MScFE 600 FINANCIAL DATA

Group Work Project # 1

See grading rubric here.

Below you will find four separate scenarios and each one includes a set of questions that you are required to answer in words, using code, or both words and code. The 'Submission requirements and format' section on Page 4 provides a detailed explanation on how to complete and submit the assignment through the online platform.

Groups of 3 members must complete all the Tasks and answer all the questions.

Groups of 2 members or 1 member will skip some of the questions following the requirements on Page 3.

Tasks

1. Data Quality

One of the important components of financial data is ensuring good quality data. However, it is helpful to understand what poor quality data looks like.

a. Provide an example of poor quality structured data
b. How would you recognize this poor quality? Write 3 - 4 sentences that show how the data fails to include properties of good quality data.
c. Provide an example of poor quality unstructured data
d. Unstructured data can be more difficult to assess than structured data. Just as you did in part b, write 3 - 4 sentences that show how this unstructured data fails to check requirements of good quality data.

2. Yield Curve Modeling

a. Pick government securities from a country. The country selected should be one of the countries from your group so that you can fit a Nelson-Siegel model.
b. Be sure to pick maturities ranging from short-term to long-term (e.g. 6 month maturity to 20 or 30 year maturities).
c. Fit a Nelson-Siegel model.

d. Fit a Cubic-Spline model.
e. Compare the models in terms of 1) fit and 2) interpretation.
f. Be sure to specify at the levels of model parameters (ex. Alpha1).
g. In Module 2 Lesson 4 ('Smoothing Data'), we said smoothing data can be unethical. If Nelson-Siegel is smoothing the yield curve, is this considered unethical? Why or why not?

3. Exploiting Correlation

Financial Data is meant not only to process data but to understand how meaningful factors can be used to summarize or represent the data. Let's understand the role that correlation and principal components play.

a. Generate 5 uncorrelated Gaussian random variables that simulate yield changes (they can be positive or negative with a mean close to 0 and a standard deviation that is small).
b. Run a Principal Components using EITHER the correlation OR covariance matrix.
c. Write a paragraph explaining how the variances of each component compare with each other. In this paragraph, you will address the following question: how much variance is explained by Component 1, Component 2, Component 3?
d. Produce a screeplot (see https://en.wikipedia.org/wiki/Scree_plot) of the variance explained for each component.

Now let's work with real data:

e. Collect the daily closing yields for 5 government securities, say over 6 months.
f. Be sure to compute the daily yield changes!
g. Re-run the Principal Components using EITHER the correlation or covariance matrix.
h. How do the variances of each component compare? In other words, how much variance is explained by Component 1, Component 2, Component 3, etc.?
i. Produce a screeplot of the variance explained for each component.
j. How does the screeplot from the uncorrelated data compare with the screeplot from the government data?

4. Empirical Analysis of ETFs

Pick a sector ETF (in the US, for example, XLRE)

a. Find the 30 largest holdings.
b. Get at least 6 months of data (~ 120 data points).
c. Compute the daily returns.
d. Compute the covariance matrix.
e. Compute the PCA.
f. Compute the SVD.

Now that you have calculated, presented and plotted tasks from c to f, you must explain each transformation thoroughly. Write a paragraph of 500 words at minimum that explains why returns are important, compare and contrast PCA and SVD, explain what the eigenvectors, eigenvalues, singular values etc show us for the specific data, etc,

Groups of 2 members: Complete Task 1, and then choose 2 Tasks among the remaining 2, 3, and 4.

Groups of 1 member: Complete Task 1, and then choose 1 Task among the remaining 2, 3, and 4.

Every Group Work Project Submission is composed by a report in PDF format and coding (see details in the 'Submissions requirements and format' below.

In the report, be sure to include a **list of references** and use them properly. This is mandatory for every GWP assignment you will submit in this Program.

Just compiling a list of references is not enough. **In your paper you also must show where each reference was used specifically (in-text citation)**.

Use the In-Text Citations and References Guide to learn how to add in-text citations and references.

In the Student Resource Center located in the dropdown menu under your name you find several other resources to complete your Group Work Projects properly:

- Academic Writing Guide,
- Anti-Plagiarism Guide,
- How to use LIRN, the online free library

Important note on the use of AI: Carefully read the WQU **Academic Policy on the use of AI** explaining how the use of AI tools is restricted and regulated. Severe penalties apply for excessive and improper use of AI.

Submission requirements and format

One team member submits on behalf of the entire group the following items:

1. **1 PDF document*** with written answers for Tasks 1, 2, 3, and 4.
    a. Use the available Report Template and fill out the required information on group members on the first page
2. A **zipped folder** including:
    a. .ibynb executable Jupyter notebook**
    b. 1 PDF document **with the output** from the Jupyter notebook. To include the output, RUN the code before downloading the PDF.

** **Use Google Docs to collaborate**. Start by uploading the Report Template provided in the Course Overview. Once your report is completed, click File → Download → PDF Document (.pdf) to obtain the copy for your submission.*

*** **Use Google Colab or GitHub to collaborate** in completing the executable Python program.*

*The PDF file with your report must be uploaded **separately** from the zipped folder that includes any other types of files. This allows Turnitin to generate a similarity report.*

Grading Rubric

Your instructor will evaluate your group submission using the following rubric:

Quantitative Analysis (open-ended questions) 60 Points	Technical and Non-technical Reports 40 Points	Writing and Formatting 20 Points
The group is able to apply results, formulas, and their knowledge of theory to real-life finance scenarios by doing the following: ● Providing all the necessary information to support their arguments. ● Presenting arguments that reflect group discussion and research. ● Using authoritative references to support a position and provide updated information ● Concluding with practical takeaways for more insightful financial decision-making	Technical Reports contain 3 parts: 1) summary of key results; 2) interpretation of results; and 3) the recommended course of action that can reasonably follow from those results and interpretations. **Note**: Technical reports will include the technicalities of models, such as names, methods of estimation, parameter values, etc. and exclude generalities about the work done. It should NOT include the names of Python code that was used.	A submission that looks professional should include: ● The axes labels and scales in graphs. ● No significant grammar errors or typos. ● Organized, clear structure, and easy to read document. ● Proper citations and bibliography using MLA format.
	Non-technical Reports contain 3 parts: 1) clear explanation of results; 2) the recommended course of action that follows; and 3) the identification of factors that impact each portfolio. **Note**: AVOID all references to model names, algorithms, unnecessary details, and focus on the investment decision.	

Revised: December 10, 2024