

This problem set consists of 3 pages. Try to answer all questions. For each question, each sub-question carries the same weight.

All questions should have an answer in plain text – in some cases possibly very short. For many questions it is also relevant to show the R code you have used. You can write your answer in any program you like, as long as it can produce a pdf file. A word processor such as Word is often suitable. Be sure to make a clear distinction between plain text and code, e.g. by using different fonts.

In the grading, most emphasis is put on a correct and clear answer. Some weight is also put on clean and efficient coding.

Data for questions I. and III. can be accessed at

<https://www.dropbox.com/scl/fo/87ei6yapb4iq13zj49ivu/h?rlkey=wxkv9tquqt7fq0527mh24ik8t&dl=0>

I. Exchange rates (weight 50 %)

The file `euro.csv` has the exchange rate between the Euro and the Norwegian krone for the last 10 year. There are two columns, `t` with dates in to format YYYY-MM-DD (so e.g. December 14th 2023 is 2023-12-14).

- Read the data into a tibble `euro`. Plot the exchange rate in a time series plot, i.e. a plot with `time` on the x-axis and the `euro` at that time on the y-axis.
- Make a table or list of the average monthly exchange rate for every month in 2022.
- For every year between 2014 and 2022 find the day with the highest and lowest exchange rate. Show a table with each year as a row, high and low dates and values as columns.

The Excel file `nb_rates.xlsx` contains Norges Bank's policy rate from 1986 to 2023. The variable `date` contains the dates on which rate decisions were made and the variable `policy_rate` the rate that was decided.

- Make a times series plot of the policy rates, i.e. a plot with `time` on the x-axis and the `policy_rate` at that time on the y-axis.
NB: Changes in the policy rate are discrete decisions in time. Your plot should reflect this.
- Merge the euro data and the rate data together, making sure to have a value for the policy rate for each day we have data on the exchange rate. Regress the exchange rate on the policy rate and comment briefly on the results.

Norges Bank has an API to access exchange rate data. One way to access it is through a web link of the format <https://data.norges-bank.no/api/data/EXR/B.EUR+USD.NOK.SP?format=csv&startPeriod=2023-01-01&endPeriod=2023-01-31&locale=en&bom=include>, which would return a csv format file (but semi-colon separated).

The given call return Euro (EUR) and US Dollar (USD) rates from January 1st to January 31st, 2023. The most relevant variables are `TIME_PERIOD`, the date, `OBS_VALUE`, the observed exchange rate, and `BASE_CUR`, the currency the exchange rate applies to.

- Explain briefly what an API is. Use the API to download the exchange rate between NOK and Danish kroner (DKK) and Swedish kroner (SEK) for the period January 2nd 2020 and December 30th 2022. Show the time series of the two exchange rates in a time series plot where you normalize the rate on January 2nd 2020 to 100 for both currencies. Explain what we can read out of the graph.

- g) Write a function `ex_rate(currency, from_date, to_date)` that given a currency (3 letter code, e.g. EUR) and a start and end date in R's date format returns a tibble with the corresponding exchange rate. Use the function to download the exchange rate for US Dollars (USD) between June 1st and August 15th 2018.

II. The Solow growth model (weight 20 %)

The growth model due to Solow (1956) can be expressed in per capita terms (without population growth, with Cobb-Douglas technology) as production being given by

$$y_t = Ak_t^\alpha$$

Assume moreover initially a constant savings rate s and a depreciation rate δ . Then capital per person evolves according to

$$k_{t+1} = sAk_t^\alpha + (1 - \delta)k_t.$$

We are going to use the following parameters

A	10
α	0.3
s	0.2
δ	0.1

- a) This Solow model has a steady state, that is, a level of k where the level stays constant over time. This level must satisfy

$$k = sAk^\alpha + (1 - \delta)k.$$

Solving this expression for k , we find the steady state level

$$k_{ss} = \left(\frac{sA}{\delta} \right)^{\frac{1}{1-\alpha}}.$$

Write a script that computes the steady state with the parameters given above.

- b) We now want to simulate the model. Start with an initial capital stock of $k_0 = 1$. Compute the evolution of k_t for the first $T = 30$ periods.
- c) Consider now an initial capital stock $k_0 = 5$. Show the results from this simulation together with the simulation from question b) in a graph. Indicate also the steady state computed in question a).
- d) Show in a pedagogical graph how the convergence to the steady state shown in questions b) and c) depend on the savings rate s .

III. Predicting corruption (weight 30 %)

In this data we are going to study the Corruption Perception Index provided by Transparency International together with some other data from various sources. The index is ranging from 0 to 10 and measures the absence of corruption. Hence the higher the index is, the less corruption there is. The data are provided in the csv file corruption.csv. We have the following variables

country	Name of country
iso3c	3 letter ISO code
region	Region
cpi	Corruption perception index
gdp	GDP per capita, PPP adjusted
population	Population (thousands)
school_enroll	Secondary school enrollment (% gross)
area	Area in square km
democracy	Democracy index (VDEM polyarchy)

- Read the data into a tibble. As `region` is a text variable, it should be converted to a factor variable. Split the data into an 80 % training and a 20 % testing sample keeping the regional composition equal in the two sets. Explain why it may be useful to maintain regional balance in the two samples.
- Define a random forest model with the default parameters, but setting `mtry` to 3. Define a recipe where you take into account that `country` and `iso3c` are identification variables, not predictors. Train the model on the training data defined above.
- Use the model from b) to predict corruption in the testing sample. Explain why this should be done in the testing sample and not in the training sample. Show some conventional metrics of the quality of the prediction and explain what they show. Show the true CPI values and the predicted values in a scatter plot. Comment briefly on the plot.
- Define a 10-fold cross validation split for the training data. Tune the parameter `mtry` using cross validation and show the RMSE for values of `mtry` ranging from 1 to 6. Explain what this process does and why this is a meaningful way to tune `mtry`.
- Use the optimal value of `mtry` from question d) to retrain the model. Run new predictions and compare to the predictions from question c). What was the effect of tuning the model?