Assignment 2

Deadline: 03rd April 2020, 18:00h

General information

- This assignment is part of the overall assessment of this course and, therefore, your answer counts for the final grade.
- This assignment has to be solved and submitted by every student individually before the above mentioned deadline closes. Group work and group submissions are not allowed.
- The assignment has to be submitted via DTU Inside. Use the entry `Assignments` in the course menu and upload your files to the corresponding assignment. In case of technical problems with DTU Inside, please send your files to `dngk@dtu.dk` before the deadline.
- The submission must consist of a pdf-document containing the answers to the tasks below. Furthermore, program code and scripts have to be uploaded as well.

Overall problem setting for this assignment

You want to invest in real estate in four different areas $I = \{$ `Zip2000`, `Zip2800`, `Zip7400`, `Zip8900` $\}$ in Denmark. The historic data of prices per m^2 for the four areas is given in Figure 1. In this assignment you are using Linear Time Series Analysis and Stochastic Programming to help you with making the investment decisions.

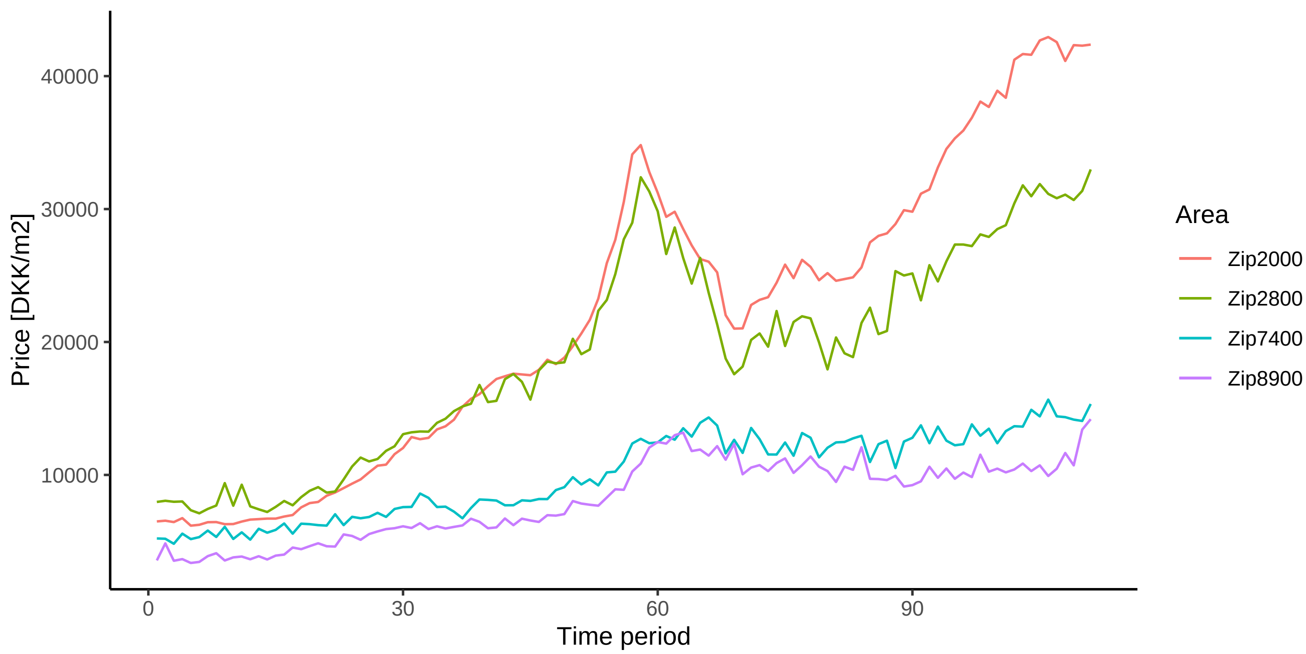


Figure 1: Historic prices in the four investment areas.

Task 1 - Linear Time Series Analysis (25 points)

As a first step to optimizing the investment decisions, you need to fit a linear time series model for predicting the price one period ahead. In this task you need to focus only on area `Zip2000`. To fit the model use the historical data from `prices.csv`. Use an ARIMA model in R.

Document your process of fitting the model and all decisions with plots and argumentations. This means:

- Describe which transformations or differentiation you do based on plots and justify your conclusion.
- Argue the values of p and q based on ACF and PACF plots.
- Evaluate the quality based on the distribution of the residuals.
- State the final time series model.
- Upload your code together with your report. Your Rscript file should have the name `forecast-<studentnumber>.R` and contain your name and studentnumber as comment in the file.

It is allowed to use the `auto.arima` function, but you need to argue the results according to the steps mentioned above.

Task 2 - Scenario generation (25 points)

Use your time series model from Task 1 to generate scenarios for the prices in area `Zip2000` for the next time period as shown in Lecture 07 - Scenario generation. Perform the following steps and document each intermediate result with plots and descriptions in your report.

1. Generate 100 scenarios from your time series model.
2. Cluster the scenarios to 10 representative scenarios using the partitioning around medoids (pam) method (In R: function `pam` in the package `cluster`)).
3. Determine the probability of the scenarios.
4. Upload your code together with your report (should be part of the Script from Task 1).

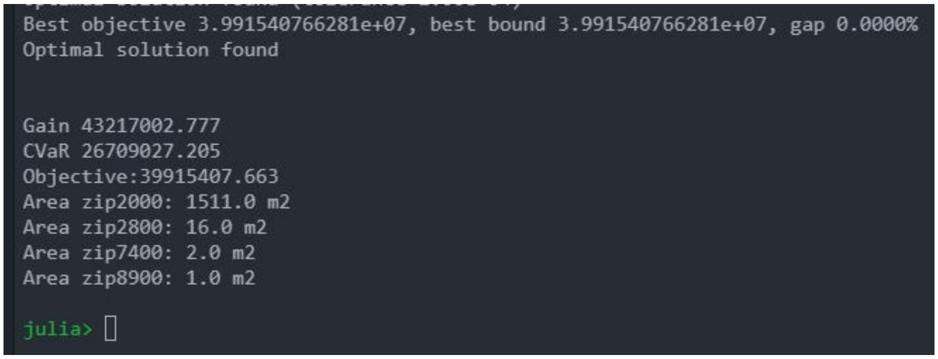
Task 3 - Stochastic programming formulation (25 points)

To optimize your investment decisions, you formulate a **general** linear two-stage stochastic program for the following problem setting. We are now considering all four areas $I = \{$ `Zip2000`, `Zip2800`, `Zip7400`, `Zip8900` $\}$ for investments. Therefore, you should use the scenarios S from in the `scenarios.csv` provided already as data input in the `real_estate.jl` file (and not your own scenarios generated in Task 1 and 2). The file contains data for all four areas.
The initial price for buying one m^2 in each area i is given by p_i^{INIT} [DKK/m^2]. The future prices are uncertain and given per scenario s by the parameter $P_{i,s}$ [DKK/m^2]. You have an initial budget of B [DKK] that you want to spend.
Furthermore, consider the following points in your model:

- The model should decide how many m^2 you will buy in each area. Use integer values to have full m^2.
- The initial budget B has to be spent fully.
- You want to maximize the expected financial gain of your investment in the next period (financial gain $=$ value in the future $-$ initial investment).
- To include risk management, you also want to optimize the Conditional Value-at-Risk (CVaR) at a confidence level of $\alpha = 0.9$.
- In the objective function, the expected gain of your assets is weighted with $(1 - \beta)$ and the CVaR is weighted with β. $\beta = 0.2$.

Perform the following tasks. For each task describe all variables and constraints that you use and define additional parameters and sets if needed. Use the notation for parameters and sets given in this task.

1. Formulate a general linear two-stage stochastic program for the problem setting. Report the mathematical formulation including descriptions.
2. Formulate the expected value version of the stochastic program by optimizing for the expected scenario. The CVaR is not used in this version. Report the mathematical formulation including descriptions.



```
Best objective 3.991540766281e+07, best bound 3.991540766281e+07, gap 0.0000%
Optimal solution found


Gain 43217002.777
CVaR 26709027.205
Objective:39915407.663
Area zip2000: 1511.0 m2
Area zip2800: 16.0 m2
Area zip7400: 2.0 m2
Area zip8900: 1.0 m2

julia>
```

Figure 2: Example ouput (no real solution values!)

3. Use the file real_estate.jl as a basis for implementing the two models. Comment the program code. Furthermore, you have to add the following output, so that your program prints out the following last rows:

   Gain: <Expected gain>

   CVaR: <CVaR value (only for stochastic model)>

   Objective: <objective value>

   Area <i>: <Amount of m^2 bought>

   ...

   Area <i>: <Amount of m^2 bought>

   An example is given in Fig 2.

   Solve both models and obtain the solution values when considering the 12 scenarios given in the data. Report this information for both models also in your report.

   Submit both files along with your report and name them real_estate_stochastic-<studentnumber>.jl and real_estate_expected-<studentnumber>.jl, respectively. Please also add your name and student number as a comment at the top of the file.

Task 4 - Out-of-sample test (25 points)

Perform the following tasks:

1. Carry out an out-of-sample test for the stochastic and expected value solutions. Use the solutions solution_stochastic for the stochastic program and solution_expected for the expected value problem given in out_of_sample.jl for the evaluation (do not use your own solution from Task 3). For the evaluation, use the 100 samples given in file samples.csv. There are already included as input in out_of_sample.jl.

   Rename the file to out_of_sample-<studentnumber>.jl and submit it along with your report.

   Submit a csv file sample-solutions-<studentnumber>.csv containing the results per sample and model. The file should have the columns SampleID, StochasticSolution and ExpectedValueSolution. Use one row per sample.

   Plot the results of the out-of-sample test (use the plot you think describes an out-of-sample test best) and describe them. Which model performs better in the eyes of an risk averse investor, and why?

2. Describe in your own words the steps you need to do to perform an out-of-sample test. (Tip: Maybe a pseudo-code or step-wise description would be nice.)

3. Why is an out-of-sample test important to evaluate a model and why is computing the EVPI and VSS not enough? Answer in a couple of sentences in your own words.