A COMPARISON OF CAPM AND FAMA-FRENCH THREE-FACTOR MODEL UNDER MACHINE LEARNING APPROACHING

Bui Thanh Khoa
Faculty of Commerce and Tourism, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

Tran Trong Huynh
Department of Mathematics, FPT University, Hanoi, Vietnam

ABSTRACT

With the economy experiencing rapid growth in recent years, more individuals have started venturing into the stock market. Precisely forecasting the rate of return can mitigate investment risks for stock investors and significantly enhance their investment returns. The Capital Asset Pricing Model (CAPM) and the 3-factor Fama-French model (FF3) are widely recognized in academic and practical settings. This model comparison provides frameworks to analyze the relationship between portfolio risk and return in inefficient markets, contributing to applied data science in finance behavior. This research utilized the Support Vector Regression (SVR) algorithm to forecast the returns of a diversified portfolio in the Hanoi stock market (HNX) from 2010 to 2022. Initially, this study calculated the factors and subsequently constructed diversified portfolios. Subsequently, the explanatory power of the CAPM and FF3 models were compared using the Ordinary Least Squares (OLS) algorithm. Finally, this research incorporated the SVR algorithm within the FF3 framework to develop a predictive model. The research findings demonstrate that the FF3 model provides a superior explanation to the CAPM model. Additionally, the study reveals that the SVR algorithm outperforms the OLS algorithm in terms of efficiency, as it yields lower Root Mean Square Error (RMSE) values. Nevertheless, despite its advantages, the FF3 model still falls short regarding explanatory factors. Consequently, the next research direction entails replacing the FF3 model with a more comprehensive multi-factor model, anticipating obtaining an enhanced predictive model.

Keywords: Fama-French 3-factor model; CAPM; return; SVR; OLS; machine learning

DOI: https://doi.org/10.15549/jeecar.v10i7.1402

INTRODUCTION

The capital Asset Pricing Model (CAPM) establishes the link between portfolio risk and return. Because of its accessibility and simplicity, the CAPM model has become a vital resource for asset management in recent years. The CAPM model classifies a stock's total risk as idiosyncratic and systematic. Portfolio diversification may help reduce exposure to individual risks, but it has little effect on systemic risk (Zaimovic et al., 2021; Adaramola et al., 2011). CAPM presupposes that finding the optimum risk-return profile for a portfolio is feasible.

Furthermore, the ideal portfolio must include all assets, with each asset's value-weighted, to accomplish the above since adding a new asset increases the portfolio's diversification. The efficient frontier is the set of all possible optimum portfolios, one for each possible rate of return. Because unsystematic risk may be spread over several investments, a portfolio's overall risk can be represented by a single number, beta. However, there is debate about whether CAPM should be used in actual investment settings. Some experimental CAPM tests found a linear connection between anticipated return and portfolio risk but with a slope coefficient angle too flat to be considered meaningful (Khoa & Huynh, 2022; López Prol & Kim, 2022). Dhankar (2019) analyzed data for 158 equities traded on the Bombay Stock Exchange using a battery of tests spanning from 1991 through 2002, which closely corresponds to the timeframe after the liberalization and implementation of capital market reforms; consequently, CAPM is invalid as it did not apply to the Indian stock market over the period under consideration. Peng (2021) examined data from the UK market between December 2016 and December 2019 and concluded that, experimentally, CAPM is not appropriate because it makes too many assumptions that are hard to satisfy. Adding to the CAPM's original two elements, the 3-factor Fama-French model (FF3) suggested a three-factor model considering both size and value. FF3 is a flawed model for anticipated returns since it only accounts for some variance in average returns due to profitability and investment (Khoa et al., 2022). When volatile, the market obscures the value impact (Fama & French, 2021); therefore, ensuring a linear link between the explanatory variables and the outcome is challenging. FF3 is the extended model of CAPM because FF3 added two factors, size, and value. In traditional econometrics, FF3 and CAPM are often estimated by the ordinary Least Squares (OLS); however, OLS is not as effective as compared with SVR in the variables' nonlinear relationship (Li & Li, 2021).

Econometrics has seen the emergence of Machine Learning (ML) algorithms as a viable alternative in recent years. The Support Vector Regression (SVR) algorithm is highly efficient for forecasting continuous variables (Khoa & Huynh, 2023). SVR's power lies in its ability to effectively use the nonlinear connection between variables (Zheng et al., 2021). Chiu et al. (2020) estimated future gold prices using Least Squares Support Vector Regression (LSSVR) by using the spot price of gold and the Opinion score calculated by mining data from news articles published in Taiwan between January 1, 2016, and December 31, 2017. Gold, silver, platinum, palladium, and the opinion score are the independent variables, whereas the one-period delayed variable is part of the period. Furthermore, MAPE was used as the benchmark for quality evaluation. The Wilcoxon and Friedman tests were to examine the performance of two prediction models and pointed out that the LSSVR method when a Genetic Algorithm optimizes its parameters, achieves lower MAPE values and increased prediction accuracy. These studies proved that the SVR algorithm is crucial in constructing predictive models. Khoa and Huynh (2021) mentioned how inefficient the Ho Chi Minh Stock Exchange (HOSE) is, which makes CAPM invalid on HOSE for the time being. The same characteristics between HOSE and HNX motivated this research to center on the advantages of the FF3 theoretical model and the SVR algorithm (FF3SVR). The combination is expected to bring higher efficiency than the original models, as this research tested the monthly HNX market cycle from January 2010 to December 2022. This research primarily aims to compare the FF3SVR model with the CAPM and the FF3 to assess the FF3SVR's efficacy. The OLS technique was calculated t. The F-test is then used to determine how successful they were. Notable results from this research include the following:

- Validate the Fama-French three-factor model's superior performance in the HNX market over the capital asset pricing model (CAPM).
- Construct a return-rate forecasting model using the FF3 framework and the SVR algorithm and verify that the resultant model is very efficient.

LITERATURE REVIEW

CAPM and 3-factor Fama-French model

Using a linear connection between anticipated return and risk, the Capital Asset Pricing Model (CAPM) may be used to calculate the optimal portfolio allocation. There are two types of risk in CAPM: systematic and unsystematic. Portfolio diversity may help reduce nonsystematic risk but

cannot eliminate systematic risk. Inflation, interest rates, and economic cycles are all systemic risks that broadly impact the market. Unsystematic risks, however, are unique to each business and include potential corporate leadership, strategy, and culture changes. Since investors may lessen their exposure to company-specific risk by diversification, CAPM states that only market risk should be compensated for (Peng, 2021).

In the CAPM model, here is how the regression equation looks:

$$r_{it} - r_{ft} = \alpha_i + \beta_i Mkt_t + \varepsilon_{it}$$

Some research found that size impacts economic years (Horvath & Wang, 2021; Hu et al., 2019). For example, tiny businesses may reap more profits than their larger counterparts. According to the findings, the CAPM fails to explain the scale impact adequately. Fama and French (1992) discovered other stock-related impacts beyond the well-known value-added effects. Investment returns are higher for value firms (those with a high B/M ratio compared to market value) than for growth equities (those with a low B/M). The market is one element in the three-factor model proposed by Fama and French (later called a Fama-French 3-factor model). As soon as it was released, the 3-factor model supplanted the CAPM as the dominant financial explanation.

Size, market, and B/M are the three components of the Fama–French model. More specifically, the market factor, represented by the portfolio's outstanding returns, is subtracted from the small firm's return (SMB) to determine the size factor. The B/M factor for the Mkt item is arrived at by subtracting the high-value stocks from the low-value stocks (HML). For a time series, the regression equation looks like this:

$$r_{it} - r_{ft} = \alpha_i + \beta_1 Mkt_t + \beta_2 SMB_t + \beta_3 HML + \varepsilon_{it}$$

Where:

r_{it} = total return of a stock or portfolio i at time t.

r_{ft}= Risk-Free Interest Rate.

Mkt_t = The market's excess rate of return.

SMB_t= Size premium (Small Minus Big).

HML_t = value premium (High Minus Low).

$\beta_{1,2,3}$ = factor coefficients.

α_i = Intercept

ε_{it} = error term

Using thousands of randomly chosen company listings on the US market, Fama and French tested their model. They discovered that when valuation and regulatory parameters were added to the system along with the beta element, the model could account for 89 percent of returns in a diversified portfolio of companies. When an investor accounts for the 89 percent of a portfolio's return that can be explained in terms of the general market, they may create a portfolio in which the average projected return is commensurate to the relative risk they have accepted.

Multiple further trials confirmed the FF3 model's performance better than the CAPM. CAPM and FF3 models were tested in several Asian markets, including those in China, Taiwan, Malaysia, Korea, and Singapore; the results indicated that the FF3 model better captured the link between expected return (Khoa et al., 2023); and market volatility (He et al., 2015). Additionally, the FF3 model's adjusted coefficient of determination is greater than the single-factor models, and the Market Factor is the model's most influential variable. The Indian market was the test environment for Sehrawat et al. (2020), which spanned between 2003 and 2019. Adjusted R^2 coefficients for CAPM vary from 64% to 91%, lower than those for FF3. This result indicates that CAPM is less effective than FF3 (64 percent to 93 percent).

Support Vector Regression

Relying on the success of SVM (a classification algorithm) in the classification setting, some research developed SVM for real value prediction, the so-called SVR (Parbat & Chakraborty, 2020; Zhong et al., 2019). In the field of machine learning, SVM is often used for the analysis of data in many dimensions. To solve the SVM classification problem, we must determine which hyperplane is most effective for classifying the training data. This hyperplane will be used to plot the test set data. If you want to find the optimal hyperplane, you need to optimize, and you do so by making the distance between the support vectors bigger (points closest to the subclass hyperplane). As an alternative to identifying the subclass hyperplane, SVR provides a -insensitive loss function to compute a hyperplane such that the predicted value is within an interval, termed -SVR model, for the regression issue. The linear -

SVR problem entails approximating a function such that the insensitive tube is as flat as possible. When minimizing w, any value will do.

$$\min_{w} \|w\|^2 \text{subject to} |y_i - w^T x_i - b| \leq \varepsilon \qquad (Eq.1)$$

Little external perturbations may make extreme changes in Eq.(1). The presence of noise will be more prominent, resulting in a more severe error in the report. The boundary value is shifted to the outside using an offset parameter to dampen the effect of extreme data points. Eq. (2) provides the answer to the -SVR problem:

$$\min_{w} \frac{1}{2}\|w\|^2 + C\sum_{1}^{n}(\xi_i + \xi_i^*) \text{ subject to} - (\varepsilon + \xi_i^*) \leq y_i - w^T x_i - b \leq \varepsilon + \xi \qquad (Eq.2)$$

A regularization parameter C > 0 determines the trade-off between the flatness of function f and the prediction errors. As the compensatory factors decrease and the s increases, the flatness is reduced for a particular value of C. The sensitivity-free loss function is given by Eq.(3):

$$L\big(y, f(x)\big) = L(\xi) = max(0, |\xi| - \varepsilon) \qquad (Eq.3)$$

When the connection between the variables is not linear, a linear function may result in significant mistakes if used. To fix the nonlinear issue, -SVR projects the original data onto a higher-dimensional space where a kernel assures the linear connection. Standard kernels include the Radial kernel, Linear kernel, and Polynomial kernel, all of which are useful for problems with a similar structure to the SVM (Benkraiem & Zopounidis, 2021; Awad & Khanna, 2015).

METHODOLOGY

Ha Noi Stock Exchange (HNX) information is gathered monthly, starting in January 2010 and ending in December 2022. Size of the Market, Stock Price (M), and Book Value (B) are the Data Points Collected (B). Following the methodology of (Dirkx & Peter, 2020). HML and SMB factors are determined. Return on the HNX-Index minus the interest rate on a one-year government bond is this formula's market element (Mkt) (Fama & French, 2015). The portfolios are rearranged every six months after being sorted by beta from low to high and then divided into ten equally weighted portfolios (Fama & French, 2021). In this case, this study gives equal consideration to the portfolios' respective rates of return. Table 1 provides an overview of the factors such as the Return Rate of Portfolios, Excess rate of return on portfolios, Yields on 1-year government bonds, the rate of return of the HNX-Index, the excess rate of return of the market portfolio, size factor, value factor. The size factor and value factor were calculated from the financial statements of businesses listed in HNX. The data set fits the variables in three methods (CAPM, FF3, FF3SVR).

Table 1: Variable Description

Variable name	Description
p_{it}	Return Rate of Portfolios
$P_{it} = p_{it} - r_{ft}$	Excess rate return on portfolios
r_{ft}	Yields on 1-year government bonds
Rm	The rate of return of the HNX-Index
$Mkt_t = Rm - r_{ft}$	The excess rate of return of the market portfolio
SMB_t	Size factor
SML_t	Value factor

Source: author's work

The Capital Asset Pricing Model (CAPM) is based on several key assumptions that form the model's foundation. Firstly, it assumes that investors are rational and risk-averse, seeking to maximize their wealth. Secondly, it assumes a single-period investment horizon, meaning all investments are evaluated over the same time frame. Thirdly, it assumes that investors have access to the same information and can freely borrow or lend at a risk-free rate. Fourthly, it assumes that the market is efficient, implying that all relevant information is reflected in stock prices. Lastly, it assumes that asset returns are normally distributed and that investors hold well-diversified portfolios. FF3 is extended from CAPM, so its assumption is the same as that of

CAPM. FF3SVR is an FF3 model calculated based on the SVR algorithm instead of OLS.

In practice in Vietnam, these assumptions are difficult to meet. First, lenders and borrowers have different interest rates, and both are based on the risk-free rate. Second, stock trading has transaction costs when investors make purchases through brokers. Securities in Vietnam are traded in batches of hundreds; therefore, the investor can not deal with highly divisible securities into small parcels. Finally, there are no the same homogeneous expectations from investors in Vietnam. Therefore, CAPM and FF3 violate some assumptions, so they operate inefficiently in Vietnam.

The equation for the CAPM regression model looks like:

Eq (4): $P_{it} - r_{ft} = \alpha_i + \beta_i Mkt_t + \varepsilon_{it}$ (Eq. 4)

The equation from the FF3 regression model in Eq (5):

$$P_{it} - r_{ft} = \alpha_i + \beta_1 Mkt_t + \beta_2 SMB_t + \beta_3 HML + \varepsilon_{it} \quad \text{(Eq. 5)}$$

This research compared the time series regression models of CAPM and FF3 using the adjusted coefficient R^2. Eq.4 and Eq.5 were used to estimate the coefficients of the CAPM, FF3, respectively. FF3SVR's coefficients were calculated using Eq.5 and SVR algorithm. After, metrics were compared to find the most efficient model (Gharaibeh et al., 2022).

The study divided the data set into two parts at the ratio of 70:30 to train/test the model. Specifically, the data from January 2010 - December 2018 are used for the training set, and January 2019 – December 2022 are used for the testing set. In addition, the study used Root Mean Square Erro (RMSE) as in previous studies by Nguyen et al. (2021). For the SVR algorithm, the study used the SVR algorithm with a linear kernel, a cost of 0.5, and epsilon = 0.5, as in the previous study by Khoa et al. (2021).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2} \quad \text{(Eq.6)}$$

Where: $\hat{Y}_i$ is the forecast value and Y_i is the actual value.

Finally, when comparing three predictive models, the F test determines whether the null hypothesis, H0, that "there is no difference across models," is true.

RESULT AND DISCUSSION

Result

The HNX equities were separated into ten groups, each representing a different beta level. There are 156 observations, one for each month, from January 2010 to December 2022. The excess of expected return is outstanding in various volatile portfolios, ranging from -0.369 to 0.287. The P4 portfolio has much to do with banks and technology, so it has a high fertility rate.

Table 2: Descriptive statistics

Statistic	Mean	St. Dev.	Min	Max
Mkt	-0.005	0.92	-2.257	4.499
HML	0.112	11.986	-24.628	40.792
SMB	0.9	10.254	-23.729	28.775
Rf	0.052	0.036	0.003	0.135
P1	0.014	3.982	-10.114	10.778
P2	0.193	4.532	-9.737	14.116
P3	-0.148	4.428	-12.572	12.094
P4	0.287	4.708	-12.293	11.694
P5	0.178	4.658	-13.663	12.577
P6	-0.394	6.033	-22.025	18.893
P7	-0.242	5.851	-23.072	18.272
P8	-0.106	5.341	-22.285	13.753
P9	-0.04	5.623	-19.167	12.942
P10	-0.369	5.696	-24.343	15.851

Source: author's work

Due to the impact of the COVID-19 pandemic and recent crises, the stock market tends to go down. Besides, banks' deposit interest rates increased sharply, so companies faced difficulties in capital sources. As a result, some portfolios, P3, P6, and P10, have negative average excess returns. Descriptive statistics were tabulated in Table 2.

As a rule, the risk-free rate is uniform and tilted downward. The risk-free interest rate has been hovering around zero for all of 2019 but has been particularly low at the start of the year. At this point, businesses had many issues with production and supply due to the Covid-19 outbreak. At this juncture, the government issued several stimulus measures, such as interest rate preferences. In addition, the market's volatility at this early stage means that returns are unpredictable. Following the global financial crisis of 2010-2013, the central bank instituted several measures to revive the economy, including the free trading of securities and stock market regulation. There was substantial inflation back then, and several stimulus measures have helped lower the risk-free interest rate to about 8%. (Previously up to 15-20 percent due to the impact of inflation). After the global financial crisis, there has been a general stabilization of market risks.

Standard deviations for the components' average returns ranged from 3.982 to 5.851. The Mkt factor typically returns -0.005 and a risk-free rate of 0.052, implying that investing in the market portfolio will be inefficient compared to investing in bonds. Fig. 1 depicts the changes in the parameters and the risk-free rate.

This study used time series regression to analyze data from CAPM and FF3. See Table III and Table IV for the breakdown of the data. The adjusted R^2 for the CAPM ranges from 0.049 to 0.107, whereas the adjusted R^2 for the FF3 model is higher, at 0.189 to 0.406%. The FF3 model is more effective than the CAPM by these calculations. Several authors have found similar findings in prior research, including Sehrawat et al. (2020); Fama and French (1993). Expressly, a positive linear connection between anticipated return and risk directory is implied by the fact that all beta estimators in CAPM are statistically significant and positive. The CAPM predicts that the Intercept will not be statistically significant in seven out of ten categories.

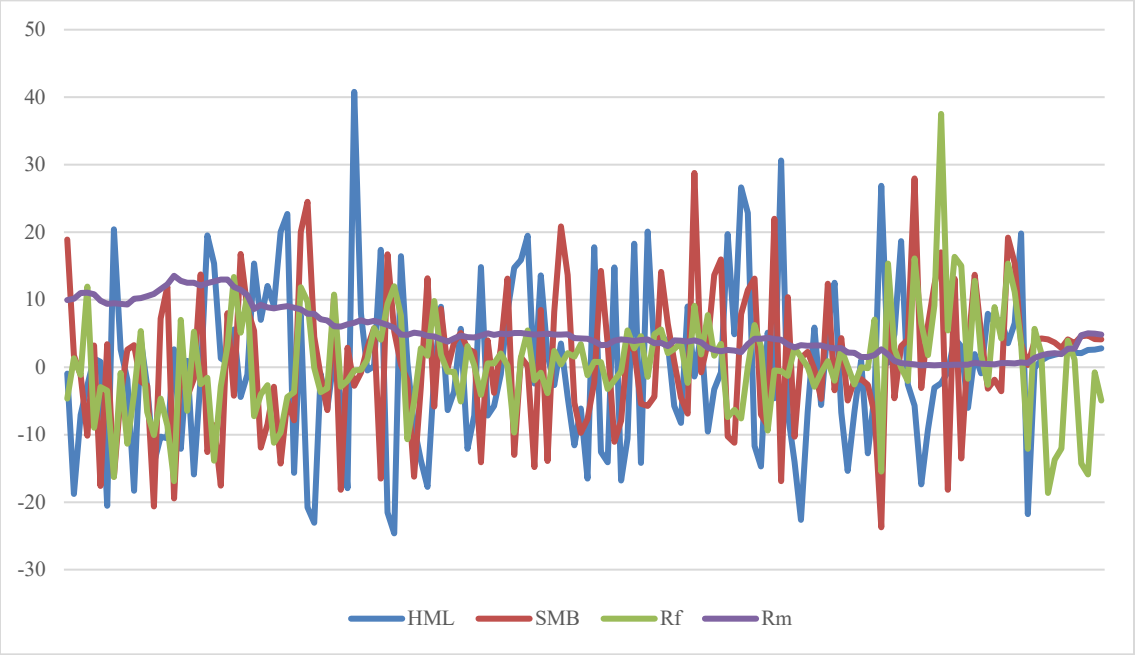

Figure 1. Factor returns and the risk-free rate

Source: author's work

All the calculated Mkt coefficients for the FF3 model have positive signs and are statistically significant. Based on the value impact (HML), these coefficients have opposite signs for

portfolios. Significant computed coefficients for the size factor suggest a scale effect. Companies with a smaller market capitalization often get a more significant risk premium because of their smaller size. This finding agrees with similar research (Basu, 1983). Table 3 and Table 4 pointed out the CAPM and FF3 regression results.

Table 3: CAPM regression results

Portfolio	Intercept	Mkt	Adjusted R^2
P1	0.021	1.415*	0.107
P2	0.199*	1.198*	0.059
P3	-0.141	1.191**	0.061
P4	0.294	1.45***	0.08
P5	0.187*	1.64***	0.105
P6	-0.387	1.452**	0.049
P7	-0.238*	0.693**	0.012
P8	-0.103	0.52*	0.008
P9	-0.03	1.868**	0.093
P10	-0.36	1.7***	0.075

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4: FF3 regression results

P	Intercept	Mkt	HML	SMB	Adjusted R^2
P1	-0.081	1.138**	0.08**	0.102*	0.189
P2	0.023*	0.779**	0.161*	0.174	0.307
P3	-0.337	0.521*	0.103*	0.2***	0.268
P4	0.08	0.77**	0.133	0.218*	0.324
P5	-0.019	1.12***	0.175**	0.204*	0.406
P6	-0.238	1.403*	-0.283*	-0.13	0.353
P7	-0.041	0.899**	-0.276*	-0.183*	0.348
P8	-0.023*	0.254*	-0.24	-0.061	0.282
P9	-0.001	1.387**	-0.228	-0.007*	0.308
P10	-0.204	1.762***	-0.254**	-0.141*	0.359

Note: *p<0.1; **p<0.05; ***p<0.01

Source: author's work

While compared to the CAPM model, the FF3 model is more effective when employing the same method. The FF3 model's mean error is 2.947, but only 3.147 for the CAPM model using the OLS method. The average error for FF3 when using the SVR algorithm is 2.674, the lowest of all tested methods. Table 4 shows the time series regression analysis findings consistent with this finding. The FF3 model outperforms the CAPM regarding R^2, a measure of model fit. This result is why the theoretical framework and the estimate method are crucial components of predictive modeling theory. The results of the forecast error are summarized in Table 5.

Table 5: RMSE error of the models

Portfolio	CAPM	FF3	FF3SVR
p1	2.48	2.94	2.47
p2	3.76	3.05	2.35
p3	2.68	2.89	2.87
p4	2.93	2.97	2.44
p5	3.38	2.85	2.89
p6	3.56	3.15	2.62
p7	2.92	2.92	2.92
p8	3.51	3.01	3.09
p9	3.48	2.88	2.36
p10	2.77	2.81	2.73
Average	3.147	2.947	2.674

Source: author's work

Despite Table 5 disapproval of H0, there is no discernible variation among the models; in particular, the FF3 model using the SVR method has outperformed the lowest average RMSE. A formal statistical test is required to verify this finding, however. Using a One-Way Analysis of Variance (F Test), we examine this hypothesis (ANOVA). There are three forecasting models (CAPM, FF3, and SVRFF3) that are relevant to this discussion (together referred to as the "model"). In this scenario, 30 numbers represent a different category's root-mean-squared error (RMSE) value. Table 6 provides a summary of the analyses of variance. According to Table 6, F_{stat}=MSSB/MSSW = 6.176 and $F_{critical}$ = 3.354, indicating that $F_{stat} > F_{critical}$ (or P = 0.006 < 0.05). This study cannot accept H0; the mean RMSE error varies between models. Considering these findings, it can be concluded that the FF3 model combined with the SVR algorithm has shown promising results.

Table 6: ANOVA result

Source of Variation	SS	df	MS	F	P-value	F critical
Between Groups	1.128	2	0.564	6.176	0.006	3.354
Within Groups	2.465	27	0.091			
Total	3.593	29				

Source: author's work

Discussion

As with any endeavor, there is always a give-and-take between potential gain and potential loss, making risk management crucial for investors. The larger the potential reward, the greater the associated danger. Diversification may mitigate unsystematic risk, while systematic risk cannot. The CAPM model establishes a numeric value for the linear connection between the anticipated return and the systematic risk of a diversified portfolio. Based on the experimental data, it may be concluded that there is a linear connection between all categories. The CAPM theoretical framework predicts a value of 0 for the Intercept, meaning that the regression model has no statistical significance. There is a discrepancy between the experiment and the CAPM prediction in three out of ten categories. That means seven groups agree with CAPM, and four disagree. Findings supporting CAPM align with those of several prior research (Sehrawat et al., 2020; Pei, 2019). In contrast to earlier research, which found R^2 values between 0.64 and 0.92 indicative of CAPM's explanatory ability (Sehrawat et al., 2020), the current study's R^2 values varied from 0.049 to 0.107.

Because of this deficiency in explanatory variables, the CAPM cannot adequately explain why some portfolios have higher or lower projected returns than others. The CAPM serves as the basis for the FF3 model, which then incorporates two additional variables, one each for the scale and value impacts. The projected returns of the portfolios are statistically affected by these factors, and the coefficient R^2 - -adjusted determination is also greatly improved compared to CAPM, with values ranging from 0.189 to 0.406 based on experimental data from the HNX market. Previous research, like that of Sehrawat et al. (2020), Chui and Wei (1998), Fama and French (1993), has shown that the three-factor model is more explanatory than CAPM; therefore, our finding is in line with that literature.

Phong and Hoang (2012) analyzed the use of the first FF3 model in the Vietnamese stock market from 2007 to 2011. The authors found a positive slope between size and stock returns. The results contradict the conclusions of previous research and are likely indicative of the stock market in Vietnam. Nguyen et al. (2019)'s study of the Vietnam market from 2010 to 2017 reveals a strong foreign ownership impact, in which more foreign ownership boosts stock liquidity, profitability, and size and exposes investors to greater risk. This finding demonstrates that the FF3 model's assumed linear connection is insecure or lacks necessary explanatory elements. Our research showed that the FF3 model only had a relatively poor coefficient of determination from 2010 to 2022, ranging from 0.19 to 0.41.

This research constructs a model to foretell the return of diversified portfolios using the benefits of machine learning techniques and the

underlying theoretical framework of CAPM and FF3 valuation models. Results in Tables 5 and 6 show that the combination performed effectively in the trial. Since FF3 has been shown to perform better than CAPM in statistical comparisons, adding SVR makes the combined model even more effective. Gogas et al. (2018) pointed out that the SVR and FF3 model combination is superior to CAPM and is supported by these findings. However, Gogas et al. (2018) research lacks support for the combination model.

Predicting stock market prices and trends is considered a formidable undertaking due to the inherently chaotic nature of the financial markets. The stock market can be a nonlinear, non-parametric, and noisy system deterministic, dominated by factors such as liquidity, stock availability, human behavior, news affecting the market, speculative activities, and international monetary fluctuations. Given its significance as an emerging sector of the economy and the involvement of many stakeholders, researchers and experts are interested in exploring this domain and elucidating the underlying chaotic system for trend pattern recognition. Using the advantages of machine learning algorithms, combined with the underlying theoretical framework of CAPM and FF3 valuation models, we build a model to predict the return of diversified portfolios. The combination has worked well in the experiment; the results in Tables 5 and 6 support this argument. The FF3 model has been statistically proven to be more efficient than CAPM, so the combined model FF3SVR is more efficient than OLSCAPM as a corollary. This result is consistent with the study of Gogas et al. (2018) that the combined model of SVR and FF3 is more effective than CAPM. However, a limitation of Gogas et al. (2018) is that some necessary tests have not been performed to support the conclusion. We overcame it using the ANOVA table to demonstrate the effectiveness through the F-test.

Due to the inherent chaos of the financial markets, predicting stock market prices and trends is often regarded as an arduous task. Market liquidity, stock availability, human behavior, news impacting the market, speculative activity, and international monetary swings are all examples of deterministic nonlinear, non-parametric, and noisy system features that may be applied to the stock market. Experts and scholars are eager to learn more about this field and clarify the chaotic system behind trend pattern identification because of its potential impact on the economy and the wide range of parties involved. This research constructed a model to forecast the profit of diversified portfolios by combining the theoretical framework of the CAPM and the FF3 valuation models with the benefits of machine learning methods. The combination has been effective in the trial, as shown in Tables 5 and 6. Since FF3 has been shown to perform better than CAPM in statistical tests, adding SVR makes the combined model more effective than OLSCAPM. Gogas et al. (2018) find that the SVR+FF3 model outperforms these findings supporting CAPM. However, not all relevant tests have been conducted to back up the findings of Gogas et al. (2018). This study got around this by utilizing the ANOVA table and the F-test to prove its efficacy.

CONCLUSION AND RECOMMENDATION

Each year, at the end of June, the market capitalization, Book to Market ratio, and other measures are used to assign a stock a rating. Combining size sorting with book-to-market ratio (B/M ratio) sorting may provide six categories from which the HML and SMB factors may be derived. In place of Mkt, this research uses the market index (HNX-index), or more accurately, the difference between the return of the HNX-index and the yield on a 1-year government bond. This research provides empirical evidence that the FF3 model outperforms the CAPM on the HNX exchange. Further, portfolio returns may be accurately predicted when the FF3 algorithm uses the SVR model. This research is unusual since it uses a mixed empirical model to analyze the HNX market and shows how effective this method is. Based on its superior predictive power, the FF3SVR model should be used by investors and risk managers in place of the more common CAPM and FF3 models.

The CAPM shows how a portfolio's projected return may be linearly related to its risk. However, it is challenging to make CAPM empirically feasible due to the excessive usage of assumptions. The CAPM-derived three-factor Fama-French model, produced by adding size and value explanation components, has outperformed CAPM. This result was also shown in the HNX market by the research. Predictive

modeling has benefited from using a mixture of many machine learning methods, particularly SVR. Based on these findings, the research suggests replacing the CAPM and FF3 models with the combination model of FF3 and SVR for forecasting portfolio returns.

The FF3 model has been proven to have gaps in its ability to explain predicted returns. A comprehensive model, such as the Fama-French 5-factor model, is required because of this. Predicting portfolio returns using machine learning algorithms like SVR or Long Short-Term Memory Recurrent Neural Network is an extension that uses a 5-factor model.

REFERENCES

Adaramola, M. S., Paul, S. S., & Oyedepo, S. O. (2011). Assessment of electricity generation and energy cost of wind energy conversion systems in north-central Nigeria. *Energy Conversion and Management, 52*(12), 3363-3368. https://doi.org/10.1016/j.enconman.2011.07.007

Awad, M., & Khanna, R. (2015). *Efficient Learning Machines*. New York: Springer nature. doi: 10.1007/978-1-4302-5990-9

Basu, S. (1983). The relationship between earnings yield, market value and return for NYSE common stocks. *Journal of financial economics, 12*(1), 129-156. https://doi.org/10.1016/0304-405x(83)90031-4

Benkraiem, R., & Zopounidis, C. (2021). Preface: Regression methods based on OR techniques and computational aspects in management, economics and finance. *Annals of Operations Research, 306*(1-2), 1-6. https://doi.org/10.1007/s10479-021-04306-7

Chiu, C., Lee, C.-H., & Yuan, F.-C. (2020). Using Market Sentiment Analysis and Genetic Algorithm-Based Least Squares Support Vector Regression to Predict Gold Prices. *International Journal of Computational Intelligence Systems, 13*(1), 234-246. https://doi.org/10.2991/ijcis.d.200214.002

Chui, A. C. W., & Wei, K. C. J. (1998). Book-to-market, firm size, and the turn-of-the-year effect: Evidence from Pacific-Basin emerging markets. *Pacific-basin finance journal, 6*(3-4), 275-293. https://doi.org/10.1016/s0927-538x(98)00013-4

Dhankar, R. S. (2019). Capital Asset Pricing Model: An Overview. *Risk-Return Relationship and Portfolio Management*, 3-17. https://doi.org/10.1007/978-81-322-3950-5_1

Dirkx, P., & Peter, F. J. (2020). The Fama-French Five-Factor Model Plus Momentum: Evidence for the German Market. *Schmalenbach Business Review, 72*(4), 661-684. https://doi.org/10.1007/s41464-020-00105-y

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The journal of finance, 47*(2), 427-465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics, 33*(1), 3-56. https://doi.org/10.1016/0304-405x(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics, 116*(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., & French, K. R. (2021). The Value Premium. *The Review of Asset Pricing Studies, 11*(1), 105-121. https://doi.org/10.1093/rapstu/raaa021

Gharaibeh, O. K., Kharabsheh, B., & Al-Quran, A. Z. (2022). Momentum strategies and market state in Moroccan industries. *Cogent Business & Management, 9*(1), 2135217. https://doi.org/10.1080/23311975.2022.2135217

Gogas, P., Papadimitriou, T., & Karagkiozis, D. (2018). The Fama 3 and Fama 5 factor models under a machine learning framework. Retrieved from https://ideas.repec.org/p/rim/rimwps/18-05.html

He, Z., Zhu, J., & Zhu, X. (2015). Multi-factor volatility and stock returns. *Journal of Banking & Finance, 61*, S132-S149. https://doi.org/10.1016/j.jbankfin.2015.09.013

Horvath, D., & Wang, Y. L. (2021). The examination of Fama-French Model during the Covid-19. *Financ Res Lett, 41*, 101848. https://doi.org/10.1016/j.frl.2020.101848

Hu, G. X., Chen, C., Shao, Y., & Wang, J. (2019).

Fama-French in China: Size and Value Factors in Chinese Stock Returns. *International Review of Finance, 19*(1), 3-44. https://doi.org/10.1111/irfi.12177

Khoa, B. T., & Huynh, T. T. (2021). Is It Possible to Earn Abnormal Return in an Inefficient Market? An Approach Based on Machine Learning in Stock Trading. *Comput Intell Neurosci, 2021*, 2917577. https://doi.org/10.1155/2021/2917577

Khoa, B. T., & Huynh, T. T. (2022, 21-22 Jan. 2022). *Factors Affecting Forecast Accuracy of Individual Stocks: SVR Algorithm Under CAPM Framework.* Paper presented at the 2022 International Conference for Advancement in Technology (ICONAT), Goa, India.https://doi.org/10.1109/ICONAT53423.2022.9725916.

Khoa, B. T., & Huynh, T. T. (2023). The value premium and uncertainty: An approach by support vector regression algorithm. *Cogent Economics & Finance, 11*(1), 2191459. https://doi.org/10.1080/23322039.2023.2191459

Khoa, B. T., Huynh, T. T., & Qin, L. (2022). Long Short-Term Memory Recurrent Neural Network for Predicting the Return of Rate Underframe the Fama-French 5 Factor. *Discrete Dynamics in Nature and Society, 2022*, 1-8. https://doi.org/10.1155/2022/3936122

Khoa, B. T., Huynh, T. T., & Thang, L. D. (2023). Effectiveness of OLS and SVR in Return Prediction: Fama-French Three-factor Model and CAPM Framework. *Industrial Engineering & Management Systems, 22*(1), 73-84. https://doi.org/10.7232/iems.2023.22.1.073

Khoa, B. T., Son, P. T., & Huynh, T. T. (2021). The Relationship Between the Rate of Return and Risk in Fama-French Five-Factor Model: A Machine Learning Algorithms Approach. *Journal of System and Management Sciences, 11*(4), 47-64. https://doi.org/10.33168/jsms.2021.0403

Li, X., & Li, S. (2021). Large-Scale Landslide Displacement Rate Prediction Based on Multi-Factor Support Vector Regression Machine. *Applied Sciences, 11*(4), 1381. https://doi.org/10.3390/app11041381

López Prol, J., & Kim, K. (2022). Risk-return performance of optimized ESG equity portfolios in the NYSE. *Finance Research Letters, 50*, 103312. https://doi.org/10.1016/j.frl.2022.103312

Nguyen, A. P., Nguyen, H. A., Ho, T. H. M., & Ngo, P. T. (2019). Risk and returns of different foreign ownership portfolios: Evidence from Vietnam stock market. *Cogent Economics & Finance*.

Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., . . . Shen, Y.-S. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering, 2021*, 1-15. https://doi.org/10.1155/2021/4832864

Parbat, D., & Chakraborty, M. (2020). A python based support vector regression model for prediction of COVID19 cases in India. *Chaos Solitons Fractals, 138*, 109942. https://doi.org/10.1016/j.chaos.2020.109942

Pei, G. (2019). *Empirical Study on the Listed Banks of China: Based on the CAPM Model.* Paper presented at the 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS).https://doi.org/10.1109/ICITBS.2019.00159.

Peng, S. (2021). *The Validity of CAPM: A Critical and Conclusive Study with Empirical Evidence from the UK Security Market.* Paper presented at the 2021 3rd International Conference on Economic Management and Cultural Industry (ICEMCI 2021).https://doi.org/10.2991/assehr.k.211209.367.

Phong, N. A., & Hoang, T. V. (2012). 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, 95*(8), 114-120.

Sehrawat, N., Kumar, A., Kumar Nigam, N., Singh, K., & Goyal, K. (2020). Test of capital market integration using Fama-French three-factor model: empirical evidence from India. *Investment Management and Financial Innovations, 17*(2), 113-127. https://doi.org/10.21511/imfi.17(2).2020.10

Zaimovic, A., Omanovic, A., & Arnaut-Berilo, A. (2021). How Many Stocks Are Sufficient for Equity Portfolio Diversification? A Review of the Literature. *Journal of Risk and Financial*

Management, 14(11), 551. https://doi.org/10.3390/jrfm14110551

Zheng, J., Wang, Y., Li, S., & Chen, H. (2021). The Stock Index Prediction Based on SVR Model with Bat Optimization Algorithm. *Algorithms, 14*(10), 299. https://doi.org/10.3390/a14100299

Zhong, H., Wang, J., Jia, H., Mu, Y., & Lv, S. (2019). Vector field-based support vector regression for building energy consumption prediction. *Applied Energy, 242*, 403-414. https://doi.org/10.1016/j.apenergy.2019.03.078

ABOUT THE AUTHORS

Bui Thanh Khoa(Corresponding author), email: buithanhkhoa@iuh.edu.vn

Bui Thanh Khoa earned his Master's degree in Business Economics from the Université Toulouse 1 Capitole in France in 2012 and a Doctorate in Business Administration from the Ho Chi Minh City Open University in Vietnam in 2020. His articles are indexed in the SCOPUS and ISI databases. He is a member of the Advisory International Editorial Board of Jurnal the Messenger, an ISI system journal, as well as the editorial boards of Scopus-indexed journals such as Journal of System and Management Sciences; Advances in Operations Research; Journal of Logistics, Informatics and Service Science, as well as the International Journal of Technology Transfer and Commercialisation from Inderscience Publisher. His study interests are methodology, electronic commerce, organizational behavior, machine learning, and consumer behavior.

Tran Trong Huynh is a lecturer at FPT University. He obtained a Master's degree in Mathematics in 2013 at Ho Chi Minh City University of Education and in Finance in 2020 at the University of Economics Ho Chi Minh City. His research interests include finance, applied mathematics, data science, econometrics, and machine learning.