Contents

1 Introduction

1.1 Motivation.

An observed stochastic process may consist of many different processes changing in time. As an example we can think of network traffic. The process may be well determined and understood as long as no problem occurs, but suddenly, when there are errors somewhere in the network, the process may get a quite different behavior. The problem is to be able to determine when the process follows a new model, and also what kind of a new model. Different kinds of errors in the network may result in different new processes. In such a situation, recognition of what kind of model the process has switched to, may also tell us what kind of an error we are dealing with.

This problem can be solved by letting the actual time series follow a Dynamic Linear Model where the parameters depend on an unobservable underlying stochastic process. To simplify the calculations, this unobservable process is, in this note, taken to be a Markov chain.

These models are very complicated, and we need the tool of stochastic simulation to analyze them. Then the problem lies in the fact that a huge number of calculations has to be done, but the advantage is the pretty reasonable mathematics involved. Faster and faster computers make these techniques more and more practical to use.

1.2 Statistical model.

We will consider an extension of the Dynamic Linear Model (DLM). First, a normal DLM is formulated, and then we motivate the extensions which lead to the model considered in this note.

1.2.1 DLM and the Kalman filter.

An ordinary DLM may be formulated as

$$X_t = \Phi_t X_{t-1} + \omega_t \tag{1}$$

$$Y_t = H_t X_t + \epsilon_t \tag{2}$$

where

X_t is an unobservable process with dimension p
Y_t is an observable process with dimension q

(1) is called the system equation and (2) the observation equation.
$\{\omega_t\}$ and $\{\epsilon_t\}$ are white noise processes with covariance matrices

$$\begin{aligned} \text{Var } \omega_t &= Q_t \\ \text{Var } \epsilon_t &= R_t \end{aligned} \tag{3}$$

The parameters of the model are Φ_t, Q_t, H_t and R_t , $t = 1, ..., n$.
We assume

1. $\text{Corr}(\omega_t, \omega_s) = 0$; $t \neq s$
2. $\text{Corr}(\epsilon_t, \epsilon_s) = 0$; $t \neq s$
3. ω_t and ϵ_t are uncorrelated

From established theory (Harrison and Stevens (1976)) we know that the best linear estimator, $\hat{X}_{t|t}$, for X_t given $y_1, ..., y_t$ is found by the Kalman filter recursions:

$$\begin{aligned} \hat{X}_{t|t-1} &= \Phi_t \hat{X}_{t-1|t-1} \\ \hat{X}_{t|t} &= K_t Z_t + \hat{X}_{t|t-1} \end{aligned} \tag{4}$$

where

$$Z_t = Y_t - H_t \hat{X}_{t|t-1} \tag{5}$$

The so-called Kalman gain, K_t, is found by the following recursions:

$$\begin{aligned} P_t &= \Phi_t S_{t-1} \Phi_t' + Q_t &&= \text{Var } (\hat{X}_{t|t-1} - X_t) \\ f_t &= H_t P_t H_t' + R_t &&= \text{Var } Z_t \\ K_t &= P_t H_t' f_t^{-1} \\ S_t &= (I - K_t H_t) P_t &&= \text{Var } (\hat{X}_{t|t} - X_t) \end{aligned}$$

When we assume white noise processes for ω_t and ϵ_t, the best linear estimator will be equal to the expected value, i.e. $\hat{X}_{t|t} = \text{E } (X_t | y_1, ..., y_t)$. The Kalman filter can also be expanded to give the best linear estimator $\hat{X}_{t|n}$ for X_t given all data $y_1, ..., y_n$. With white noise processes we again have $\hat{X}_{t|n} = \text{E } (X_t | y_1, ..., y_n)$.

1.2.2 DLM controlled by a hidden Markov chain.

In many situations it is more reasonable to believe that the behavior of the stochastic process changes in time. It may consist of many dependent shorter processes with different parameters. We can state this as a DLM where the parameters vary in time. The variation of the parameters is here expressed by a hidden Markov chain controlling which one of many possible models the system follows at a given time. In this note we will only let the parameters controlling the system process vary, and they will be controlled by the same hidden Markov chain. But in theory all the parameters may be controlled by unequal hidden processes.

Let $\{C_t\}$ be a Markov chain with transition matrix P controlling the parameters of the system process, i.e. $\Phi_t = \Phi(C_t)$ and $Q_t = Q(C_t)$. The number of possible states of the Markov chain is finite, say K. In practical situations K will often be small, e.g. equal to 2.

The model is then given by

$$\begin{aligned} X_t &= \Phi(C_t)X_{t-1} + \omega_t & (6) \\ Y_t &= HX_t + \epsilon_t & (7) \end{aligned}$$

where

$$\begin{aligned} \text{Var}\ \omega_t &= Q(C_t) \\ \text{Var}\ \epsilon_t &= R \end{aligned}$$

and $\{C_t\}$ is an unobservable Markov chain with transition matrix P where $C_t \in \{1, 2, ..., K\}$.

If $\{C_t\}$ had been known this would just be a normal Kalman filter situation. Our purpose is to simulate the $\{C_t\}$ chain and use the Kalman filter with the simulated series plugged in, to give $\hat{X}_{t|n}$.

First, we will consider the K possible parameter values for $\Phi(C_t)$ and $Q(C_t)$ as known, and then we will consider the situation where the parameters have to be estimated.

1.3 Simulation method.

We want to simulate realizations $(c_1, ..., c_n)$ of the hidden Markov chain based on data $y_1, ..., y_n$. This can be done by sampling from the distribution of $(C_1, ..., C_n | y_1, ..., y_n)$. The number of possible outcomes is K^n. The Metropolis algorithm is well suited for an intelligent way of performing such a sampling.

In this situation we start with a random state $i = (c_{1,i}, ..., c_{n,i})$ and calculate

$$\pi_i = P\left(C_1 = c_{1,i}, C_2 = c_{2,i}, ..., C_n = c_{n,i} | y_1, ..., y_n\right) \tag{8}$$

A new state $j = (c_{1,j}, ..., c_{n,j})$ is chosen, π_j is calculated and the Metropolis algorithm tells us to move from i to j with probability $min(1, \frac{\pi_j}{\pi_i})$.

It can be shown (Hastings (1970)) that we can walk sequently through the series and change only one component at the time. Then we just deal with transitions from

$$\begin{array}{rcl} i & = & (c_1, ..., c_t, ..., c_n) \\ \text{to } j & = & (c_1, ..., c_t^{'}, ..., c_n) \end{array}$$

When we reach $t = n$ the procedure is repeated. This is done ν times. The Metropolis algorithm ensures us that when $\nu \rightarrow \infty$ we end up with a state which may be considered as sampled from the distribution of $(C_1, ..., C_n | y_1, ..., y_n)$.

2 How to perform the simulations

Our aim is to develop a simulation algorithm to discover the hidden Markov chain $\{C_t\}$. An important point is to develop an algorithm of order n. General algorithms of order n^2 are not too difficult to develop, but they will often be too slow for practical use.

We will simulate B realizations of $\{C_t\}$ and finally use the "best" combination of these as the estimated hidden Markov chain. Here we simply pick the one of the K states that has occurred most times for each t, as the final value of c_t. Many other and maybe better solutions can be made.

2.1 One realization of the hidden Markov chain.

The main point is to be able to get a realization of the hidden Markov chain, that is to sample from the distribution $f(c_1, ..., c_n | y_1, ..., y_n)$. As we saw in section 1, such a sampling can be done by using the Metropolis algorithm. We then need to find the proportion between $f(c_1, ..., c_t^{'}, ..., c_n | y_1, ..., y_n)$ and $f(c_1, ..., c_t, ..., c_n | y_1, ..., y_n)$ where $c_t^{'}$ is one of the possible values c_t may be changed to.

By using Bayes' theorem, we get

$$f(c_1, ..., c_t, ..., c_n | y_1, ..., y_n) = \frac{f(y_1, ..., y_n | c_1, ..., c_t, ..., c_n) P(c_1, ..., c_t, ..., c_n)}{f(y_1, ..., y_n)} \tag{9}$$

and then

$$\frac{f(c_1, ..., c_t^{'}, ..., c_n | y_1, ..., y_n)}{f(c_1, ..., c_t, ..., c_n | y_1, ..., y_n)} = \frac{f(y_1, ..., y_n | c_1, ..., c_t^{'}, ..., c_n)}{f(y_1, ..., y_n | c_1, ..., c_t, ..., c_n)} \frac{P(c_1, ..., c_t^{'}, ..., c_n)}{P(c_1, ..., c_t, ..., c_n)} \tag{10}$$

Notice that the complicated factor $f(y_1, ..., y_n)$ has disappeared. Because of the Markov dependency, the last part of (10) is reduced to

$$\frac{P(c_{t+1} | c_t^{'}) P(c_t^{'} | c_{t-1})}{P(c_{t+1} | c_t) P(c_t | c_{t-1})} \tag{11}$$

and if the transition matrix P is known, this part is simple.

Our main problem is then to calculate the density $f(y_1, ..., y_n | c_1, ..., c_n)$ at a given time step. We will do this by developing a recursive technique where we in a systematic way move from $t = 1$ to $t = n$ and change $c_1, .., c_n$ according to the result of the Metropolis algorithm.

To do this we have to allow one backward recursion from $t = n, ..., 1$ for each time n is reached. In this backward recursion we find $\tilde{X}_t$ which is the optimal estimator for X_t given $y_t, ..., y_n$. This is equal to E $(X_t|y_t, ..., y_n)$ since we only consider white noise. To find $\tilde{X}_t$ we will maximize $f(y_t, ..., y_n|x_t)$ as a function of x_t. In Teigland (1992) it is shown that

$$f(y_t, ..., y_n|x_t) = f(\tilde{x}_{t+1}, y_t|x_t)\ f(y_{t+1}, ..., y_n|\tilde{x}_{t+1}) \tag{12}$$

and since we consider a backwards recursion, where $\tilde{X}_{t+1}$ is known when we calculate $\tilde{X}_t$, we only have to maximize the first part of (12). Let us consider $\tilde{X}_{t+1}$ and Y_t as two independent measurements of X_t:

$$\begin{aligned} \tilde{X}_{t+1} &= \Phi_{t+1} X_t + (\omega_{t+1} + \tilde{\epsilon}_{t+1}) \qquad (13) \\ Y_t &= H X_t + \epsilon_t \qquad (14) \end{aligned}$$

where $\tilde{\epsilon}_{t+1} = \text{Var } \tilde{X}_{t+1} = \tilde{R}_{t+1}$. The variance of Y_t is R_t and the variance of $\tilde{X}_{t+1}$ is given by $\tilde{V}_{t+1} = Q_{t+1} + \tilde{R}_{t+1}$.

The estimate of X_t will then become:

$$\tilde{X}_t = \left(\Phi'_{t+1} \tilde{V}_{t+1}^{-1} \Phi_{t+1} + H' R_t^{-1} H\right)^{+} \left(\Phi'_{t+1} \tilde{V}_{t+1}^{-1} \tilde{X}_{t+1} + H' R_t^{-1} y_t\right) \tag{15}$$

and

$$\begin{aligned} \tilde{R}_t = \text{Var } \tilde{\epsilon}_t &= \text{Var } \tilde{X}_t \\ &= \left(\Phi'_{t+1} \tilde{V}_{t+1}^{-1} \Phi_{t+1} + H' R_t^{-1} H\right)^{+} \\ &\quad \left(\Phi'_{t+1} \tilde{V}_{t+1}^{-1} \tilde{R}_{t+1} \tilde{V}_{t+1}^{-1'} \Phi_{t+1} + H' R_t^{-1'} H\right) \\ &\quad \left(\Phi'_{t+1} \tilde{V}_{t+1}^{-1} \Phi_{t+1} + H' R_t^{-1} H\right)^{+'} \qquad (16) \end{aligned}$$

Some of these matrices may be singular, so we have to use general inverse matrices which we denote by the symbol "+".

To start the recursion, we have

$$\begin{aligned} \tilde{X}_n &= \left(H' R_n^{-1} H\right)^{+} \left(H' R_n^{-1} y_n\right) \qquad (17) \\ \tilde{R}_n &= \left(H' R_n^{-1} H\right)^{+} \left(H' R_n^{-1'} H\right) \left(H' R_n^{-1} H\right)^{+'} \qquad (18) \end{aligned}$$

Now we can consider the optimal estimator $\tilde{X}_t$ for X_t given data $y_t, ..., y_n$ as known when we return to the problem of calculating $f(y_1, ..., y_n|c_1, ..., c_n)$

recursively. We will drop the conditional probability in our notation, and write $f(y_1, ..., y_n|c_1, ..., c_n)$ just as $f(y_1, ..., y_n)$. We can split this density in two parts:

$$f(y_1, ..., y_n) = f(y_1, ..., y_t, \tilde{x}_{t+1}) f(y_{t+1}, ..., y_n|\tilde{x}_{t+1}) \tag{19}$$

Notice that the last part of (19) is irrelevant when we consider possible changes from c_{t+1} to c'_{t+1} in the Metropolis algorithm since $\tilde{x}_{t+1}$ is known.

The first part of (19) may be further split into:

$$f(y_1, ..., y_t, \tilde{x}_{t+1}) = f(y_1)f(y_2|y_1)....f(y_t|y_1, ..., y_{t-1})f(\tilde{x}_{t+1}|y_1, ..., y_t) \tag{20}$$

Only the last factor will depend on c_{t+1}, so what we have to calculate is the density $f(\tilde{x}_{t+1}|y_1, ..., y_t)$ for both c_{t+1} and c'_{t+1}.

At time $t+1$, $c_1, ..., c_t$ are found. All the parameters up to this time are then given, and the Kalman filter can be used recursively to get $\hat{X}_{t|t}, P_t, S_t, f_t$ and K_t.

Since $(\tilde{X}_{t+1}|y_1, ..., y_t)$ is a linear combination of normallly distributed variables, it has to be normally distributed itself, so we only have to find the expectation and variance of this variable to calculate $f(\tilde{x}_{t+1}|y_1, ..., y_t)$ explicitly.

$\tilde{X}_{t+1}$ can be thought of as an extra observation of X_{t+1} white noise $\tilde{\epsilon}_{t+1}$:

$$\begin{aligned}\tilde{X}_{t+1} &= X_{t+1} + \tilde{\epsilon}_{t+1} \\ &= \Phi_{t+1} X_t + (\omega_{t+1} + \tilde{\epsilon}_{t+1}) \end{aligned} \tag{21}$$

Here, ω_{t+1} and $\tilde{\epsilon}_{t+1}$ are independent. They both have expectation equal to zero, and their variances are known. Then we can easily find the expectation of $(\tilde{X}_{t+1}|y_1, ..., y_t)$:

$$\begin{aligned}\text{E }(\tilde{X}_{t+1}|y_1, ..., y_t) &= \text{E }\{\Phi_{t+1} X_t + (\omega_{t+1} + \tilde{\epsilon}_{t+1})|y_1, ..., y_t\} \\ &= \Phi_{t+1} \text{ E }(X_t|y_1, ..., y_t) + 0 \\ &= \Phi_{t+1} \hat{X}_{t|t} \end{aligned} \tag{22}$$

Let us define

$$\text{Var }\left(\tilde{X}_{t+1}|y_1, ..., y_t\right) = \tilde{f}_{t+1} \tag{23}$$

Now, for time $t+1$, we have a system which can be considered as a normal DLM with system and observation equations:

$$\begin{aligned} X_{t+1} &= \Phi_{t+1} X_t + \omega_{t+1} &(24)\\ \text{``}Y_{t+1}\text{''} &= \tilde{X}_{t+1} = X_{t+1} + \tilde{\epsilon}_{t+1} &(25)\end{aligned}$$

By using the Kalman filter one step on this system, we get the value

$$\tilde{P}_{t+1} = \Phi_{t+1} S_t \Phi'_{t+1} + Q_{t+1} \tag{26}$$

and then

$$\tilde{f}_{t+1} = \tilde{H}_{t+1} \tilde{P}_{t+1} \tilde{H}'_{t+1} + \tilde{R}_{t+1} \tag{27}$$

where $\tilde{H}_{t+1}$ is equal to the identity matrix. By combining (26) and (27) we get the result

$$\tilde{f}_{t+1} = \tilde{P}_{t+1} + \tilde{R}_{t+1} = \Phi_{t+1} S_t \Phi'_{t+1} + Q_{t+1} + \tilde{R}_{t+1} \tag{28}$$

The procedure can be summarized as follows:
The expectation (22) and the variance (28) are calculated for both c_{t+1} and c'_{t+1}, that is with different Φ_{t+1} and Q_{t+1}. Then $f(\tilde{x}_{t+1}|y_1,...,y_t)$ is found for both. Since all other terms are equal, the proportion between $f(y_1,...,y_n|c_1,...,c'_{t+1},...,c_n)$ and $f(y_1,...,y_n|c_1,...,c_{t+1},...,c_n)$ is easily calculated, and (10) is used to decide if c_{t+1} will be changed to c'_{t+1} or remain unchanged. Finally, the Kalman filter is used one further step to give $f_{t+1}, K_{t+1}, S_{t+1}$ and $\hat{X}_{t+1|t+1}$ which are needed in the next step of the algorithm.

When $t = n$ is reached, we perform a new backwards recursion to give new $\tilde{X}_t$ and $\tilde{R}_t$, $t = n,...,1$, before we again move from $t = 1,...,n$ to decide new values $c_1,...,c_n$. This is repeated ν times, and when $\nu \to \infty$, we have a sample $c_1,...,c_n$ from the distribution of $(C_1,...,C_n|y_1,...,y_n)$. In practical situations it is impossible to let $\nu \to \infty$, but in many models we seem to get acceptable results with about 100 repetitions.

3 Parameter estimation

In many of the situations considered in this note, the parameters of the different models will be fixed. The problem is to discriminate between possible models with given parameters. However, we may think of situations where the parameters are not explicitly known. As an example consider situations with two possible models, one with little noise and one with much noise, but where the actual values of the variances are unknown. Then we can try to estimate these values.

It seems to be an extremely difficult task to implement parameter estimation in our stochastic simulation method. The updating of the Markov chain is based on known parameters. Therefore, we need a recursive technique where we from initial values for both the parameters and $c_1, ..., c_n$ first get a new Markov chain. From these simulated values the estimates of the parameters are updated before we simulate a new chain. In this way we finally get a vector $(c_1, ..., c_n)$ sampled from the distribution $f_\theta(c_1, ..., c_n | y_1, ..., y_n)$, where $\theta = (\Phi_1, ..., \Phi_K, Q_1, ..., Q_K, H, R)$, at the same time as well, the parameter estimates are updated to increase the likelihood at each step.

We repeat the whole procedure B times, and will then have B simulated $\{c_t\}$-s and B estimates of the parameter values. The final estimates are then naturally given as the mean of these B values.

To solve the problem of updating the estimates of the parameters, we use the EM-algorithm. In this note we only give an outline of a solution. Details are given in Teigland (1992). The results look promising when we don't have to estimate too many of the parameters at the same time.

3.1 The EM-algorithm

Let θ be the parameter vector, Y the vector of observations with distribution $f_\theta(y)$, and X the unobservable vector with distribution $P_\theta(x)$. X and Y are stochastic dependent and $f_\theta(y|x)$ is known. To follow the procedure above, we need an expression for $\frac{\partial \log f_\theta(y)}{\partial \theta}$ based on the distribution of X given $Y = y$ to get maximum likelihood estimators at each step with the $c_1, ..., c_n$ given. If we let E^* denote the expectation of X given $Y = y$, the following result can be shown:

$$\frac{\partial \log f_\theta(y)}{\partial \theta} = \mathrm{E}^* \left\{ \frac{\partial \log f_\theta(y|X)}{\partial \theta} \right\} + \mathrm{E}^* \left\{ \frac{\partial \log P_\theta(X)}{\partial \theta} \right\} \tag{29}$$

In our situation the parameters of interest will occur either in the distribution of X or in the relation between X an Y, but not in both. This means that one of the parts of (29) will equal 0.

3.2 Estimation of the parameters in the model

By combining the result in (29) and the Kalman filter, it is shown in Teigland(1992) that we get the following formulae for updating our parameters:

$$\hat{H} = \left[\sum_{t=1}^{n} y_t \hat{X}'_{t|n}\right] \left[\sum_{t=1}^{n}\{\hat{X}_{t|n}\hat{X}'_{t|n} + P_{t|n}\}\right]^{-1} \tag{30}$$

$$\hat{R} = \frac{1}{n}\sum_{t=1}^{n}\left\{\ (y_t - H\hat{X}_{t|n})(y_t - H\hat{X}_{t|n})' + HP_{t|n}H'\ \right\} \tag{31}$$

$$\hat{\Phi}_k = \left[\sum_{t;c_t=k}(\hat{X}_{t-1|n}\hat{X}'_{t|n} + L_{t-1}P_{t|n})\right] \left[\sum_{t;c_t=k}(\hat{X}_{t-1|n}\hat{X}'_{t-1|n} + P_{t-1|n})\right]^{-1} \tag{32}$$

$$\begin{aligned}\hat{Q}_k \quad = \quad & \frac{1}{(\text{number of } c'_t s = k)} \sum_{t;c_t=k}\{\ (\hat{X}_{t|n}\hat{X}'_{t|n} + P_{t|n}) \\ & -\Phi_t(\hat{X}_{t-1|n}\hat{X}'_{t|n} + L_{t-1}P_{t|n}) \\ & -[\Phi_t(\hat{X}_{t-1|n}\hat{X}'_{t|n} + L_{t-1}P_{t|n})]' \\ & +\Phi_t(\hat{X}_{t-1|n}\hat{X}'_{t-1|n} + P_{t-1|n})\Phi'_t\ \}\end{aligned} \tag{33}$$

4 Examples

In this section some examples will be presented to illustrate how the stochastic simulation method presented, performs in various situations. Five examples are presented, and in each example data is simulated by using the corresponding dynamic linear model. This provides us with knowledge of the true underlying Markov process, and thus the evaluation of the results from the estimation becomes easy.

The parameters in the DLM are assumed known in four of the examples. The fifth example contains estimation of parameters in two of the preceding examples. In all examples 100 simulations of the Markov chain are calculated (B=100), and each realization is a result of 100 repetitions in the Metropolis algorithm (ν=100).

4.1 Example 1: A simple model

The first example is a simple model with the observable variable, Y_t, equal to the unobservable, X_t, with noise added. The variance of ϵ_t is $R_t = 0.05$. The DLM can be written:

$$\begin{aligned} X_t &= \Phi_t X_{t-1} + \omega_t \\ Y_t &= X_t + \epsilon_t \end{aligned}$$

The Markov process consists of two possible states. Both Φ_t and the variance, W_t, in the system model depends upon the state of the Markov chain:

$$\begin{array}{ll} \Phi_t(1) = 0.8 & W_t(1) = 0.01 \\ \Phi_t(2) = 1.2 & W_t(2) = 0.5 \end{array}$$

The probability of transitions between the two different states is set to 0.1 and thus the probability of remaining in the same state is 0.9:

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

In figure 1 the observed variable is plotted together with the true and estimated Markov processes. The estimated Markov process is equal to the true at all points in time with one exception, the starting point. It is typical for the model that the endpoints cause problems. The algorithm for deciding C_t depends upon previous estimates for C_{t-1} and C_{t+1}, and these do not both exist at the endpoints.

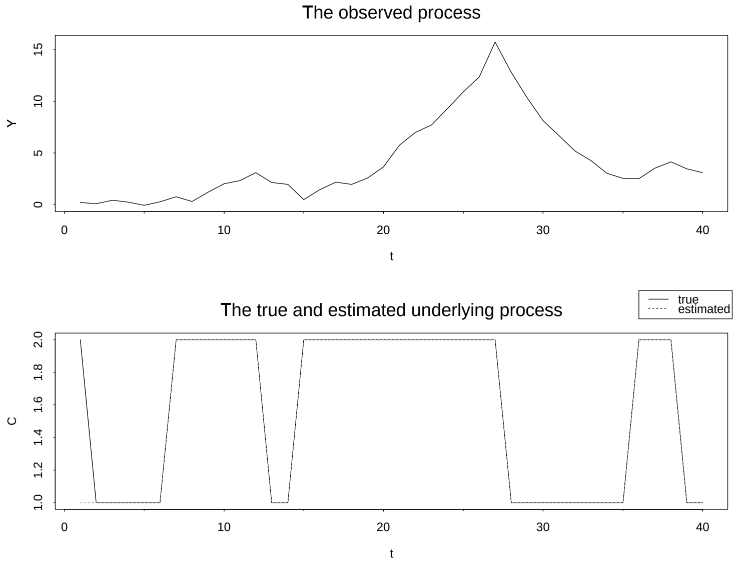

Figure 1: The upper plot shows the observed data Y_t, and the lower plot shows the true and estimated Markov processes for the model in example 1. The estimated Markov process is equal to the true at all points in time with one exception, the starting point.

4.2 Example 2: An extension of the simple model

To see if the number of states in the Markov chain has an influence on the ability of discrimination between states, example 1 is extended to consist of a Markov chain with 3 possible states. Φ_t and W_t now have a third possible value:

$$\begin{array}{ll} \Phi_t(1) = 0.8 & W_t(1) = 0.01 \\ \Phi_t(2) = 1.2 & W_t(2) = 0.5 \\ \Phi_t(3) = 2.0 & W_t(3) = 0.2 \end{array}$$

The probability of transition between different states is still 0.1 and thus the probability of remaining in the same state is 0.8:

$$P = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

The simulation method performs well also with three possible states. There is one point in time where the estimated process differs from the true process: At time 8 the estimated chain jumps up to class 3, one point in time too late. The observed process and the true and estimated Markov processes are shown in figure 2.

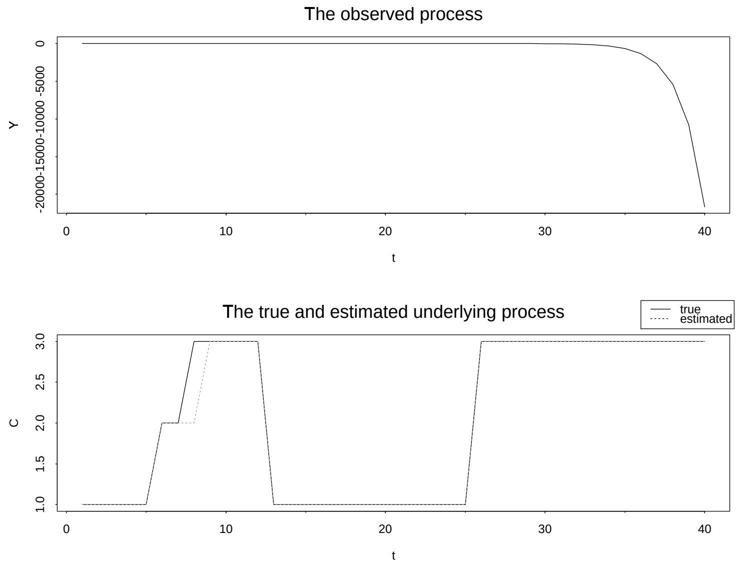

Figure 2: The upper plot shows the observed data Y_t, and the lower plot shows the true and estimated Markov processes for the model in example 2. There is one point in time where the estimated process differs from the true process.

4.3 Example 3: Difference in variance

In the DLM model in this example, only the variance term depends upon the underlying model. In this case the model discriminates between the two different possible states by using only the variance. As in the previous examples, the observable variable equals the unobservable with noise added. The variance of ϵ_t is 0.05 and $\Phi_t = 1$ for all states. The DLM can be written:

$$\begin{aligned} X_t &= X_{t-1} + \omega_t \\ Y_t &= X_t + \epsilon_t \end{aligned}$$

The two levels of the variance of ω_t are:

$$\begin{aligned} W_t(1) &= 0.01 \\ W_t(2) &= 1 \end{aligned}$$

The transition between states is assumed to have probability of 0.05, and thus the probability of remaining in the same state is 0.95:

$$P = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$$

The stochastic simulation method discriminates the two states well. The only difficulties are that the estimated process jumps too late at two of the jumps in the Markov process. However, this looks reasonable from the variance in the data set and the conservatism in the transition matrix. See figure 3 for the observed data and the true and estimated Markov chains.

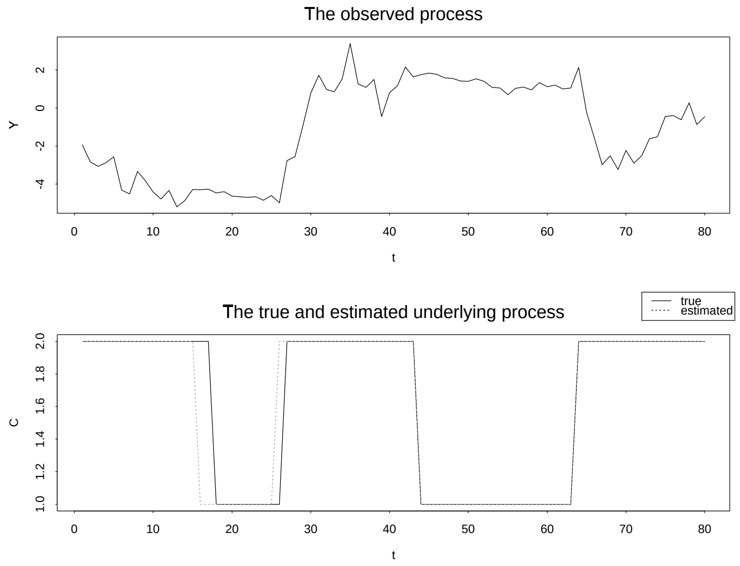

Figure 3: The upper plot shows the observed data Y_t, and the lower plot shows the true and estimated Markov processes for the model in example 3.

4.4 Example 4: AR(3)

This example is an autoregressive process of order 3, AR(3). The AR(3) is assumed to have coefficients which depends on an underlying process with two possible states. This can be written as a DLM as follows:

$$\begin{aligned} X_t &= \Phi_t X_{t-1} + \omega_t \\ Y_t &= X_t \end{aligned}$$

and the coefficients:

$$\Phi_t(1) = \begin{bmatrix} 1 & 0.01 & 0.01 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\Phi_t(2) = \begin{bmatrix} 0.2 & 0.2 & 0.2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Notice that there is no noise in the observation equation, $R_t = 0$. The noise term in the system equation is set to depend on the underlying process:

$$W_t(1) = \begin{bmatrix} 1 \\ 0.000001 \\ 0.000001 \end{bmatrix}$$

$$W_t(2) = \begin{bmatrix} 0.01 \\ 0.000001 \\ 0.000001 \end{bmatrix}$$

The transition between states is assumed to have a probability of 0.05, and thus it is a probability of 0.95 to remain in the same state:

$$P = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$$

The stochastic simulation method performs well also for this difficult model. The results are plotted in figure 4. The method has some small problems at the beginning of the Markov chain. These problems are due to the fact that the data values lie around 0. The estimated process also jumps one point in time too late at the last jump of the Markov process. But the overall impression of the performance of the method is very good.

4.5 Example 5: Estimation

Often some of the parameters in the DLM are unknown. Here, two examples of the stochastic simulation method with estimation are presented. These are example 1 with unknown Φ and example 3 with unknown W.

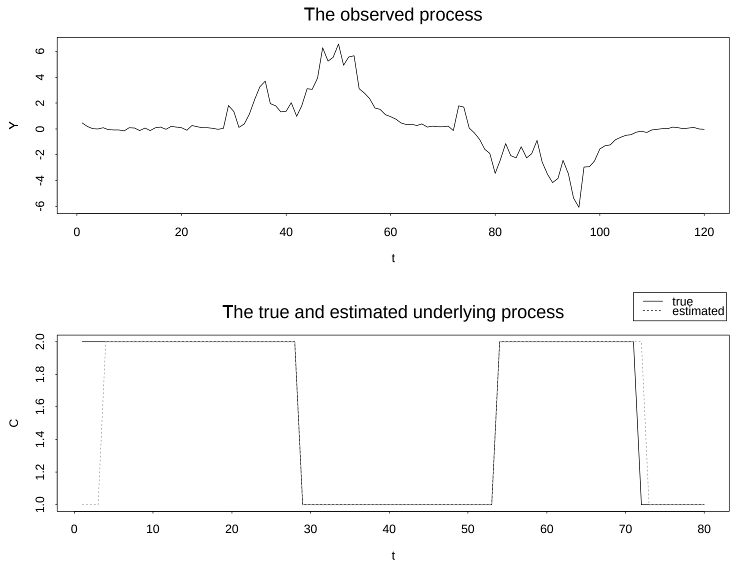

Figure 4: The upper plot shows the observed data Y_t, and the lower plot shows the true and estimated Markov processes for the model in example 4.

Example 1 The simulation method for the same data set as in example 1 is run. The only difference is that Φ_t is assumed to be unknown with initial values

$$\Phi_t(1) = 1.0$$
$$\Phi_t(2) = 2.0$$

From example 1 we have that the true value of Φ_t is 0.8 and 1.2 respectively. The estimation method performs well in this situation. The estimated Markov process is equal to the result in example 1, and the parameter is estimated to

$$\hat{\Phi}_t(1) = 0.803$$
$$\hat{\Phi}_t(2) = 1.196$$

which is close to the true values.

Example 3 Here the simulation method is run on the same data set as in example 3. In example 3 only the variance in the system equation depended on the states in the Markov process. Now, this variance dependence is assumed unknown, thus W_t has to be estimated. The starting values used are:

$$W(1) = 1.0$$
$$W(2) = 3.0$$

The true values are given in example 3 and are 0.01 and 1.0 respectively.

Also in this situation the estimation method performs well. The results of the estimation are:

$$\hat{W}_t(1) = 0.015$$
$$\hat{W}_t(2) = 1.038$$

Booth estimates agree well with the true values. Further the estimation of the Markov chain is equally good as in example 3, where all parameters are known.

References

Harrison, P.J. and Stevens, C.F (1976): "Bayesian Forecasting." *Journal of Royal Statistical Society, Series B, 38.*

Hastings, W.K. (1970): "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika 57.*

Teigland, A. (1992): "Blandingsprosesser i dynamisk lineære modeller. Noen bidrag til statistisk metodikk ved bruk av stokastisk simulering." *Cand.scient thesis, University of Oslo* (in Norwegian).