High Dimensional Data Analysis Assignment 2

Department of Econometrics and Business Statistics, Monash University

Due Date: 5th October 2020 at 3PM

1 Data

This assignment uses data on a set of 147 stocks from five major financial indices (NDX, DJI, FTSE, GDAXI, HSI). The data was sourced from Yahoo Finance.

Data on the following variables are available:

- **Name**
- **Symbol**
- **Market**
- **Sector**
- **Industry**

The following variables are all measures related to the value of the stock:

- **Market cap (intra-day)**
- **Enterprise value**
- **Trailing P/E**
- **Forward P/E**
- **PEG ratio (5-yr expected)**
- **Price/sales (ttm)**
- **Price/book (mrq)**
- **Enterprise value/revenue**
- **Enterprise value/EBITDA**

The following variables are all measures related to environmental, social and governance risks:

- **Total ESG risk score**
- **Environmental Risk Score**
- **Social Risk Score**
- **Governance Risk Score**

The full dataset can be found on Moodle under the name *stocks.csv*, and the description of the variables therein can be found in Investopedia.

2 Task

Your are required to conduct some preliminary analysis on the data. The only mandatory requirement is that you **MUST** use principal component analysis (PCA). In addition, you may also use the other techniques covered in the unit such as cluster analysis and multidimensional scaling, but each of these is optional. You do not need to use all the variables or all the observations. You must summarise your results in a report of no more than 1500 words. Your R code and additional work not crucial to the analysis can be included in an Appendix (this will not count towards the word limit).

3 Guidance

To assist you, a list of questions are provided below. These are designed to prompt you to think about the analysis and will influence the grading of the assignment. However, this list is non-exhaustive - you do not necessarily need to answer all of these questions, and if you can think of issues not listed here then you are encouraged to address them.

- Is the data clean? Are there missing values, outliers or other data credibility issues?
- Can you derive any insights from the data from simple exploratory analysis including summary statistics and basic plots?
- Can the data be easily visualised?
- How can you profile the principal components? Do they have some interpretation in terms of the data itself?
- Does the report contain enough information to be reproduced by somebody with knowledge of the techniques used?
- Are all plots clearly presented and correctly explained?
- Is the analysis robust to minor changes in the methodology?
- Are any assumptions made for the analysis or in drawing conclusions. If so, are these clearly explained?
- Does the report focus on a small number of interesting features of the analysis or does the report simply list everything that was attempted (the former is preferable to the latter)?
- Are the limitations of the analysis clearly discussed?

4 Submission

The assignment is a **group assignment**. The maximum group size is four people. You may form groups with students from different tutorial groups and from different unit codes. A single soft copy should be submitted with a group assignment cover page added to the front. All assignments should be submitted via Moodle.