Mathematics and Programming for FinTech

Question 1

Risky investments are expected to have higher returns because investors are to be compensated for taking up risks. However, is it true that risky investments have higher *realised* returns in the markets?

In this GBA, we will examine if riskier stocks provide higher returns when compared to its lower risk counterparts. We will construct volatility portfolios, i.e. portfolios that are constructed with different risk levels and evaluate the resulting portfolio returns.

Return is formulated as

$$r_i(t) = \frac{S_i(t)}{S_i(t-1)} - 1 \qquad \text{Equation (1)}$$

where $S_i(t)$ and $S_i(t-1)$ are the prices of stock *i* on the day *t* and *t*-1 respectively. Risk is measured using the sample standard deviation of the daily returns, i.e.

$$\sigma_i = \sqrt{\frac{\sum_{k=1}^{k=N} (r_i(k) - \bar{r}_i)^2}{N-1}} \qquad \text{Equation (2)}$$

$\bar{r}_i$ is the mean return of stock *i* and *N* is the number of returns datapoints.

You are given the daily return data of 2000 stocks, from 1994 – 2021 in the file named *stock_returns.csv*.

```
fn = r'stock_returns.zip'
breakpoints = [10, 30, 50, 70, 90] # percentile
```

Figure 1: Python inputs to analysis

Use the pandas module to implement the portfolio construction described in Appendix A. The breakpoints indicated in Figure 1 state that you are to construct 6 portfolios, where the first portfolio (to be labelled as pf_01) consists of all stocks whose volatilities are between the 0 and 10th percentile, the second portfolio (pf_02) consists of all stocks whose volatilities are between the 10th and 30th percentile. The last portfolio (pf_06) will consist of all stocks whose volatilities are greater than the 90th percentile.

You are to present your results using the bar chart shown in Appendix B.

If necessary, state clearly any assumptions you have made in your implementation.

You are to conclude clearly if risker portfolios resulted in higher realised returns is observed in your analysis, using the data given.

For marking and evaluation purposes, you are to document/print the following information in your Jupyter notebook submission:

(a) the name of the main developer for each section of code/function.

(b) the daily returns of s_328 and s_1583 on the last trading date of Apr and Dec 2020.

(c) the volatilities and monthly returns of s_0005, s_1000, s_1875 and s_2000 in Mar, Jun, Sep, and Dec 2019.

(d) the portfolio allocation of stocks s_0123, s_0582, s_1409 and s_1990 in Apr and Oct 2005.

(e) the date (in YYMM format) where the highest realized monthly returns is observed for all the portfolios.

(f) the annualized return and volatility of all portfolios presented in 2 decimal points (i.e. 3.24%) for the entire analysis period.

(100 marks)

You are reminded to practice good programming habits in your implementation and your code must be driven entirely off the inputs given in Figure 1.

Submission Requirements

You are required to submit:

- A Jupyter Notebook (IPYNB) script that fulfill the requirements of the GBA
- A pdf version of the same Jupyter Notebook script for Turnitin evaluation

---- END OF ASSIGNMENT ----

Appendix A

You are given the daily return information of 2000 stocks and using this information, you are to determine the volatility of all the stocks for month *k*, at the last trading date in the month, using the daily return information in the month. In month *k*, the volatility is calculated using Equation (2) and the stocks are then ranked according to their volatility level. The stocks are then sorted into different portfolios, dependent on the breakpoints provided by the user in Figure 1. An equal weighed (EW) portfolio is created using the stocks.

The stocks will be held for a month. In month (k + 1), the return of EW portfolio is realized. The process of sorting stock according to their volatility is repeated and new portfolios will be created. Again, the portfolio will be held for a month.

This process will repeat till the end of the data.

A numerical example is provided for 10 stocks in the universe. To simplify the example, the 10 stocks are assumed to only trade for 3 days in a month. Three portfolios are created with the following breakpoints:

breakpoints = [30, 70]

Daily return of stocks in the universe is presented in Table A1.

	Stock Universe Daily Returns									
	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**
02-Jan-20	-0.15	-0.22	-0.13	-0.20	-0.12	0.24	-0.06	0.09	0.20	-0.12
03-Jan-20	0.10	0.14	-0.21	0.04	0.18	0.06	-0.08	0.18	0.12	-0.06
31-Jan-20	0.06	0.13	-0.21	0.01	-0.12	0.16	-0.12	-0.25	0.24	-0.04
03-Feb-20	0.01	-0.10	0.01	0.00	0.19	-0.12	0.07	-0.14	0.02	-0.03
04-Feb-20	-0.14	-0.16	-0.05	0.18	-0.06	0.07	0.03	-0.20	-0.07	0.20
28-Feb-20	-0.13	-0.04	-0.04	-0.12	0.14	-0.03	0.16	0.06	0.18	-0.15
02-Mar-20	-0.22	0.15	0.10	-0.22	0.19	0.05	-0.18	-0.14	0.14	0.13
03-Mar-20	0.13	-0.23	-0.08	0.03	-0.04	-0.17	-0.25	0.24	-0.13	-0.21
31-Mar-20	-0.07	0.05	-0.16	0.05	0.18	-0.01	-0.11	0.15	0.21	0.17

Table A1: Daily return data of all stocks in the universe

On the <u>31-Jan-20</u>, the volatilities of all the stocks in the universe is calculated and the stock with the lowest volatility is *g*, whose volatility is 3.1%. The stock with the highest volatility is *h*, whose volatility is 22.7%.

The stocks are sorted, and stocks *c*, *g* and *j* are allocated to the first portfolio which corresponds to the first breakpoint of 30th percentile and the portfolio is constructed with equal allocation from the stocks.

Stocks *b*, *e* and *h* are allocated to the third portfolio which corresponds to the breakpoint greater than 70th percentile. Again, the allocation of each stock in the portfolio is 33.33% each.

All the portfolios were held for a month. For stock c in the first portfolio, daily returns of 0.01, -0.05 and -0.04 were experienced in Feb-20. The monthly return for stock *c* in Feb-20 is – 7.9% (=1.01*0.95*0.96 - 1).

Hence, the monthly return for the first portfolio in Feb-20 is 6.3%.

The stock sorting and portfolio construction is repeated monthly, using the daily return of the stock universe. With the data given, the monthly returns of the three portfolios are summarized in Table A2.

	1	2	3
28-Feb-20	6.3%	-4.3%	-9.0%
31-Mar-20	-22.4%	5.8%	3.8%

Table A2: Time series of monthly returns of portfolios

With the given daily return data, 3 portfolios can be constructed on the 31-Mar-20. However, the monthly return of the portfolios constructed on the 31-Mar-20 cannot be evaluated as the data for Apr-20 is not available.

The annualized return of the portfolio can be calculated as follows:

$$r = 12\prod_{t=1}^{t=N}(1+r_t)^{\frac{1}{N}} - 1 \qquad \text{Equation (A1)}$$

where r_t is the time series of the portfolio monthly returns and N is the number of terms in the time series.

The annualized volatility of the portfolio is calculated as follows:

$$\sigma = \sqrt{12}\,\sigma_r \qquad \text{Equation (A2)}$$

where σ_r is the standard deviation of the time series of the monthly portfolio returns.

Appendix B

The results of the analysis are to be represented in accordance with Figure B1 for easy visualization. Note that dummy numbers are used in the generation of Figure B1.

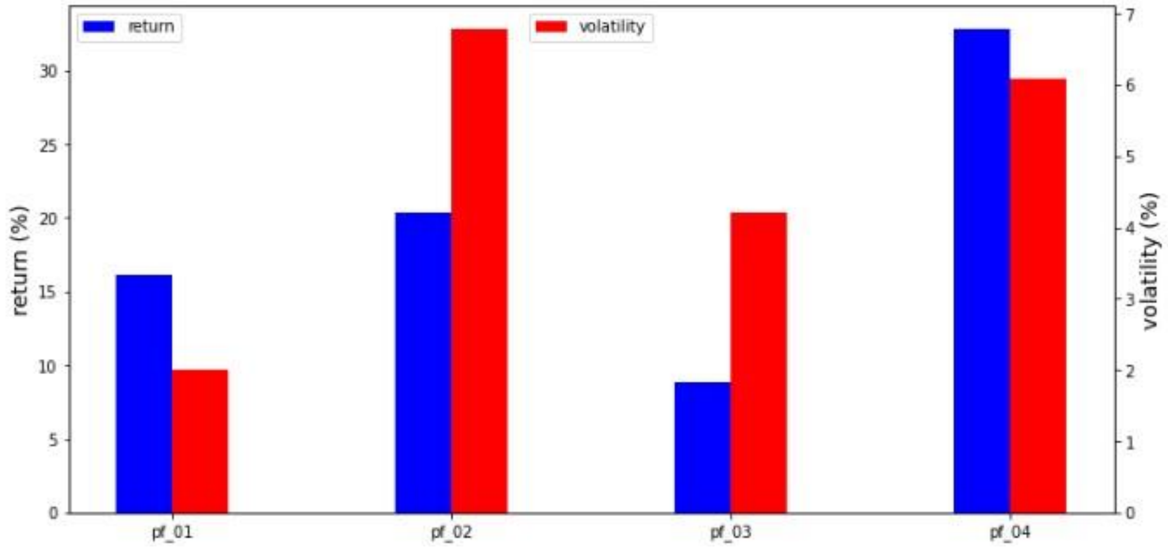

Figure B1: Visualization of analysis results