

## **List of projects**

Please find below a list of projects. If you prefer working on a different project (which is related to the course), you can write to me with a description of the project you are interested in. For example, the list below does not include projects on factor investing / smart beta / smart alpha strategies, such projects would be perfectly fine.

You can work on your project in groups of 1 to 5 people. Expectations will of course be dependent on the number of people in the group. I share with you a file where you can add your name in front of the project you are interested in.

For each project, you will be asked to present your findings in one of the two last weeks of class. Each presentation will last around 30 minutes (based on the number of people in the group), followed by 10 minutes of questions / discussion. Everyone is required to attend all presentations. Participation in the discussion is encouraged.

You also need to write a report, for each project, with 1) a brief literature review on the topic, 2) the description of what you have done, your modelling choices and results. The length of a report should not exceed 20 pages. There is no need to put code in your report, the code should be sent separately.

You have the choice of the programming language you use. I recommend using Python, R or Matlab.

Please feel free to send me emails in case you have questions.

### **Projects on text analysis**

- 1) Topic analysis related to ESG matters, based on website snapshots

Research question: How has companies' information disclosure on ESG matters evolved with time?

- a. Data extraction and cleaning
  - i. Extract a time series of website snapshots from the Wayback Machine (internet archive) for selected companies (start with a few large companies in the S&P 500, then if time allows, increase the set of companies to cover different sectors, sizes and countries)
  - ii. Extract ESG-related website contents, if any
  - iii. Data analysis and summary statistics
- b. Model choice
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods (LDA, ...)
  - iii. Selection of method(s)
- c. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
- d. Analysis of results
  - i. Critical analysis of results

## 2) Building an ESG rating from sustainability reports

Research question: How has companies' information disclosure in annual sustainability / Corporate Social Responsibility evolved with time?

- a. Data extraction and cleaning
  - i. Extract sustainability reports from [www.sustainability-reports.com](http://www.sustainability-reports.com)
  - ii. Data analysis and summary statistics
- b. Model choice
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods (weighted average of criteria, ...)
  - iii. Selection of method(s)
- c. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
- d. Analysis of results
  - i. Critical analysis of results

## 3) Sentiment analysis related to ESG matters, based on 10K reports and proxy statements

Research question: Has companies' awareness on ESG matters evolved with time?

- a. Data extraction and cleaning
  - i. Extract 10K reports and proxy statements from SEC EDGAR database.
  - ii. Data analysis and summary statistics
- b. Model choice
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods (bag of words, ...)
  - iii. Selection of methods (at least one simple and one more sophisticated one)
- c. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
- d. Analysis of results
  - i. Critical analysis of results
  - ii. Compare with ESG rating

## 4) Sentiment analysis of earning call transcripts

Research question: Can one elaborate a trading strategy based on earning calls, by buying stocks with the most positive earning calls and selling stocks with the most negative earning calls?

- a. Data extraction and cleaning

- i. Extract earning call transcripts from <https://seekingalpha.com/earnings/earnings-call-transcripts>.
  - ii. Data analysis and summary statistics
- b. Model choice for tone analysis
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods (bag of words, ...)
  - iii. Selection of methods (at least one simple and one more sophisticated one)
- c. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
  - iii. Testing
  - iv. Implementation of trading strategy (focus on the tone in the Q&A section)
- d. Analysis of results
  - i. Critical analysis of results

#### 5) Mutual fund deviations from their benchmarks

Research question: Do fund managers with more assertive/aggressive language deviate more from their benchmark?

- a. Data extraction and cleaning
  - i. Extract certified shareholder reports from <https://www.sec.gov/edgar/search/#/category=custom&forms=N-CSR%252CN-CSRS>.
  - ii. Extract benchmark and find benchmark returns from CRSP.
  - iii. Calculate tracking error and tracking error volatility
  - iv. Data analysis and summary statistics
- b. Model choice for tone analysis
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods
  - iii. Selection of methods (at least one simple and one more sophisticated one)
- c. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
  - iii. Testing
  - iv. Correlation analysis between tracking error and tone
- d. Analysis of results
  - i. Critical analysis of results

#### 6) Sentiment analysis of Tesla-related tweets and predictive power on returns

Research question: Can one elaborate a trading strategy based on Tesla-related tweets?

- a. Data extraction and cleaning
  - i. Extract tweets from accounts of Elon Musk and Tesla.
  - ii. Extract Tesla returns

- iii. Data analysis and summary statistics
- b. Model choice
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods (bag of words, ...)
  - iii. Selection of methods (at least one simple and one more sophisticated one)
- c. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
- d. Analysis of results
  - i. Critical analysis of results

7) Sentiment analysis of Bitcoin-related tweets and predictive power on Bitcoin value (BTCUSD)

Research question: Can one elaborate a trading strategy based on Bitcoin-related tweets?

- a. Data extraction and cleaning
  - i. Extract tweets from relevant accounts, e.g., Bitcoin.org @bitcoin, Andreas M. Antonopoulos @aantonop (Bitcoin and blockchain expert), Vitalik Buterin @VitalikButerin (creator of Ethereum), Rhythm @Rhythmtrader (crypto-enthusiastic), whale\_alert @whale\_alert (live tracker of crypto-transactions), Coindesk @coindesk (media platform for investors in cryptos), Bloomberg Crypto @crypto etc.
  - ii. Extract Bitcoin-USD rates
  - iii. Data analysis and summary statistics
- b. Model choice
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods (bag of words, ...)
  - iii. Selection of methods (at least one simple and one more sophisticated one)
- c. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
- d. Analysis of results
  - i. Critical analysis of results

8) Sentiment analysis on r/Wallstreetbets

Research question: Can one elaborate a trading strategy based on reddit comments?

- a. Data extraction and cleaning
  - i. Extract data from r/Wallstreetbets on several dates
  - ii. Data analysis and summary statistics
- b. Model choice
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods (bag of words, ...)
  - iii. Selection of methods (at least one simple and one more sophisticated one)

- c. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
- d. Analysis of results
  - i. Critical analysis of results

#### 9) Spread of information through social media

Research question: How is information disseminated on social media (Twitter)?

Reference: [https://www.nber.org/system/files/working\\_papers/w24631/w24631.pdf](https://www.nber.org/system/files/working_papers/w24631/w24631.pdf)

- a. Choice of information setup
  - i. Choose an event, e.g., Brexit, last US presidential election, covid...
- b. Data extraction and cleaning
  - i. Extract data from Twitter
  - ii. Data analysis and summary statistics
- c. Model choice
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods of sentiment analysis
  - iii. Selection of methods
- d. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
- e. Analysis of results
  - i. Critical analysis of results

#### 10) Sentiment analysis on inflation after covid

Research question: How did expectations change about after-covid inflation?

- a. Choice of relevant data (Google searches, Twitter...)
  - i. Choose dataset that is in your opinion informative on inflation expectations
- b. Data extraction and cleaning
  - i. Extract data
  - ii. Data analysis and summary statistics
- c. Model choice
  - i. Review of literature
  - ii. Analysis of strengths and weakness of different methods of sentiment analysis
  - iii. Selection of methods
- d. Implementation
  - i. Data preprocessing
  - ii. Implementation of chosen model(s)
- e. Analysis of results
  - i. Critical analysis of results

## Projects on cryptocurrencies

### 11) Analysis of cryptocurrencies

Research question: What is the risk and return of trading cryptos?

- a. Description of main cryptocurrencies
  - i. Brief description of the Blockchain technology they are based on
  - ii. Differences between the main cryptocurrencies (mechanisms, supply...)
  - iii. Description of the markets (liquidity)
- b. Data analysis
  - i. Stylized facts and statistical tests on time series (univariate and multivariate tests: normality, stationarity, cointegration...)
  - ii. Comparison with equity market
  - iii. Comparison with commodities (e.g., gold)
- c. Analysis of results
  - i. Discussion on crypto trading: speculation or hedging?

### 12) Volatility of cryptocurrencies

Research question: Can we model the volatility of cryptos with GARCH models?

- a. Data analysis
  - i. Statistical analysis of BTCUSD time series and time series of volatility
- b. Model selection
  - i. Review of literature
  - ii. Choice of a set of models to implement
- c. Analysis of results
  - i. Comparison of the performance of implemented models
  - ii. Analysis of the strengths and weaknesses of the implemented models

## Projects on forecasting

### 13) Nowcasting of GDP

Research question: What is the best method to nowcast the GDP?

- a. Data analysis
  - i. Analysis of GDP time series
  - ii. Description of the problem of nowcasting
- b. Model choice
  - i. Literature review
  - ii. Choice and implementation of model
- c. Analysis of results
  - i. Critical analysis of the model's performance

#### 14) Forecasting of inflation

Research question: Can you propose a method to forecast inflation?

- a. Data analysis
  - i. Analysis of inflation time series
- b. Model choice
  - i. Literature review
  - ii. Choice and implementation of model
- c. Analysis of results
  - i. Critical analysis of the model's performance

#### 15) Momentum strategies.

Research question: How can momentum strategies be built using GARCH volatility modelling?

Reference: [https://www.nber.org/system/files/working\\_papers/w20439/w20439.pdf](https://www.nber.org/system/files/working_papers/w20439/w20439.pdf)

- a. Data analysis
  - i. Description of data
- b. Model choice
  - i. Literature review
  - ii. Choice and implementation of model
- c. Analysis of results
  - i. Critical analysis of the model's performance