# UNIVERSITY OF PRETORIA
## DEPARTMENT OF MATHEMATICS AND APPLIED MATHEMATICS

### WTW 801 - Big Data : Financial Engineering & Financial Mathematics

External Examiners:    Prof. T. Verster (North-West University, Potchestroom)
Internal Examiners:    Dr R Kufakunesu and Dr G van Zyl.(University of Pretoria)

| Surname: **McDonald** | Initials: **CA** |
|---|---|

| Student number: **u16040725** |
|---|

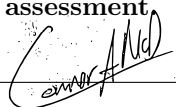| Telephone number: **0711547818** |
|---|

### GRADES:

| TOTAL |
|---|
| |

## Conditions

(1) The deadline for submission is 4 December 2021.

(2) After finishing the project you should upload your .pdf file via the provided ClickUp link.

(3) On the cover page, you should indicate your student details.

(4) Answer all questions.

(5) All sources (including the prescribed texts for this module) which you consult, should be clearly referenced in your project- this should include references to the internet.

(6) Use Python, Matlab, R or any other software. You need to explain your steps and you may put simplifying assumptions where necessary. You may submit your project in a typed version (using for instance Latex or any other word processing software).

(7) Attach the signed plagiarism form on your project.

(9) Your final report should be a PDF document of less than 20 pages, with a reference list.

(10) Please also remember to add the pseudo/code that you used.

### Plagiarism Pledge

The University of Pretoria commits itself to produce academic work of integrity. I affirm that I am aware of and have read the Rules and Policies of the University, more specifically the Disciplinary Procedure and the Tests and Examinations Rules, which prohibit any un-ethical, dishonest or improper conduct during tests, assignments, examinations and/or any other forms of assessment. I am aware that no student or any other person may assist or attempt to assist another student, or obtain help, or attempt to obtain help from another student or any other person during tests, assessments, assignments, examinations and/or any other forms of assessment.

Signature:_____

# The Tasks

1. Consider an asset universe of at least 20 shares (companies) in a common country, with at least 1000 returns observations, for an observation period ending on 31 December 2020. List your data source, data frequency, shares and observation period. Choose a data set for which you can determine the market size (usually calculated as number of shares times price per share) of each of the companies.

   (a)  i. Comment on how "clean" the data set is. Are there many zeroes or blanks? Many repeated values?

        ii. Use Principal Component Analysis (PCA) to reduce the 20 shares to a smaller number $m$ of factors that explain the most variance of the returns. (Choose $m$ with $3 \leq m \leq 10$.) Plot the variances explained (squares of principal values) in decreasing order, for all 20 factors, before they are reduced to the smaller set. Motivate your choice of $m$ and state which percentage of total returns variance is described by your set of $m$ factors.

        iii. Do any of the set of $m$ factors chosen above seem to be related to any economically meaningful quantity such as company size, foreign exposure or anything else?

        iv. Determine the portfolio, investing in your set of $m$ factors, which has the highest return of all portfolios that have a volatility of either 10%, 20% or 30%, choosing a volatility constraint that is compatible with your data set.

        v. Having found the optimal portfolio of factors, how would you approach the practical problem of investing in those factors, which are not in themselves trading on the market? Give some example calculations.

   (b) Did your optimal portfolio, for the period ending 2020, of the question above beat the market benchmark during the period 1 January 2021-30 June 2021? Compare the variance and average return with that of your 20-asset cap-weighted market as benchmark. Which of your set of $m$ factors is this difference to be attributed to?

   (c) Repeat (ii) to (iv) of Question 1a, for the period 1 January 2020 to 31 October 2021. Comment on which factors were more relevant during the Covid-19 period, than before.

   Marks: 5+15+5+5+8+2+10=50 marks.