Assignment 3: ARMA and seasonal Processes, and ARX transfer function models

Instructions: The assignment report is to be handed in via DTU Learn "FeedbackFruits" latest at April 21st at 23:59. You are allowed to hand in in groups of 1 to 4 persons. You must hand in a single pdf file presenting the results using text, math, tables and plots, do not include code in the report! Shortened software result output is ok to include, to save some time formatting in tables. Arrange the report in sections and subsections according to the questions in this document. Please indicate your student numbers on the report.

NOTE, that this time there is no peer-review. The teachers evaluate the reports and the assessment will count for the grade.

NOTE, that the report should not be too long! Include only one (max two figures) (i.e. one figure can have multiple plots) per question and make the text concise! Long and unprecise reports are not good!

All the additional material needed is provided in the `assignment3_2025.zip` file.

1 Stability

Continue the first part of Assignment 2.

Let the process $\{X_t\}$ be an AR(2) given by

$$X_t + \phi_1 X_{t-1} + \phi_2 X_{t-2} = \epsilon_t$$

where $\{\epsilon_t\}$ is a white noise process with $\sigma_\varepsilon = 1$.

Answer the following:

1.1. Simulate 5 realizations of the process up to $n = 200$ observations with the coefficient values $\phi_1 = -0.6$ and $\phi_2 = 0.5$. Plot them in one plot (remember to set a seed when you do simulations and remember the sign of the AR coefficients is flipped in some implementations).

1.2. Calculate the empirical ACF of the simulations and plot them together with $\rho(k)$, up to lag 30. Comment on the result.

1.3. For the next questions redo plots of simulations and ACFs with different values of the coefficients. For each comment shortly on the outcome, especially focus on stationarity and compare the results across the simulations. Use now $\phi_1 = -0.6$ and $\phi_2 = -0.3$.

1.4. Now use $\phi_1 = 0.6$ and $\phi_2 = -0.3$.

1.5. Now use $\phi_1 = -0.7$ and $\phi_2 = -0.3$

1.6. $\phi_1 = -0.75$ and $\phi_2 = -0.3$

1.7. Would you recommend always plotting the time series data or does it provide sufficient information just to examine the ACF?

2 Predicting monthly solar power

In renewable energy systems forecasting is really important to plan the operation of systems. The forecasting horizon can both be short, but also longer, as in the case below, where the next twelve months electricity generation on a solar PV plant is to be predicted.

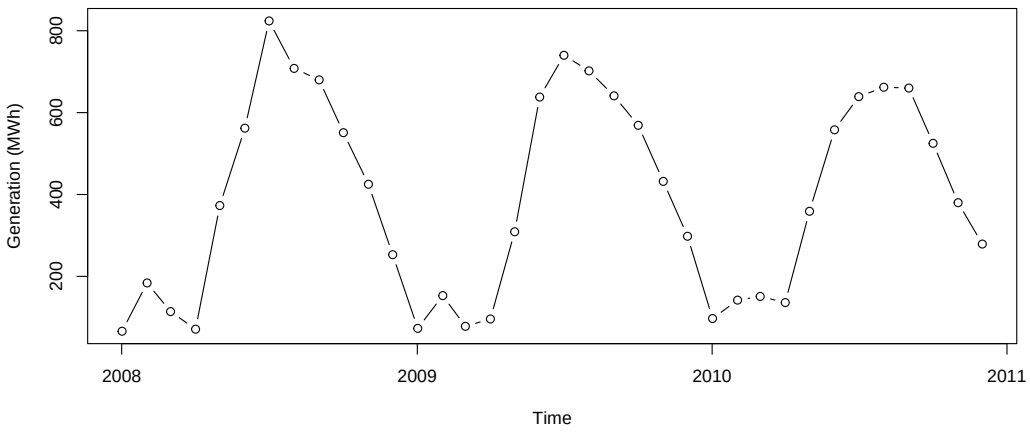
Based on historical data the following seasonal AR model has been identified a particular plant

$$(1 + \phi_1 B)(1 + \Phi_1 B^{12})(\log(Y_t) - \mu) = \varepsilon_t$$

where Y_t is the monthly energy from the plant in MWh, $\{\varepsilon_t\}$ is a white-noise process with variance σ_ε^2. The parameters $\phi_1 = -0.38$, $\Phi_1 = -0.94$ and $\mu = 5.72$ are assumed to be known. Based on 36 observations, it is found that $\sigma_\varepsilon^2 = 0.22^2$.

The file `datasolar.csv` holds the measurements of monthly electricity generation from the plant available.

A plot of the time series is



Answer the following:

2.1. Introduce $X_t = \log(Y_t) - \mu$ and re-write the model to calculate the residuals $\hat{\varepsilon}_{t+1|t}$. Do a model validation by checking the assumptions of i.i.d. errors.

2.2. With the specified model calculate $\hat{Y}_{t+k|t}$ for $t = 36$ and $k = 1, \ldots, 12$, i.e. predict the power for the following twelve months. Provide the values in a table and plot them extending the observed time series (note, remember to transform back to power).

2.3. Calculate 95% prediction intervals for the twelve months ahead and add them to the plot. Use Eq. (5.149)-(5.151) and note that you can do the calculation it without including the seasonal part! So just use the AR(1) part of the model.

2.4. Comment: would you trust the forecast? Do you think the prediction intervals have correct width (all the time)?

3 An ARX model for the heating of a box

The final part is about identifying a suitable ARX model for predicting the hourly heating of a test box (a small building). The box has a window in the south facing wall.

In an experiment times series were recorded of the variables:

- P_{h} (`Ph` in data) the heat from electrical heaters (W).
- T_{delta} (`Tdelta` in data) the difference between the internal and external temperature (°C)
- G_{v} (`Gv` in data) the vertical solar radiation onto the box side with a window (W/m^2)

The data consists of average hourly values – in total 231 hours. It's available in the `box_data_60min.csv` file.

A picture of the box, its heaters and the measurement setup:

In the left plot the box is seen from the outside – the experiment was in Belgium during the winter. Notice the pyranometer, which measures G_{v}. The internal temperature was measured in the middle of the box and the external temperature right next to the box. In the right plot the inside is seen: The two heaters on the floor and various sensors.

In the experiment the internal air temperature was kept constant with a thermostatic control of the heating – hence the heating change depending on the weather conditions.

The objective is to find a suitable model, which makes good predictions of the heating.

Lags have been generated and included in the data with a naming syntax such that, e.g.: `Ph.l1` is the heating lagged one step, `Ph.l2` is heating lagged two steps and so fourth.

Answer the following:

3.1. Read the data and plot the three non-lagged time series (`Ph`,`Tdelta`,`Gv`). Describe the time series and if you can see some dependencies between the variables.

3.2. Split the data into a train and test set, such that "2013-02-06 00:00" is the last data point in the training set (i.e. $t_{\mathrm{hour}} = 1, ..., 167$ is the training set). From now on, work only on the training set, except where explicitly told to use the test set.

3.3. Investigate the variables and their relations: e.g. with scatter, auto-correlation and cross-correlation plots. Most focus on P_{h}. Highlight key aspects of the dynamics and interrelationships among the variables. What can be seen directly and what cannot?

3.4. Estimate the impulse response from T_{delta} and G_{v} to P_{h} make it up to lag 10. Present it for both variables in plots and comment.

3.5. Fit the linear regression model

$$P_{\mathrm{h},t} = \omega_1 T_{\mathrm{delta},t} + \omega_2 G_{\mathrm{v},t} + \varepsilon_t$$

The error is assumed $\varepsilon_t \sim N(0, \sigma^2)$ and i.i.d.

Analyse the estimation result, one-step prediction, residuals with plots, ACF and CCF. Comment with focus on the potential need for a model which includes a transfer function.

3.6. Fit the first order ARX model

$$P_{\mathrm{h},t} = -\phi_1 P_{\mathrm{h},t-1} + \omega_1 T_{\mathrm{delta},t} + \omega_2 G_{\mathrm{v},t} + \varepsilon_t$$

Note the sign of ϕ in the equation, as you know, it's just to keep the same notation throughout, in practice just estimate the coefficients and only if you need to consider their value remember potential sign flip.

Analyse the one-step predictions and residuals as above. Comment, was an improvement achieved?

3.7. Increase the model order of the ARX. The second order model is

$$P_{\mathrm{h},t} = -\phi_1 P_{\mathrm{h},t-1} - \phi_2 P_{\mathrm{h},t-2} + \omega_{1,0} T_{\mathrm{delta},t} + \omega_{1,1} T_{\mathrm{delta},t-1} + \omega_{2,0} G_{\mathrm{v},t} + \omega_{2,1} G_{\mathrm{v},t-1} + \varepsilon_t$$

hence keep the same number of lags for all inputs.

Plot BIC and AIC vs. the increasing model order. Explain the difference between the two curves, and consider what model order you would select based on this?

3.8. Make the one-step predictions through the test period and calculate the Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{64} \sum_{t=168}^{231} \varepsilon_{t|t-1}^2}$$

again plot vs. the model order. Does this result indicate the same model order as the BIC and AIC?

3.9. Make a multi-step prediction (simulation) throughout the entire period, i.e. from beginning of the training period to the end of the test period. Use the model order that you find most suited. Use the observed time series of the inputs (T_{delta} and G_{v}), and the AR lags calculated iteratively.

Is the model good at predicting the heating in such a multi-step prediction setting? Could such multi-step predictions be carried out in a real-time operational setting?

3.10. Make a summarizing conclusion on your findings. You are also welcome to make more analysis, try different relevant predictions, etc.