NHH

HOME EXAMINATION BAN432

Fall, 2019

Start: 12.11.2019, 09.00

End: 19.11.2019, 14.00

THE HOME EXAMINATION SHOULD BE SUBMITTED IN WISEFLOW

You can find information on how to submit your paper here:
https://www.nhh.no/en/for-students/examinations/digital-assessment/

Your candidate number will be announced on StudentWeb. The candidate number should be noted on all pages (not your name or student number). In case of group examinations, the candidate numbers of all group members should be noted.

SUPPLEMENTARY REGULATIONS TO HOME EXAMINATIONS

https://www.nhh.no/en/for-students/regulations/regulations-for-examinations-at-nhh-full-time-programmes/supplementary-regulations-to-the-regulations-for-examinations-at-nhh/

All pages, including front page: 3

Number of attachments: 2 files:

ipos_2nd_qtr_2008_2019.zip

ipos_2nd_qtr_2008_2019_nouns_adj.zip

BAN432, fall 2019 – Final Project

Formalities

This final project will be handed out on 12.11.2019 at 9:00 and has to be submitted no later than 19.11.2019 at 14:00.

In addition to the required hand-in files, you will present your findings on the 25/26th Nov 2019. Presentations will be strictly 5 minutes long, followed by a 15 minute question and answer session. The latter will refer to both your submitted report as well as your presentation. If your group has not done so, you can sign up on Canvas for individual presentation slots. All members of the group have to sign up at the same time slot. This exam is group-based. All group members will receive the same grade for the final project. The grade will be based on your submitted report, the presentation, and the question and answer session.

Please submit the following three documents

- **Coding file (.R)** Please describe your general coding approach for each task. You do not need to explain the used functions.
- **Presentation (.pdf)** It is not allowed to change the slides between your submission and the actual presentation. We will provide your submitted presentation on the computer in the room where you give the presentation. Make sure that you use your 5 minutes presentation time wisely. It is important that you provide an economic rational for choices you made during this project.
- **Report (.pdf)** In the report, you should present, analyse, and interpret your results. You should provide numeric and written answers to all questions asked in the exam. Do not discuss your coding in the pdf as this should be found in the R file. Please keep your answers short and precise. Only focus on questions specifically asked in the outline of the project.

Your task

In this report, you try to address following research question (RQ) using textual analysis:

What are the emerging and successful technologies during the last years?

Using textual analysis to answer this questions has an empirical appeal, namely letting the data speak by itself. Other approaches, such as industry classifications, usually require human intervention and hence are subject to human judgement.

In order to answer your research question, you will look at documents surrounding Initial Public Offerings (IPOs). The economic rational is that firms which go public have successfully created a business which has growth potential in the eyes of investors. In other words, if a new successful industry emerges, many firms in this field will go public. Hence, comparing textual information of these firms to firms that went public a decade ago, will inform us about new emerging technologies.

To solve the task, you will work with a sample of S-1 and S-1/A forms that the issuer files to the SEC shortly before the IPO date. The prospectus describes, among others, the proposed business and the risk factors. The sample that we provide covers documents that were filed in the 2nd quarter of the years 2008 to 2019, but not all S-1 forms from that period are part of the sample. Typically an issuer files a S-1 form and several S-1/A forms before an IPO. For our sample, only the first filing per CIK in the above mentioned period is considered.

Only basic cleaning steps have been applied to the documents. The data comes in two versions:

ipos_2nd_qtr_2008_2019.zip contains the raw text of 2356 S-1 and S-1/A forms from the above mentioned period. Only the attachments are stripped.

`ipos_2nd_qtr_2008_2019_nouns_adj.zip` contains the same documents as in the file above, but for each document only nouns and adjectives are kept.

The file names of the individual text documents is constructed as follows: `{CIK}-{year}-{month}-{day}.txt`, where the date refers to the day the document was filed to the SEC.

You can choose to base your analysis on one of the files or both.

Comments and suggestions

- Do not be discouraged if data is fairly messy, it usually is.

- The corpora are fairly big which might create some challenges. We have following words of advice:
  - You could select a sub-sample of the data. Even if you decide to use the whole data, we would recommend to try your code on a sub-sample before you apply it to the whole set of data.
  - You could merge individual filings for different frequencies.
  - We would advice to be careful with how many documents you hold in memory. You might want to use an approach that loops over all or a subset of documents, only extracts the needed items, and subsequently deletes the textual data from memory.
  - If you intend to construct Document-Term-Matrices, we strongly advice you to first construct Document-Term-Matrices for subsamples of the data and later merge by using `c(dtm.1, dtm.2)`. It takes unreasonable time to make a dtm on all documents in one go.

- Make sure that you do not pick more than two approaches to answer the RQ. Your empirical approach has to be creative and convincing. In particular, the reader has to be convinced that your approach actually can pick up emerging industries. We would like to point out that just using an off-the-shelf model, even if it is econometrically very involved, might not be the most convincing approach. Make sure that you tailor the tools we covered in the course to the exact RQ. You might want to combine several tools into one approach.

- As we have seen in several lectures, the actual preprocessing steps matter. Make sure that you are considerate with your choices but also that you understand the impact of each preprocessing step on the final outcome.

- Both guest lectures might be relevant for the task.

- If you have questions during the exam, do not visit us in our office but write an email. If we answer your request, we will do so on Canvas so that information is public and distributed fairly.

- Good luck!