**Quantifying “Risk” of Mutual Funds/ETFs**

Alejandro de la Borbolla (aborbo@mit.edu, ID: 903436910)  
Adrian Oesch (adrianoe@mit.edu, ID:971024139  
Tim Felix Schmedding (felix877@mit.edu, ID:910325402)  
Harsh Kumar ( )

**Introduction**

**Risk Profile Analysis Based on Asset Class and Fund Design**

In this section, we explore how the type of fund and the asset classes it invests in affect a fund’s risk profile. We analyze the universe of large mutual funds in the United States from 2010 to 2024. We use data from CRSP Mutual Fund Database obtained from the WRDS portal.

The database has information on more than 50,000 funds over this period. However, most of them are small funds with poor data quality. We restrict our sample to funds that have total net assets above 10 billion dollars during at least one year. Additionally, we only analyze funds that have complete information on total assets and on portfolio composition across asset classes. This reduces our sample to 698 funds, out of which 161 are exchange traded funds (ETFs).

We use information on monthly returns provided by CRSP that include the change in the net asset value and dividends paid out by the fund. We calculate the standard deviation of these returns. Graph 1 shows the histogram of the standard deviation of the returns across all 698 funds. The distribution seems to be bimodal. In the following analysis, we will try to uncover the reason behind this type of distribution to better understand the sources of risk in these funds.

**Graph 1**  
A graph of a number of columns

Description automatically generated

We first analyze if there is a difference in the risk profile between index and non-index funds. A fund is considered an index fund if it utilizes indexes as its primary filter for the purchase and sale of securities. This includes pure and non-pure index funds, as well as enhanced index funds. Graph 2 shows the overlapping histograms of index and non-index funds’ return volatility. Most of the funds in our sample are index funds. There is no clear difference in the standard deviation of both types of funds.

**Graph 2**  
A graph of a number of columns

Description automatically generated

In Graph 3 we explore if ETFs have a different risk profile than regular mutual funds. The histogram of the volatility of returns shows that there is not a clear difference between them.

**Graph 3** A graph of a number of columns

Description automatically generated

In Graph 4 we split the sample into funds that can be accessed by retail investors and those targeted to institutional investors. Again, the histograms have a similar shape.

**Graph 4**A graph of a number of columns

Description automatically generated with medium confidence

Finally, we explore the risk of the funds based on the asset classes they invest in. We only have information of the allocation of each fund at a broad level. On average, funds in our sample invest 65% of their portfolio in common stocks, 12% in government bonds, 9% in corporate bonds, and the rest in other asset classes such as mortgage backed securities, municipal bonds, cash, etc. In Graph 5, we plot the histogram of the return volatility of funds that invest more than 50% of their assets in common stocks and those that invest less than 50% in stocks. The graph shows that this characteristic explains the bimodal distribution of funds’ return volatility. Funds that invest mostly in stocks have riskier returns than those that invest in other asset classes.

**Graph 5**A graph of a number of columns

Description automatically generated

To understand how different fund characteristics affect the fund’s risk profile, we estimate regressions to explain the fund’s returns standard deviation. The regression takes the form:

where is the standard deviation of fund’s *i* monthly returns from 2010-2024. For the first model, is a vector of explanatory variables that includes the percentage of the funds assets allocated to i) stocks, ii) government bonds, and iii) corporate bonds. Since the rest of the asset classes are omitted to avoid multicollinearity, the coefficients of these variables should be interpreted as the effect of investing on one of these asset classes relative to the omitted asset classes.

For the second model, besides the asset allocation variables, includes

* Natural logarithm of Total Net Assets to understand the effect of a fund’s size
* Retail fund dummy that shows the effect of retail funds relative to institutional funds.
* Index fund dummy that shows the effect of a fund that tracks an index relative to the rest of funds.
* ETF dummy that shows the effect of exchange traded funds relative to the rest of funds.
* Concentration index that is a Herfindahl index that is calculated as the sum of squares of the percentage allocation to each asset class. If one fund invest 100% of its assets to common stock, the index will reach the maximum level of . If it has 50% in stocks and 50% in mortgage backed securities, the index will be . Therefore, the higher the index, the higher the concentration.

For the third model, we include the mean return of the fund over our sample period. Since excess returns are a compensation for risk, we would expect a positive relationship. However, it is unclear if the rest of the independent variables will retain explanatory power after controlling for the average stock return.

The results are shown in Table 1. The R-square of the first column shows that the asset allocation variables explain nearly two thirds of the total variance of returns standard deviation across funds. Consistent with the histogram, investing in stocks increases a fund’s risk, while investing in government or corporate bonds reduces risk relative to the rest of the asset classes. All the effects are highly significant. A one percentage point increase in the stock weight in the portfolio increases the returns standard deviation in 0.029 percentage points. Interestingly, the effects of corporate and government bonds are similar to each other. This might be explained by the fact that corporate bonds have lower duration, which offsets their higher credit risk.

The second column shows that adding additional variables has a small effect on the R-square of the model. The results show that funds with a large size, measured by total net assets, have slightly lower risk. This might be due to increased diversification, but considering that all funds analyzed are large funds, this explanation is not completely compelling. Additionally, ETFs have a slightly higher return volatility compared to the rest of funds. Being a retail fund or an index fund does not have a significant effect on risk. This is especially interesting for index funds given that one might expect actively managed funds to be more risky than passive index funds given the very different investment styles. Finally, our concentration index does not have a significant coefficient. It is important to mention that this only measures concentration at the asset class level, and not at the securities level.

Finally, the third column incorporates the fund’s mean return. As expected, its coefficient is positive and highly significant. However, even if it dampens the effects of other variables such as the stock allocation, in terms of sign and statistical significance, the results for the rest of the variables are similar. This suggests that a fund’s return volatility is not solely explained by the average return that fund has delivered in the past. Therefore, as investors, we can choose funds with lower risk and the same expected returns based on the characteristics analyzed.

**Table 1. Regression Results**

|  |  |  |  |
| --- | --- | --- | --- |
| Dependent variable:  Return Standard Deviation (%) | Model 1 | Model 2 | Model 3 |
| Stocks (%) | 0.029\*\*\*  (15.944) | 0.028\*\*\*  (14.916) | 0.016\*\*\*  (7.618) |
| Government Bonds (%) | -0.0085\*\*\*  (-3.063) | -0.010\*\*\*  (-4.185) | -0.0091\*\*\*  (-4.035) |
| Corporate Bonds (%) | -0.0090\*\*\*  (-3.582) | -0.0093\*\*\*  (-3.152) | -0.0121\*\*\*  (-4.422) |
| Log Total Net Assets |  | -0.135\*\*\*  (-2.724) | -0.111\*\*  (-2.429) |
| Retail Fund Dummy |  | -0.040  (-0.444) | -0.096  (-1.157) |
| Index Fund Dummy |  | 0.045  (0.398) | 0.093  (0.898) |
| ETF Dummy |  | 0.271\*\*  (2.139) | 0.240\*\*  (2.053) |
| Concentration Index |  | 0.000  (0.338) | 0.000  (0.361) |
| Mean Return (%) |  |  | 1.473\*\*\*  (10.954) |
| Constant | 1.976\*\*\*  (12.227) | 3.207\*\*\*  (6.349) | 2.795\*\*\*  (5.974) |
| R-square | 0.667 | 0.678 | 0.726 |
| Adjusted R-square | 0.665 | 0.674 | 0.722 |
| No. Observations | 698 | 698 | 698 |

There are many other fund characteristics that influence a fund’s risk profile such as the fund’s investment style. However, the investment style is difficult to identify. The CRSP database has Lipper objective codes. According to the documentation, the codes are assigned based on the language that the fund uses in its prospectus to describe how it intends to invest. The problem is that the codes are not consistent across time and there are many categories, which is the reason we do not include them in the regression analysis. Despite these problems, in Graph 6 we plot the histogram of the standard deviation of returns for growth and value funds. A fund is considered a growth fund if it has the Lipper code “Large-Cap Growth Funds” in at least one period in our sample. It is considered a value fund if it has the code “Large-Cap Value Funds” in at least one year. These funds are a small subsample of our universe of funds. There are only 63 growth funds and 31 value funds. The histograms show that value funds have a lower return volatility than growth funds.

**Graph 6**