{t|t-1}\right)$. To update in the most optimal, we have to assess the uncertainty of the internal state and compare this with new measurements and the associated uncertainty of the sensor. This weighting is expressed in the Kalman gain:

$$\boldsymbol{K}_t = \boldsymbol{\Sigma}^{xx}_{t|t-1}\boldsymbol{C}^T\left(\boldsymbol{\Sigma}^{yy}_{t|t-1}\right)^{-1} \tag{4.3}$$

$\boldsymbol{\Sigma}^{xx}_{t|t-1}\boldsymbol{C}^T$ is the estimated variance of the states when we use the internal model. This is normalised by the estimated variance of the sensor, $\boldsymbol{\Sigma}^{yy}_{t|t-1}$. If the sensor cannot be trusted at all, then we see directly $\widehat{X}_{t|t} \approx \widehat{X}_{t|t-1}$. If the sensor is highly accurate then $\boldsymbol{\Sigma}_2 \approx 0$. Consider equation (10.78) [4]:

$$\boldsymbol{\Sigma}^{yy}_{t|t} = C\boldsymbol{\Sigma}^{xx}_{t|t-1}C^\intercal + \boldsymbol{\Sigma}_2 \tag{4.4}$$

With this update and $\boldsymbol{\Sigma}_2 \approx 0$, we see that the Kalman gain (4.3) becomes

$$\begin{aligned}\boldsymbol{K}_t &= \boldsymbol{\Sigma}^{xx}_{t|t-1}\boldsymbol{C}^T\left(C\boldsymbol{\Sigma}^{xx}_{t|t-1}C^\intercal\right)^{-1}\\ &= C^\intercal\end{aligned} \tag{4.5}$$

This would change the update of equation (4.2) to become

$$\begin{aligned}\widehat{X}_{t|t} &= \widehat{X}_{t|t-1} + C^\intercal\left(Y_t - C\widehat{X}_{t|t-1}\right)\\ &= C^\intercal Y_t\end{aligned} \tag{4.6}$$

$C^\intercal$ is the inverse of $C$, hence it will now take us from the observations to the internal state $X$. Explicitly, if the sensor is perfect, then we will directly change the internal state s.t. it matches the observations perfectly.
From a pragmatic view, we can see how the Kalman filter is useful. To fully grasp theoretical details consult [4].

### 4.1.2 Part 4 prelude - The extended Kalman filter

The Kalman filter relies heavily on the result that any linear combination of normally distributed random variables is normal. However, many real-world systems are non-linear. Consider a generic non-linear system

$$\begin{aligned}X_{t+1} &= f(X_t, U_t, v_t)\\ Y_t &= h(X_t, e_t)\end{aligned} \tag{4.7}$$

wherein we consider $v_t$ and $e_t$ to be additive noise with $\mathrm{E}\left[\boldsymbol{v}\boldsymbol{v}^T\right] = \boldsymbol{R}_{1,t}$, $\mathrm{E}\left[\boldsymbol{e}\boldsymbol{e}^T\right] = \boldsymbol{R}_{2,t}$, and $\mathrm{E}\left[\boldsymbol{v}\boldsymbol{e}^T\right] = \boldsymbol{0}$.
The simplest solution to make the machinery work is to make a Taylor approximation of order one for each of our functions around the current estimate.
Consider the prediction step. Here we linearize the dynamics around the current estimate of the state $x = \mu_t$, the control input $u = u_t$, and the noise equal to 0 $v = 0$:

$$X_{t+1} \approx f(\mu_{t-1}, u_t, 0) + \left.\frac{\partial f}{\partial x}\right|_{\mu_{t-1},u_t,0}(x-\mu_{t-1}) + \left.\frac{\partial f}{\partial u}\right|_{\mu_{t-1},u_t,0}(u-u_t) + \left.\frac{\partial f}{\partial v}\right|_{\mu_{t-1},u_t,0}(n-0) \tag{4.8}$$

Likewise, for the update step, we make a Taylor approximation about the predicted state $x = \hat{\mu}_t$ and zero noise $e = 0$

$$Y_t = h(x,v) \approx h\left(\overline{\mu_t}, 0\right) + \left.\frac{\partial h}{\partial x}\right|_{\overline{\mu_t},0}(x-\overline{\mu_t}) + \left.\frac{\partial h}{\partial e}\right|_{\overline{\mu_t},0}(e-0) \tag{4.9}$$

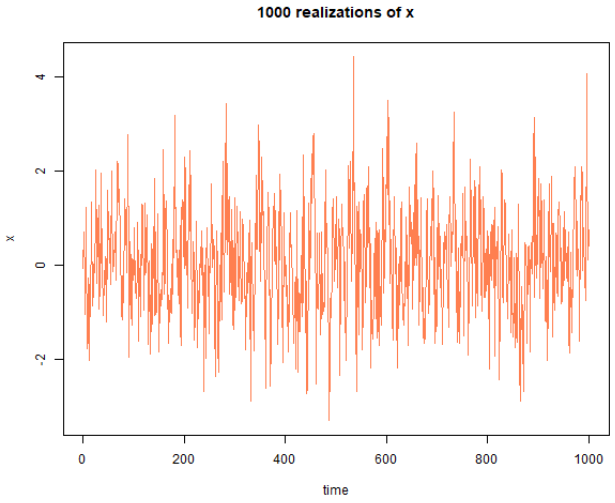
With these linearizations, we can now use the Kalman filter as before. We will not derive the new update equations but in can be found in e.g. [5].

## 4.2 Part 4a

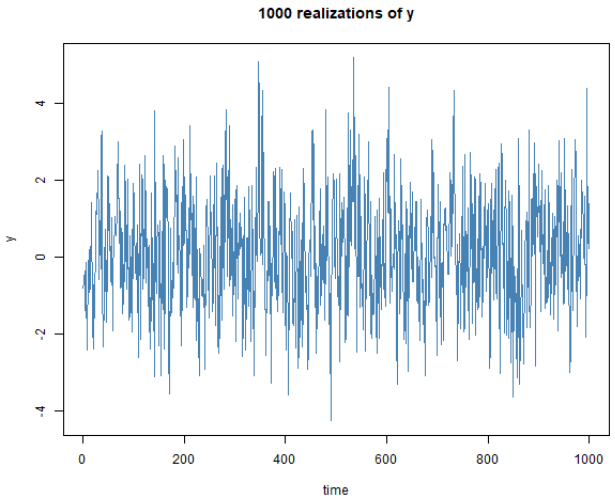
In this exercise, we consider the system

$$\begin{aligned} x_{t+1} &= ax_t + v_t \\ y_t &= x_t + e_t \end{aligned} \tag{4.10}$$

here $a$ is a parameter we want to estimate and $v_t$, $e_t$ are mutually uncorrelated white noise process with with variances $\sigma_v^2$ and $\sigma_e^2$. In the following, we will consider the specific system with $a = 0.4$ and $\sigma_v^2 = \sigma_e^2 = 1$. We will create 20 simulations with 1000 steps in each of the simulations, and we sample the initial conditions $x_0 \sim \mathcal{N}(0, 1)$. One of these 20 simulations looks like this:




**(a)** One simulation with 1000 realizations of $x$




**(b)** One simulation with 1000 realizations of $y$

**Figure 7** – One simulation with 1000 realizations of the system

**Rewrite to state space form** We introduce $a$ as a state and hence introduce a new state space vector $\mathbf{Z}_t \in \mathbb{R}^2$ where $\mathbf{Z}_t = \begin{bmatrix} x_t & a \end{bmatrix}^\intercal$. The system equation is

$$\mathbf{Z}_{t+1} = \begin{bmatrix} x_{t+1} \\ a_{t+1} \end{bmatrix} = \begin{bmatrix} a_t & 0 \\ 0 & a_t \end{bmatrix} \begin{bmatrix} x_t \\ a_t \end{bmatrix} + \begin{bmatrix} v_t \\ 0 \end{bmatrix} \tag{4.11}$$

and the observation equation is

$$Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ a_t \end{bmatrix} + e_t \tag{4.12}$$

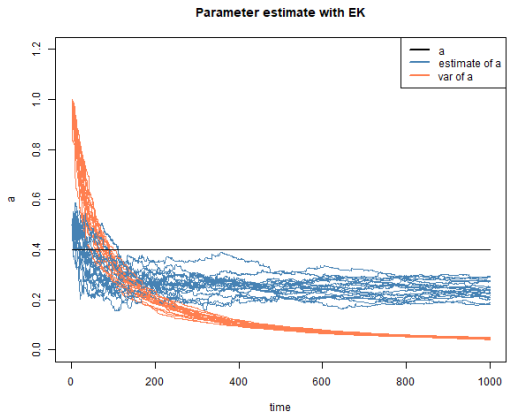
With this setup, we can now estimate $a$ using the extended Kalman filter. We will now estimate $a$ using the extended Kalman filter for each of the 20 simulations.

## 4.3 Part 4b

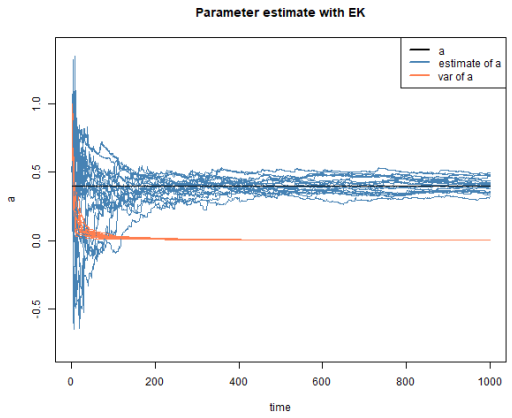
We now consider 8 scenarios with different initial parameter settings for the system. We will try $a_0 = 0.5$ and $a_0 = -0.5$ and for each of them try the following settings:

1. $\tilde{\sigma}_v^2 = 10$, $\hat{\sigma}_{a,0}^2 = 1$
2. $\tilde{\sigma}_v^2 = 1$, $\hat{\sigma}_{a,0}^2 = 1$
3. $\tilde{\sigma}_v^2 = 10$, $\hat{\sigma}_{a,0}^2 = 10$
4. $\tilde{\sigma}_v^2 = 1$, $\hat{\sigma}_{a,0}^2 = 10$

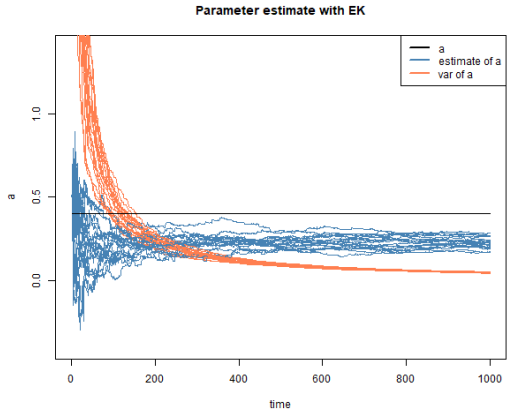
Where $\tilde{\sigma}_v^2$ is the assumed variance of the system noise while $\hat{\sigma}_{a,0}^2 = 1$ is the assumed initial variance of the estimate of $a$. In the following, we let $\hat{\sigma}_{a,i}^2 = 1$ denote the estimate at point $i \in [0, 1000]$.
In the following, we will plot the conditional mean of $a$ alongside the conditional variance of the system for each time point $t$. The plots are divided into two group; one with $a = 0.5$, one with $a = -0.5$



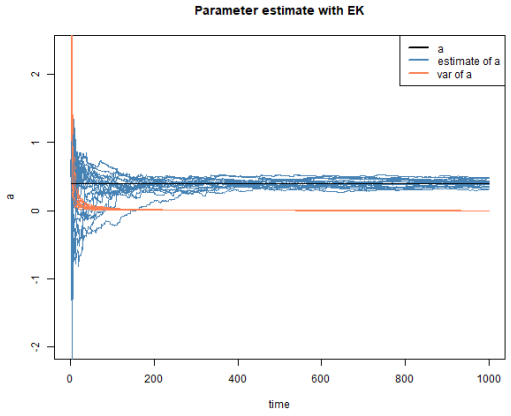

**(a)** $a = 0.5$, $\tilde{\sigma}_v^2 = 10$, $\hat{\sigma}_{a,0}^2 = 1$




**(b)** $a = 0.5$, $\tilde{\sigma}_v^2 = 1$, $\hat{\sigma}_{a,0}^2 = 1$

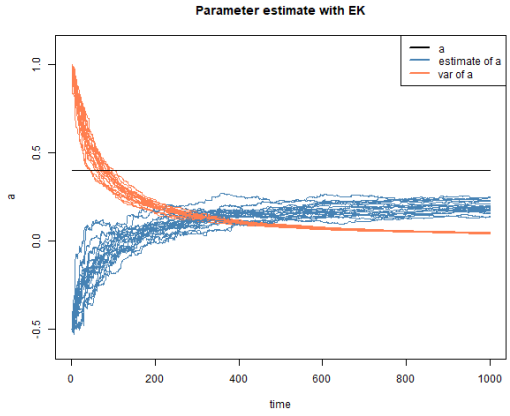



**(c)** $a = 0.5$, $\tilde{\sigma}_v^2 = 10$, $\hat{\sigma}_{a,0}^2 = 10$

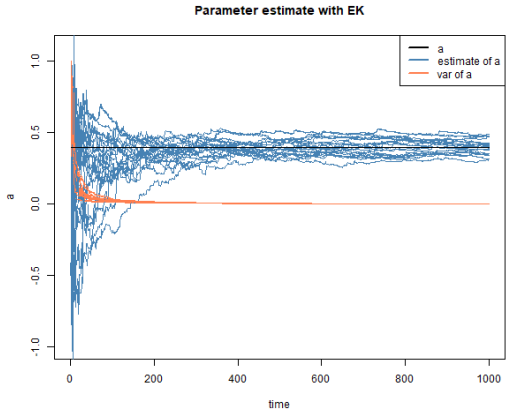



**(d)** $a = 0.5$, $\tilde{\sigma}_v^2 = 1$, $\hat{\sigma}_{a,0}^2 = 10$

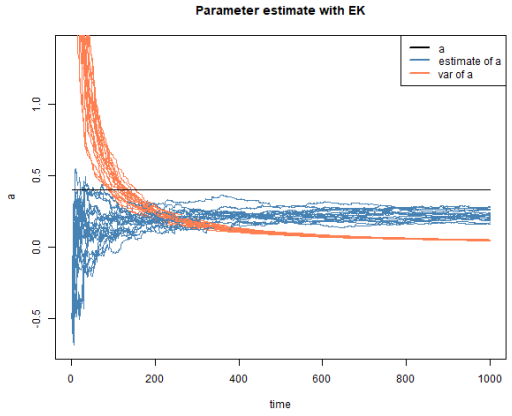
**Figure 8** – Parameter estimations using the extended Kalman filter for different initial parameter settings. All above have $a = 0.5$. *Notice that the y-axis has different scales*



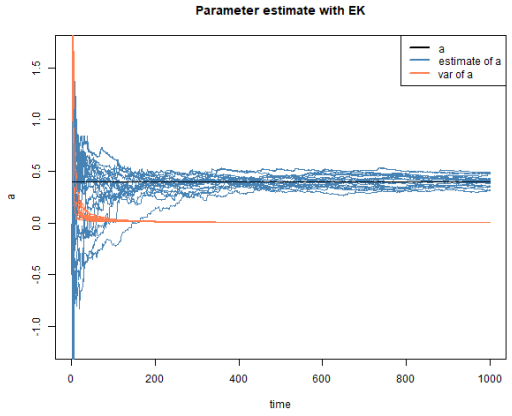

**(a)** $a = -0.5$, $\tilde{\sigma}_v^2 = 10$, $\hat{\sigma}_{a,0}^2 = 1$




**(b)** $a = -0.5$, $\tilde{\sigma}_v^2 = 1$, $\hat{\sigma}_{a,0}^2 = 1$




**(c)** $a = -0.5$, $\tilde{\sigma}_v^2 = 10$, $\hat{\sigma}_{a,0}^2 = 10$




**(d)** $a = 0.5$, $\tilde{\sigma}_v^2 = 1$, $\hat{\sigma}_{a,0}^2 = 10$

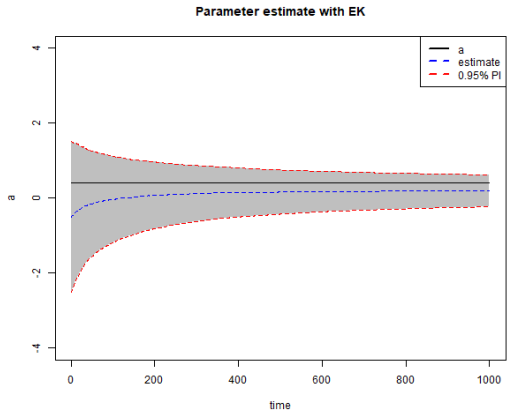
**Figure 9** – Parameter estimations using the extended Kalman filter for different initial parameter settings. All above have $a = -0.5$. *Notice that the y-axis has different scales*

From the plots above we make the following observations

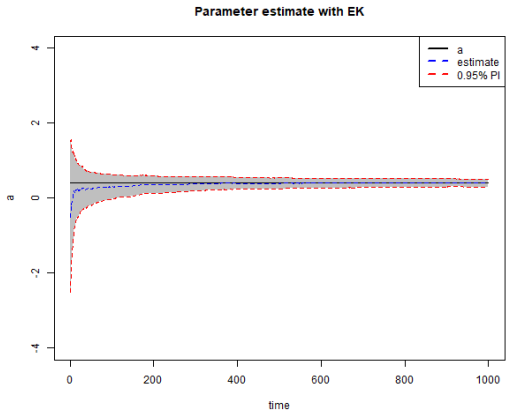
- Comparing figures with $a = 0.5$ in figure 8 and $a = -0.5$ in figure 9, we see that the initial guess of our estimate does not affect the asymptotic convergence.
- Consider the 1. scenario, figure 8a and 9a, where $\tilde{\sigma}_v^2 = 10$ and $\hat{\sigma}_{a,0}^2 = 1$. Here the estimate of $a$ has a negative bias with estimates below 0.4 even asymptotically. The uncertainty of $a$ is large in the beginning but is monotonically decreasing with the number of iterations even though the estimate of $a$ is still far from 0.4. It seems that the estimated $a$ is too optimistic. Further, the initial uncertainty of $a$ might also be set too low.
- In scenario 2 with $\tilde{\sigma}_v^2 = \hat{\sigma}_{a,0}^2 = 1$, figure 8b and 9b, we see a very fast convergence for the estimate of $a$ and a fast decrease in the variance. In comparison with scenario 1, we see that with a low system noise, it seems that the filter performs way better. Here the initial variance estimate of $a$ seems more adequate.
- For scenario 3 with $\tilde{\sigma}_v^2 = \hat{\sigma}_{a,0}^2 = 10$, figure 8c and 9c, we again see a systematic negative biased for conditional mean estimate of $a$. The variance is again decreasing over time even though the estimate of $a$ is biased. Even with a larger $\hat{\sigma}_{a,0}$ than in scenario 1, the filter still wrongfully display asymptotic decrease in the variance of the estimate.
- In scenario 4 where $\tilde{\sigma}_v^2 = 1$ and $\hat{\sigma}_{a,0}^2 = 10$ we see faster convergence of the estimators an decrease in variance of $a$, see figure 8d and 9d. Because the initial variation of $a$ is high, the very first points have a large estimated variance as well, however, it seems that scenario 3 has the same asymptotic convergence properties as scenario 2 with $\tilde{\sigma}_v^2 = \hat{\sigma}_{a,0}^2 = 1$. From this particular seems that as long as the system noise is low, we can set $\hat{\sigma}_{a,0}$ large and we will still have fast convergence towards the right conditional mean and monotonically decreasing conditional variance.

The aforementioned bias for scenario 1 and 3 is a direct consequence of the wrong assumption that $\tilde{\sigma}_v^2 = 10$ which is far from the real $\sigma_v^2 = 1$. As mention in [6] p. 41, the bias in

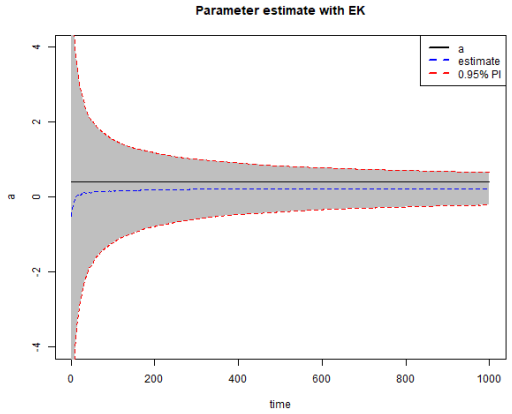
the parameter estimates can arise for any system if $\tilde{\sigma}_v^2$ is very far from $\sigma_v^2$ and this is what we see in these particular cases with scenario 1 and 3.
In a real situation, we would not know the true value of $a$ nor $\sigma_v^2$ hence here we would focus much more on the estimate of $a$ and how certain we are about that value. Consider therefore the filter with $a = -0.5$, $\hat{a}$ and $\hat{\sigma}_a^2$. We now take the average of the parameter estimates for each of the 20 simulations and filters and display the variance as a 95%-confidence interval.



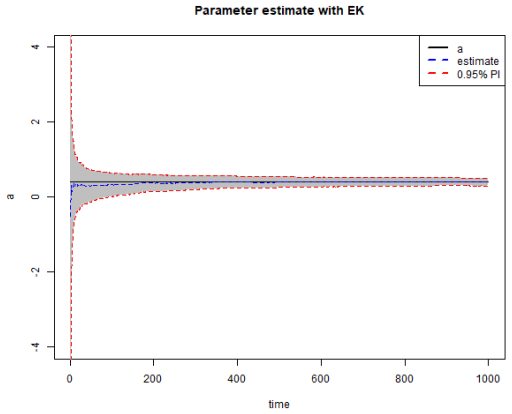

**(a)** $a = -0.5$, $\tilde{\sigma}_v^2 = 10$, $\hat{\sigma}_{a,0}^2 = 1$




**(b)** $a = -0.5$, $\tilde{\sigma}_v^2 = 1$, $\hat{\sigma}_{a,0}^2 = 1$




**(c)** $a = -0.5$, $\tilde{\sigma}_v^2 = 10$, $\hat{\sigma}_{a,0}^2 = 10$




**(d)** $a = 0.5$, $\tilde{\sigma}_v^2 = 1$, $\hat{\sigma}_{a,0}^2 = 10$

**Figure 10** – Parameter estimations using the extended Kalman filter for different initial parameter settings. All above have $a = -0.5$. *Notice the y-axis is now fixed*

The parameter estimates at $t = 1000$ are listed in the table below:

| | $a_0$ | $\sigma_v^2$ | $\hat{\sigma}_{a,0}^2$ | $\hat{a}_{1000}$ | $\hat{\sigma}_{a,1000}^2$ | 95% Pred. |
|---|---|---|---|---|---|---|
| 1a | 0.5 | 10 | 1 | 0.2379 | 0.0451 | $[-0.1867, 0.6625]$ |
| 2a | 0.5 | 1 | 1 | 0.4044 | 0.0025 | $[0.3039, 0.5049]$ |
| 3a | 0.5 | 10 | 10 | 0.2268 | 0.047 | $[-0.2067, 0.6603]$ |
| 4a | 0.5 | 1 | 10 | 0.4032 | 0.0025 | $[0.3026, 0.5038]$ |
| 1 b | $-0.5$ | 10 | 1 | 0.1928 | 0.0451 | $[-0.232, 0.6176]$ |
| 2 b | $-0.5$ | 1 | 1 | 0.4005 | 0.0026 | $[0.2991, 0.5019]$ |
| 3 b | $-0.5$ | 10 | 10 | 0.2221 | 0.047 | $[-0.2115, 0.6556]$ |
| 4 b | $-0.5$ | 1 | 10 | 0.4025 | 0.0025 | $[0.3018, 0.5033]$ |

(4.13)

The general comments are the same and the bias is still obvious for $\sigma_v^2$, however, in all cases we see that the estimates are still within the 95% prediction interval. There is no doubt that for scenario 2 and 4, figure 10b and 10d, the prediction interval is much tighter which is made very explicit with this fixed $y$-axis for all of the scenarios. We see again that it is of great importance to have adequate values for the system variance.

# 5 References

[1] H. Madsen and J. Holst, "Modelling non-linear and non-stationary time serie," *IMM*, 2016.

[2] D. Doric, "M - estimates of setar model parameters," 2002.

[3] D. Tjøstheim, "Some doubly stochastic time series models," *Time Series Anal 7, 51–72.*, 1986.

[4] H. Madsen, *Time series analysis*. Chapman and Hall/CRC, 2007.

[5] J. Durbin and S. J. Koopman, *Time series analysis by state space methods*. Oxford university press, 2012.

[6] L. Ljung, "Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems," *IEEE Transactions on Automatic Control*, vol. 24, no. 1, pp. 36–50, 1979.