Analyzing the data from beverage companies to evaluate the validity of the Fama-French three-factor model

Wenhui Xue*

Faculty of arts and humanities & Culture Media, London, King's college London

*Corresponding author: xue.wenhui@kcl.ac.uk

Abstract. The Fama-Fench 3-factor model (FF3) is one of the most commonly used models for valuing companies. It is curious to see whether this model can reasonably capture the changes in the treacherous stock market and accurately predict the expected returns of a company in practice. In this article, machine learning is adopted to analyze data from the past five years in terms of the 10-year t-bill return as the risk free rate and the S&P 500 rate of return as the market rate of return to compare and analyses the rate of return of Nestle, Danone, Unilever, Coca-Cola and PepsiCo over the past five years. In the comparative analysis, this essay randomly selected the rate of return for 80% of the days in the past five years as the training set and used the rate of return for the remaining 20% of the days as the test set. Besides, this study uses regression analysis to obtain a theoretical model of FF3, which is then brought to the test set data for testing. In this article, the null hypothesis is that the error is equal to 0 and the alternative hypothesis is that the error is not equal to 0. According to the analysis, hypothesis validation is adopted to check the accuracy of FF3 by analyzing the results of the hypothesis test and using the FF3 derived from the training set to predict whether the error between the resulting data and the actual data is 0.

Keywords: Fama-French three-factor model; Nestlé; Market risk; return of stock investment.

1. Introduction

Contemporarily, valuation of stock remains interesting issues to investigate and various models have been proposed which have been verified based on empirical analysis [1-4]. A great amount of economists and investors have conducted many times of attempt to forecast the price fluctuations. Among various models, CAPM is an idealized model to show how market prices and determine the expected return on capital, and estimating the cost of equity would lead to subjective results [1, 2, 5, 6]. Therefore, a more complex and accurate model is introduced: The Fama French Three-Factor model [1, 7-10]. The Fama French Three-Factor model is established to estimate expected excess return over the risk-free rate for a particular stock. It is an asset pricing model that expands on CAPM by adding three factors in the following to the market risk in CAPM. The Fama French model states that the expected return can be explained by three factors: market factor, size factor, and book-to-market factor. The general equation is:

$$R(t) - RF(t) = \alpha + \beta[RM(t) - RF(t)] + sSMB(t) + hHML(t) + e(t) \quad (1)$$

In this equation, $\boldsymbol{RM(t) - RF(t)}$ is the return on market index minus risk-free interest rate. Small minus Big (**SMB**) and High minus Low (**HML**) are two key components and factors. The coefficient in front of the factors illustrates the sensitivity of the stock return relative to each factor. First, β measures the market risk sensitivity. Positive beta means the stock has a positive correlation with the market return. Second, s measures the tilt toward small stocks since the model assumes that small stocks tend to have higher return than the stocks of large and mature companies. Lastly, h measures the strategy's tendency to be tilted toward stocks with a high B/M since the higher book-to-market ratio, the larger the return the stock will have. The intercept term α measures the excess return.

At the beginning of the establishing the FF3, it is found that small companies tend to outperform the larger ones in the long run, and value companies tend to outperform growth companies. And this model is more often used by professional analysts rather than average investors, but investors can still

merge the premise of the model into their strategies [8]. In the original version of the model, the factors applied were specifically to four countries, which are the U.S., Canada, the U.K. and Japan; however, with the increasing needs for the model and further adjustment made, it gradually become applicable for other areas including Eroupe and the Asia-Pacific region [3, 4].

When the model is first established, it is widely accepted and used by professionals to estimate the expected return of stocks. However, many studies have proved that the Fama French model cannot accurately measure the expected return of a stock [9-12]. Therefore, Nestle (NSRGY) and its related beverage industry data are chosen to test whether the model is valid with historical data. The comparable companies chosen were 5 beverage industry, which ranked Top 10 in the world based on their sales revenue: Unilever (UL), PepsiCo (PEP), Coca-Cola (KO), Conagra (CAG), Danone (BN.PA). These companies are chosen since their sales revenue and market capitalization are comparable to those of Nestle. The rest part of the paper is organized as follows. The Sec. 2 will introduce the origination of the data. The Sec. 3 will demonstrate the results and give explanations.

2. Data & Method

2.1 Data

The data incorporated was obtained from Yahoo Finance from July 2017 to July 2022 regarding 10-year U.S. treasury bond rate (since the default risk of U.S. government is low), each company's historical stock prices, market capitalization, and book-to market ratio, and S&P500 historical index price.

Since the goal is to estimate the annual return of Nestle stock but the 10-year U.S. treasury bond rates are provided daily, the average daily rate is annualized and incorporated as the risk-free into the model. The rate of change in each company's stock prices are computed to get the actual return on stock which is later used to compare with the expected return. The index price in t is also used to divide by the index price in t-1 to get the market return which is used to capture the market risk in the model. The market capitalization of the 6 companies in total are ranked to distinguish between small stocks and large stocks. The book-to market ratio are ranked to distinguish between high B/P stocks and low B/P stocks.

2.2 Model

The datasets are separated into two groups, i.e., train sets (accounted about 80%) and test sets (accounted about 20%). The model is developed based on the equation $R(t) - RF(t) = \alpha + \beta[RM(t) - RF(t)] + sSMB(t) + hHML(t) + e(t)$. $RM(t)$ incorporated the index return. $RF(t)$ uses the annualized 10-year U.S. treasury bond rates. According to the ranks obtained for the size factor and book-to market factor, difference of return on stocks was computed to be used as $SMB(t)\ and\ hHML(t)$. After the model is operated, a hypothesis testing based on each error term aiming to test whether the results of the machine learning are accurate through the significance level of error will be performed.

3. Emprical analysis

According to hypothesis testing, if the error is not significant, which means that the trained model is suitable for empirical testing of 20% data, it implies that it would rely on the Fama French model to predict the expected return of stocks. However, if the error is tested to be significant according to hypothesis testing, which means the model fails to pass the empirical test, then one would conduct debugging process according to the selected data and processing methods, further analyze the particularity of the selected industries and stocks, and summarize the reasons for the failure. The Table. 1 lists the capitalization of different corporations.

In order to determine whether the comparable stock has a larger or small size compared with Nestle, the data regarding each company's market capitalization from Yahoo Finance was incorporated. As

shown in the Table. 1, Nestle has a larger market capitalization than each of the 5 comparable companies. Therefore, it is concluded that the return of Nestle stock should be modelled in as the large-cap inputs whereas each of the other companies' return would be used as small-cap inputs for the size factor.

After that, the SMB factor in Fama French model can be computed. To determine whether the comparable stock has a higher or lower book-to-price ratio compared with Nestle, the data regarding each company's book-to-price ratio from Yahoo Finance and the result is incorporated and shown in Table. 2. The 6 companies' book-to-price ratio are different. If a comparable stock has a higher book-to-price ratio than Nestle, the return of the comparable stock would be inputted into the model as return for high book-to-price ratio. If a comparable stock has a lower book-to-price ratio than Nestle, the return would be used as return for low book-to-price ratio. After that, the HML factor in the Fama French model can be computed.

Table. 1 The capitalization of different corporations.

Capitalization of Nestle	336.434 Billion
Capitalization of Unilevel	123.799 Billion
Capitalization of PepsiCo	242.715 Billion
Capitalization of Coca-Cola	275.349 Billion
Capitalization of Conagra Brand	16.314 Billion
Capitalization of Danone	34.589 Billion

Table. 2 The Book/price ratio of different corporations.

Book/price ratio of Nestle	1/7.28
Book/price ratio of Unilevel	1/6.03
Book/price ratio of PepsiCo	1/13.15
Book/price ratio of Coca-Cola	1/12.02
Book/price ratio of Conagra Brand	1/1.86
Book/price ratio of Danone	1/1.99

Subsequently, the first model was constructed the model based on the return of Nestle (NSRGY) and Unilever (UL). The linear regression model can be written as:

$$R(t) - RF(t) = -2.002e^{-5} + 2.509e^{-1}[RM(t) - RF(t)] - 3.741e^{-1}[SMB(t) + hHML(t)] \quad (2)$$

In this model comparing Nestle (NSRGY) and Unilever (UL), Nestle is both the small-cap and high-BP-ratio stock, so the size factor and the book-to-price ratio factor are combined automatically. Due to the nearly identical situation of both companies, the coefficients generated are close to zero. The metrics values are: $mean\ error = 0.0005647976, sd\ error = 0.00957467, test - statistic = 0.93642$, where $p\ value = 0.35$. Based on the hypothesis testing at 95% confidence level, it failed to reject the hypothesis of true mean being not zero since the p-value is 0.35>0.05. This implies that the established model can be used to accurately forecast the rest 20% data of return. The results are summarized in Table. 3.

Table. 3 The regression results of model 1.

	Estimate	Std.Error	T value	P-value
β_0	-2.002e-05	2.392e-04	-0.084	0.933
β_1	2.509e-01	2.054e-02	12.218	<2e-16
β_2	-3.741e-01	1.311e-02	-28.547	<2e-16

Afterwards, the second model was constructed the model based on the return of Nestle (NSRGY) and Coca-Cola (KO). The linear regression model can be written as:

$$R(t) - RF(t) = 7.144e^{-5} + 1.421e^{-1}[RM(t) - RF(t)] + 4.807e^{3}SMB(t) + 4.807e^{3}hHML(t) \quad (3)$$

The metrics are $mean\ error = 0.000494971, sd\ error = 0.0112631, test - statistic = 0.69762$ with $p\ value = 0.4861$. Based on the hypothesis testing at 95% confidence level, it failed to reject the hypothesis of true mean being not zero since the p-value is 0.4861>0.05. This implies that the established model can be used to accurately forecast the rest 20% data of return. The summary of the models is presented in Table. 4.

Table. 4 The regression results of model 2.

	Estimate	Std.Error	T value	P-value
β_0	7.144e-05	3.154e-04	0.226	0.821
β_1	1.421e-01	3.381e-02	4.204	2.86e-05
β_2	4.807e-01	1.047e-02	0.046	0.963
β_3	4.807e-01	1.047e-02	0.046	0.963

The 3rd model was constructed the model based on the return of Nestle (NSRGY) and PepsiCo (PEP). The linear regression model can be written as:

$$R(t) - RF(t) = 1.239e^{-4} + 2.426e^{-1}[RM(t) - RF(t)] + 1.861e^{5}SMB(t) + 1.861e^{5}hHML(t) \quad (4)$$

where the metrics are $mean\ error = 0.000574965, sd\ error = 0.010881, test - statistic = 0.83878$ with $p\ value = 0.4024$ Based on the hypothesis testing at 95% confidence level, it failed to reject the hypothesis of true mean is not zero since the p-value is 0.4024>0.05. This implies that the established model can be used to accurately forecast the rest 20% data of return. The results are listed in Table. 5.

Table. 5 The regression results of model 3.

	Estimate	Std.Error	T value	P-value
β_0	1.239e-04	3.212e-04	0.386	0.6999
β_1	2.426e-01	3.446e-02	7.040	3.57e-12
β_2	1.861e+05	1.047e+05	1.733	0.0833
β_3	1.861e+05	1.047e+05	1.733	0.0833

The 4th model was constructed the model based on the return of Nestle (NSRGY) and Conagra (CAG). The linear regression model can be written as:

$$R(t) - RF(t) = -3.840e^{-4} + 3.99e^{-11}[RM(t) - RF(t)] - 5e^{-1}SMB(t) - 5e^{-1}hHML(t) \quad (5)$$

where the metrics are $mean\ error = -1.740515e-05, sd\ error = 0.0151427, test - statistic = -0.018246$ with $p\ value = 0.9855$. Based on the hypothesis testing at 95% confidence level, it rejects the hypothesis of true mean is not zero since the p-value is 0.9855>0.05. This implies that the established model cannot be used to accurately forecast the rest 20% data of return. The results are presented in Table. 6.

Table. 6 The regression results of model 4.

	Estimate	Std.Error	T value	P-value
β_0	-3.840e-04	1.410e-12	-2.723e+08	<2e-16
β_1	3.990e-11	1.295e-10	3.080e-01	0.758
β_2	-5.000e-01	7.082e-11	-7.060e+09	<2e-16
β_3	-5.000e-01	9.228e-11	-5.418e+09	<2e-16

Table. 7 The regression results of model 5.

	Estimate	Std.Error	T value	P-value
β_0	-6.416e-05	2.559e-04	-0.251	0.802
β_1	1.510e-01	2.258e-02	6.685	3.84e-11
β_2	-3.766e-01	1.555e-02	-24.220	<2e-16

The 5th model was constructed the model based on the return of Nestle (NSRGY) and Danone (BN.PA). The linear regression model can be written as:

$$R(t) - RF(t) = -6.416e^{-5} + 1.510e^{-1}[RM(t) - RF(t)] - 3.766e^{-1}[SMB(t) + hHML(t)] \quad (6)$$

In this model comparing Nestle (NSRGY) and anone (BN.PA), Nestle is both the small-cap and high-BP-ratio stock, so the size factor and the book-to-price ratio factor are combined automatically.

where the metrics are $mean\ error = 0.0005810549$, $sd\ error = 0.008907057$, $test - statistic = 1.0356$ with $p\ value = 0.3014$. Based on the hypothesis testing at 95% confidence level, it failed to reject the hypothesis of true mean is not zero since the p-value is 0.3014>0.05. This implies that the established model can be used to accurately forecast the rest 20% data of return.

4. Limitations

As a matter of fact, there are some defects and limitations for the results of this paper. To be specific, the risk-free rate applied in this study is derived and retrieved from the average rate of 10-year T-bills. Besides, the S&P500 index was used to estimate the market return rate. However, it is clearly not enough to reach to an accurate answer since there are many more markets besides the U.S. market. In addition, the quantity of corporations chosen in the sample is still a relatively small amount, and Nestle is the largest company in its industry, making it less comparable to other companies in the field. Hence, biases exist.

5. Summary

In conclusion, this paper evaluates the validation of FF3 in the context of the data from beverage industry. Based on the above analysis, it can be concluded that the FF3 is valid for the stock returns selected. This means that the three factors incorporated in the model can be proved contributive to stock returns. Also, this analysis process also showed the usefulness of machine learning such that the trained model has no significant error and can be used to conduct similar analysis in the future. However, there are some drawbacks and shortcoming s of this model, e.g., larger standard errors. Moreover, it is noticed that the coefficient and intercept in every model are close to zero which implies that the explained variable has low sensitivity to these factors. Although the validation of FF3 is verified, it is still questionable how well do these three factors explain the return of a stock. Therefore, the factors to include other fundamentals or macroeconomic factors could be expanded to fully analyze the return of stocks, and longer range of time and data are needed to evaluate the complete industry statistics to get more accurate results. Overall, these results offer a guideline for industry evaluation in terms of FF3.

References

[1] Fama E F, French K R. A Five-factor Asset Pricing Model. Journal of Financial Economics, 2014, 116(1).

[2] Capm K, Jel C G. The use of CAPM and Fama and French Three Factor Model: portfolios selection. 2012.

[3] CFI TEAM. (2022, February 18). Fama-French three-factor model. Corporate Finance Institute. Retrieved August 13, 2022, from https://corporatefinanceinstitute.com/resources/knowledge/finance/fama-french-three-factor-model/

[4] CFI TEAM. (2022, May 8). Capital Asset Pricing Model (CAPM). Corporate Finance Institute. Retrieved August 13, 2022, from https://corporatefinanceinstitute.com/resources/knowledge/finance/what-is-capm-formula/

[5] Hayes, A. (2022, July 8). What is the fama and French three factor model? Investopedia. Retrieved August 13, 2022, from https://www.investopedia.com/terms/f/famaandfrenchthreefactormodel.asp.

[6] Mullins, D. W. (2014, August 1). Does the capital asset pricing model work? Harvard Business Review. Retrieved August 13, 2022, from https://hbr.org/1982/01/does-the-capital-asset-pricing-model-work

[7] Reed, E. (2021, May 12). How does the fama french 3 factor model work? SmartAsset. Retrieved August 13, 2022, from https://smartasset.com/investing/fama-french-3-factor-model

[8] Allen D E, Mcaleer M J. Drawbacks in the 3-Factor Approach of Fama and French. Econometric Institute Research Papers, 2019.

[9] Phong N A, Hoang T V. Applying Fama and French Three Factors Model and Capital Asset Pricing Model in the Stock Exchange of Vietnam. International Research Journal of Finance and Economics, 2012, 95.

[10] Rehman A. Asset Pricing Dynamics in Karachi Stock Exchange: A Comparative Study of CAPM, ICAPM, Fama – French and Momentum Factor Models, 2011.

[11] Piela, Katarzyna. Evaluation of the CAPM and the Fama-French Asset Pricing Models. Lap Lambert Academic Publishing, 2013.

[12] Karp A, Vuuren G V. The Capital Asset Pricing Model And Fama-French Three Factor Model In An Emerging Market Environment. International Business & Economics Research Journal (IBER), 2017, 16(4):231.