

# Predicting Mutual Fund Prices: A Linear Regression Analysis

J. Scotty Solomon

**Abstract**—This essay examines the effectiveness of Linear Regression models in predicting mutual fund prices, focusing on 800 actively managed funds in the United States market. The study reveals that, on average, these funds exhibit low returns, questioning their reliability as investments. Using standard linear regression, as well as polynomial and lagged linear regression models, the research finds that lagged linear regression outperforms others, indicating varying predictability among different funds. However, linear models struggle to capture the complex fluctuations in fund prices over time.

**Index Terms**— Retirement Planning, Investment Funds, Actively Managed Mutual Funds, Historical Financial Data

## 1 INTRODUCTION

The United States' approach to retirement planning relies almost entirely on the stock market.<sup>1</sup> Retirement investing typically involves individuals consistently contributing to investment funds, gradually accumulating a growing number of shares in a fund over time. Upon reaching retirement age, it is advisable for individuals to withdraw a portion of their investments to support their living expenses, allowing the remaining funds to continue accruing gains. Experts suggest that each American needs at least one million dollars to retire.

Most retirement plans include some type of investment in actively managed mutual funds, though they're not strictly required. Actively managed mutual funds are curated by investment managers. The performance of mutual funds is often compared to the performance of the S&P 500. There are several types of actively managed mutual funds. *Equity funds* are made up entirely of stocks. They usually have the most risk associated with them. However, they have the capacity for the highest (and lowest) return. *Fixed income funds* are made up entirely of bonds, meaning that they have less risk than equity funds, but they typically have lower returns. *Mixed funds* or *hybrid funds* invest in a combination of stocks and bonds. They can have varying degrees of risk depending on the split between bonds and stocks that the fund employs.

The S&P 500 (Standard and Poor's 500) is a stock market index that acts as an indicator of the performance of the United States' stock market. It is maintained by the S&P Dow Jones Indices and consists of 500 of the largest publicly traded companies in the United States. In a broad sense, these companies are selected to represent the performance of the U.S. economy. It is often used as a benchmark or representative sample of the stock market's overall performance. Investors typically compare their portfolios to the S&P 500 to understand how their investment returns compare to the broader market. The S&P 500 typically experiences varying annual returns, with historical averages suggesting an average annual increase of approximately ten percent, though it can fluctuate significantly from year to year.

*Actively managed mutual funds* are funds that are

specifically managed by professional portfolio managers who make decisions about which stocks or bonds to invest in. Most actively managed mutual funds attempt to outperform the S&P 500, since the S&P 500 is typically used as a comparison index. In contrast, *index funds* (or *passive mutual funds*) typically attempt to replicate the S&P 500 by holding a portfolio of stocks that matches the index's composition. Actively managed mutual funds tend to have higher fees associated with them, whereas index funds tend to have lower fees due to minimal maintenance costs.

The current research objective is to determine if mutual fund prices can be predicted via observing repetitive patterns within the fluctuation of the fund prices themselves. These predictions would be made without any reference to current market conditions.

### 1.1 Related Work

*Luck versus Skill in the Cross-Section of Mutual Fund Returns* examines the performance of actively managed mutual funds. The journal questions whether the higher cost associated with actively managed mutual funds justify their returns and concludes that the net returns of actively managed funds are not worth it when compared to index funds due to expense ratios.

*Price and Performance in the Mutual Fund Industry* explores the fact that not only do investors continue to invest in actively managed mutual funds despite poor performance when compared to index funds, but poor performing funds also tend to have higher fees. While the study does not come to a single conclusion of why this might take place, it suggests that funds with worse performance might have less sensitive investors who stick with them despite the bad performance, which incentivizes the funds to raise their fees. Additionally, investors with less of a concern of short-term losses may choose to stick with funds despite their low performance in the hopes of the fund eventually performing well.

*Incentive Fees and Mutual Funds* defends both the fees charged by actively managed mutual funds. The journal article argues that incentive fees attract better fund managers than funds that do not offer incentives. The journal argues that funds with incentive fees tend to have lower beta scores, meaning that the fund prices tend to fluctuate less than the market benchmark while those that do not use incentive fees have beta scores higher than 1.

<sup>1</sup> There is typically an emphasis on a diversified portfolio within retirement planning, with most investment funds including bonds or other fixed-income investments to balance the risk of stocks and produce stability. However, on their own, bonds alone typically do not generate sufficient returns for individuals to retire comfortably. Thus, investment in the stock market through vehicles such as a 401(k) or brokerage account is often deemed necessary for the vast majority of Americans.

*Evaluating Mutual Fund Performance* explores different metrics that can be used to evaluate mutual funds.

*Should Investors Avoid All Actively Managed Funds?* explores whether actively managed mutual funds should be used by investors at all. The journal fails to explore the cost analysis of actively managed funds and instead uses Bayesian reasoning to determine that actively managed funds are viable.

*What is Retirement Planning? Steps, Stages, and What to Consider* discusses the key ideas behind retirement planning in the United States.

## 2 EXPLORATORY DATA ANALYSIS

Accessing free historical financial data for mutual funds or stocks is a somewhat challenging task. Search results often return performance graphs, but the numerical data itself is not easily available to the public. Companies like Morningstar lock historical data behind paywalls, and even with a paid account, data is displayed in such a way that prevents web scraping while also not allowing users to download historical data in any meaningful way.

The initial approach to data collection was to web scrape historical data from Yahoo Finance. The website allows users to navigate the historical data page of a fund and select filters. Upon further investigation, Yahoo Finance allows users to request this historical data into a csv, which is likely a remnant of the Yahoo Finance Developer API that has been discontinued. With this functionality, csv files were requested via a Python script that requests a csv for a fund and sets a filter for historical data for the first of every month. The historical data of each fund dates from March of 2010 to September of 2023, with funds of a shorter lifespan beginning the month after its creation. An example of an entry from a retrieved csv can be found at Table 1.

Table 1

Date	Open	Close
01-03-10	11.05	11.34

Table 2. Example data entry for a requested csv from Yahoo Finance. This example comes from the AABFX fund.

The URL that requests a csv requires the ticker symbol for the fund. In order to make requests for specific tickers, a data set created by Jeroen Bouma, a financial risk analyst, was used. This data set contained the ticker symbols of various countries, as well as demographic information, such as the fund manager, the country the fund is based in, and the type of fund. A collection of 800 actively managed mutual funds were randomly selected from the total list of U.S. based funds. An example of an entry can be seen below:

Table 2

Symbol	Category	Market
AAAAX	Equities	U.S. Market

Table 1. Example data entry from the list of mutual funds created by Jeroen Bouma.

After early evaluation, the representative sample suggested that, on average, the returns from actively managed mutual funds are quite low. From the 800 funds

selected, the average appreciation from March of 2010 to September of 2023 was 1.1%, with the lowest appreciation being a decrease of 80.4%. This translates to the median lifetime appreciation in a fund's price being only thirteen cents. Further analysis suggests that year to year returns are typically between a decrease of ten percent and an increase of ten percent, with most funds congregating around an appreciation rate of 0%. See Figure 1 for the breakdown of median yearly returns. This is typically because year-to-year returns of managed funds tend to fluctuate, with some good years and some bad years.

Based on these results, actively managed do seem to be a questionable investment. However, it is important to note that higher-performing actively managed funds do exist and can perform well. As seen in Figure 1, a small selection of funds tend to appreciate ten percent or more each year. Still, these well-performing funds tend to be few and far between.

Figure 1

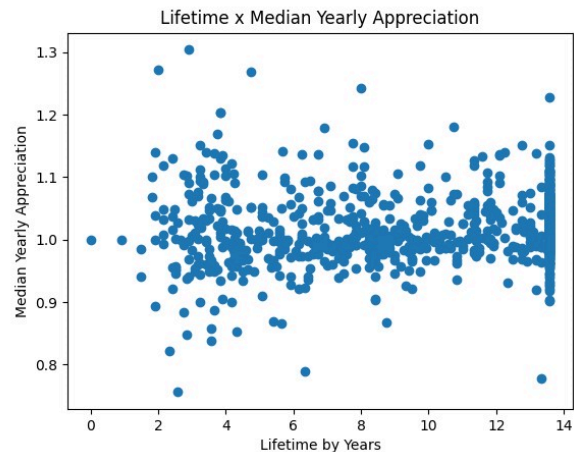


Fig. 1. Depicts the relation between median yearly appreciation and fund lifetime.

## 3 METHODOLOGY

Linear Regression was used as the primary predictive model for this project. The models were trained using the Sklearn Python module. Because the current date is used as an input, fund data was not randomly divided between training and testing data. Instead, the first eighty percent of the fund's historical data was used for training data and the remaining twenty percent of the fund's historical data was used for testing.

Four models were used throughout the project. The first three models were constructed for each individual fund. The first model used linear regression with the date as an input and the fund value as an output. The second model used polynomial regression with a degree of two. The third model used lagged linear regression, with the previous month's price being used as an input for the current month. The last model was a single model that was used across all the funds. This model used the date, monthly pricing, and net appreciation as inputs and the appreciation for the next month as an output. The mean squared error and mean absolute error were the primary metric used for evaluating model performance. Mean squared error was chosen as a metric, since this compares

the difference between the predicted fund value and the actual fund values. Mean absolute error was chosen, since it indicates the average dollar amount that each prediction was off by.

## 4 RESULTS AND DISCUSSION

Of the four models, polynomial regression performed the worst. The median squared value was 8.09, with the highest value going to 23044.6. The median mean absolute error for each model was \$2.57, with the highest error being \$148.29. Although an error of \$2.57 is not very high, it is important to note that most fund prices are between \$15 to \$25, meaning that a median error of \$2.57 per share vastly inflates the project gains an investor might achieve. Linear Regression performed better; however, it still suffers from similar issues, with the median mean square error across all models being 4.53 with a maximum of 10448.64, and the median mean absolute error being \$1.92 with a maximum of \$99.98.

Lagged linear regression seems to be the best among the individualized models. The median mean squared error was 0.34 with the high being 260.84 and the median mean absolute error being \$0.47 with a high of \$14.17. Although \$14.17 is a relatively high margin of error, it is much better than the marginal errors when compared to polynomial regression and standard linear regression. Still, these results suggest that some fund prices are simply easier to predict than others, typically because the funds themselves already have a relatively simple linear pattern. These easier patterns usually take the form of funds that consistently increase or decrease by a small, marginal amount. Funds that tend to fluctuate in price are not as easy to predict and are therefore not ideal for the lagged linear models.

One fault of applying linear regression is the model assumes that the fund prices follow a linear pattern, which isn't necessarily the case. Although most funds have rising and falling prices, over time the value of the fund itself tends to appreciate, even marginally. Linear models struggle to replicate this fluctuation in price. Instead, the models tend to predict minimal returns and do not predict any sort of fluctuation where a price goes down and then back up. The only instance in which a reduction in price is predicted is when the fund has already been decreasing in value, so the model predicts a consistent downward trend. Creating models to use across multiple funds might seem that it would help to circumvent this issue; however, the issue at large with linear regression would not be addressed, since there would still be an assumption of a linear relationship.

## REFERENCES

- [1] Baks, Klaas P., et al. "Should Investors Avoid All Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation." *The Journal of Finance*, vol. 56, no. 1, 2001, pp. 45–85.
- [2] Carhart, Mark M. "On Persistence in Mutual Fund Performance." *The Journal of Finance*, vol. 52, no. 1, 1997, pp. 57–82.
- [3] Elton, Edwin J., et al. "Incentive Fees and Mutual Funds." *The Journal of Finance*, vol. 58, no. 2, 2003, pp. 779–804.
- [4] Fama, Eugene F., and FRENCH, Kenneth. "Luck versus Skill in the Cross-Section of Mutual Fund Returns." *The Journal of*

*Finance*, vol. 65, no. 5, 2010, pp. 1915–47.

- [5] Gil-Bazo, Javier, and Pablo Ruiz-Verdú. "The Relation between Price and Performance in the Mutual Fund Industry." *The Journal of Finance*, vol. 64, no. 5, 2009, pp. 2153–83.
- [6] Investor.gov "Understanding Fees" *Investor.gov*. Retrieved from <https://www.investor.gov/introduction-investing/getting-started/understanding-fees>
- [7] Kagan, J. "What Is Retirement Planning? Steps, Stages, and What to Consider," Investopedia. 2023. Retrieved from <https://www.investopedia.com/terms/r/retirement-planning.asp#:~:text=Retirement%20planning%20refers%20to%20financial,money%20with%20certain%20tax%20advantages.>
- [8] Kothari, S. P., and Jerold B. Warner. "Evaluating Mutual Fund Performance." *The Journal of Finance*, vol. 56, no. 5, 2001, pp. 1985–2010.