**Zhang Qihan’s Personal Project Report**

**Monte Carlo based Portfolio Optimisation**

**Project Title: Monte Carlo Risk Assessment of Stock Portfolio**

**Table of Contents**

1. Introduction

* Nature and Objective of Project

1. Background

* Methodology
* Assumptions made

1. Further Improvement to be made

* Feature Engineering
* ARIMA time series modelling
* Isolation Forest
* Long Short Term Memory Models (LSTM)

***Segment 1:***

***Introduction***

This project became as a prove of knowledge and as a personal interest of mine with the intention to delve into the world of quantitative finance from the simplest possible project I can embark on with my limited skillset.

This project is a Monte Carlo based portfolio optimisation product that seeks to quantify the risk and adjusted returns of a given set of stocks chosen for an equities-based portfolio, for which these metrics are quantified with a given portfolio’s Beta (level of volatility in relation to the wider market), and Sharpe ratio (level of adjusted returns in relation to the wider market).

This report will discuss the methodology, functions, limitations and areas of improvement in future projects, for which such improvements will be added within the appendix in future edits of this report.

***Segment 2:***

***Methodology***

Much like any simulation, this Monte Carlo Simulation (MCS) essentially optimises a portfolio under a few major albeit impractical assumptions on the stock market (to be discussed in a later section in this segment). The intended output of this project would be set as the “optimal” weightage in percentage for each stock in a portfolio. The project will further output such weightages and its associated Sharpe Ratio (SR) and Beta, expected equity curve volatility and returns at 3 distinct risk appetites, high low and adequate, noting that these levels are arbitrary and carry minimal “benchmark” other than a comparison between each portfolio within the simulation.

The summarised methodology of project are as follows:

1. Determining Stock Tickers and Time Range of analysis

* These metrics are set as inputs by the user, given the simplicity of the project, time range might have miniscule effects on the overall accuracy of the simulation, such shortfalls will be evaluated in the later sections.

1. Fetching and cleaning Historical Data:

* This is done through yahoo finance, where the only data required is the adjusted close of the given stocks, which is required to mathematically derive returns, covariance and standard deviation of each stock in relation to another.

1. Running Monte Carlo Simulation

* A Monte Carlo simulation is done by generating n numbers of random portfolios with differing weightages. This simulation follows a normal distribution of the given data. (This is an assumption to be discussed in a later segment)

1. Calculating Portfolio Metrics

* Portfolio return is derived through the weighted sum of returns of each individual stock
* Portfolio Volatility is the standard deviation of the portfolio’s returns, representing risk
* Sharpe Ratio is the portfolio’s risk-adjusted return, calculated as the excess return over the risk-free rate divided by volatility.
* Beta measures the portfolios sensitivity against the wider market movements, calculated as covariance of portfolio returns with market returns (SPY) divided by the variance of market returns.

1. Aggregate Results

* Are the results derived from the simulations which include portfolio returns, volatility, SR, Betas, which are then aggregated and stored.

1. Visualising Results

* The program generates visualisations of the simulation results, such as scatter plots of portfolio returns against volatility. This would also allow for simple decision making in terms of mean variance dominance of any particular portfolio.

1. Displaying Metrics (output)

* Metrics such as SR and Betas for each stock at each risk appetite is displayed to users for further analysis.

***Assumptions made***

This project assumes a few major criteria’s which are impractical in the real world, however, this project serves as a basis to build new strategies upon.

Here are the major assumptions made and their associated issues and ramifications:

1. Normal distribution of returns

* The project assumes the individual stocks and the market index follows a normal distribution. Which is never the case as a normal distribution assumes a fair market where stocks are traded at a fair value.

1. Stationarity of Returns

* This assumes that statistical properties of returns remain constant over time.

1. Efficient market hypothesis

* This assumes that the market is efficient which is not the case. Under such hypothesis, it is challenging to consistently outperform the market through active trading or stock selection without first implementing other strategies which can better identify anomalies.

1. Risk Free Rate

* Risk free rate is represented by the US government bond yield, serving as a bench mark for risk adjusted returns.

1. Linear relationship between stocks and markets

* Linear relationship between stocks and market index. This is necessary to calculate beta which measures degree of change (gradient of a linear system)

1. Constant Correlation structure

* Assumes constant correlation between stocks with minimal change, which is seldom the case.

1. No Transaction Cost or Taxes

* This project does not account for taxes and brokerage costs, neither does it consider brokerage leakage or other inefficiencies in the market.

***Segment 3***

***Future Improvements to be made***

At the point of creation of this documentation, there are 3 main features that are currently in the works of implementation along side this project in further improving risk management. They are as follows:

1. ARIMA time series model

* This is being used along-side the alternative strategy of LSTM model for time series forecasting. Essentially providing a better alternative to the linear nature of the current project by forecasting multiple scenarios of future stock prices. Furthermore, the residuals provided by ARIMA (difference between observed and predicted returns) can be analysed to identify patterns and autocorrelation that may not have been previously captured, helping refine the simulation model with “less” data. This can then allow for better simulation that closer resembles historical data.

1. Isolation Forest Algorithm

* This machine learning algorithm would serve to detect anomalies in returns, particularly in data points that deviate significantly from expected behaviour. This eliminates unexpected events such as rug pulls, market crashes or data errors from brokerages.
* This algorithm will further serve to identify and filter extreme scenarios within MCS which may not be a good representation of typical market scenarios.

1. Long Short Term Memory (LSTM)

* This serves as a recurrent neural network that is suited for sequential data modelling. This neural network serves as a more reliable way of predicting prices compared to ARIMA as LSTM models generally have better ability in finding correlation between complex patterns in stock prices.
* Additionally, LSTM models can better capture volatility and anomalies.
* However, LSTM tends to lose accuracy given repeated prompts with limited data due to overfitting.
* (At this current juncture, I do not have the confidence nor expertise to create an LSTM model from scratch, requiring extensive help from GPT 4)

Thank you for reading through my personally curated documentation/report of my NUS year 1 Summer 2024 projects. This document will be constantly updated.