Enhancing the Fama-French Model with Sector ETFs and Search Analytics

Christopher Ambrus, Jung Park, James Schimmel, Aryaman Utkarsh Gangwar

Abstract

The Fama-French three-factor model has long served as a benchmark for understanding stock returns, incorporating market risk, size, and value factors. Our study extends this model, to improve the predictive power of the model, by introducing additional variables that capture the climate of the digital age and sector-specific influences. We examine three stocks: Beyond Inc (BYON), a consumer furniture goods company; Eli Lilly and Company (LLY), a pharmaceutical giant; and NVIDIA Corporation (NVDA), a technology and semiconductor company. These stocks are paired with their respective Exchange-Traded Funds (ETFs): iShares Cohen & Steers REIT ETF (ICF) for BYON, Health Care Select Sector SPDR Fund (XLV) for LLY, and iShares Semiconductor ETF (SOXX) for NVDA. We also analyze Google search analytics as a proxy for retail investor interest, interpolating the data from a monthly to a daily scale to align with the retail trading rhythm.

Our empirical findings reveal a nuanced landscape. While Google search analytics offered an intriguing glimpse into public interest, they did not yield statistically significant predictors of stock performance within our model. Conversely, the inclusion of sector-specific ETFs as factors demonstrated significance in certain tested time intervals, suggesting that these financial instruments may indeed hold predictive power over the stocks in question. This underscores a critical insight: retail investor interest, as gauged by search volume, should not be used as a standalone indicator for making buy or sell decisions on stocks. Instead, our results advocate for a more sophisticated approach that considers industry trends and broader market indicators. This study not only enriches the asset pricing literature but also serves as a cautionary tale against the oversimplification of complex market dynamics.

Introduction

The landscape of financial economics has been profoundly shaped by the development of asset pricing models, which seek to explain the behavior of stock returns through various risk factors. Among these, the Fama-French three-factor model stands as a seminal contribution, offering a framework that extends beyond the Capital Asset Pricing Model (CAPM) by incorporating size and value factors alongside market risk. This model, introduced by Eugene Fama and Kenneth French in 1992, has become a cornerstone for both academic research and practical investment strategies. The original Fama-French model posits that in addition to the market excess return, the size of firms and their book-to-market ratios are significant predictors of stock returns. The size factor captures the tendency of small-cap stocks to outperform large-cap stocks, while the value factor reflects the higher returns often associated with high book-to-market firms compared to their low book-to-market counterparts. These factors have been empirically validated across time and markets, affirming their robustness and enduring relevance.

However, the financial markets are dynamic and ever-evolving, influenced by a myriad of factors that extend beyond the scope of traditional models. The advent of the digital age has ushered in new forms of data and measures of investor sentiment, such as Google search analytics, which may offer additional insights into market movements. Moreover, the rise of exchange-traded Funds (ETFs) has provided investors with tools to gain exposure to specific sectors, potentially introducing new dimensions to the risk-return profile of stocks.

In light of these developments, our project aims to expand the Fama-French model by integrating novel factors that reflect the current financial landscape. We focus on three publicly traded companies—Beyond Inc (BYON), Eli Lilly and Company (LLY), and NVIDIA Corporation (NVDA)—each representing distinct sectors of the economy. These companies are analyzed in conjunction with their related ETFs: ICF for BYON, XLV for LLY, and SOXX for NVDA. These ETFs are chosen for their relevance to the respective industries of the stocks, providing a sector-specific lens through which to view the companies' performance.

We explore the potential of Google search analytics as a proxy for retail investor interest. The premise is that increased online searches for a company may indicate heightened investor attention, which could translate to movements in the stock's price. To this end, we interpolate Google's monthly search data to a daily frequency, aligning it with the daily trading data of the stocks. Our analysis employs a five-factor model that includes the original Fama-French factors—market, size, and value—alongside the ETF and Google search analytics factors. This approach allows us to test the hypothesis that these additional factors can provide significant explanatory power for stock returns. We conduct our study over various time intervals to assess the consistency and robustness of our findings.

The results of our investigation present a complex picture. While the ETF factors exhibit significance in some of the tested time intervals, suggesting that they may indeed capture industry-specific risks that affect stock returns, the Google search analytics factor does not demonstrate statistical significance. This finding challenges the notion that retail investor interest, as measured by search volume, is a reliable predictor of stock performance. It implies that decisions to buy or sell stocks should not be based solely on the level of public interest, as reflected in online search behavior.

Our study contributes to the asset pricing literature by examining the validity of incorporating modern data sources and financial instruments into traditional models. It also offers practical implications for investors, cautioning against the overreliance on measures of retail investor sentiment and underscoring the importance of industry trends and fundamental factors. As the financial markets continue to evolve, so too must our models and strategies, adapting to incorporate new sources of information and risk. In the following sections, we delve deeper into the methodology of our analysis, the specifics of the data used, and the detailed results of our empirical tests. We discuss the implications of our findings for the field of finance. Through this exploration, we aim to shed light on the multifaceted nature of stock returns and the factors that drive them in today's interconnected and data-driven world.

Literature review

The Fama-French Three factors have long been researched with extensions and comments for the same. David E. Allen and Micheal McAleer (2018) applied rolling OLS regressions to explore the relationship between the three factors. The results showed that these factors suffered from endogeneity when combined in OLS. The findings indicate that the application of these factors in a linear regression framework, as proposed by Fama and French (2018), to evaluate factor significance, presents challenges. This is because the precision of the estimated standard errors is greatly influenced by the accuracy of the model specification.[1] In their study, Ziyang Ji and Victor Chang (2020) proposed an innovative approach to augment the Fama-French three-factor model by incorporating sentiment as a novel independent variable. They harnessed online data that mirrored investor sentiment, which they then mined and quantified to create a sentiment index. When comparing the traditional three-factor model with their new sentiment-inclusive four-factor model, it was evident that the addition of the sentiment factor boosted the model's explanatory power. Interestingly, their findings also suggested that the size effect and the book-to-market effect were not prevalent in the Chinese blockchain industry. [2]

Allen and McAleer (2018) conducted a statistical examination of the Fama/French three-factor model using monthly data. They utilized rolling OLS regressions to investigate the interplay between the three factors, drawing on data from July 1926 to June 2018. Their findings indicate that when these factors are integrated into an OLS regression analysis, as proposed by Fama and French (2018), they may be subject to endogeneity. Further, their use of Ramsey's RESET tests in conjunction with OLS regression analysis revealed a non-linear correlation among the three series, with significant squared and cubed terms. [1]

Jin Xu and Shaojun Zhang(2014) provided empirical insights into the effectiveness of the three-factor model in accounting for fluctuations in Chinese stock returns. They discovered that unique characteristics inherent to China significantly impact the three factors and subsequently alter the predictive power of the three-factor model.[3] Liao and Shen (2008) utilized the Fama-French three-factor model to analyze the impact on stock prices following the completion of the split-share structure reform by Chinese listed companies, a process that began in April 2005. They constructed the size factor by dividing stocks into small and large categories based on the median market value of their tradable shares, calculated as the product of the number of tradable shares at the start of each year and the share price. For the value factor, they classified stocks into three groups according to their BE/ME ratio, which is the net assets per share divided by the share price. The intersection of the two size categories and the three BE/ME groups resulted in six portfolios. The returns of these portfolios were value-weighted by the market value of the tradable shares, implicitly assuming that the portfolios consisted solely of tradable shares.

Mobin Anwar and Sanjay Kumar conducted an empirical analysis to understand the effectiveness of the Fama and French three-factor model in the context of the Indian stock market. Their study spanned from April 1, 2009, to March 31, 2016. Their findings revealed that while the Fama and French three-factor model was unsuccessful in explaining individual asset returns, it was able to account for portfolio asset returns when these were categorized based on size and value. They also observed a notable impact of market risk premium, size premium, and value premium on asset returns.[4] Daniel and Titman (1997) followed a similar methodology to Fama and French's 1993 approach, constructing portfolios for analysis. They sought to examine the effects observed in January versus other months, with a particular focus on seasonal variations. Their observations indicated that both the book-to-market effect and the size effect for large firms were exclusively

January phenomena. Ultimately, they favored a model based on characteristics over the multifactor model proposed by Fama and French.[5]

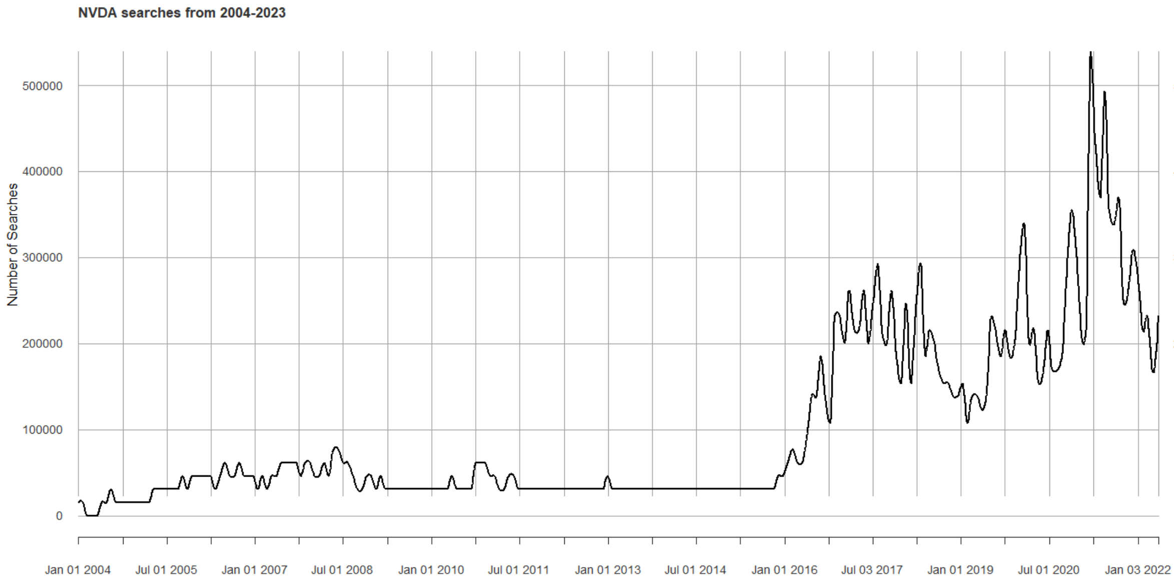
Taneja (2010) conducted a study on the Fama and French three-factor model in the context of the Indian stock market. His findings suggested that this model provides a more comprehensive explanation of average stock returns compared to the traditional CAPM. He also noted a strong correlation between size and value premiums, leading him to propose a two-factor model for the Indian stock market.[6] In the same year, Czapkiewicz and Skalna concluded that the SMB and HML factors satisfactorily account for average stock returns. They found that the three-factor model explained 65% of the variance (R2 = 65%). Their study also revealed that smaller firms perform better than larger ones and that the market tends to undervalue stocks with high value, while accurately predicting the returns of small-sized firms. Durand and colleagues (2011) suggested that changes in the investor's fear gauge play a crucial role in determining the size premium and value premium. They argued that an increase in the investor's fear gauge could lead to financial instability.

Fama and French(2012) surveyed four regions (North America, Europe, Japan, and Asia Pacific) and discovered that, except for Japan, value premiums tend to decrease with size in terms of average stock returns. Their joint test across these four regions did not yield strong results, and even local models struggled to explain the dependent variable. In a subsequent study in 2014, Fama and French expanded the three-factor model by introducing two additional factors: RMW (Robust [profitability] Minus Weak) and CMA (Conservative [investment] Minus Aggressive). They found that firms with a low beta had positive average stock returns. This suggests that firms that invest aggressively tend to experience negative returns, while those that invest conservatively see positive returns. Firms with a high beta, which are more aggressive, exhibit negative slopes for RMW and CMA, behaving similarly to less profitable firms. According to conventional wisdom, high beta firms are associated with positive SMB slopes.[7]

Berk and Van Binsbergen (2016) posited that investors should utilize a model that accurately prices risk. They further argued that expanding the traditional CAPM leads to numerous anomalies, which should be interpreted as risk factors. They concluded that the traditional CAPM surpasses all its extensions in performance.[8]

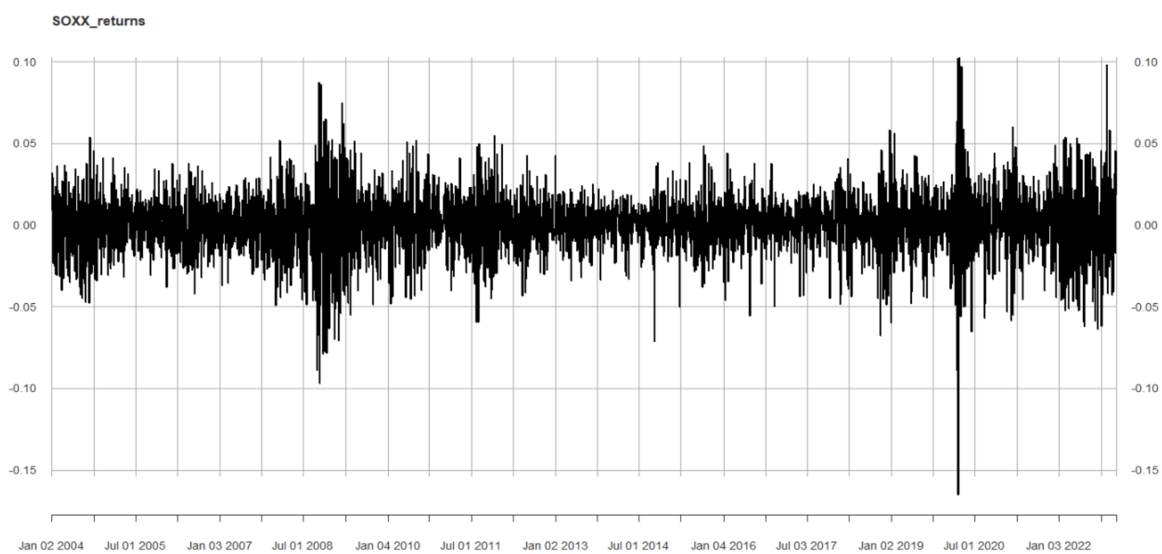
Data Description

For our regressions, we consider 3 companies all from very different sectors. For technology we look at Nvidia (NVDA), for healthcare, we use Eli Lilly (LLY), and for real estate, we use Beyond, Inc. (BYON). We collect 3 unique data sets for each stock regression. The daily price of the stock is sourced from the Yahoo Finance API, returning daily closing prices from January 1, 2004, through January 1, 2024. For each stock, we collect a measure of retail investor sentiment. For this, we use the Google Trends API to collect data for the number of searches for the stock ticker. For Nvidia, this means we take the number of total searches for NVDA. The data used is monthly over the time frame of January 1, 2004, through January 1, 2024.

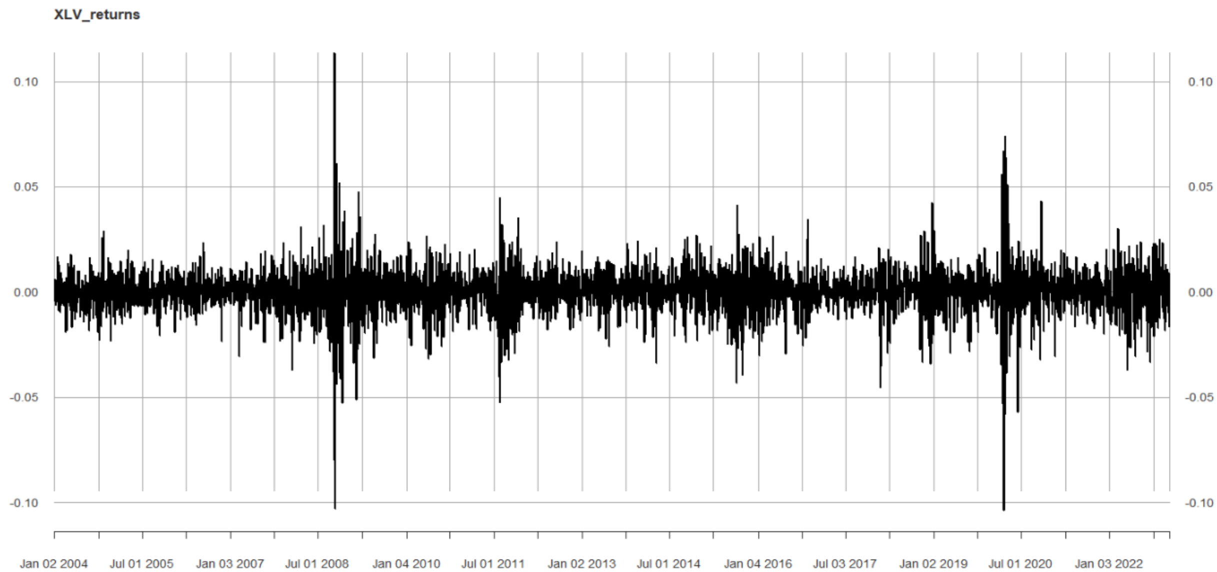


Graph 1. Nvidia search volume from 2004-2023

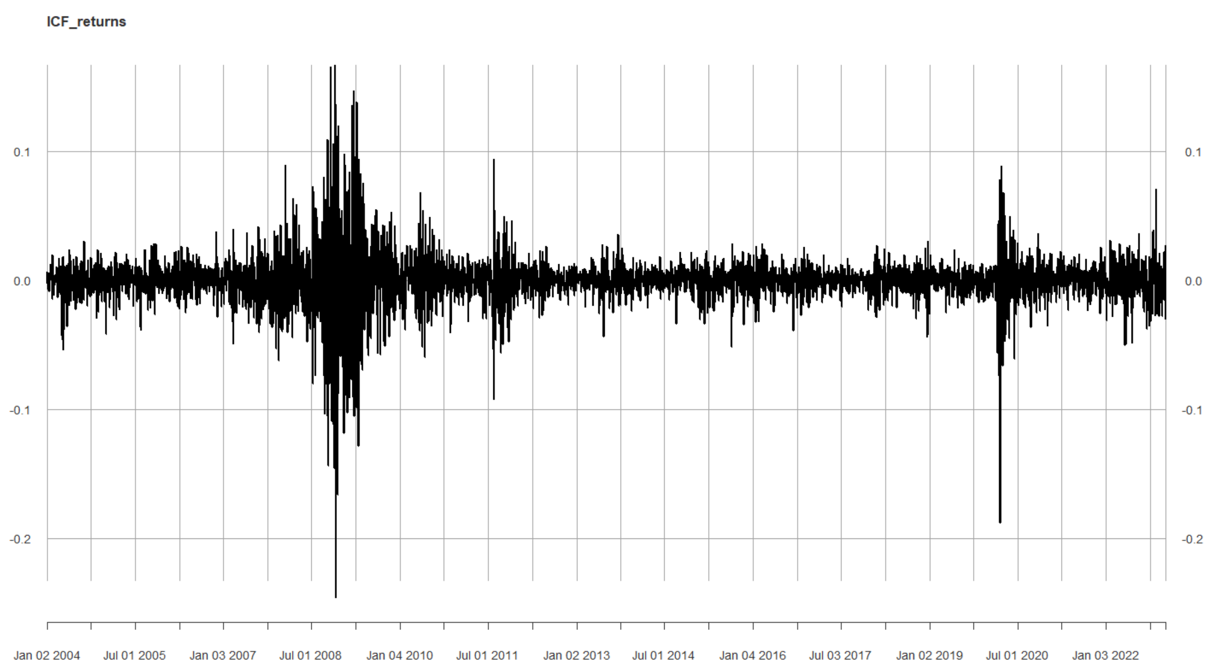
The last factor for each regression is a unique exchange-traded fund (ETF) that represents the sector of the individual stock. The technology ETF we use to approximate the sector performance is the iShares Semiconductor ETF (SOXX). For healthcare, we use the Health Care Select Sector SPDR Fund (XLV). For housing, we use the iShares Cohen & Steers REIT ETF (ICF).



Graph 2. Returns of SOXX ETF from 2004-2023



Graph. 3 XLV ETF returns from 2003-2004



Graph.4 ICF ETF

When looking at the returns data for the ETFs we see that they each spiked around the time of the US housing market bubble burst in 2008 and once more in 2020 at the start of the COVID-19 outbreak. For the data on the risk-free rate, market factor, Small minus Big, and High minus Low we use data available from the Fama French website.

Data Processing
We take the daily data for the stock prices and ETF prices and convert them into the log returns. We then subtract the risk-free rate from the Fama French data to get the excess returns for each of these measures. For the retail investor sentiment, the publicly available data is in monthly form so we interpolate the data down to daily searches, limiting the time frame to only trading days in the month to ensure it matches the 252 trading days the stocks and ETFs have over the year.

Over the time frame of January 1, 2004, through January 1, 2024, we end up with 5033 data entries for each stock and ETF. To fairly evaluate each year fully and keep the timeframe synced we use the first 19 full years of data for our regressions.

The relationship between prices and explanatory variables changes heavily over time. To measure this effect we break up our time frame into rolling time frames. We use both discrete and continuous time frames. The first uses 5-year blocks, breaking the data into 2004-2009, 2009-20014, and 2014-2019. For the continuous time frame, we roll over half the time frame, so the first period is 2004-2009, the second period is June 2006- June 2011, and so on.

Preliminary Data Analysis
When interpolating the Google search data, we used Modified Akima cubic Hermite interpolation. The problem with polynomial interpolation or cubic spline interpolation is that sometimes interpolation can lead to negative numbers, which is not feasible. This happened when there was a zero search followed by a positive search and then a zero search again. Thus, by using cubic Hermite interpolation, we could prevent this from happening. While this prevented interpolation from going below zero, it sometimes showed straight lines between two base points, which is also not an accurate representation of our data. By using Modified Akima Cubic Hermite, we obtained smoother interpolation.

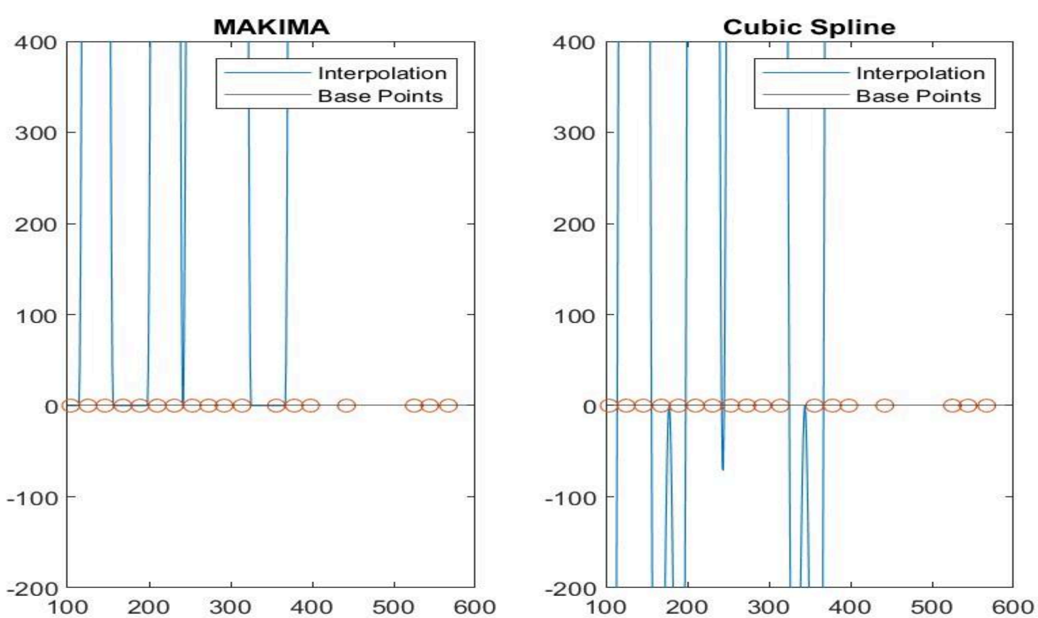


Figure1. Interpolation results of MAKIMA and cubic spline methods

Model Building

Our model is a multivariate linear regression model with 5 regressor variables. The variables are a market factor, size factor, value factor, ETF factor, and retail interest factor. Our dependent variable is a daily excess stock return. We use daily frequency for all of our data. After constructing a linear regression model, we check if our model satisfies all the assumptions of the linear regression model. We want to ensure linearity between independent and dependent variables, meaning there exists a linear relationship between them. Next, we check for multicollinearity between independent variables. If we observe multicollinearity, since we only have 5 independent variables, we can simply transform the data to reduce the multicollinearity effect. Then, we can check for homoscedasticity and normality of residuals. Further explanation will be discussed in the model validation section. Lastly, we ensure the independence of each observation. We will assume that stock returns are independent and identically distributed.

Model Validation

In time series and regression analysis, validating the predictive models is as crucial as constructing them. For our project, which aims to refine the Fama-French three-factor model by adding new factors, we employed a rigorous validation process to ensure the robustness and reliability of our findings. The techniques we used are rooted in statistical best practices and are tailored to address the unique challenges posed by financial time series data.

Residual Analysis: A fundamental step in our validation was the examination of residuals—the differences between the observed values and those predicted by our model. By plotting these residuals, we could visually inspect for any patterns or systematic deviations that might indicate model inadequacies. Additionally, we utilized the partial autocorrelation function (PACF) to detect any lingering correlations in the residuals. The absence of significant correlations in the PACF plots would suggest that our model has successfully captured the information in the data.

Ljung-Box Test: To statistically test for autocorrelation, we applied the Ljung-Box test to the residuals and squared residuals of our models. This test is particularly useful for identifying any remaining dependencies in the time series data that our model may not have accounted for. A non-significant p-value from this test would indicate that the residuals are independently distributed, affirming the model's adequacy.

Rolling Window Analysis: Recognizing the dynamic nature of financial markets, we implemented a rolling window analysis to evaluate the stability of our model over time. This technique involves re-estimating the model parameters over different subsets of the data, akin to a simulation of real-world forecasting. By observing how the model's performance varies across these windows, we can assess its consistency and adaptability to changing market conditions.

Through these validation techniques, we aimed to ensure that our extended Fama-French model is not only statistically sound but also practically relevant. The insights from this process informed our interpretation of the results and guided our recommendations for the model's application in investment decision-making.

Results:

The simple linear regression results for NVDA, LLY, and BYO for the full-time frame are shown below.

NVDA

```
lm(formula = NVDA_Excess_returns ~ marketFactor + SMB + HML +
    SOXX_Excess_returns + NVDA_searches)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.39005 -0.00986 -0.00076  0.00926  0.21281 

Coefficients:
                            Estimate        Std. Error t value             Pr(>|t|)    
(Intercept)          0.0015687280274  0.0004211615375   3.725             0.000198 ***
marketFactor        -0.0000516652220  0.0002510321726  -0.206             0.836947    
SMB                  0.0001359044798  0.0005009201437   0.271             0.786165    
HML                  0.0001504153561  0.0003830840015   0.393             0.694600    
SOXX_Excess_returns  1.2328294025854  0.0152118987964  81.044 < 0.0000000000000002 ***
NVDA_searches       -0.0000000003036  0.0000000029650  -0.102             0.918436    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02053 on 4782 degrees of freedom
Multiple R-squared:  0.5818,	Adjusted R-squared:  0.5814 
F-statistic:  1331 on 5 and 4782 DF,  p-value: < 0.00000000000000022
```

AIC: - 23617.36

LLY

```
lm(formula = LLY_Excess_returns ~ marketFactor + SMB + HML +
    ICF_Excess_returns + LLY_searches)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.110285 -0.007612  0.000164  0.007432  0.138218 

Coefficients:
                       Estimate    Std. Error t value             Pr(>|t|)    
(Intercept)        -0.0079053196  0.0007081027 -11.164 < 0.0000000000000002 ***
marketFactor       -0.0001293022  0.0001781519  -0.726               0.4680    
SMB                 0.0006003746  0.0003543472   1.694               0.0903 .  
HML                 0.0006391117  0.0002714889   2.354               0.0186 *  
ICF_Excess_returns  0.3883470385  0.0103704978  37.447 < 0.0000000000000002 ***
LLY_searches        0.0000001837  0.0000000230   7.989  0.00000000000000169 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01451 on 4781 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.2479,	Adjusted R-squared:  0.2471 
F-statistic: 315.2 on 5 and 4781 DF,  p-value: < 0.00000000000000022
```

AIC: -26930.86

BYON

```
lm(formula = BYON_Excess_returns ~ marketFactor + SMB + HML +
    XLV_Excess_returns + BYON_searches)

Residuals:
     Min       1Q   Median       3Q      Max
-0.51994 -0.02005 -0.00128  0.01754  0.32809

Coefficients:
                        Estimate    Std. Error t value           Pr(>|t|)
(Intercept)         0.0018172616  0.0013424835   1.354             0.1759
marketFactor        0.0000579967  0.0005391606   0.108             0.9143
SMB                -0.0002643359  0.0010778028  -0.245             0.8063
HML                -0.0018365565  0.0008253266  -2.225             0.0261 *
XLV_Excess_returns  1.2218585801  0.0517172587  23.626 <0.0000000000000002 ***
BYON_searches      -0.0000001922  0.0000001891  -1.016             0.3095
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04421 on 4781 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.1061,    Adjusted R-squared:  0.1052
F-statistic: 113.5 on 5 and 4781 DF,  p-value: < 0.00000000000000022
```

AIC: -16266.92

For NVDA we see that only the intercept and superconductor ETF are statistically significant. We go on to see that for each stock the sector ETF is statistically significant at the 99.99% level in the regression model. In the regression of LLY stock, we do see that retail investor interest is statistically significant at the 99.99% level.

For the robustness of the results, we run the Box-Jung test on the residuals to check for autocorrelation. For all three regressions, the P-value for the test was above 0.05, indicating that there is no significant autocorrelation in any of the residuals of these models.

The root mean squared error of our model for NVDA, LLY, and BYON are 0.0205, 0.0145, and 0.0442, respectively. This means that our model predictions are within 1.5 to 4.5 percent of the actual return.

We go on to the discrete rolling time frame regressions. We see from the results below that for NVDA and BYON in all but one time frame the Fama French factors are insignificant. The regression of LLY is the only one that shows significance for retail investor interest but only in

Table 1: NVDA Discrete Time Frame Regression Results

Timeframe	Intercept	Market Factor	SMB	HML	SOXX	Retail Investor Interest
2004-2009	0.004020084 *	0.00071421901	0.00000409484	0.00042109638	1.33579356972 ***	0.00000002235
2009-2014	-0.003827673	-0.00083535641	0.00074916695	0.00063355393	1.21171486973 ***	0.00000010267
2014-2019	0.001342655	-0.00075315501	0.00073352879	0.00153743808	1.26363130775 ***	0.000000001986
2019-2023	0.001346130	0.00010239316	-0.00004349016	-0.00029586145	1.21317815050 ***	-0.0000000001543

*: p-value < 0.05 **: p-value < 0.01 ***: p-value < 0.001

Average AIC: -5997.10875

Table 2: LLY Discrete Time Frame Regression Results

Timeframe	Intercept	Market Factor	SMB	HML	XLV	Retail Investor Interest
2004-2009	0.0542541892 ***	-0.0015992153 ***	0.0035877181 ***	0.0022403047 **	-0.0017921828 ***	-0.0000005145 ***
2009-2014	-0.0071736131 *	0.0000120472	-0.0001804906	-0.0013719694 *	-0.0000567995	0.0000003734
2014-2019	0.0211819166 ***	0.0003044643	-0.0005239859	-0.0004589466	-0.0002517774 ***	-0.0000001235
2019-2023	-0.0191637733 **	-0.0015836283 ***	0.0018440322 *	-0.0003835519	0.0000062772	0.0000003774

*: p-value < 0.05 **: p-value < 0.01 ***: p-value < 0.001

Average AIC: -6625.388

Table 3: BYON Discrete Time Frame Regression Results

Timeframe	Intercept	Market Factor	SMB	HML	ICF	Retail Investor Interest
2004-2009	0.0122166337	-0.0001219743	0.0048815384 ***	-0.0022764389	-0.0006217540 **	-0.0000001271
2009-2014	0.0274354542	-0.0006828637	-0.0024562017	-0.0023094049	-0.0004230452	-0.0000014702
2014-2019	-0.01055513957	-0.00117911078	-0.00013096445	-0.00344118896	0.00015599771	0.00000005984
2019-2023	0.040645725	-0.000444487	-0.000126009	-0.000889571	-0.000975514 **	0.000002440

*: p-value < 0.05 **: p-value < 0.01 ***: p-value < 0.001

Average AIC: -4048.471

the first time period. We do also see that across all regressions as the time moves from 2004 towards 2023 the coefficient of the sector ETF and retail investor interest decreases. We once again perform the Box-Jung test on residuals to check for autocorrelation in the model. The BYON 2019-2023 timeframe is the only regression that shows statistically significant evidence of autocorrelation.

We see across the board that the AIC scores for the discrete time frames are much larger than that of the full time frame model. This indicates that these models are worse overall compared to the full time frame.

We then perform continuous time frame regressions where the first half of the previous time frame remains in the data pool. While the SOXX ETF remains significant over all time periods

Table 4: NVDA Rolling Time Frame Regression Results

Timeframe	Intercept	Market Factor	SMB	HML	SOXX	Retail Investor Interest
2004-2009	0.0040 *	0.0007	0.0000	0.0004	1.3358 ***	2.235e-8
06/2006- 06/2011	-0.0019	-0.0001	-0.0002	0.0004	1.3300 ***	9.653e-8
2009 - 2014	-0.0038	-0.0008	0.0007	0.0006	1.2117 ***	1.0267e-7
06/2011-06/2016	-0.0057 *	-0.0010	-0.0000	0.0007	1.0569 ***	1.9214e-7*
2014-2019	0.0013	-0.0008	0.0007	0.0015	1.2636 ***	1.986e-9
06/2016-06/2021	0.0003	-0.0001	0.0001	0.0001	1.2167 ***	6.700e-9
2019-2023	0.0013	0.0001	-0.0000	-0.0003	1.2132 ***	-1.543e-10

*: p-value < 0.05 **: p-value < 0.01 ***: p-value < 0.001

Average AIC: -6757.9

We see that for both LLY and BYON, the ETF was significant from 2004-2011 but lost significance over time. Interestingly we see that for retail investor interest as time goes on for NVDA the coefficient gets smaller, it remains around the same size for LLY, and increases in size for BYON.

Also of note is the average AIC of the rolling regressions is lower for all three companies when compared to their discrete counterparts. While the scores are still larger than the full-time period versions it does indicate that a rolling time period creates a better model for pricing. There is potential that different-sized windows or rolling methods could perhaps outperform a full-time period model.

Table 5: LLY Rolling Time Frame Regression Results

Timeframe	Intercept	Market Factor	SMB	HML	XLV	Retail Investor Interest
2004-2009	0.0543***	-0.0016***	0.0036***	0.0022**	-0.0018***	-5.15e-07***
06/2006- 06/2011	-0.00078	-0.0015***	0.0029***	0.0013*	-0.0013***	1.55e-06***
2009 - 2014	-0.0072*	0.000012	-0.00018	-0.0014*	-5.68e-05	3.73e-07*
06/2011-06/2016	0.00070	-0.00016	0.000018	-0.00036	-1.87e-05	2.87e-08
2014-2019	0.0212***	0.00030	-0.00052	-0.00046	-0.00025**	-1.24e-07
06/2016-06/2021	-0.0238***	-0.0017***	0.0013.	0.000014	0.000044	4.32e-07**
2019-2023	-0.019**	-0.0016***	0.0018*	-0.00038	6.28e-06	3.77e-07

*: p-value < 0.05 **: p-value < 0.01 ***: p-value < 0.001

Average AIC: -6727.85

Table 6: BYON Rolling Time Frame Regression Results

Timeframe	Intercept	Market Factor	SMB	HML	ICF	Retail Investor Interest
2004-2009	0.0122	-0.000122	0.0049*	-0.0023	-0.000622**	-1.27e-07
06/2006- 06/2011	0.0182	-9.64e-05	0.0022	-0.0027	-0.00067***	-2.10e-07
2009 - 2014	0.0274*	-0.000683	-0.0025	-0.0023	-0.000423	-1.47e-06*
06/2011-06/2016	0.0238	-0.00176*	-0.0024	-0.0040*	-0.000314	-1.88e-06
2014-2019	-0.0106	-0.00118	-0.000131	-0.00344	0.000156**	5.98e-08
06/2016-06/2021	0.0271	0.000665	0.00161	-0.00207	-0.00112*	6.60e-06*

2019-2023	0.0406**	-0.000444	-0.000126	-0.000890	-0.000976**	2.44e-06

*: p-value < 0.05 **: p-value < 0.01 ***: p-value < 0.001

Average AIC: -4269.025

Conclusion

In conclusion, our study sought to enhance the predictive power of the Fama-French three-factor model by incorporating additional variables that capture the nuances of the modern financial landscape. Through an empirical analysis encompassing three diverse stocks—Beyond Inc (BYON), Eli Lilly and Company (LLY), and NVIDIA Corporation (NVDA)—paired with their respective sector ETFs and Google search analytics data, we aimed to provide insights into the factors driving stock returns in today's dynamic market environment.

While the inclusion of sector-specific ETFs as factors demonstrated statistical significance in certain time intervals, suggesting that these instruments may indeed capture industry-specific risks affecting stock returns, Google search analytics did not yield statistically significant predictors of stock performance within our model. This challenges the notion that retail investor interest, as measured by search volume, is a reliable indicator of stock performance.

Furthermore, our model validation process, which included residual analysis, the Ljung-Box test for autocorrelation, and rolling window analysis, confirmed the robustness of our extended Fama-French model.

Future Scope

Looking ahead, there are several potential avenues for further research. One possibility is to transform the retail interest factor, such as by taking the percentage change of search numbers, to see if this enhances the predictive power of the model.

Another potential area of exploration is the inclusion of other types of data in the model. For instance, social media sentiment analysis or news sentiment analysis could be used to capture investor sentiment more accurately. Additionally, the model could be tested on different markets or sectors to assess its applicability and robustness in different contexts.

Finally, the model could be extended to include more factors, such as macroeconomic indicators or firm-specific characteristics, to capture a broader range of influences on stock returns. By exploring these avenues, future research can continue to enhance our understanding of stock returns and the factors that drive them, contributing to the ongoing evolution of financial modeling in the digital age.

References

1)Allen, D. E. (2018, December 9). 'Choosing Factors' by Fama and French (2018): A Comment. Search eLibrary:: SSRN. Retrieved April 25, 2024, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3272608

2) Ji, Z. (n.d.). Empirical Research on the Fama-French Three-Factor Model and a Sentiment-Related Four-Factor Model in the Chinese Blockchain Industry. MDPI. Retrieved April 25, 2024, from https://www.mdpi.com/2071-1050/12/12/5170

3) Jin Xu. (2014). The Fama-French Three Factors in the Chinese Stock Market. https://www.researchgate.net/publication/314575048_The_Fama-French_Three_Factors_in_the_Chinese_Stock_Market

4) Anwar, M., & Kumar, S. (2018). Three-Factor Model of Asset Pricing: Empirical Evidence from the Indian Stock Market. The IUP Journal of Applied Finance, 24(3), 16-34
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3275917

5) Daniel, K., Titman, S., & Wei, K. C. J. (1999). Explaining the Cross-Section of Stock Returns in Japan: Factors or Characteristics? NBER Working Paper No. 7246. National Bureau of Economic Research
https://www.nber.org/system/files/working_papers/w7246/w7246.pdf

6) Taneja, N. (2010). Revisiting Fama French Three-Factor Model in Indian Stock Market. The IUP Journal of Applied Finance, 16(4), 5-19
https://www.investopedia.com/terms/f/famaandfrenchthreefactormodel.asp

7) Fama, E. F., & French, K. R. (2011). Size, Value, and Momentum in International Stock Returns. Fama-Miller Working Paper, Tuck School of Business Working Paper No. 2011-85, Chicago Booth Research Paper No. 11-10
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1720139

8) Berk, J. B., & Van Binsbergen, J. H. (2014). Assessing Asset Pricing Models Using Revealed Preference. NBER Working Paper No. 20435. National Bureau of Economic Research https://www.nber.org/system/files/working_papers/w20435/revisions/w20435.rev0.pdf