

# ***Research on the Application and Expansion of Fama-French Model in Healthcare Industry (ETF)***

**Jiawen Hao<sup>1,a,\*</sup>**

<sup>1</sup>*Adam Smith Business School, University of Glasgow, Glasgow, Scotland, UK*  
*a. 2721945h@student.gla.ac.uk*

*\*corresponding author*

**Abstract:** Research about the Fama-French (F-F) model has long attracted significant academic attention. This study explores the application and extension of the F-F model within the healthcare sector. Using the Vanguard Health Care ETF (VHT) as a representative of the healthcare sector, this research examines its monthly returns over the period from 2014 to 2024. A comparison between the Fama-French three-factor (FF3) and five-factor models (FF5) reveals that the FF5 model is more suitable for the healthcare industry, the impact on ETF returns increased by 3.82%. To identify a model better suited for the healthcare sector, this study extends the FF5 model by incorporating an additional 11 factors. Through stepwise regression and significance testing, the final model with the highest adjusted  $R^2$  value and strongest explanatory power for ETF returns was constructed. The independent variables in this Extension model include market excess returns ( $R_m - R_f$ ), HML, CMA, the momentum factor (MOM), inflation rate (INFR), employment change rate (EMP), and the PPI for medical laboratories (PPILAB). Compared to the original FF5 model, which had an adj  $R^2$  of 0.6980, the Adj  $R^2$  of the Extension model increased to 0.7359, demonstrating a better fit for the healthcare industry. As a result, with the help of this study, investors will be better equipped to use the F-F Model and make informed investments in the healthcare sector.

**Keywords:** Healthcare sector, Fama-French model, investment returns, Extension factors.

## **1. Introduction**

The healthcare industry has consistently played a critical role in the global economy and financial markets. Since the 2010s, the generation born in the baby boomer period (1946 ~1964) has gradually entered old age, leading to a growing demand for healthcare services due to the aging of this demographic [1]. This trend has fueled the continuous growth of the healthcare sector and solidified its importance within investment portfolios. As a sector heavily reliant on technological development and innovation, the healthcare industry is favored by investors for its stable growth potential and resilience to economic cycles. Despite the challenges of the COVID-19 pandemic, when the stock market faced significant challenges, the S&P 1500 index showed that the healthcare sector was one of the best-performing industries at the time [2]. As a result, the unique market characteristics of the healthcare industry make studying its investment returns particularly crucial.

Exchange-traded funds (ETFs), as a diversified financial investment tool, provide investors with broad industry exposure by incorporating a range of stocks or assets [3]. Sector-specific ETFs issued within the healthcare field are composed of stocks from various sub-industries, which not only reflect

the overall performance of the sector but also reveal key factors influencing returns in the healthcare industry. Therefore, healthcare sector ETFs offer strong support for studying investment returns.

The Fama-French (F-F) model can be applied to research on investment returns. This model, an extension of the Capital Asset Pricing Model (CAPM), is commonly used to explain differences of returns in the financial market [4]. Currently, the research on the F-F model predominantly focuses on stock markets in various regions, only with fewer studies specifically targeting certain industries. Ekaputra and Sutrisno found that the Fama-French three-factor (FF3) model was more applicable than the five-factor (FF5) model in the stock markets of Singapore and Indonesia [5]. Fatima and Haroo analyzed the influence of CAPM and FF3 models on mutual fund performance in Pakistan [6]. Although some studies have examined the application of the F-F model in specific markets. For example, Korenak's research identified that, in emerging market ETFs, market excess returns and small-cap companies positively contributed to performance [7]. The degree of explanation of the F-F model varies across regions and sectors. When focusing on the healthcare sector, the complexity and uniqueness of the industry—such as R&D costs, workforce dynamics, and healthcare levels—may cause the applicability of the FF3 and FF5 models to differ in their effectiveness.

The aim of this study is to assess the F-F model's suitability in healthcare sector ETFs, analyzing its effectiveness in interpreting this sector's excess returns. Additionally, it seeks to add more extension factors into the model, identifying key drivers of returns and developing a new model more suitable for the healthcare market. Through this approach, the study intends to reveal the limitations of traditional factor models in the healthcare sector, and offer investors more effective tools to evaluate the risks and returns of healthcare sector ETFs.

## 2. Data and Methodology

### 2.1. Data and Variables

#### 2.1.1. Variable definition

This study includes several risk factors of the F-F model, with data sourced from the Kenneth R. French Data Library [8]. It also includes economic and investment market-related factors such as INTR, INFR, UR, and VIX, with data obtained from the Fred database. Additionally, EMP, CONS, MCLEVEL, PPILAB, and EPI represent healthcare sector-specific factors, with data downloaded from the U.S. Bureau of Labor Statistics. Data on investor confidence (UNVCONFI) was collected from the "Sentiment Survey Historical Data" available through The American Association of Individual Investors (AAII) [9]. A summary of all variables utilized in this research is provided in Table 1.

Table 1: Variables and descriptions.

Variables	Descriptions
Ri	Total return of Health Care ETF (VHT)
Rf	Risk-free rate of return (One-month Treasury bill rate)
Rm	Total market portfolio return
Rm-Rf	Market excess returns; Market risk premium
SMB	Size premium, returns of small-cap minus the returns of large-cap firms.
HML	Value premium, high book-to-market (B/M) equity returns minus low book-to-market (B/M) equity return
RMW	RMW (Robust Minus Weak), Returns of profitable businesses less returns of unprofitable businesses

Table 1: (continued).

CMA	CMA, return differences between conservative and aggressive investment styles
MOM	Momentum factor, which captures the momentum effect of an asset
INTR	Federal funds rate
INFR	10-Year Break-even Inflation Rate in the U.S.
UR	The total unemployment rate in the U.S.
VIX	CBOE Volatility Index (VIX), which represents the market's estimation of the recent volatility based on stock index option prices.
EMP	All employees in the US health care industry, percentage change
CONS	U.S. healthcare industry total construction spending, percentage change
INVCONFI	Investor confidence: Bullish, Percent
MCLEVEL	Average U.S. city Medical care, all city consumers, 1-month percentage change
PPILAB	Producer Price Index (PPI) Industry Data for Medical Laboratory - Medical Laboratory Services, June 1994 Index = 100
EPI	Monthly export price index percent change, Pharmaceutical products

### 2.1.2. Data Collection

The study selects the Vanguard Health Care ETF (VHT) as the research subject. VHT is an ETF focused on the healthcare sector, and compared to other same industry ETFs such as XLE and IBB, it covers a broader range of fields, including 418 stocks from various sub-sectors such as pharmaceuticals, medical devices, biotechnology, and health insurance [10]. Therefore, VHT is used to represent the healthcare market in this study, and analyze its ETF return variations.

A multiple regression model will be used, with a monthly frequency ( $t=1$  month), to analyze the excess returns of VHT ( $R_i - R_f$ ) over a ten-year period from 2014 to 2024. The study targets to find the key factors affecting its returns. The monthly stock prices of VHT will be used to compute the ETF returns. Yahoo Finance was used to download ETF stock data [11].

## 2.2. Methodology

### 2.2.1. Model

The Fama-French three-factor (FF3) model is widely used to explain the returns of the US stock market, which incorporates several risk factors that affect market returns [4]. Proposed by Fama and French in 1993, this model aimed to improve the explanatory strength of the CAPM concerning excess stock market returns. In addition to the original market risk premium factor ( $R_m - R_f$ ), the size factor (SMB) and the B/M ratio factor (HML) are two further components added in the model. These factors account for stock return variations beyond what CAPM predicts, and are effective factors influencing stock market returns and investment risk. The formula is presented as follows in Equation (1).

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + s_iSMB_t + h_iHML_t + \varepsilon_{it} \quad (1)$$

A further development of the FF3 model, the Fama-French five-factor model (FF5) was unveiled by Fama and French in 2015 [12]. To further enhance the model's impact on stock return, the FF5 model considers firm characteristics. As shown in Equation (2), the model adds two new independent

variables: the profitability factor (RMW) and the investment style factor (CMA). By capturing both market and firm-specific characteristics more comprehensively, the FF5 model improves the accuracy of asset pricing, making it more effective at explaining return anomalies compared to the FF3 model.

$$R_{it} - Rf_t = \alpha_i + \beta_i(Rm_t - Rf_t) + s_iSMB_t + h_iHML_t + w_iWML_t + \varepsilon_{it} \quad (2)$$

### 2.2.2. Research method

This study improves and extends the F-F factor models to explore the effective factors affecting the returns of the healthcare industry. The F-F model is based on the entire stock market in US, which consists of stocks listed on several stock exchanges, including the NASDAQ, AMEX, and NYSE. This research will first examine the applicability of the FF3 and FF5 models within the VHT portfolio. The model will use the ordinary least squares (OLS) method to select the model with the strongest explanatory power.

Secondly, the study extends the selected model by adding additional 11 expansion factors. Given the large number of independent variables, stepwise regression will be employed to screen and evaluate the model. Stepwise regression aims to simplify the model and optimize its predictive capability by gradually adding or removing independent variables [13]. This method typically uses the p-value significance level (5% level) for variable selection and Adjusted R-squared to assess model optimization. Stepwise regression can be performed in three main ways: forward selection (starting from no variables and progressively adding significant ones), backward elimination (starting from a full model and gradually removing insignificant variables), and bidirectional elimination (a combination of the first two methods, adjusting variables based on significance levels). This study will use bidirectional elimination to sequentially test all factors, aiming to construct the most accurate and explanatory new Extension model while reducing model complexity.

## 3. Result and Analysis

### 3.1. Descriptive statistics

The study provides a summary of the basic data for all variables, as shown in Table 2. Each variable contains 121 observations. Most of the variables are expressed in percentage (%) form. Notably, the mean value of the PPILAB variable is relatively high at 122.5582, as this dataset is presented as an index (Index Jun 1994=100). Additionally, the VIX variable exhibits considerable volatility, with a standard deviation of 6.7101%. This heightened volatility may be attributed to market uncertainty triggered by factors such as releases of economic data, earnings reports of corporations, and policy changes, which lead to sharp fluctuations in investor sentiment and, consequently, significant movements in option prices.

To prevent the model from being affected by multicollinearity, variables with high linear correlations were excluded during the initial screening process. The correlation test results for the independent variables indicate that, apart from a slight multicollinearity between PPILAB and INTR (-0.5129), no significant linear correlations were found among the other variables. Therefore, multicollinearity is not expected to significantly impact the accuracy of the regression results.

Table 2: Variables summaries.

Variable	Obs	Mean	Std. dev.	Min	Max
Ri-Rf	121	0.7600	4.2109	-9.6701	13.4445
Rm-Rf	121	0.9173	4.5331	-13.38	13.65

Table 2: (continued).

SMB	121	-0.1490	2.7723	-5.95	7.37
HML	121	-0.1557	3.7210	-13.83	12.88
RMW	121	0.3636	2.1323	-4.76	7.2
CMA	121	-0.0320	2.3485	-6.81	7.74
MOM	121	0.1217	3.9614	-16.02	9.98
INTR	121	1.3107	1.5839	0.05	5.33
INFR	121	1.9726	0.3695	0.99	2.88
UR	121	4.8835	1.7694	3.4	14.8
VIX	121	18.0820	6.7101	10.13	57.74
EMP	121	0.1496	0.9211	-9.3	2.4
CONS	121	0.4835	1.8811	-4.7	5.1
INVCONFI	121	33.4639	7.3442	18.964	52.9705
MCLEVEL	121	0.2174	0.2894	-0.6	0.9
PPILAB	121	122.5582	2.3151	119.3	126.2
EPI	121	-0.1207	0.8380	-3.8	1.9
VHT_RETURN	121	0.8603	4.1998	-9.4701	13.4445
RF	121	0.1003	0.1316	0	0.47

### 3.2. Comparison of Fama-French models

The FF3 Model (1) and FF5 Model (2) were first individually regressed using the OLS method. The regression results are shown in Table 3.

In the FF3 Model, the market excess returns ( $R_m - R_f$ ) and book-to-market factors (HML) were statistically significant ( $p < 0.05$ ), while the t-test of size factor (SMB) was not significant. Additionally, a large portion of the ETF's excess return was explained by the market excess returns. In the FF5 Model, SMB remained statistically insignificant, but the significance of the HML improved in this regression. Moreover, the newly added profitability factor (RMW) was not applicable to the VHT stock portfolio, as its p-value was less than 0.05. Therefore, neither model fully explained the VHT's returns.

The regression results indicate that the Adj  $R^2$  of the FF5 Model was 0.6980, higher than the FF3 Model (Adj R-square = 0.6598). This comparison confirms that the FF5 Model provides a stronger explanation for VHT's excess returns and is more suitable for healthcare ETFs. Therefore, the following research will focus on extending the FF5 Model by adding other risk factors.

Table 3: Comparison of Fama-French models.

	(1)	(2)
	Fama-French Three Factor Model (FF3)	Fama-French Five Factor Model (FF5)
Rm-Rf	0.7511*** (14.49)	0.8079*** (15.63)
SMB	0.0139 (0.16)	0.0357 (0.37)

Table 3: (continued).

HML	-0.1438*	-0.3555***
	(-2.38)	(-4.53)
RMW		-0.0669
		(-0.56)
CMA		0.5173***
		(4.09)
_cons	.0005	0.0097*
	(0.22)	(0.04)
N	121	121
Adj R-squared	0.6598	0.6980
t statistics in parentheses		
* p<0.05, ** p<0.01, *** p<0.001		

### 3.3. Extension model

After screening out the optimal Fama-French model, the study conducted a stepwise regression on the FF5 Model five factors and 11 additional factors, resulting in a more refined and explanatory model for the healthcare ETF (VHT) excess returns. The regression results for the FF5 Model (2) and the Extension Model (3) are compared in Table 4.

The comparison reveals that Rm-Rf, HML, and CMA remain statistically significant in the expanded model and are included in the new model, while the SMB and RMW were excluded. Moreover, the coefficients of these factors were minimally affected in the new model, with slopes remaining relatively stable. Among the 11 additional factors, the momentum factor (MOM), inflation rate (INFR), change in employment (EMP), and Producer Price Index for medical laboratories (PPILAB) were retained in the expanded model based on significance testing.

The R-squared test shows that the extended model is significantly better at explaining the excess returns of ETFs. Compared to the FF5 Model, the Adj R-square increased from 0.6980 to 0.7359. Therefore, the Extension Model provides a substantial enhancement over the FF5 model, it is more applicable to the healthcare market.

Extension Model equation:

$$R_{it} - R_{ft} = -30.9695 + 0.8540(Rm_t - R_{ft}) - 0.3316HML_t + 0.4694CMA_t + 0.1452MOM_t + 1.3987INFR_t - 0.4998EMP_t + 0.2302PPILAB_t \quad (3)$$

Table 4: Comparison of Fama-French Five Factor model and Extension Model.

	(2)	(3)
	Fama-French Five Factor Model (FF5)	Extension Model
Rm-Rf	0.8079***	0.8540***
	(15.63)	(17.09)
SMB	0.0357	
	(0.37)	
HML	-0.3555***	-0.3316***
	(-4.53)	(-4.08)
RMW	-0.0669	

Table 4: (continued).

	(-0.56)	
CMA	0.5173***	0.4694***
	(4.09)	(3.96)
MOM		0.1452*
		(2.45)
INFR		1.3987 *
		(2.30)
EMP		-0.4998*
		(-2.26)
PPILAB		0.2302*
		(2.48)
_cons	0.0097*	-30.9695*
	(0.04)	(-2.61)
N	121	121
Adj R-squared	0.6980	0.7359

#### 4. Discussion

In the Extension Model, seven factors significantly impact ETF excess returns. Among these, the  $R_m - R_f$  and the CMA factors exhibit a strong positive influence on healthcare ETF returns, the coefficients are 0.8540 and 0.4694. Conversely, the HML factor and employment change (EMP) show a negative relationship with the dependent variable, with coefficients of -0.3316 and -0.4998, respectively. The inflation rate (INFR) shows a strong correlation, the coefficient is 1.3987. Additionally, the Momentum Factor (MOM) and Medical Laboratory PPI (PPILAB) have coefficients of 0.1452 and 0.2302.

The healthcare sector is often viewed as a defensive sector, favored by investors for its stable demand and growth potential [14]. Market excess returns ( $R_m - R_f$ ) reflects the overall performance of the market. When the market performs well, investors typically allocate more funds to the healthcare sector to achieve risk diversification and secure stable returns (positive coefficient). Therefore, changes in market excess returns can effectively predict the returns of healthcare sector ETFs.

In the Extension model, the inverse relationship for HML indicates that growth stocks (low B/M ratio) outperform value stocks (high B/M ratio). This negative correlation may arise because many companies in the healthcare sector, especially in pharmaceuticals and biotechnology, generally have higher growth potential and R&D investments. These characteristics are consistent with growth stocks rather than value stocks. Thus, HML exerts an effective negative impact on the excess returns of VHT.

The return difference between conservative, low-investment businesses and aggressive, high-investment businesses is measured by the CMA factor. Given the stable demand and revenue characteristics of the healthcare sector, many healthcare companies tend to adopt more conservative investment strategies to maintain business and financial stability. In the expanded model, the positive coefficient of CMA reflects this investment characteristic, indicating that conservative investment strategies positively impact the returns of healthcare sector ETFs.

The positive coefficient of MOM indicates that assets (or sectors) with strong past performance are likely to maintain that momentum and continue performing well in the future. As a unique field driven by significant R&D and innovation, the healthcare industry often sees major technological

breakthroughs or drug approvals from pharmaceutical or clinical companies, which can influence future investment directions. These developments attract more investor attention, further boosting future performance. Therefore, it is essential to consider the importance of the momentum factor when analyzing market returns.

The inelastic demand in the healthcare industry allows it to respond to rising costs during periods of high inflation by increasing the prices of its products and services. This price adjustment mechanism, which passes costs onto consumers, can even enhance the industry's profitability during inflationary periods. Consequently, the inflation rate exhibits a strong positive correlation with market excess returns.

An increase in the percentage change in the number of employees (EMP) signals rising costs. In the healthcare industry, expanding the workforce in clinics, hospitals, and pharmaceutical companies often leads to higher expenses related to salary payments, admissions, employee training, and welfare. If business performance does not improve correspondingly, profit margins may be severely compressed. Therefore, the EMP factor in the extended model shows a significant negative correlation with ETF returns.

Medical laboratories are specialized institutions or facilities that assist in diagnosing, treating, and controlling diseases, serving as a key role in the proper functioning of the healthcare system and patient health management [15]. Innovations in laboratory testing techniques and analytical accuracy typically result in a rise in the Producer Price Index (PPI). An increase in PPI positively impacts dividends, often indicating an enhancement in the company's profitability [16]. As a result, investor confidence in related industries or companies grows, driving stock and sector returns higher. In the consumer market, technological advancements increase demand for services, which in turn reflects on industry performance and has a significant impact on corporate earnings.

## 5. Conclusion

The ten-year study of VHT market returns reveals that the FF5 Model is more applicable to the healthcare sector than the FF3 Model, and has a stronger effect on the investment returns of healthcare ETFs. However, after expanding the model, the inclusion of new factors further enhanced the explanatory power of the FF5 Model on ETF returns, with the Adjusted  $R^2$  increasing by 3.79%. Through stepwise regression, seven significant factors—Rm-Rf, HML, CMA, MOM, INFR, EMP, and PPILAB—were selected for inclusion in the extended model.

Factors such as Rm-Rf, HML, and CMA not only affect the overall financial market returns but also have a significant impact on healthcare sector earnings. MOM shows a positive relationship with the returns of the healthcare sector based on asset performance trends. As a key indicator of national economic dynamics, INFR influences healthcare sector returns from a macroeconomic perspective. Additionally, EMP also impacts ETF returns due to its close correlation with costs. PPILAB positively influences healthcare sector earnings by driving innovation and meeting demand.

In summary, while the Fama-French model proves useful in the healthcare industry, it has certain limitations. Investors can more accurately assess healthcare sector returns and risks by focusing on the seven factors mentioned. However, 26.41% (1-0.7359) of the variations in VHT returns remain unexplained. Future research could explore additional factors to incorporate into the extended model, providing a more comprehensive explanation of investment returns in the healthcare industry.

## References

- [1] Cohn, D., & Taylor, P. (2010). *Baby Boomers approach 65*. Retrieved from <https://s3.amazonaws.com/static.pseupdate.mior.ca/media/links/Baby%20Boomers%20Approach%2065%20glumly.pdf>
- [2] Mazur, M., Dang, M., & Vega, M. (2021). *COVID-19 and the march 2020 stock market crash. Evidence from S&P 500*. *Finance research letters*, 38, 101690.

- [3] Kuziak, K., & Górka, J. (2023). *Dependence analysis for the energy sector based on energy ETFs*. *Energies*, 16(3), 1329.
- [4] Fama, E. F., & French, K. R. (1996). *Multifactor explanations of asset pricing anomalies*. *The Journal of Finance*, 51(1), 55.
- [5] Ekaputra, I. A., & Sutrisno, B. (2020). *Empirical tests of the Fama-French five-factor model in Indonesia and Singapore*. *Afro-Asian J of Finance and Accounting*, 10(1), 85.
- [6] Fatima, S. Z., & Haroon, M. (2018). *Performance Evaluation of Pakistan's Mutual Fund through CAPM and Fama French 3-Factor Model*. *Archives of Business Research*, 6(3).
- [7] Korenak, B. (2022). *The Fama-French Five-Factor Model: Evidence of the Emerging Markets Exchange Traded Funds Performance*. In *Conference Proceedings Determinants Of Regional Development* (No. 3, pp. 56-67).
- [8] Kenneth R. French - Data Library. (n.d.). Retrieved from [https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)
- [9] American Association of Individual Investors. (n.d.). *Sentiment Survey Historical Data*. Retrieved from [https://www.aaai.com/sentimentsurvey/sent\\_results](https://www.aaai.com/sentimentsurvey/sent_results)
- [10] VettaFi. (2024). *Vanguard Health Care ETF*. Retrieved from <https://etfdb.com/etf/VHT/#etf-ticker-profile>.
- [11] Yahoo is part of the Yahoo family of brands. (n.d). *Vanguard Health Care Index Fund ETF Shares (VHT)*. Retrieved from <https://finance.yahoo.com/quote/VHT/profile/>
- [12] Fama, E. F., & French, K. R. (2015). *A five-factor asset pricing model*. *Journal of Financial Economics*, 116(1), 1–22.
- [13] Johnsson, T. (1992). *A procedure for stepwise regression analysis*. *Statistical Papers*, 33(1), 21-29.
- [14] Chen, H., Estes, J., & Pratt, W. (2018). *Investing in the healthcare sector: mutual funds or ETFs*. *Managerial Finance*, 44(4), 495-508.
- [15] Baker, F. J., & Silvertown, R. E. (2014). *Introduction to medical laboratory technology*. Butterworth-Heinemann.
- [16] Sheikh, M., Bokaei, M. N., Alijani, H., Saadatmand, M., Fard, S. M. H., & Sharahi, I. C. (2011). *Investigating relationship between consumer price index and producer price index and dividend per share*. *Australian Journal of Business and Management Research*, 1, 121-128.