

Factoring Consumer Staples

Kyle Walsh[†] and Siddharth Kantamneni[‡]

[†]kew96@cornell.edu

[‡]skk82@cornell.edu

Abstract

In this paper, we attempt to determine if the Fama-French 3-Factor Model can be used with Markowitz Portfolio Optimization on the Consumer Staples sector. The Fama-French model is typically applied to the broader market, and we wanted to explore Consumer Staples because of the market crash due to the global pandemic. We implemented a Weighted Least Squares regressor to calculate assets' factor loadings and Markowitz Optimization for selecting maximum Sharpe Ratio portfolios. While this framework does achieve positive excess returns over the past fifteen years, the volatility makes it difficult to add to any portfolio. Sensitivity analysis did provide some insight into how parameters affect the model but further testing is required. Achieving a positive excess return gives us hope for this framework but necessitates future research into new factors and different techniques for calculating factor loadings.

1 Introduction

1.1 Motivation

This project has been motivated by a desire to replicate the famed Fama-French Three-Factor model in a novel way. Wanting to stay consistent with the original theory, we decided to use the Consumer Staples sector as the entire universe, along with a risk-free asset. This allowed us to implement the original model with unknown outcomes.

Additionally, we wanted to see if the predictions from the model could be used to create portfolios with excess returns relative to the risk-free rate. If the model was truly able to predict excess

returns, then one could ostensibly find sources of alpha among the consumer staples sector. Moreover, if our model was unable to do so, it may motivate finding additional or superseding factors to make the model more robust.

1.2 Background

Factor models are an extremely popular and rigorously explored investment model. Factor models seek to explain a stock's excess returns relative to a risk-free rate by computing coefficients based on the return of external "factors" which can vary widely. In general, there are two types of factor models: macroeconomic and fundamental. Macroeconomic models rely on factors that typically reflect the broader state of the economy. On the other hand, fundamental models use measurements that are specific to individual companies.

1.2.1 The Capital Asset Pricing Model

Proposed in 1972 by Dr. Fischer Black, The Capital Asset Pricing Model (CAPM) was one of the first factor models. As a macroeconomic model, CAPM relies on a univariate regression of an asset's excess return on the market excess return, both over a risk-free interest rate. This method would pave the way for further using linear regressions to decompose the excess returns of various assets.

$$\mathbb{E}[R_i] = r_f + \beta_i (\mathbb{E}[R_m] - r_f) \quad (1)$$

As the "market factor", $\mathbb{E}[R_m] - r_f$ provides a general opinion on market direction. While β_i is the regression coefficient on the market's excess return, it also represents an asset's market sensitivity. Often times, β_i is referred to as a factor loading.

$$\beta_i = \frac{\text{Cov}(R_i, R_m)}{\text{Var}(R_m)} \quad (2)$$

1.2.2 Fama-French Three-Factor Model

Following the intuition of Black, in 1992, Dr. Eugene Fama and Dr. Kenneth French proposed an updated factor model, hoping to have more explanatory power when comparing to an asset's excess return. In addition to the market's excess return, Fama and French used two additional

factors to represent a company's size and value. A company's market capitalization is often used as a facsimile for size. On the other hand, the book-to-market ratio often represents the value of a company, although other measures are able to be substituted, accordingly. These factors were chosen due to the phenomenon that small businesses tended to outperform large businesses and that value stocks (i.e. high book-to-market) would outperform growth stocks (low book-to-market).

As neither size nor value actually represent an asset's return, Fama-French proposed ranking the companies according to these measures - small to large for size and high to low for value. They then took the returns of the top assets and subtracted the returns of the bottom assets. These return-differences are now referred to as SMB (small minus big) for the size factor, and HML (high minus low) for the value factor. This model is a macroeconomic model because we use the market return on HML and SMB instead of the specific values of capitalization or book-to-market for each company.

2 Methodology

2.1 Assumptions

Assumptions underlying the Fama-French Three-Factor Model are all assumed to hold here. This includes the assumptions underlying linear regression. We also assume no trading costs, infinite liquidity/fractional trading, and no price impacts.

2.2 Data

The data was mainly collected from the Wharton Research Database Services (WRDS) along with investing.com. Within WRDS, the data was sourced from The Center for Research in Security Prices (CRSP) and Compustat.

2.2.1 Investment Universe

We started by including all current stocks in the S&P 500 Consumer Staples sector. Due to the limited size of this set, and to avoid look-ahead bias on assets that entered into the S&P 500's Consumer Staples during our timeline, we further pulled the six largest consumer staples focused ETFs' holdings over our timeline and added all stocks to the universe. In total, we generated

602 unique tickers which reduced to 350 after collecting the required data and eliminating those for which we did not have data points in each of the data sets. Many of these lost tickers represent other assets that are not classified as equities and, therefore, do not fit the parameters for our model.

2.2.2 Cleaning

As with most quantitative projects, the data cleaning process was much more involved than we hoped. While all data was collected from reputable sources, there was still substantial missing data for many companies in most of the required data fields. In general, missing data was handled with simplistic methods with specifics covered in the remaining paragraphs of this section. Furthermore, we only imputed a maximum of two periods (six months) of missing data as our investment universe is large enough to afford excluding incorrigible assets from consideration.

Although consistent adjusted-returns were provided for most assets, we still had to impute share price for many assets over our period in order to calculate the market capitalization of each company. For this data we assumed linearity for all missing points, including at the end of a period, and imputed accordingly.

For the same reason, we also had to impute missing shares outstanding. Unlike with prices, shares outstanding seems to remain constant over most time periods. For this reason, we used the most recent, previous shares outstanding. We only used back-looking data points because we make the assumption that the database will be updated whenever the number of shares outstanding is changed in either direction.

2.2.3 Factors

After cleaning the data, the factors were computed. To create the market excess factor, we subtracted the 3-Month US T-Bill (assumed to be our risk free rate) from the S&P 500 Consumer Staples Index. Since our focus was on this Sector, we rationalized that we should have our factors be restricted to this sector as well. For computing the SMB factor, we computed the average returns of stocks that were either in the lower 10th percentile or upper 90th percentile in terms of market capitalization, and subtracted the average return of the 90th percentile stocks from that of the 10th percentile stocks, achieving our Small Minus Big "return." This process was repeated for every date in our data set. For HML we followed the same process, except we use the percentiles of book-to-market ratio instead of market capitalization and subtracted the average return of the 10th

percentile stocks from that of the 90th percentile stocks. Once we had all of our factors computed, we were ready to start computing factor loadings for each stock.

2.3 Computation of Model

To ensure our model would not suffer from look-ahead bias, we chose to compute factors at every date beginning in 2005 Q1. We would only use data up to the period prior to the current period to compute the factor loadings for each stock. This simulates the model updating with the most up-to-date data at each period, further increasing model accuracy.

Actual computation of the factors utilized Weighted-Least-Squares (WLS) using a linear weighting system, with recent data being given the most weight. We then compared the excess returns of the stock in question against each of the factors using the WLS regression. This model returned our β values, or factor loadings, for each of the factors and assets.

We now focused on calculating the alpha values for each asset, the portion of the expected excess return not explained by the model. There are two methods for obtaining this alpha value, which affected our model noticeably. One method includes a column of constants among the covariates of the WLS model, meaning that alpha serves as the intercept for the excess return of the assets. This method will be referred to as the "alternative model" in following sections. The other method does not include a constant in the covariates and calculates alpha as in Equation 3 and errors in Equation 4. This second model will be referred to as the "base model" in subsequent sections due to its consistency with the theoretical model.

$$\alpha_i := \sum_{t=1}^T w(t) * (r_i(t) - \sum_{j=1}^m \beta_{i,j} f_j(t)) \quad (3)$$

$$e_i(t) = r_i(t) - \alpha_i - \sum_{j=1}^m \beta_{i,j} f_j(t) \quad (4)$$

After computing the factor loadings and alpha, we must then compute the risk matrix for all of the stocks in a given time period. First, we must calculate the factor covariance matrix using Equation 5. We then use Equation 6 to get the risk matrix where Δ is a diagonal matrix whose i^{th} diagonal is $\sum_{t=1}^T w(t) e_i(t)^2$

$$F = \Sigma_{t=1}^T w(t)(f(t) - \Sigma_{s=1}^T w(s)f(s))(f(t) - \Sigma_{s=1}^T w(s)f(s))^T \quad (5)$$

$$V = BFB^T + \Delta \quad (6)$$

2.4 Markowitz Optimization

In order to choose our optimal portfolio weights, we used Markowitz optimization techniques. As parameters for the objective function, we pass the expected excess returns and risk matrix we generated from the factor loadings in the previous section. These expected excess returns are first sorted, descending, according to expected return. We then only use the top and bottom $n/2$ returns (for a total of n assets). As our objective function, we aim to maximize the portfolio's total expected return subject to a specified portfolio variance and market exposure. On top of these two common constraints, we also included constraints to ensure that we are not overexposed to any one asset by letting the maximum single position be one half of the total invested amount. These constraints further enforce that we only have long positions in the $n/2$ assets with the highest expected excess return and short the remaining assets. This model can be defined mathematically in Equation 7.

$$\begin{aligned} \min_w \quad & \mu^T w \\ \text{s.t.} \quad & w^T V w \leq \sigma^2 \\ & e^T w = 0 \\ & 0 \leq w_i \leq 0.5 \quad \forall i \in \{1, \dots, n/2\} \\ & -0.5 \leq w_i \leq 0 \quad \forall i \in \{n/2, \dots, n\} \end{aligned} \quad (7)$$

Our choice of being market neutral results from two initial findings. The first was that we noticed a strong inclination towards being long assets, effectively driving our short positions to be nearly zero while our long positions were nearly at the limit of their bounds. Along similar lines, by ensuring we are market neutral, we force the model to be more selective with our positions and eliminate any potential market drift that would be captured if we were long on average.

3 Results

3.1 Performance

As stated in section 2.3, we evaluated two models that differed only in the way that the alpha parameter is calculated. With a focus on following the original theory, we chose to use the method of calculating alpha using the results from WLS as our base model. For completeness, we have provided the results of both models with Table 1 containing the average values and Table 2 containing the standard deviations of the values from Table 1.

	Base	Alternative
Actual Return	4.76%	5.57%
Expected Return	33.07%	39.05%
Sharpe Ratio	0.34	0.35

Table 1: Model metrics' expectations

	Base	Alternative
Actual Volatility	16.99%	18.67%
Expected Volatility	25.22%	24.45%
Sharpe Volatility	20.61%	20.50%

Table 2: Model metrics' standard deviations

Note that "Return" refers to excess market return

As we can see in Table 1 and Table 2, while the alternative model does have a higher actual excess return, the base model performs only marginally worse than the alternative model. When comparing the excess returns over time in Figure 1, we can see that they typically track well with each other over time, where the alternative model's higher volatility results from larger and more frequent excess return spikes.

We now turn our attention to the base model. We can see in Figure 2 that the actual excess returns move relatively well with the expected excess return and this is represented by a correlation coefficient of 0.56. Unfortunately, the expected excess return given by the model is not representative of the excess return that will be achieved.

3.2 Sensitivity Analysis

In order to test our results against some variations in input assumptions, we performed a basic sensitivity analysis. We only looked at variations of two parameters: the basket size and the total market exposure. Overall, the results were promising on multiple fronts. When varying the basket size, we saw that too many or too few assets experienced reduced returns with an optimal size being between 10 and 50. On a similar note, when varying the market exposure, as we leave our

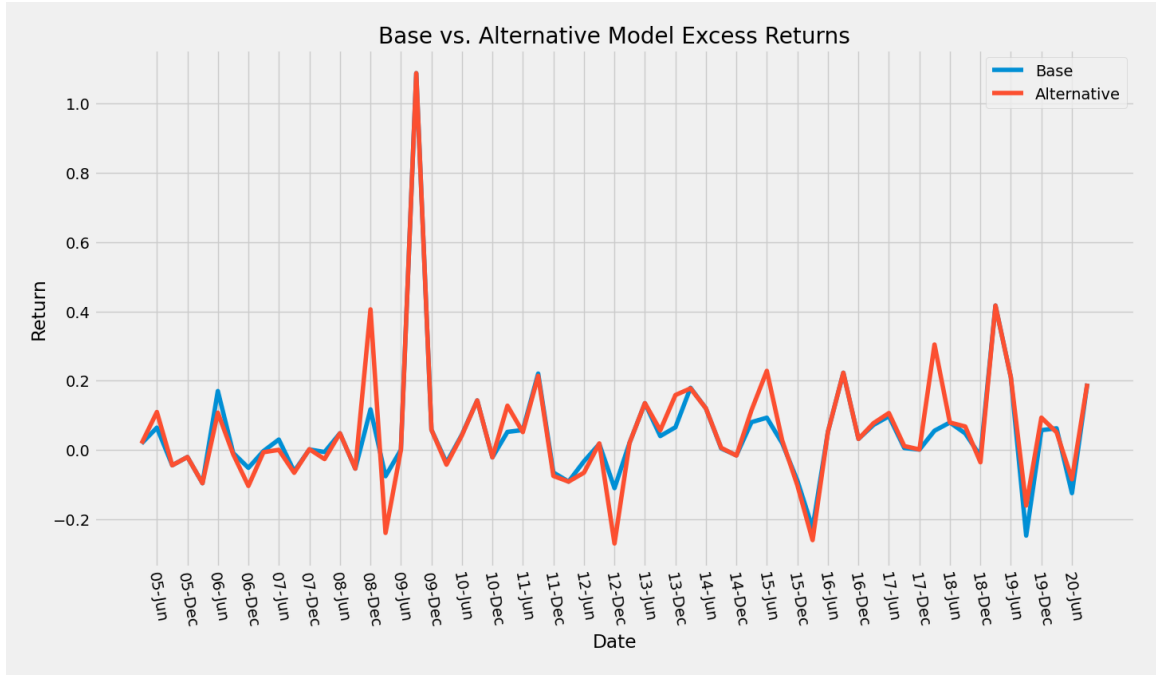


Figure 1: Base and alternative model's actual excess return from 2005 through 2020

neutral position, we see the portfolio experience higher excess returns with varying degrees of volatility. These results are summarized more concretely in Table 3 and Table 4 respectively.

	Return	Volatility
4 Assets	5.34%	21.26%
10 Assets	6.51%	21.80%
20 Assets	4.76%	16.99%
50 Assets	6.26%	23.01%
100 Assets	5.07%	24.71%

Table 3: Basket size

	Return	Volatility
-1.0 Exposure	8.66%	32.57%
-0.5 Exposure	7.57%	23.56%
0.0 Exposure	4.76%	16.99%
0.5 Exposure	6.70%	28.21%
1.0 Exposure	10.68%	34.26%

Table 4: Market exposure

Note that "Return" refers to excess market return

4 Conclusion

The Fama French 3 Factor Model, when employed in conjunction with Markowitz Optimization, shows potential for being able to create baskets of stocks with excess returns on a quarterly basis. Interestingly, when using the alternative model (where alpha is computed inside the linear regression) returns and risk are marginally higher and lower respectively. Obviously, the Sharpe

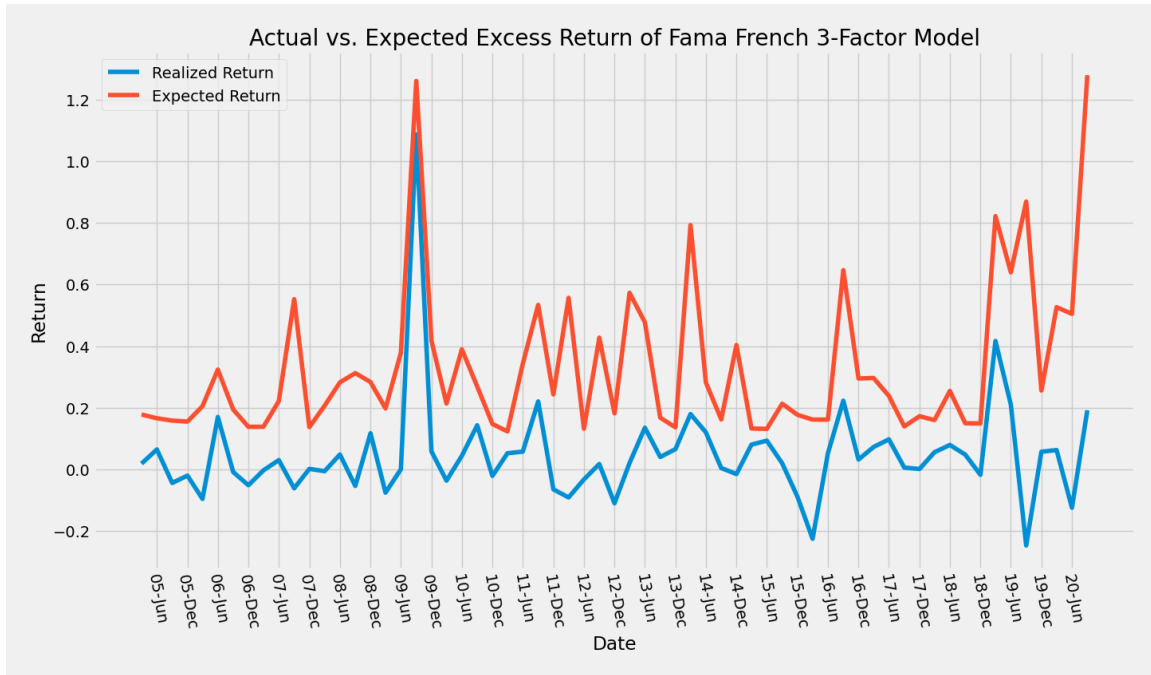


Figure 2: Actual and expected excess return from 2005 through 2020

Ratio and underlying risk of said portfolios are not suitable enough to warrant investing in such a strategy at this time. However, this could mean that assumptions underlying the calculation of alpha may not hold as the two portfolios are not the same. Ultimately, even with the predictions from the Fama French model, our Markowitz optimizer was unable to produce consistent results. This may be due to several confounding features. Firstly, the assumptions underlying the linear relationship between factors and excess return may not hold in the Consumer Staples sector. Secondly, our method of creating factor returns could have been erroneous, thus producing incorrect factors at each time period. Lastly, the errors of each model may not have had 0 expectation, resulting in heavily biased models.

5 Future Research

Exploration into other factors such as momentum and profitability could prove insightful, especially since Fama and French themselves created models using those factors. We would also like to measure the effect of other economic metrics such as unemployment rate and how these play into returns in the Consumer Staples industry. In addition to traditional WLS, there are other regression

and general machine learning techniques that could provide some improvement such as Bayesian Regression. Lastly, a dive into relaxing some of the assumptions in the models such as friction-less trading and infinite liquidity could allow us to the model into a more practical setting.

References

- Black, Fischer (1972). "Capital Market Equilibrium with Restricted Borrowing". In: *The Journal of Business* 45-3, pp. 444–455.
- CRSP and Compustat (2021). URL: <https://wrds-www-wharton-upenn-edu.proxy.library.cornell.edu/> (visited on 04/16/2021).
- Fama, E. and French, Kenneth (1993). "Common Risk Factors in the Returns on Stocks and Bonds". In: *Journal of Financial Economics* 33, pp. 3–56.
- Fama, Eugene and French, Kenneth (2004). "The Capital Asset Pricing Model: Theory and Evidence". In: *Journal of Economic Perspectives* 18-3, pp. 25–46.
- S&P 500 Consumer Staples (SPLRCS) (2021). URL: <https://www.investing.com/indices/s-p-500-consumer-staples> (visited on 04/08/2021).
- Sharpe, William (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk". In: *The Journal of Finance* 19-3, pp. 425–442.
- United States 3-Month Bond Yield (2021). URL: <https://www.investing.com/rates-bonds/u.s.-3-month-bond-yield> (visited on 04/08/2021).