MATH 6380Q Project 1 Peer Review

06. WANG, Meilan and LIU, Di. Finance Data PCA, Parallel Analysis.

[ Reviewer 1 ]

1. Finance Data PCA, Parallel Analysis

Summary:

By employing PCA, the author tried to extract some interesting pattern among different industries’ stock price normalized increase.

Strength of the project:

The normalization which transform the daily stock price into price increasing rate is reasonable and could give further insights to the data.

Weakness of the project:

As shown from the figure, the first two PCs only explained 8.5% of variation in the dataset. Therefore, the trends identified by projecting the data to these two PCs may not be significant at all. And how can the “upstream” and “downstream” clusters be accounted for? What could each cluster possibly represent? The above questions are rather central to the analysis but remained unanswered.

|  |  |
| --- | --- |
| Evaluation on Clarity and quality of writing (1-5): | 3 |
| Evaluation on Technical Quality (1-5): | 3 |
| Overall rating: | 3 |
| Confidence on your assessment: | 1 |

[ Reviewer 2 ]

In poster six, “Finance Data PCA, Parallel Analysis”, authors carried out PCA and Parallel Analysis over a time series of stock closed prices in SNP’500. The dataset contains 1258-by-452 matrix with closed prices of 452 stocks in SNP’500 for workdays in 4 years. They made dimension reduction by PCA and confirm if the number of eigenvalues they chose is in the reasonable range with Parallel Analysis. The conclusion is that the Energy and IT sector is more distant than downstream and daily-life related industry sectors. The principal component 1 could be upstream and downstream of business chain, while the principle component 2 could be the distance to daily-life.

The strength of this poster is authors successfully make data dimension reduction by PCA and use Parallel Analysis to test if the number of eigenvalues they chose is in the reasonable range.

The weakness of this poster is there is just one model and no comparison.

Evaluation on Clarity and quality of writing: 4. The report is clearly written and the poster is well organized. And there is a good use of figures.

Evaluation on Technical Quality: 4.

Overall rating: 3.

Confidence on my assessment: 2.

[ Reviewer 3 ]

06. WANG, Meilan and LIU, Di. Finance Data PCA, Parallel Analysis.

The authors conducted PCA and parallel analysis on the SNP500 data and found out some interpretation.

One good thing is that they know to do a pre-analysis on the data by calculating ‘log()’. However, this variable may be of little meaning since in finance industry people always tend to use return() or log-return (log())).

To be honest, I cannot make it clear what the ‘downstream’ and ‘upstream’ mean in this poster. The organization is fine and simple.

Technically, for each plot of subgroup data, the meaning of different colors should be explained. Also, the processed data of log() is of little meaning. Last but not least, you should combine the results with some examples. The strengths and weaknesses are missing.

The overall rating is 3 and my confidence is 2.

[ Reviewer 4 ]

This report performs good PCA analysis on the Finance Data. It has many figures containing rich information and easy to understand.

Evaluation on Clarity and quality of writing: 5

Evaluation on Technical Quality:3

Overall rating:4

Confidence on your assessment:2

[ Reviewer 6 ]

5.1 Summary

For Wang Meilan and Liu Di’s work, they applied Principle Component Analysis and Parallel analysis on the SNP500 dataset. They investigated the dataset by normalizing the data and then plot the PC1-PC2 figure to determine the outliers, and infer the representation of first principal component and second principal component.

5.2 Strength and Weakness

Wang and Liu have done an excellent job on analyzing the dataset. Each step they have done are reasonable. The PC1 versus PC2 figure are divided by each industries, which is easy for us to get information. The weakness of their report can be clearer and deeper.

5.3 Score

5.3.1 Clarity and Quality of Writing

The arrangement and structure of the poster is great, I believe they have spent a lot of effort on this report. I didn’t find any typo. I will give them 5/5 on this aspect.

5.3.2 Technical Quality

Reasonable steps for getting the analysis conclusion. Each plot are useful and construct a great explanation of their work. It’s good that they mentioned the future improvement of their work. If the conclusion can be deeper this report will be much better. I will give them 4/5 on this aspect.

5.3.3 Overall

The overall score for this poster is 4.5/5.

[ Reviewer 7 ]

* **Summary of this report:** PCA and Parallel Analysis over a time series of stock closed prices in SNP’500 are carried out. It is found that the Energy and IT sector is more distant than downstream and daily-life related industry sectors. Moreover, the first principal component could be upstream and downstream of business chain, while the second principal component could be the distance to daily-life.
* **Describe the strengths of the report:** This report clearly shows thatthe first two components contributes over 8.5% of explained variance ratio. The analysis are well thought out and convincing.
* **Describe the weaknesses of the report:** Some legends of the figures in this report are to small to be clearly recognized.
* **Evaluation on Clarity and quality of writing (1-5): 3**

Details of typos: 1) In order to know more about the dataset, or modelling (modeling) with the dataset. 2) Then we conducted the PCA to find out the components have (the) most significant impact. While (the) first two component(s) contributes over 8.5% of explained variance ratio, we project the original data on the space of PC1 x PC2. 3)we could conducted (conduct) PCA using time as (a) feature, or fit PCA within sector companies; With PCA result, we could build a predicting model with less (fewer) features.

* **Evaluation on Technical Quality (1-5): 5**

The PCA is correctly used. Moever, the report is well reported.

* **Overal ratings: 4**
* **Confidence on your assessment: 1**

[ Reviewer 8 ]

• Summary of the report.

Apply PCA and parallel analysis on SNP500 data.

• Describe the strengths of the report.

Clearly written.

• Describe the weaknesses of the report.

Lack of deeper analysis. The methods (basically just PCA) might be a little bit too simple.

• Evaluation on Clarity and quality of writing (1-5): Is the report clearly written? Is there a good use of examples and ﬁgures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? Please make suggestions to improve the clarity of the paper, and provide details of typos.

4

• Evaluation on Technical Quality (1-5): Are the results technically sound? Are there obvious ﬂaws in the reasoning? Are claims well-supported by theoretical analysis or experimental results? Are the experiments well thought out and convincing? Will it be possible for other researchers to replicate these results? Is the evaluation appropriate? Did the authors clearly assess both the strengths and weaknesses of their approach? Are relevant papers cited, discussed, and compared to the presented work?

2

Lack of deeper analysis after applying the methods.

• Overall rating: (5- My vote as the best-report. 4- A good report. 3- An average one. 2below average. 1- a poorly written one).

3-

• Conﬁdence on your assessment (1-3) (3- I have carefully read the paper and checked the results, 2- I just browse the paper without checking the details, 1- My assessment can be wrong)

2+

[ Reviewer 9 ]

1. **Summary**

The authors conducted PCA and parallel analysis over a time series of stock closed prices in SNP’500. By performing PCA analysis, it is found that Energy sector and IT sector are more distant from traditional industries.

1. **Strength of the report**

The report is clearly written and well organized. And the conclusion is reached from applied PCA analysis.

1. **Weakness of the report**

No serious weakness is found.

1. **Evaluation of clarity and quality of writing**

The report is clearly written and well organized. The figures are clearly explained. Minor grammar issue like” The Energy sector and IT sector is more distant” should be “The Energy sector and IT sector are more distant”.

1. **Evaluation on technical quality**

The results are technically sound and no obvious flaws in the reasoning are found. The report provides the methods for analyzing similar dataset.

1. **Overall rating**

An overall rating of 4 is given.

1. **Confidence on your assessment**

2- I just browse the paper without checking the details

[ Reviewer 11 ]

Summary: The authors do PCA and Parallel Analysis over a time series of stock closed prices in SNP500 dataset.

Strengths: Clearly states the method and results.

Weaknesses: The images are too small for a poster. The report does not have much explanations of the images.

Writing: (4) The report is well-written but the images are too small for a poster.

Technical Quality: (4) They describe the experiments clearly but the results (images) do not have much explanations. It is hard to see how they get the conclusion.

Overall rating: (4)

Confidence on my assessment: (2)

[ Reviewer 12 ]

**Summary:** This project uses PCA to analyze SNP 500 dataset. It first normalizes the data and then use PCA to find the principal components. It finally uses first and second principal component map to try to distinguish different types of companies and give some reasons.

**Strengths:** The overall logic is clear and easy to understand. And the visualization is neat. Apart from only showing the results, this project also tries to give some reasonable explanations for the results.

**Weaknesses:** Normalization in PCA is not a necessary choice and I think it would be better if this project can compare the results given by data with and without normalization. Besides, some of the figures are not clear. Figure 5 is a little bit strange and there is no explanation about it in the poster.

**Evaluation on Clarity and quality of writing:** 4

**Evaluation on Technical Quality:** 3

**Overall rating:** 4

**Confidence on your assessment:** 2