

# Advanced Time Series Analysis: Computer Exercise 1

*Anders Launer Bæk (s160159)*

*10 Oktober 2017*

Sparring partners:

- Anja Liljedahl Christensen (s162876)
- Marie Mørk (s112770)

## Part 1

There are simulated  $n = 1000$  where  $\epsilon_t \sim \mathcal{N}(0, 1)$ .  $\epsilon_t$  is used as noise input for all simulations in part one.

The equations below are the used parameters though out this exercise. Let us call eq. 1 and eq. 2 parameter set one ( $par_1$ ), eq. 3 and eq. 4 parameter set two ( $par_2$ ).

$$a_0 = [0.125, -0.125] \quad (1)$$

$$a_1 = [0.6, -0.4] \quad (2)$$

$$a_{02} = [0.25, -0.25] \quad (3)$$

$$a_{12} = [-0.6, 0.8] \quad (4)$$

## SETAR(2,1,1)

The Self-Exciting Threshold AR (SETAR) model is given by eq. 5.

$$X_t = a_0^{(J_t)} + \sum_{i=1}^{k(J_t)} a_i^{(J_t)} X_{t-i} + \epsilon^{(J_t)} \quad (5)$$

where  $J_t$  are regime processes. The complete model are defined in eq. 6.

$$X_t = \begin{cases} a_{0,1} + a_{1,1}X_{t-1} + \epsilon_t & \text{for } X_{t-1} \leq 0 \\ a_{0,2} + a_{1,2}X_{t-1} + \epsilon_t & \text{for } X_{t-1} > 0 \end{cases} \quad (6)$$

The model  $X_t$  (eq. 6) has been simulated with two different set of parameters (eq. 1 - eq. 4) and its simulations are plotted in fig. 1.

Fig. 1 shows the plot of the SETAR(2,1,1) model with the two different parameter sets.

- For both model it is possible to differentiate between the regimes and their transitions.
- It is also possible to see the inverse properties of the slop for the two models.
- Both models are using different offsets where the transition are most separated in the model which is using  $par_2$ .

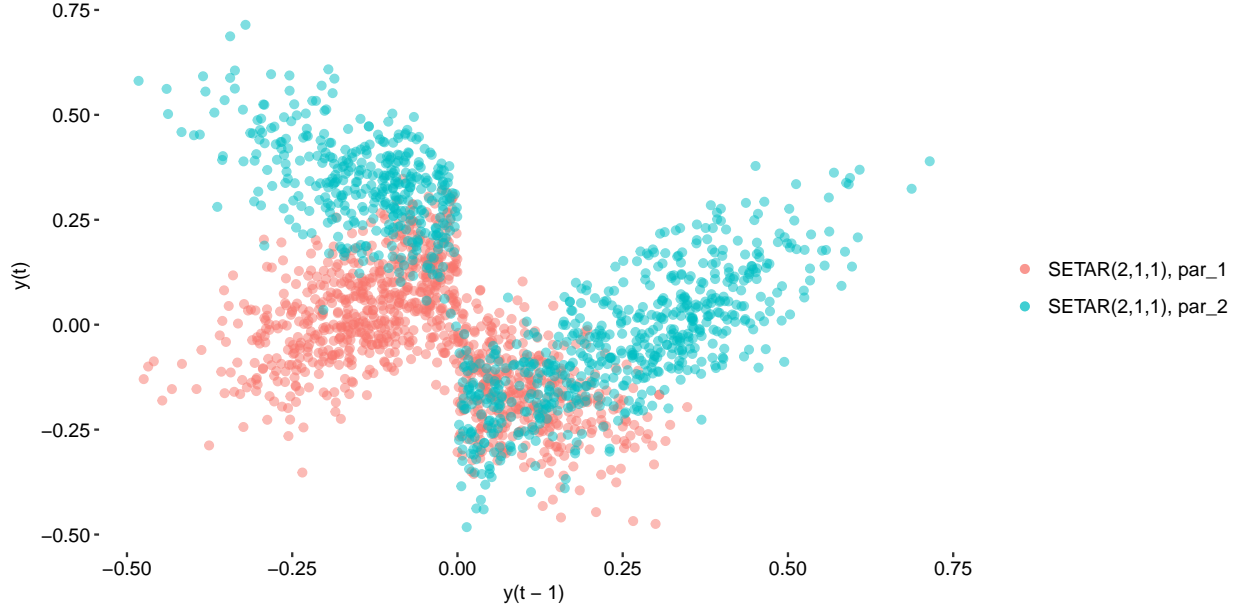


Figure 1: Two simulated SETAR(2,1,1) models using  $par_1$  and  $par_2$ .

### IGAR(2,1)

The IGAR model are given by the same equation as the SETAR model (eq. 5) but using an external parameter to switch between regimes. The external shift parameter is in this case a random variable  $p_t \sim \mathcal{U}(0, 1)$ .

The complete simulated IGAR model is given in eq. 7.

$$X_t = \begin{cases} a_{0,1} + a_{1,1}X_{t-1} + \epsilon_t & \text{for } p \leq 0.5 \\ a_{0,2} + a_{1,2}X_{t-1} + \epsilon_t & \text{for } p > 0.5 \end{cases} \quad (7)$$

Fig. 2 shows the plot of the IGAR(2,1) model with the two different parameter sets.

- The IGAR model using a given external parameter to switch between regimes which is different from the SETAR model.
- The shift threshold is  $p \leq 0.5$  or  $p > 0.5$  (eq. 7) which supports the distribution of the data points in fig. 2. The data point are more less equally distributed in both regimes for both IGAR models.

### MMAR(2,1)

The simulated MMAR model has same properties as the IGAR model in eq. 7. The main difference are the properties of the transition parameter  $p$ . The transition parameters between regimes are given by the transition matrix  $P$  in eq. 8.

$$P = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix} \begin{matrix} \in R_1 \\ \in R_2 \end{matrix} \quad (8)$$

Fig. 3 shows the plot of the MMAR(2,1) model with the two different parameter sets.

- $P$  is useful for setting different thresholds for shifting between regimes.

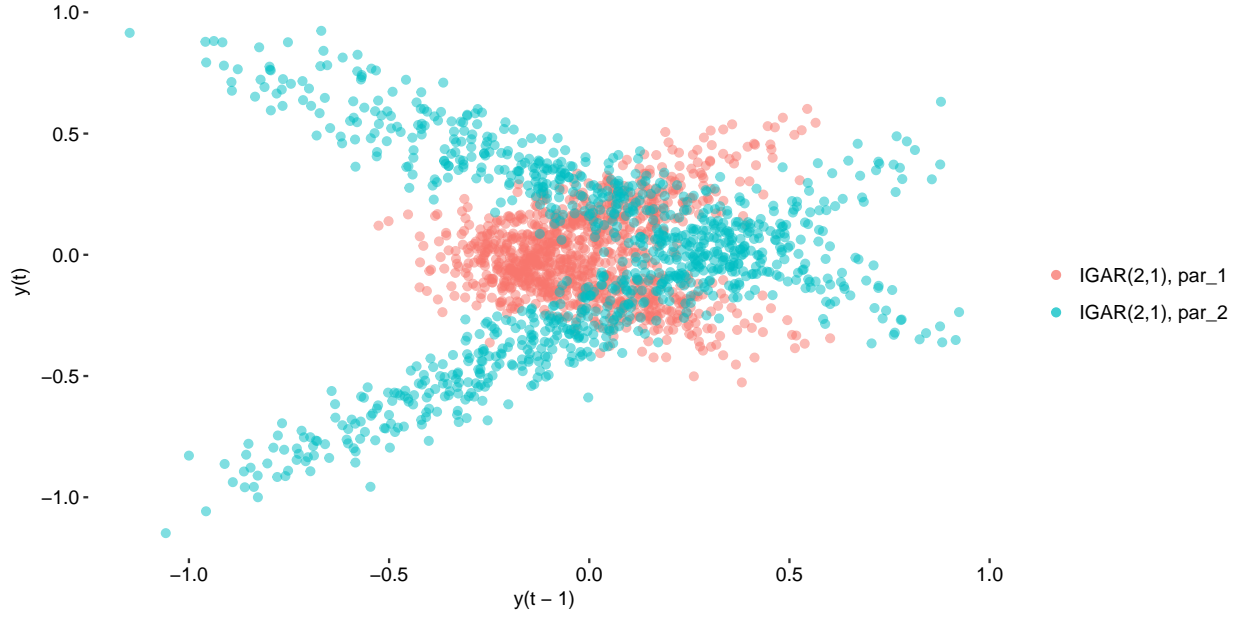


Figure 2: Two simulated IGAR(2,1) models using  $par_1$  and  $par_2$ .

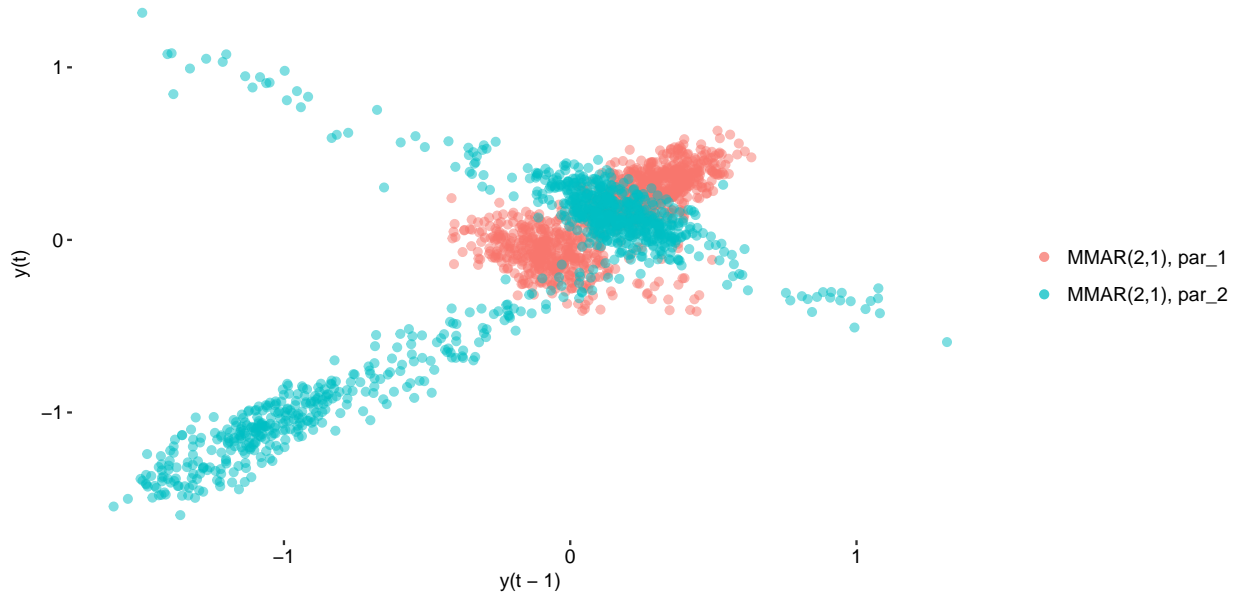


Figure 3: Two simulated MMAR(2,1) models using  $par_1$  and  $par_2$ .

- The diagonal in  $P$  is, in this case, determine external parameter for being this the current regime. The off-diagonal are external parameter for shifting to a new regime.
- The same external parameter  $p$  is used as input.
- It is possible to see a larger distribution in the "lines" compared to the IGAR(2,1) model. This is due to the transition matrix and because the model will be in the same regime for longer periods.

### **Common for above models**

- The main difference between the three models are the properties for shifting to a new regime. The shift in the SETAR model depends on the previous value of the model. And the shifting in the IGAR model and in the MMAR model is activated by an external parameter.
- The main difference between the IGAR model and the MMAR model is that it is possible to determine different thresholds for shifting between different regimes. Hereby it is possible to model "logic" transition between stages.

## Part 2

Using the same SETAR model with  $par_1$  from part 1, eq. 6.

### Compute the theoretical mean

The theoretical mean, is given by eq. 9.

$$M(x) = E \{X_{t+1}|X_t = x\} \quad (9)$$

By the fact that the noise are Gaussian distributed, then  $\epsilon_t = 0$  must be true and it is possible to rewrite the SETAR(2,1,1) model (eq. 6) to the theoretical mean in eq. 10.

$$M_t = \begin{cases} a_{0,1} + a_{1,1}X_{t-1} & \text{for } X_{t-1} \leq 0 \\ a_{0,2} + a_{1,2}X_{t-1} & \text{for } X_{t-1} > 0 \end{cases} \quad (10)$$

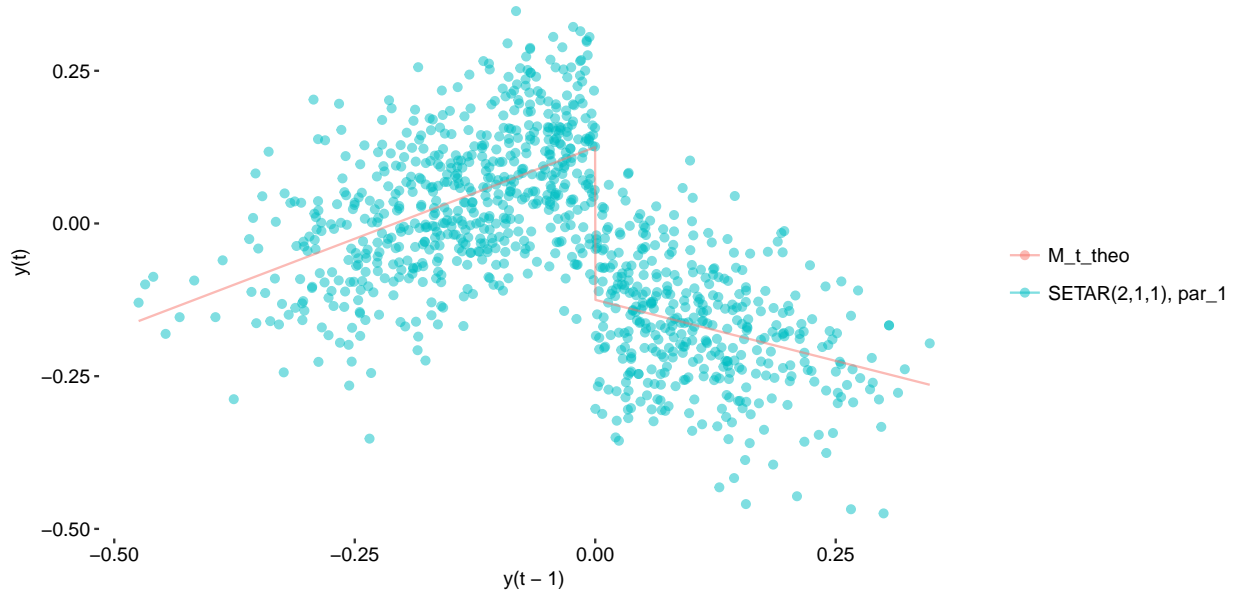


Figure 4: SETAR(2,1,1) with theoretical mean.

Fig. 4 shows the plot of the theoretical mean ( $M(x)$ ) for the parameters set one. The plot looks as expected and I do not have any further comments.

### Estimate the mean

I have chosen to use the function `lm()` to estimate the mean of the SETAR(2,1,1) model with the two selected bandwidths  $bw_n = (0.2, 0.7)$ . `lm()` is set to use a local second order polynomial regression.

The `lm()` uses the the weights from the Epanechnikov kernel (function from sample code) to do the local estimate for the given bandwidth.

Fig. 5 shows the plot of the estimated means with two different bandwidths.

- The conceptual interpretation of the bandwidth is a measure for how many samples which should be used in the local fit of the second order polynomial.

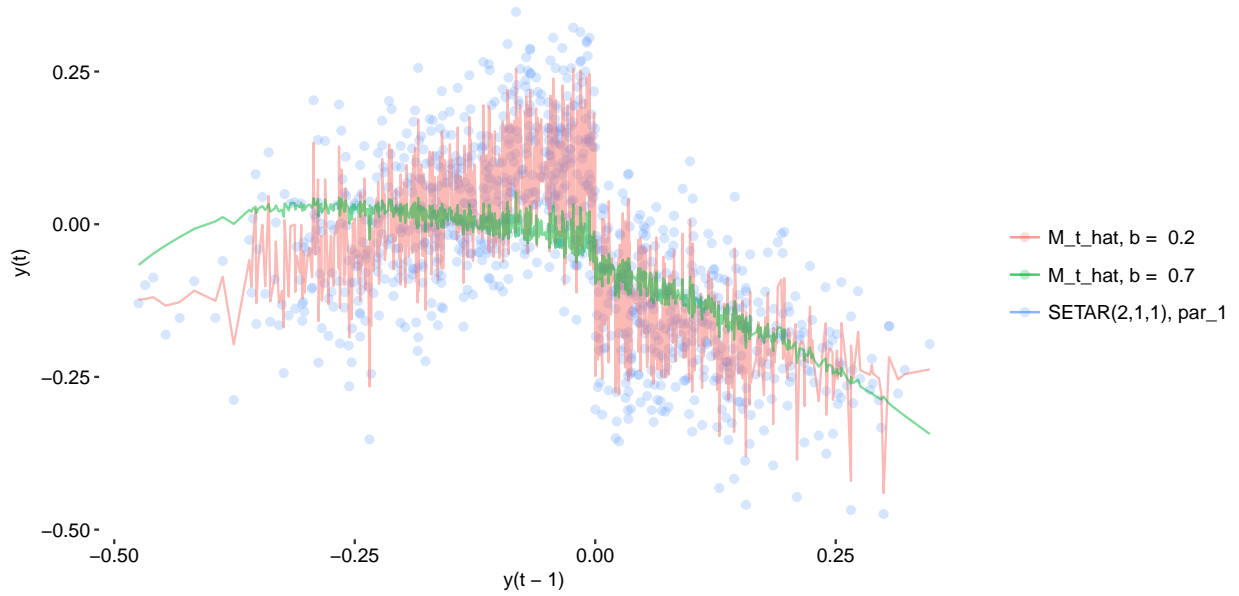


Figure 5: Plots of the estimated mean with different bandwidths.

- A lower bandwidth will increase the variance but decrease the bias.
- A higher bandwidth will decrease the variance but increase the bias.
- The best selection of the bandwidth can be found by using cross validation<sup>1</sup> and evaluate the residuals with respect to the nature of the problem.
- If boundary estimation is essential for the problem then a lower bandwidth will perform best.

---

<sup>1</sup>Sec. 2.3.6 Modelling Non-Linear and Non-Stationary Time Series

### Part 3

The `cumulativeMeans.R` script have been used to calculate the cumulative conditional mean.

The estimated cumulative conditional mean is based on  $X_t$  from eq. 6 and the theoretical cumulative conditional mean is based on  $M_t$  from eq. 10.

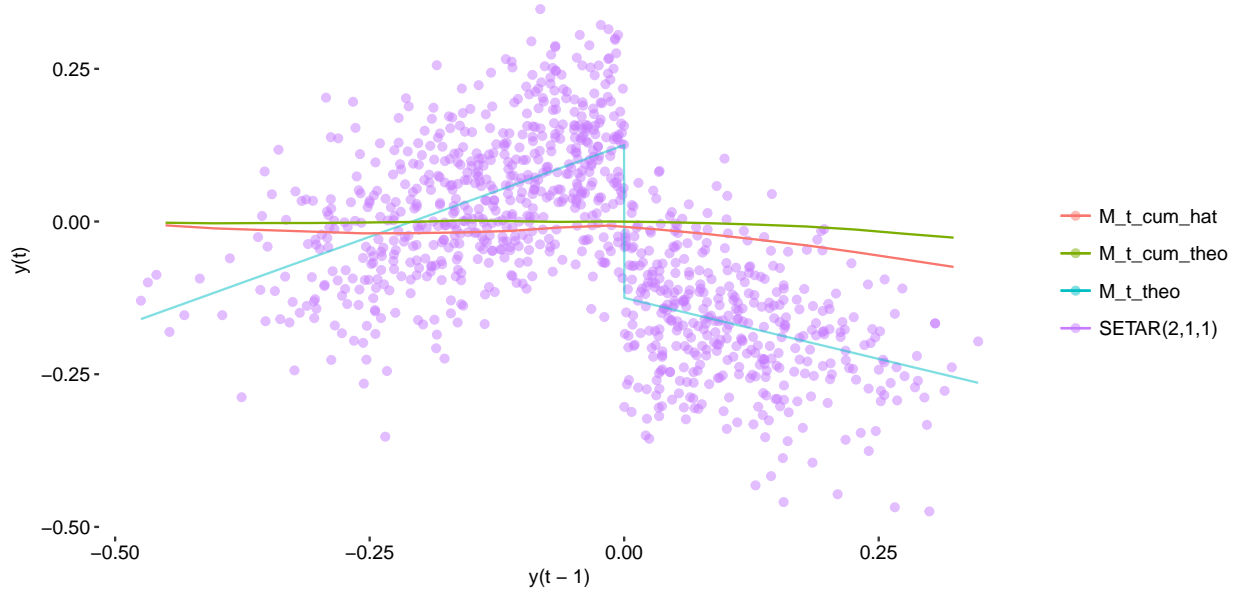


Figure 6: Plot of the theoretical cumulative conditional mean, estimated cumulative conditional mean, the theoretical mean and the SETAR(2,1,1) model.

It has been chosen to use the same number of bins in for the theoretical cumulative conditional mean and the estimated cumulative conditional mean.

Fig. 6 shows the plot of the theoretical cumulative conditional mean, the estimated cumulative conditional mean, the theoretical mean and the SETAR(2,1,1) model.

- Due the data it was possible to separate the data into 17 bins in order to satisfy the minimum of at least five observations in each bin.
- Common for both of the cumulative conditional means is the difficulties in the adoption to the regime shift.
- Despite using the same number of bins in each of the cumulative conditional means, is it possible to see the different breakpoints/bin widths. This must be due to the distribution of the data. The cumulative conditional mean is sensitive to the change in bandwidth (bin width)<sup>2</sup>.

---

<sup>2</sup>P. 70 Modelling Non-Linear and Non-Stationary Time Series

## Part 4

The conditional parametric model we wish to identify is given in eq. 11.

$$Y_t = \mu + g(X_{t-1})Y_{t-1} + \epsilon_t \quad (11)$$

where  $X_t$  is the input and  $Y_t$  is the output. The input is given by  $X_t \sim \mathcal{U}(0.01, 0.99)$ ,  $\mu = 0$  and  $g(x)$  is defined in eq. 12.

$$g(x) = 4x(1 - x) \quad (12)$$

### Local regression approach

There has been simulated a process of eq. 11 and I would like to find the dependence of  $Y_t$  on  $X_{t-1}$  and  $Y_{t-1}$ . The dependences will be discovered using local regression and a contour plot.

I have chosen to use the `lm()` with the Epanechnikov kernel with the bandwidth is chosen to 0.15. I assume there is a local linear relation between  $X_{t-1}$  and  $Y_{t-1}$ , therefore the first order regression: `lm(Y_t ~ X_{t-1} + Y_{t-1}, weights=w[ok], data = data[ok, ])`. There will only be considered samples which have weights greater than zero, as in the example code.

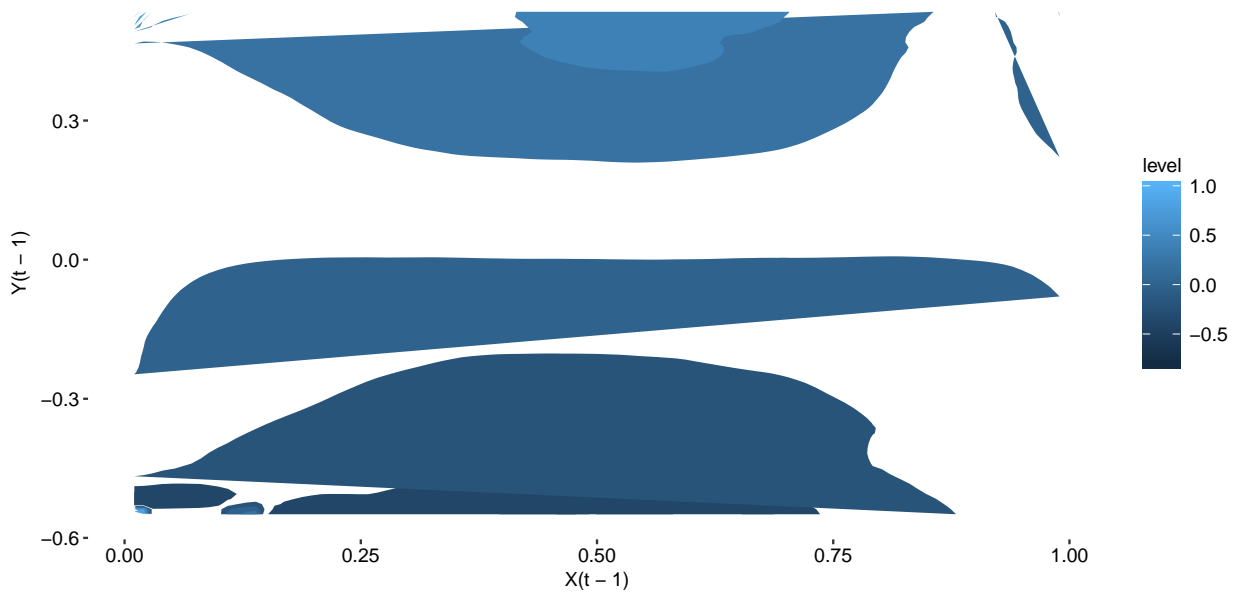


Figure 7: Contour plot of the non-parametric model approach.

Instead of the 3D visualisation I have chosen to visualise  $Y_t$  dependence on  $X_{t-1}$  and  $Y_{t-1}$  in a contour plot. Fig. 7

- The interesting part of the contour plot is for  $X_{t-1}$  where  $Y_{t-1} = [-0.3, 0]$  illustrates a slightly skewed concave function for negative values of  $Y_{t-1}$ .
- The contours illustrates how  $Y_t$  dependence on  $X_{t-1}$  and  $Y_{t-1}$  and illustrates how the external input  $X_{t-1}$  and the previous state  $Y_{t-1}$  determines the new state  $Y_t$ . There is an affine relation between  $X_{t-1}$  and  $Y_{t-1}$  in the interested region as illustrated in the plot.



## Conditional parametric model approach

I have used the following function call in the conditional parametric approach: `loess(Y_t ~ X_{t-1} + Y_{t-1}, span = bw, degree = 1, data = data, parametric = "Y_t")`.

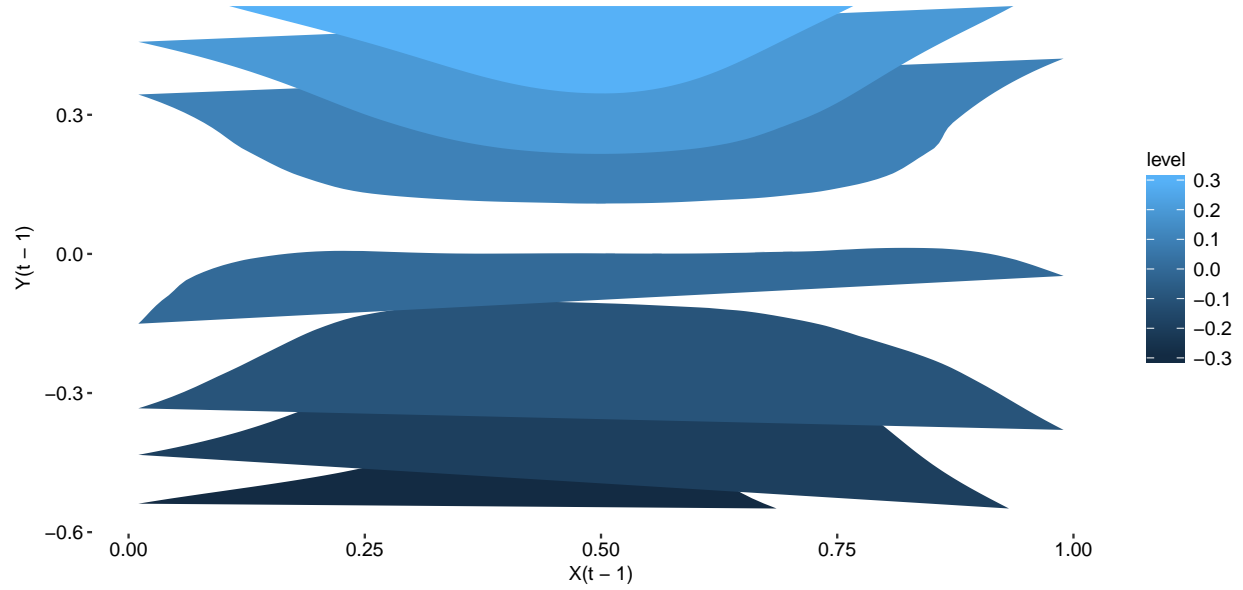


Figure 8: Contour plot of the conditional parametric model approach.

Fig. 8 shows the contour plot of the conditional parametric model approach.

- The contour plot shows the same concave contour in the same region of the plot.
- There are more contours when using the conditional parametric model approach. The contour forms, outside the region of interest, are more like Gaussian, which I assume is caused by the "least-squares" fitting property of the 'loess()' function.

## Part 5

Native function `acf()` is the tool for identification of the model order and measures the degree of linear dependency in the time series.

I have tried to apply the `acf()` on the SETAR(2,1,1) model from part one. There were some degree of linear dependency measured.

I will only focus at the first 10 lags in order to reduce computation time in the `ldf()` function.

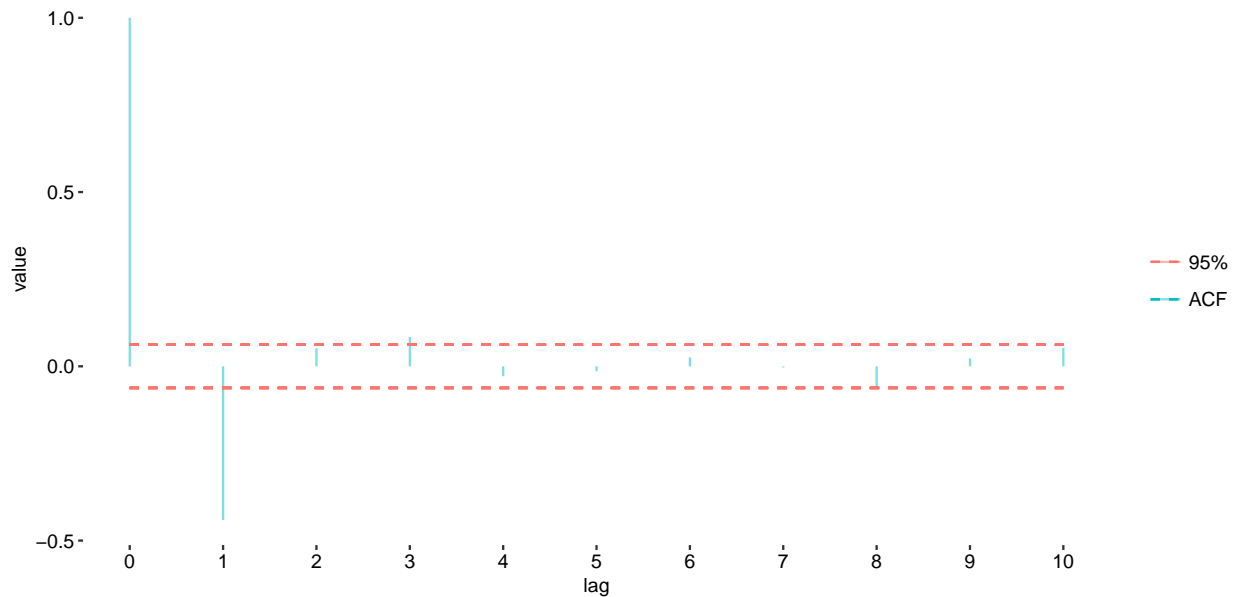


Figure 9: ACF of the SETAR(2,1,1) model from part one.

Fig. 9 shows the ACF of the model  $Y_t$  from eq. 11.

- The ACF from the SETAR(2,1,1) model returns a negative large spike in lag one. This indicates a  $MA(1)$  process.
- The properties of the SETAR model are not linear and therefore is the ACF not appropriate for identification of the lag dependencies of the model<sup>3</sup>.

Fig. 10 shows the LDF of the SETAR(2,1,1) model (eq. 11).

- The LDF return 2 positive lags which are above the 0.95% confidence interval. This indicates an  $AR(2)$  process.
- The absolute value of lag one from the LDF indicates a strong associated correlated than in the lag one from the ACF.

---

<sup>3</sup>Sec. 3.3 Modelling Non-Linear and Non-Stationary Time Series

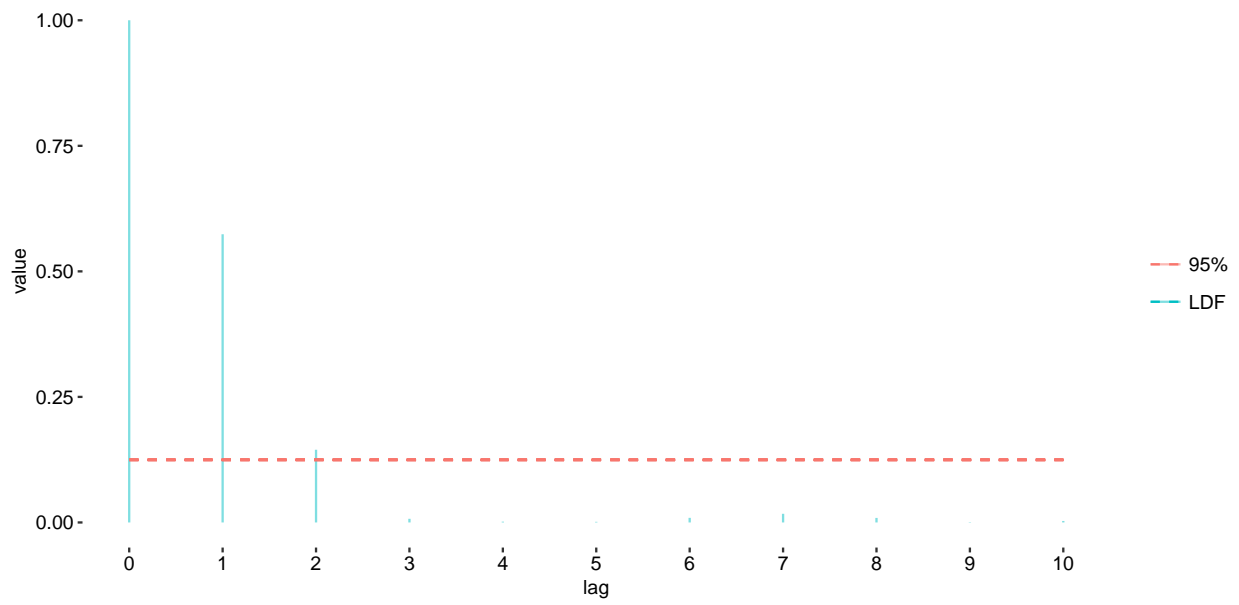


Figure 10: LDF of the SETAR(2,1,1) model from part one.