

Analysis of Growth Performance of Actively and Passively Managed Retirement Portfolios



Western Governors University

Table of Contents

A. Proposal Overview	5
A.1 Research Question or Organizational Need	5
A.2 Context and Background.....	5
A.3 and A3A Summary of Published Works and Their Relation to the Project.....	5
Review of Work 1	5
Review of Work 2.....	5
Review of Work 3.....	6
A.4 Summary of Data Analytics Solution	6
A.5 Benefits and Support of Decision-Making Process.....	7
B. Data Analytics Project Plan.....	7
B.1 Goals, Objectives, and Deliverables.....	7
B.2 Scope of Project	8
B.2.A Included in Project Scope.....	8
B.2.B Not included in Project Scope	8
B.3 Standard Methodology	8
B.4 Timeline and Milestones	8
B.5 Resources and Costs.....	9
B.6 Criteria for Success	9
C. Design of Data Analytics Solution.....	9
C.1 Hypothesis.....	9
C.2 and C.2.A Analytical Method.....	9
C.3 Tools and Environments.....	9
C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance	10
C.5 Practical Significance.....	10
C.6 Visual Communication.....	10
D. Description of Dataset.....	11
D.1 Source of Data.....	11
D.2 Appropriateness of Dataset	11
D.3 Data Collection Methods	11
D.4 Observations on Quality and Completeness of Data.....	11
D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances	11
E. Sources and References	12
E.1 Sources	12
E.2 References	12

A. Proposal Overview

A.1 Research Question or Organizational Need

When comparing historical data of a group of Target Date Retirement funds against a group of index fund Exchange Traded Funds (ETF), is there a noticeable difference in theoretical returns? When conducting simulations of contributions to these funds, do the practical results (simulation) reinforce the theoretical results?

A.2 Context and Background

Using tax-advantaged retirement accounts to invest in "Target Date Retirement Funds" or index fund ETFs can both be considered viable "hands-off" approaches when it comes to saving for retirement. The "hands-off" and "set it and forget it" philosophies entail contributing to a fund (or fund portfolio) on a consistent long-term basis. Using this approach, there is typically no selling of securities until it is time for an overall portfolio rebalance as an individual nears retirement. This practice applies the concept of having "time in the market" rather than "timing the market", under the assumption that, generally, the value of stocks will increase over time. If there is a significant difference in performance in a certain group of funds compared to the other, it may be worth exploring options within that group of funds in order to maximize potential returns over time.

A.3 and A3A Summary of Published Works and Their Relation to the Project

Review of Work 1

"Passive versus Active Fund Performance: Do Index Funds Have Skill?"

In this informative research journal, the authors (Crane & Crotty, 2018) conduct in-depth analytics of various active funds and passive funds to determine how they compare to one another. By analyzing the underlying holdings of various funds from each category and exploring hundreds of funds with decades of historical data, the authors compare the performance of both types of funds against industry benchmarks to determine which funds perform better overall and attempt to discover why certain funds tend to outperform others. This journal also further decomposes passive index funds into Exchange Traded Funds (ETFs) and open-ended mutual funds, for which there is some kind of active management involved, and subsequently fees. In section IV-A, they demonstrate that, under some statistical testing, while the worst performing index funds outperform actively managed funds, the marginal difference in performance decreases as you compare higher performing funds from each category, with a notable difference at the 75th percentile. Nonetheless, through their rigorous testing, the authors conclude that overall, index funds outperform actively managed funds to such a substantial degree that they determine the overall risk of investing in one category to always be higher in actively managed funds compared to passive index funds.

Review of Work 2

"The investor's choice between active and index funds."

In their (Mingo-Lopez & Matallin-Saez, 2018) study, the authors cite various works related to the topic at hand, performance of index funds compared to performance of mutual funds. From their research, their findings show that some specific mutual fund groups may perform better than some index funds. They also discuss contrasting conclusions, stating that an actively managed fund cannot beat an index of its own market, on average, due to the fact that the index would include the actively managed fund itself. Their research into existing literature on this subject highlights that this is still a debated and somewhat subjective topic, as a result of the complexity of the subject at hand. The authors conduct their own research and testing, utilizing a Sharpe ratio, to compare market returns against a risk-free return (typically the rate of federally guaranteed bonds as it is widely considered to be risk-free). Subsequently, they analyze gross returns followed by net returns of each type of fund. Their findings indicate that portfolio rebalancing and turnover greatly affect the returns, when adjusting for risk. They concluded that actively managed funds with higher turnover netted less returns than comparable index funds; however, actively managed funds with lower turnover could outperform their index fund counterparts. Their exploration highlighted a potential cause for overall differences in performance as well as a way to segment funds for more in-depth analysis.

Review of Work 3

“Testing the Predictability of Stock Returns”

This enlightening journal from Lanne (Lanne, 2002), attempts to gauge how predictable stock market investment returns can be through various testing techniques. The author notes that while the t-test may suggest predictability overall, the t-test tends to over-reject the null hypothesis, based on the sample empirical data. For our study, we are not necessarily attempting to predict investment returns in our project; rather, we are conducting analysis on past performance over a set period of time. This may cause us to ponder about future performance; however, past performance does not always indicate future performance. I believe that key takeaway from this article would be that there is always a certain type of unpredictability when it comes to the stock market and one should take caution when attempting to use certain variables for prediction of future returns.

A.4 Summary of Data Analytics Solution

Throughout this project, I will conduct analysis of stock market returns of actively managed funds and passively managed funds. The overall goal is to determine whether there is a significant difference in performance of these categories. I will utilize historical stock market data for various funds in each category. To present this analysis, a Jupyter Notebook will explore various attributes of each dataset, providing multiple visualizations throughout each step of the analytical process. As long as our resulting datasets follow a normal distribution, which is expected, a t-test should be sufficient to help determine whether a certain category of funds resulted in a significantly higher return compared to the other category. If the distribution of data doesn't follow the normal distribution as well as expected, other strategies can be applied, including bootstrapping, to reliably compare the different distributions.

A.5 Benefits and Support of Decision-Making Process

There are countless investment options available to those seeking to grow their savings for retirement. This fact, combined with the overall complexity of business, finance, markets, and taxes, all of which are integral to this subject, can make funds an attractive investment option compared to investing in individual securities. The concept of having experienced personnel actively manage a fund can be appealing; however, it does come at obvious cost, having higher expense ratios. Looking at nearly any historical chart of fund's performance, readers may find that "past performance is not a guarantee of future performance"; however, this historical performance is always factored in at some point which conducting analysis for investing, using both technical and fundamental analysis. This project's analysis can be a useful stepping stone in determining whether to invest in actively managed funds passively managed funds for long-term investment.

B. Data Analytics Project Plan

B.1 Goals, Objectives, and Deliverables

Goal : The goal of this project is to compare performance of various stock market securities falling into two categories, actively managed funds and passively managed funds, in order to determine whether there was a significant difference in performance over a set period of time.

- Objective 1.1: Data Retrieval
 - Deliverable 1.1.1: Downloaded and cleansed a dataset containing historical passively managed fund closing prices.
 - Deliverable 1.1.2: Downloaded and cleansed a dataset containing historical actively managed fund closing prices.
- Objective 1.2: Simulate Investments
 - Deliverable 1.2.1: Dataset containing simulated investment contributions into passively managed funds at historical closing prices.
 - Deliverable 1.2.2: Dataset containing simulated investment contributions into actively managed funds at historical closing prices.
- Objective 1.3 Conduct Statistical Analysis of Datasets
 - Deliverable 1.3.1: Descriptive summary statistics, t-test results, and visualizations of historical stock market closing prices, showing potential returns.
 - Deliverable 1.3.2: Descriptive summary statistics, t-test results, and visualizations of simulated returns.

B.2 Scope of Project

B.2.A Included in Project Scope

The scope of this project will include data downloaded from yfinance for selected ticker symbols over a defined date range, 2000-2023. Further analytics will determine a more finite date range based on the availability of data for the selected ticker symbols. Only the closing price for each symbol will be used for analytical purposes.

B.2.B Not included in Project Scope

Only the ticker symbols that have been selected for this project will be analyzed. There is an abundant amount of other ticker symbols available for each fund category; however, this project is focusing on some of the more popular options for each fund type. Since only the closing price is being utilized in this project, all other variables, including volume, open price, expense ratios, and market sentiment, will be excluded from analysis.

B.3 Standard Methodology

The learning and development methodology used for this project is the ADDIE framework. The acronym ADDIE can be expanded to Analyze, Design, Develop, Implement, and Evaluate. When conducting research for this project, I was on the Analyze step. This document itself is part of the Design process. While developing the application to retrieve the necessary datasets and apply transformations, I will be in the Develop phase of the process. The final analytics and testing process is part of the implementation phase. The evaluation will be done through additional papers and a video presentation that are part of Task 3.

B.4 Timeline and Milestones

I will abide by the following timeline for achieving the objectives of this project. The first step that must be conducted is data acquisition. Once data is collected, I can begin simulations as well as analysis. Some analysis can be conducted simultaneously as it only relies on the original datasets; however, final analytics will need to be completed after all simulated datasets are finalized.

Milestone or deliverable	Duration (hours or days)	Projected start date	Anticipated end date
Retrieve and clean datasets (Objective 1.1)	8 hours	8/29/2023	8/29/2023
Simulate investment contributions (Objective 1.2)	24 hours	8/29/2023	8/30/2023
Conduct analysis and testing, including visualizations (Objective 1.3)	48 hours	8/29/2023	8/31/2023

B.5 Resources and Costs

1. Market data from yfinance: free / 0 cost
2. Python programming language: free / 0 cost
3. Jupyter notebook and data science libraries: free / 0 cost
4. Human labor (self): free / 0 cost

There are no expected costs for this project. The data and programming tools are free for this use. I will be conducting all labor, which will incur no cost as well.

B.6 Criteria for Success

To determine whether this project is successful or not, I will not rely on whether there is a significant difference in performance between the separate groups of funds; rather, that our t-test and p-values align with associated visualizations of our data distributions. Regardless of whether the t-test shows significance, the visualizations of the different data distributions to provide further evidence of the result. In order to call this a success, accurate visualizations need to be produced and p-values will need to be generated from the t-tests that are conducted.

C. Design of Data Analytics Solution

C.1 Hypothesis

My hypothesis is that there will be a significant difference in performance between the two groups of funds, showing that index fund ETFs will generally outperform actively managed mutual funds.

C.2 and C.2.A Analytical Method

For the 'market close price' datasets and the 'simulated investment contribution' datasets, if the data follows a normal distribution, I will utilize a t-test to determine whether there is a significant difference in performance. If the analysis shows that the datasets do not follow a normal distribution, the bootstrapping method will be utilized to determine whether there is a significant difference in the mean of each dataset.

Rubric C2A: Justification of Analytical Method

When making comparisons between normal distributions, the t-test is an appropriate method to choose. It is considered to be efficient and sensitive to differences between normal distributions, making it a top choice for this distribution. In the case of non-normal distributions, bootstrapping can also be an efficient approach to compare sample sizes of different distributions since this approach does not make any assumptions about the data and does not require the data to follow a normal distribution.

C.3 Tools and Environments

For this project, I will utilize the Python programming language for data extraction, cleansing, analytics, and visualizations. I will also utilize a Jupyter notebook environment to

execute the application, which will allow me to display output and visualizations along with the code.

C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance

For my statistical testing, I will utilize a p-value to determine if there is any statistical significance. Our alpha will be a commonly used value, .05. To apply this to the aforementioned hypothesis:

- If the p-value meets or exceeds the alpha of .05, then we can conclude that there is no statistical significance in difference of performance between the two groups of funds. This would indicate accepting the null hypothesis.
- If the p-value is less than the alpha, then the conclusion would point to showing statistical significance in performance of the two fund groups that are compared. This would mean we can reject the null hypothesis.

The p-values will be determined from conducting t-tests of our datasets during the implementation of the program.

Rubric C4A: Justification of Metrics

The p-value is an appropriate means to measure the differences in central tendencies of the datasets being compared. The method of deriving a p-value may be different when using a t-test vice bootstrapping; however, the p-value that is derived will still be an appropriate method to determine statistical significance.

C.5 Practical Significance

While this study is not intended to provide any financial, tax, or legal advice or recommendations, it can be a useful starting point when determining where to invest retirement funds in order to maximize growth. Depending on the results of the process, we may find that there may not be a significant difference between the different fund types or we may conclude otherwise. These results can help guide us and inspire us to conduct further and more complex analysis in order to ensure we utilize the best approaches available to us when it comes to saving for retirement.

C.6 Visual Communication

Our project report will provide multiple visualizations of the various datasets we utilize. Some visualizations will be determined based on our findings during the implementation phase; however, at a minimum, I will be sure to include the following visualizations:

- Histograms – Layers of histograms will be included to present the various distributions in an appropriate method for visual comparison.
- Boxplots (Box and Whiskers Plot) – Boxplots will help compare various distributions side-by-side in such a manner that we can see the central tendency as well as the overall spread of each distribution.
- Bar Charts – Bar charts will be useful to help visualize returns by individual ticker symbol once minimal analysis has been conducted.

In order to produce this visualization, I will utilize Python libraries such as Matplotlib and Seaborn. To present the visualizations, I will include them in the aforementioned Jupyter notebook, which is an appropriate choice for this project.

D. Description of Dataset

D.1 Source of Data

The source that I will use to retrieve market data for this project is the Python yfinance library. This library uses Yahoo Finance stock market data, abstracting the retrieval process to facilitate ease of use. This data is free to utilize and acceptable to use for educational and research purposes.

D.2 Appropriateness of Dataset

The stock market dataset is appropriate for this project, as actual historical data is the best way to determine if there was significant differences in performance. The most critical components of our dataset would be the Date, Ticker, and Close attributes. This allows us to determine the closing price of a security on a given date. All other attributes will be excluded for our testing and analytics.

D.3 Data Collection Methods

The data will be collected by utilizing the yfinance library in Python. I will build python functions to accept parameters specifying a time range as well as ticker symbols, and will retrieve the historical market data for that ticker over the specified time range. This process will be repeated for all ticker symbols targeted by the project.

D.4 Observations on Quality and Completeness of Data

Initial analysis on the quality and completeness of this data shows that the data is already in great condition for our use. The initial data is pulled into a dataframe where the Date is the index. Since I will be aggregating all datasets for a given group of funds, I will need to reset the index to include a Ticker symbol, creating a composite index. Finally, all attributes other than the Close price will be dropped when conducting final analysis.

D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory

Compliances

The data being used is free and publicly available for education and research purposes. The data contains attributes related to the stock market for specific Ticker symbols over a specified period of time. The data also contains no personally identifiable information and there are no security, privacy, or ethical obligations of concern regarding the content of the data. The Yahoo Developer API Terms of Use does state that to accept the terms and conditions, the user must be at least 13 years old, possess a valid Yahoo account, and maintain responsibility for compliance with their terms. I have ensured that I meet all criteria for use.

D5A: Precautions

In order to prevent any rate limits that Yahoo Finance may enable, the Python yfinance library allows sessions and caching to prevent any unnecessary requests. To help prevent any requests blocks and data corruption, I will be sure to implement these features in the program.

E. Sources and References

E.1 Sources

Project Data Source: <https://pypi.org/project/yfinance/>

Yahoo Developer API Terms of Use: <https://legal.yahoo.com/us/en/yahoo/terms/product-atos/apiforydn/index.html>

E.2 References

Crane, A. D., & Crotty, K. (2018). *Passive versus Active Fund Performance: Do Index Funds Have Skill?* *Journal of Financial & Quantitative Analysis*, 53(1), 33–64.
<https://doi.org/10.1017/S0022109017000904>

de Mingo-López, D. V., & Matallín-Sáez, J. C. (2019). *The investor's choice between active and index funds*. *UCJC Business & Society Review*, 61, 38–59.
<https://doi.org/10.3232/UBR.2019.V16.N1.02>

Lanne, M. (2002). *Testing the Predictability of Stock Returns*. *Review of Economics & Statistics*, 84(3), 407–415.
<https://doi.org/10.1162/003465302320259439>