Predicting Factor Returns via Machine Learning Classification

I. Abstract

This paper examines four investment factors that we believe would lead to significant investment returns. We examine multiple macroeconomic features that we believe to be relevant to each of the ascribed factors. We apply two Machine Learning classification algorithms, Regularized Logistic Regression and Linear Discriminant Analysis (LDA), to estimate factor returns with relevant features. Our approach provides a Sharpe ratio of 0.161, with alphas of 0.0043 and 0.0046 compared to the market risk premia and the Fama-French three-factor model.

II. Introduction

The purpose of our research is to predict the sign of the next-month factor returns. Our first step is to split the total periods into two sets, the training periods and the testing periods. Our initial split divides the entire periods into 203 months and 192 months, respectively. We split the data this way, so the testing periods are divisible by our expanding window size, which is six months. Our training data contains the 2001 financial crisis, while our testing data includes the 2008 financial crisis, to ensure a sound and robust model.

As for target picking, we choose the top 5 performing factors in our initial training periods, which are the initial 203 months, in terms of the factor returns' Sharpe ratios. The factors we initially select are Free Cash Flow to Price (FCFP) under Valuation category; 1M Price High minus 1M Price Low (HL1M) and 5 Day Price Reversal (PM5D) under Momentum category; Capital Acquisition Ratio (CapAcqRatio) under Capital Efficiency category; and 1Y Change in Asset Adjusted Free Cash Flow (AstAdjChg1YFCF) under Earnings Quality category. FCFP indicates a company's ability to generate additional revenues. Low FCFP implies an undervalued company, and we believe it is an excellent factor to predict returns. The capital acquisition ratio (CapAcqRatio) reflects the firm's ability to finance capital expenditures from internal sources. A ratio of less than 100 % indicates that capital acquisitions are draining more cash from the business than generating. We use this factor to measure the profitability and cash liquidity of the investment. Adjusted Free Cash Flow means net cash provided by or used for operating activities minus capital expenditures. It also measures the cash liquidity of the investment. However, AstAdjChg1YFCF measures the yearly change in asset adjusted free cash flow, and it does not measure the net cash flow within one year, which seems more important for our analysis. Thus, we decided to drop the redundant AstAdjChg1YFCF factor.

Since we want to predict whether the factor return will be positive next month, we created a target return column to show the next-month factor returns. We then created two labels to represent the signs of next-month returns based on the target return column, with -1 for negative returns and 1 for positive ones.

Our next step is to choose certain relevant features for each of the four factors. We decide to select the features based on their regression coefficients regarding each of the factors. We select features that have significant regression coefficients and do not have a high correlation with each other. We first normalize our in-sample features by subtracting the historical mean and dividing the historical standard deviation. We run Regularized Logistic Regression to fit the in-sample target factor return labels and all the in-sample macroeconomics variables, eliminating irrelevant features with zero coefficients. After those steps, we come up with fewer than 30 features for each factor. Finally, we use the same features to predict out-of-sample target factor labels.

III. Methodology

We use two classifier methodologies, Regularized Logistic Regression and Linear Discriminant Analysis, to analyze and compare results. Then we pick the one with a higher overall accuracy score and greater robustness.

Regularized Logistic Regression

Logistic Regression displays the probability of an observed variable belonging to a specific class, given the features of that class:

$$P[Y=K|X=x] = \beta_0 + \beta^T{}_1 x$$

We usually use log-odds ratios instead of probability, which ranges from negative infinity to positive infinity. The coefficients for predicting each class are estimated based on maximizing log-likelihood functions. The maximizing log-likelihood functions generate the coefficients to maximize the observed variables' probability of belonging to the specific classes. If the number of observations is less than the number of features, Logistic Regression may not work well.

Thus, we apply penalties to irrelevant features with the Lasso Regularization, which is called Regularized Logistic Regression.

Regularized Logistic Regression is easy to implement and interpret, and very efficient to train the data. But it is also limited to the assumption of linearity between the dependent variable and the independent variables. However, Regularized Logistic Regression is more robust when the underlying variables are not normally distributed, because it takes the variables as given. When the underlying variables are normalized, Regularized Logistic Regression does not perform as well as LDA.

Linear Discriminant Analysis (LDA)

LDA is a classifier with a linear boundary, assuming that all groups have the same variance-covariance matrix under the Gaussian distribution. It estimates the conditional probability of the observation belonging to each class using the equation below. The class with the largest conditional probability would be the one to which the observation belongs. In the equation, π_k is the prior probability of class k, and $f_k(x)$ is the conditional probability of observation x belonging to class k.

$$P[G = k|X = x] = \frac{f_k(x)\pi_k}{\sum_{l=1}^{K} f_l(x)\pi_l}.$$

Under LDA, the underlying assumption is that the variance-covariance is the same for all classes, leading to bias when the variances are significantly different between class 1 and class –1. Our biased assumption about the variance-covariance could lead to unwanted outputs.

But as stated in *Linear discriminant analysis: A detailed tutorial*, Small Sample Size or under-sampled problem may be the main issue of LDA. This problem stems from the high-dimensional pattern classification task or the number of training samples available for each class compared to the sample space's dimensionality. Fortunately, we have a large sample size (more than 30 years of monthly data) and a relatively small size of features (around 25-30 features for each factor). Besides, we only have two classes to predict (whether the factor return would be positive or negative next month), so LDA works quite well in our analysis.

IV. Empirical Analysis

We use an expanding window of 6 months to produce better estimations and conduct back test. We normalize the data to produce features on the same scale without obscuring the original data, therefore assigning the same weights to each feature. As an example, for our first expanding window with LDA, we use the target factor return labels in the first 203 months as Target. We normalize the relevant macroeconomic variables in the same period and treat them as Features. We then use one of our classifiers to train the data and use one cross-validation method, GridSearchCV in Sklearn, to conduct an exhaustive search for the optimal tuning parameters (the solver methods in the case of LDA). With the optimal classifier, we use the normalized next-6-month (Month 204-209) real features to predict the next-6-month (Month 204-209) labels. In the next iteration, we include the real data (both features and labels) in the subsequent 6 months (Month 204-209) in our training data and repeat the same steps to predict the next-6-month (Month 210-215) targets.

We predict the target return labels using Regularized Logistic Regression and LDA. As mentioned above, we put the normalized training data into LDA, and we use cross-validation to find the tuning parameters that give us the best model. By fitting our testing data into the best model, we obtain the out-of-sample target label predictions. Because we want to test our model's robustness, we also return the out-of-sample accuracy score for each iteration (every 6 months). If the accuracy scores from each iteration stay relatively stable, then our model is robust. Besides, we calculate the overall accuracy scores for the entire testing period (Month 204-395). We use the scores to select between the Regularized Logistic Regression and LDA classifier. We repeat the same process for Regularized Logistic Regression, except that we do not need to find the best tuning parameters using cross-validation.

According to the accuracy table below, we can see that all the factor accuracy scores for LDA are higher than those for Regularized Logistic Regression, especially for factor 4 CapAcqRatio. From this perspective, LDA seems to be a better prediction model. The average accuracy score gap is 0.07, and for CapAcqRatio, the accuracy score gap reaches 0.14.

Figure 1: Accuracy Table

Factors	LDA Accuracy Score	LR Accuracy Score
FCFP	0.520833	0.494792
HL1M	0.526042	0.442708
PM5D	0.515625	0.479167
CapAcqRatio	0.572917	0.432292

Figure 2: LDA accuracy score with 6-month expanding window

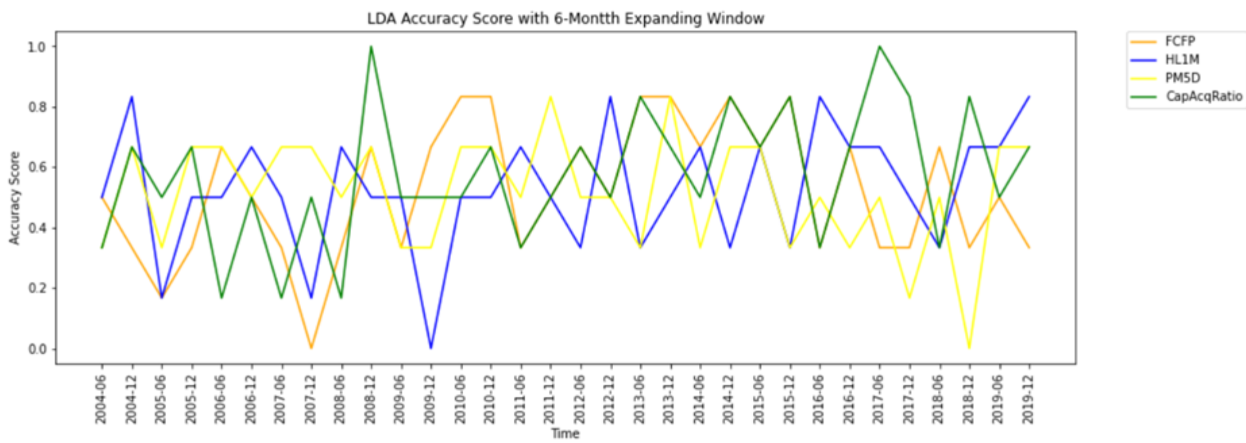
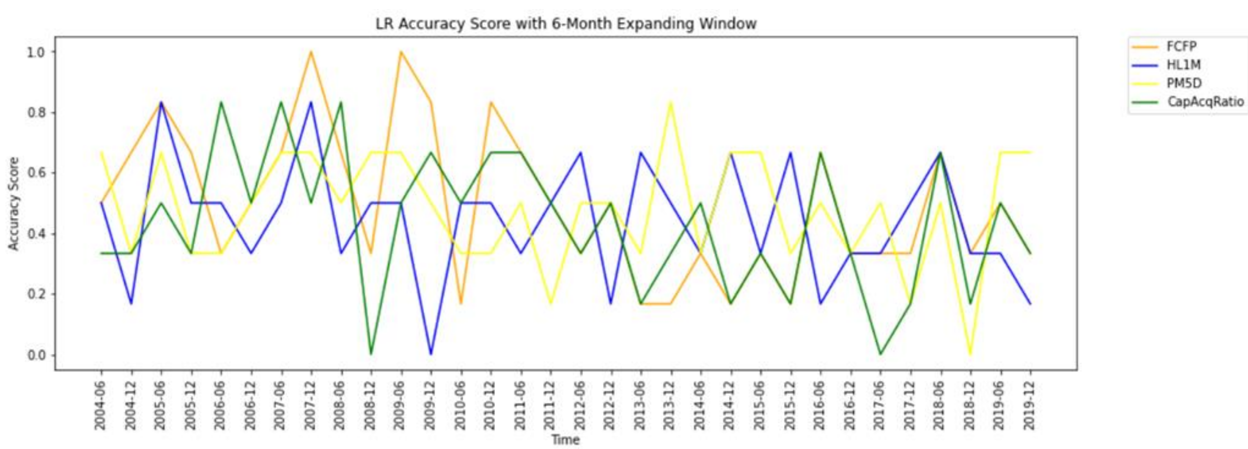


Figure 3: Regularized Logistic Regression accuracy score with 6-month expanding window



As we can see from Figures 2 and 3, the curve for FCFP, which is the orange line in the graph, is smoother for LDA than Regularized Logistic Regression. That indicates LDA has a higher level of consistency. The Global Financial Crisis started in 2007, and The National Bureau of Economic Research declared June 2009 as the end date of the U.S. recession. According to Figures 2 and 3, the accuracy scores of four factors under two prediction models experience major fluctuations during June in 2007 and June in 2009. In the Regularized Logistic Regression model, the accuracy scores decrease rapidly during the financial crisis. However, in LDA, the accuracy scores increase steadily. Although both models fluctuate more during the crisis, the result indicates that the LDA model performs better when the market is unpredictable and turbulent.

Figure 4: Logistic Regression prediction volatility

	LR Prediction Volatility
FCFP	0.241107
HL1M	0.193691
PM5D	0.185171
CapAcqRatio	0.223837

Figure 5: LDA prediction volatility

	LDA Prediction Volatility
FCFP	0.219493
HL1M	0.200301
PM5D	0.188006
CapAcqRatio	0.220233

According to Figures 4 and 5, after calculating the standard deviation of the accuracy scores in each 6-month window, the LDA's prediction volatility for factor FCFP is 2% lower than that of the Regularized Logistic Regression's. The other factors' prediction volatilities are about the same between the two methods. The average volatility for LDA prediction is 0.207008. The average volatility for Regularized Logistic Regression prediction is 0.210952, leading us to conclude that LDA is more robust than Regularized Logistic Regression. As a result, we conclude that LDA is more accurate and stable than Regularized Logistic Regression, so we choose LDA as our final model for our investment strategy.

V. Investment Strategy

We decide to invest based on the prediction of the signs of factor returns next month. We long the positive ones and short the negative ones in the subsequent month. If we predict the returns are all negative in the next month, we short all the factors and buy 3-month treasury bills to compensate for our risks and make the sum of our weights equal to 1.

We assign equal weights in our winners and losers since we would not know how the actual performance will be in our predicted month, and we want to take a relatively conservative approach. We also apply maximum weight constraints to limit the shorted weight to 10% amongst all losers and make the sum of our weights equal to 1. For a typical month, each loser factor will be assigned with a weight of 0.1/n, where n is the number of losers that month; the remaining 110% of our portfolio will be longing the winners or treasury bills. We use a weight constraint of 10% for losers because we observe more uptrends rather than downtrends in most market conditions. Although the markets typically increase over time, it is still imperative to take the opposite side and have a short position, whether for a pseudo-hedge or to maximize returns on loser factors. The maximum long weight assigned was 1.1, so under a long-short scenario, both long weight and short weight add up to 1. Similar to the loser class, each winner class gets assigned a weight of 1.1/m, where m is the number of winners each month. When we only have winners in one month, we decided to assign an equal weight of ¼=0.25 among each factor. When we only have losers in one month, we decide to assign an equal weight of -¼=-0.025 among each factor, and long 110% in 3-month treasury bills, to compensate for risks and make our weight sum to 1.

We examine our portfolio and investment strategy by computing the Sharpe Ratio and the alpha against the Fama-French three factors model. By including basic stats (mean, standard deviation, min, max, etc.) as part of our output, we have a broad view of how our strategy performs and gives us a snapshot view if the strategy is adjusted in the future (Figure 6). We evaluated using the Sharpe ratio to see how much return we can make by taking one unit of risk. Also, we want to see if our strategy outperforms the market and the famous three-factor Fama-French model (Figure 7).

Strategy Results

Figure 6: Portfolio Return Stats

	Portfolio Return Stats
count	186.000000
mean	0.004542
std	0.021425
min	-0.065500
25%	-0.009073
50%	0.004249
75%	0.016395
max	0.113403

Figure 7: portfolio mean, rf mean & Sharpe ratio

	portfolio mean	rf mean	sharpe ratio
0	0.004542	0.001082	0.161472

Figure 6 outlines the basic statistics of our results. Our investment strategy performs well with a mean return of 0.45% monthly (annualized returns of 5.4%). We also have a low portfolio standard deviation, thanks to our conservative approach. On a risk-adjusted return basis, our strategy still provides significant results with a Sharpe ratio of 0.1614; although our returns were average, they aligned well with the low risk we took.

As we can see in Figure 8, our investment strategy produces a significantly positive alpha compared to the market return premium over the risk-free rate. The t-test of 2.749 is more than our 1.96 significance threshold.

Figure 8: Alpha and Market Return minus Risk-Free Rate

	coefficients	t_test
Alpha	0.004375	2.749197
mktrf	0.033677	0.698339

The value factor we considered for the Fama-French model were BP (HML) and the size factor was LogMktCap (SMB). HML accounts for the spread in returns between value stocks and growth stocks and argues that companies with high book-to-market ratios, also known as value stocks, outperform those with lower book-to-market values, known as growth stocks (Fernando, 2020). For LogMktCap (SMB), we log the market caps of the Small minus Big to normalize the data to ensure quality analysis under normal distribution assumptions.

Figure 9: Alpha compared with Fama-French Three Factor Model

	coefficients	t_test
Alpha	0.004654	2.973041
mktrf	-0.042255	-0.827377
BP	0.038907	0.577939
LogMktCap	0.175485	2.279840

Figure 9 above provides us with the coefficients to compare the alpha we obtained with the Fama-French model and the corresponding t-tests. With a t-test value of greater than 2 (2.9730), we find that our investment strategy outperforms the Fama-French three factors.

VI. Discussions

A major limitation of our investment approach is that we ignore transaction costs. With a portfolio that is readjusted monthly among four factors, the turnover will be high and further analysis will need to be conducted to assess the implications of transaction costs fully. Besides, to construct each factor, we must build a long-short portfolio (for example, long the top bins and short the bottom bins), leading to even higher turnover in our strategy. However, we do not have data on the underlying factor portfolio turnover. A research paper conducted by Beraldi, Violi, Ferrara, Ciancio, & Pansera in 2019 concluded that "importance of accounting for real-life transaction costs in the financial planning and has shown that portfolios obtained by considering approximated cost structures are less effective in terms of overall profitability." If further research and analysis are conducted using the four-factor models presented in this paper, transaction costs should be considered. Since the alpha we arrived at was near zero (0.004), this inclusion of these transaction costs might provide a negative alpha. A possible solution to this may be to reconsider the number of factors used to construct this strategy or simply reduce the turnover of the portfolio (i.e., reconstruct every quarter instead of every month). Those methods would help reduce the number of times your portfolio will change, reducing your overall transaction costs.

Our models are not quite stable during the stress period, as shown in Figure 2 and Figure 3; thus, they are less reliable during the crisis. One approach would be to increase our short positions when crisis hits and decrease short positions when the market recovers. Thus, our models could be potentially more robust in predicting returns during market downturns and generating more profits.

We did not use non-linear classification methods in our analysis. We tried Random Forest, Gradient Boosting, and Linear Support Vector classifications, but they performed poorly in our testing period with very low accuracy. The reason why Linear Discriminant Analysis classification works better might be the large size of our sample and the simple classification problem (with only two classes).

VII. Conclusion

This paper investigates whether a four-factor long-short investment model can provide investors with significant returns that outweigh the benchmarks. After comparing the accuracy scores and volatility of two classification methods, we choose to use Linear Discriminant Analysis to classify the factor returns next month into winners and losers. Our strategy also incorporates a maximum short-weight constraint to limit the downside risk of the short leg. Based on the prediction given by LDA, our investment strategy results in positive returns. Overall, our investment strategy currently remains conservative and profitable. Further modifications to the model’s stressed period predictions and the inclusion of transaction costs will produce more forecast accuracy and higher returns.

VIII. References

Beraldi, P., Violi, A., Ferrara, M., Ciancio, C., & Pansera, B. A. (2019). Dealing with complex transaction costs in portfolio management. *Annals of Operations Research*, 1-16.

Fernando, J. (2020). Understanding High Minus Low (HML). Retrieved 3 December 2020, from https://www.investopedia.com/terms/h/high_minus_low.asp

Tharwat, Alaa & Gaber, Tarek & Ibrahim, Abdelhameed & Hassanien, Aboul Ella. (2017). Linear discriminant analysis: A detailed tutorial. Ai Communications. 30. 169-190,. 10.3233/AIC-170729.