BF 3224 / 3228 Equity Investing with Big Data

Project Overview and Goals

Use SAS and the data I share with you to develop and backtest a quantitative investment strategy. **Please work in teams of 2-3 students**. You may work with the following data, which are posted on NTU Learn:

- CRSP-daily: daily return data from 1st January 2018 to 31st December 2022
- CRSP-monthly: monthly return data from 1st January 2000 to 31st December 2022
- Factors-daily: daily factor data from 1st January 2018 to 31st December 2022
- Factors-monthly: monthly factor data from 1st January 2000 to 31st December 2022
- Industry: data assigning each stock to an industry (https://www.msci.com/our-solutions/indexes/gics)
- Robinhood-daily: data set containing for each US stock for each day from 2nd May 2018 to 16th June 2020, the number of Robinhood investors holding a particular stock. Robinhood is one of the largest stock trading platforms for retail investors in the US.
- Tweets-daily: data set containing for each US stock for each day from 31st December 2012 to 31st December 2018, the number of tweets and retweets sent regarding a stock.
- Analyst reports: data set containing information regarding analyst reports written on US stocks from 1st January 2006 through 31st October 2021. To explain some of the variables: BUYSELL (1-sell, 2-hold, 3-buy), word_length (number of words in analyst report), negword (number of negative words), posword (number of positive words), tone (posword-negword/word_length)

You can work with the daily or monthly return data. To get your "signals" dataset, you can use only one of the following three data sets or two or all three: Robinhood-daily, Tweets-daily, Analyst reports. You may use the Industry data set if you believe your strategy should work better in specific industries (or not work in certain industries). The use of the Industry data set is optional.

Deliverables and Evaluation

Write a **1-2 page report** that touches on all points listed in Appendix 1. **Email the report** and your **SAS code** to bh.hwang@ntu.edu.sg by **Monday, 17th April, at 9:00 am**. One submission per team only, please. Please cc all your team members as you email your report and code. I describe the **assessment criteria** in Appendix 2. Please note that the assessment criteria vary slightly from what is given in your syllabus. I mark all variations in bold red.

I will email you a link to a **team evaluation form**. I describe the **assessment criteria** for the team evaluation in Appendix 3. Please submit the form only if you feel your teammate's average score falls below four (as described in Appendix 3). Should you want to submit a team evaluation form, you must do so by **Monday, 17th April, at 9:00 am**.

SAS Project (Revised – March 21st, 2023)

Appendix 1: Things you need to cover in your written report

1. Who are your team members?

2. Detail how you constructed your long-short portfolio.

3. Detail your rationale for constructing your long-short portfolio. Put differently, why should we expect your strategy to generate abnormal returns? Why should your strategy work in theory?
   [most important part; go back to Week 4, slides 12-20]

4. Plot the cumulative performance of your long-short strategy (in terms of raw returns). [Hint: If you work on a computer in the computer lab, export your returns from SAS into Excel and go from there. If you use SAS OnDemand, open the SAS dataset, copy the relevant data and past them into Excel.]

5. What was the average annual return on your long leg? What was the average annual return on your short leg? What was the average annual return on your long-short portfolio? What were the corresponding Sharpe Ratios? Finally, what were the corresponding alphas with respect to the Market Model and the 4-Factor Model?
   [go back to Week 5, slides 15-22]

6. Overall, how would you assess your strategy's performance?

Appendix 2: Assessment Criteria

You will be evaluated based on the following criteria:

Traits	1 Below Expectations	2 Met Expectations	3 Above Expectations
Apply Concepts to Real-World Situations	Student cannot apply the concepts learned in class to real-world situations. **The rationale behind the investment strategy is comparatively weak/unlikely to describe the real world.**	Student can mostly apply the concepts learned in class to real-world situations. **The rationale behind the investment strategy is reasonable/likely to describe aspects of the real world.**	The student can fully and expertly apply the concepts learned in class to real-world situations and expand upon them. **The rationale behind the investment strategy is strong and creative and likely to play a significant role in the real world.**
Analysis	SAS code contains numerous mistakes and is not adequately commented on. Students incorrectly compute/describe their portfolio construction/portfolio performance.	SAS code contains only minor mistakes and is adequately commented on. Students correctly compute/describe their portfolio construction/portfolio performance.	SAS code contains no mistakes and is thoroughly commented on. Students correctly compute/describe their portfolio construction/portfolio performance.
Writing	The report is difficult to understand. Problems with sentence structure, leaving the reader unsure of the meaning.	The report can be adequately understood.	The report is well-organized, easily understood, concise, and to the point.

Team Evaluation

All group members will receive the same Group Component Score **unless** the peer evaluation indicates that the member has not contributed enough to the project. The peer evaluation is conducted in the same manner as described in Appendix 1B.

Computation of Total Numerical Score:

To compute your total numerical score for the SAS Project, I will take your scores for the various traits (ranging from 1 to 3) multiplied by the weights below and add them up. Your total numerical score will thus range from 1 to 3 (with two decimal places).

Traits	Weight
Apply Concepts to Real-World Situations	**40%**
Analysis	**35%**
Writing	**25%**

Appendix 3: Team Evaluation

All group members will receive the same Group Component Score **unless** the peer evaluation indicates that the member has not contributed enough to the project. The peer evaluation is conducted as follows:

Each member will rate the other team members with regard to five traits on a scale ranging from "1" through "7," with "7" denoting the most positive evaluation. For each member, I will compute the average rating across the five traits submitted by the other team members:

1. If a member's average rating is ≥ 4, the member will receive **100%** of the overall score awarded to the team assignment.
2. If a member's average rating is < 4 but ≥ 3, the member will receive **80%** of the overall score awarded to the team assignment.
3. If a member's average rating is < 3 but ≥ 2, the member will receive **50%** of the overall score awarded to the team assignment.
4. If a member's average rating is < 2, the member will receive **30%** of the overall score awarded to the team assignment.

Ratings will be kept confidential. I will only inform you if the peer evaluation indicates that you have not contributed enough to the project and how many points you lost as a result.

The following is a description of the five traits, along with a description of the scales:

Traits	Performance	
Roles and Responsibility Behaves professionally and fulfills responsibilities.	**Scant** Unclear about his/her role; refuses to take a role in the group; insists on working individually, and has limited coordination or communication with others.	**Substantially Developed** Always fulfills responsibilities; performs his/her role within the group with enthusiasm; demonstrates a willingness to work collaboratively.
	Evaluation: **Scant 1 2 3 4 5 6 7 Substantially Developed**	
Communication Identifies appropriate channels to coordinate and correspond with team members.	**Scant** Modes of communication are inappropriate, causing confusion and miscommunication among team members.	**Substantially Developed** Modes of communication are appropriate; maintains timely communication and correspondence with team members.
	Evaluation: **Scant 1 2 3 4 5 6 7 Substantially Developed**	
Conflict Resolution Resolves conflicts using a variety of approaches.	**Scant** Does not recognize conflicts or is unwilling to resolve conflicts.	**Substantially Developed** Consistently resolves conflicts by facilitating open discussion and compromise.
	Evaluation: **Scant 1 2 3 4 5 6 7 Substantially Developed**	
Contributions Contributes positively; effectively utilizes one's knowledge and expertise.	**Scant** Largely disinterested in working in a group and refuses to participate; observes passively or is unwilling to share information with other team members.	**Substantially Developed** Actively attends and participates in all activities and articulates ideas and opinions.
	Evaluation: **Scant 1 2 3 4 5 6 7 Substantially Developed**	
Relationship Maintains cooperative interaction with other team members regardless of individual /cultural differences and respects diverse perspectives.	**Scant** Rarely listens to others and does not acknowledge the opinions that differ from his/her own.	**Substantially Developed** Engages in respectful relationships with all other members of the team. Embraces and accepts diverse points of view without prejudice.
	Evaluation: **Scant 1 2 3 4 5 6 7 Substantially Developed**	