
Research article

Integrating LSTM with Fama-French six factor model for predicting portfolio returns: Evidence from Shenzhen stock market China

Tahir Afzal^{1,*}, Muhammad Asim Afridi² and Muhammad Naveed Jan¹

¹ Department of Management Sciences, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan

² Department of Economics, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22060, Pakistan

* **Correspondence:** Email: Tahirafzal@kust.edu.pk.

Abstract: The study aimed to examine the effectiveness of long short-term memory (LSTM) model in predicting portfolio returns employing Fama and French's six-factor model. Monthly data were collected for A-share prices for firms listed on the Shenzhen Stock Exchange, China, for the period extended over 25 years (1997–2022). Portfolios are constructed on bivariate dependent sorting using the beta and downside beta. The random forest model was employed as a surrogate for LSTM to detect the nonlinear and threshold-based conditional relationship between risk factors and portfolio returns. The findings of the study reveal that LSTM has robust predictive ability and effectively captures the relationship between risk factors and portfolio returns. The relationship is more conspicuous in high downside risk portfolios than in the medium and high beta portfolio groups. Furthermore, portfolios within the high beta group, regardless of their downside beta levels, exhibit positive returns. The study suggests that investors need to consider high-risk portfolios for stable earnings and predictive returns.

Keywords: Fama and French's six factors; LSTM; deep learning; machine learning; downside risk; portfolios; random forest; support vector machine

JEL Codes: G11, G12, G17, C8

1. Introduction and background

The traditional asset pricing models have played a foundational role in understanding market dynamics and investment strategies. However, with the evolving landscape of technological development, these models have fallen short in the accuracy required to predict the returns of stocks and portfolios. Traditional models are based on linear assumptions, whereas stock market data are inherently nonlinear, complex, and voluminous, necessitating tools that are capable of addressing these challenges. In the field of machine learning (ML), deep learning (DL) models in particular have shown promising capabilities in handling such complex, nonlinear patterns in financial data. One such model, the long short-term memory (LSTM) network (Hochreiter & Schmidhuber, 1997), a type of recurrent neural network (RNN) designed for sequential data processing, offers a highly adaptive and flexible framework for analyzing and predicting asset prices and returns (Liu & Wang, 2023; Kundlia & Verma, 2021; Feng et al., 2018).

This study focuses on the Chinese market, specifically the Shenzhen Stock Exchange (SZSE), an emerging and rapidly growing market in an evolving economic landscape presenting unique challenges and opportunities. Unlike developed markets, the SZSE is relatively immature, with a predominance of individual investors having less experience in comparison with their counterparts in developed markets (Yang & Yang, 2022).

The study integrates the theoretical foundations of the Fama–French six-factor model (FF6F), the predictive capabilities of LSTM, and the construction of portfolios based on downside risk. This hybrid approach aimed to enhance the predictive accuracy of portfolio returns.

Summary of the existing literature

The foundations of asset pricing theories, namely the modern portfolio theory (MPT) (Markowitz, 1952) and the capital asset pricing model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966), provided the initial framework for the risk and return relationship, but their reliance on simplified assumptions limits their explanatory power. To overcome these limitations, Fama and French (1992, 2015, 2018) introduced their multi-factor models by including size, value, profitability, investment, and momentum factors in the models. The explanatory power of the model improved, but their results were mixed, particularly in developing markets (Goo & Wang, 2024; Jia, 2023; Zaghoulani & Mezzez, 2020). These traditional models are based on linear assumptions, whereas stock market data are inherently nonlinear and complex, necessitating tools that are capable of addressing these challenges (Sharma & Bhalla, 2022; Ashwood, 2013).

ML models, particularly long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997), have the capacity to capture complex and nonlinear patterns in financial data. Studies show that LSTM has outperformed not only the traditional model but other ML models as well (Aminimehr et al., 2022; Petrozziello et al., 2022; Staudemeyer & Morris, 2019). Economic interpretability remains a concern for the ML models; thus, recent research suggests their integration with economically interpretable models for meaningful insights (Giglio et al., 2022).

The traditional models lack predictive accuracy, which the ML models provide, but the ML models lack economic grounding. Additionally, the downside risk remains unexplored in the sample market. This study bridges these gaps by hybridizing LSTM with the FF6F model to achieve optimal predictive accuracy while providing the necessary economic interpretability. Moreover, the study also

incorporates downside risk in the construction of portfolios to better capture the systematic risk in emerging markets like China.

1.1. Literature review

Traditional asset pricing models

The development of modern portfolio theory (MPT) and the capital asset pricing model (CAPM) represented significant progress in financial economics, which has fundamentally altered investors' approach to asset allocation and risk management. The foundation for MPT was laid down by Markowitz (1952) with the concept of mean–variance optimization in portfolios. The MPT was groundbreaking, but the problem that hindered the application of MPT as a model was that it was only an algebraic condition that happened to be mean–variance efficient in terms of the weights of the assets in a portfolio. To operationalize the MPT and make it into a testable prediction, Sharpe (1964), Lintner (1965), and Mossin (1966) presented the CAPM, which establishes a linear relation between the expected return of the assets and its market risk, measured by beta (β). While CAPM became a pioneering model, its assumption of market risk (market beta) as a single risk factor for the determination of asset returns was criticized widely for oversimplification.

Building upon this critique, Fama and French (1992) emphasized that if priced rationally, the nature of the risk is not unidimensional, and they introduced a three-factor model (FF3F) by incorporating size and value factors with the market factor. Although the introduction of the new factors improved the explanatory power of the model, a significant portion of the variation was left unexplained. Novy-Marx (2013) suggested that when a firm is profitable, it will give significantly higher returns than less profitable firms, even if the less profitable firms have higher valuation ratios. Aharoni et al. (2013) applied a firm-level test measure to prove the prediction of Fama and French (2006) and valuation theory (Miller & Modigliani, 1961). They found that the relationship between expected investments and expected returns was significantly negative, which was in line with the prediction of the valuation theory. These findings led Fama and French (2015) to extend their model to the five-factor model (FF5F) by introducing profitability and investment as factors and new risk dimensions, as these variations remained uninvestigated in the FF3F.

Jegadeesh and Titman (1993) suggested that the variance related to momentum is not captured by the CAPM. Carhart (1997) adopted it as a factor in his four-factor model, and the empirical results validated its utility in predicting returns. The skepticism around the validity and relevance of momentum as a factor has existed for a long time; however, consistent empirical evidence suggests that stocks demonstrating strong past performance manage to continue their momentum in the short term (Dirkx & Peter, 2020). Recognizing its role, Fama and French (2018) included momentum as the sixth factor in their six-factor model (FF6F). The FF6F was formulated to give a comprehensive framework of the factors that drive stock returns. The model has been tested across different markets, where its explanatory power has been both affirmed and challenged (Goo & Wang, 2024).

Table 1 summarizes the key traditional asset pricing models, features, and limitations to provide a precise and more explicit overview.

Table 1. Overview of the asset pricing models.

Model	Authors (year)	Key feature	Limitations
Modern portfolio theory (MPT)	Markowitz, 1952	Algebraic model to optimize asset weights for mean–variance efficiency; diversification of portfolios for minimization of risk	No consideration of real-world complexities and assumptions on rationality, efficiency, and data distribution
Capital asset pricing model (CAPM)	Sharpe, 1964; Lintner, 1965; Mossin, 1966	There is a linear relationship between the expected returns and beta.	Oversimplification of the risk factors and assumption of market efficiency.
Fama–French Three-Factor Model	Fama and French 1992	Size and value factors are introduced in addition to the market factor in the model.	Improved the explainability of returns but still left a significant portion of variation unexplained.
Carhart four-factor model	Carhart, 1997	Adds momentum factor to the Fama–French three-factor model	The effect of the momentum may be short-lived and varies across markets.
Fama–French five-factor model	Fama and French, 2015	Adds profitability and investment as factors to the Fama–French three-factor model.	When applied, the empirical results show that some factors are redundant or insignificant in specific markets.
Fama–French six-factor model	Fama and French, 2018	Momentum was included as the sixth factor in the model.	Overall, the results are mixed, specifically in emerging markets.

Researchers have employed different methodologies across different datasets and markets for the empirical validation of the earlier models of Fama and French (1992, 2015) and to assess their performance (Diallo et al., 2023; Miralles-Quirós et al., 2020). Despite extensive empirical validation, persistent issues in the applicability of these models have been faced by researchers, particularly in developing markets, where the behavior of the models may differ from those of developed markets (Jia, 2023; Zaghouni & Mezzez, 2020). Several studies have attempted to adapt these models by employing alternative methodologies, such as machine learning, yet they are struggling to improve the predictive power.

One significant overlooked dimension in traditional asset pricing studies is the downside risk. The downside beta quantifies stocks' sensitivity to adverse market movements and offers a more accurate measure of risk, especially for securities with high systematic risk (Galagedera, 2009), which is significantly more relevant in emerging markets, where downside risk plays a significant role in explaining the relationship between risk and return (Estrada & Serra, 2005).

Portfolios based on downside beta have been empirically found to have superior performance. Ang et al. (2002) found that portfolios based on high downside correlations perform better than those with low downside correlations. This finding was further supported by Kazmi et al. (2021), who showed that double sorted portfolios-based on the downside beta yield better risk-adjusted returns than traditional beta-based portfolios. The downside beta explains the returns more effectively than the conventional CAPM if the critical concern is downside risk (Liu, 2019). Double sorting-based portfolios based on traditional beta and downside beta were also employed by Ayub et al. (2020), further validating the downside beta's importance in asset pricing.

Given the growing recognition of asset pricing for the construction of downside risk-based portfolio, this study explored this alternative portfolio sorting method to improve returns' predictability.

Traditional left-hand side (LHS) portfolios have been constructed with multiple variations and choices of variables, with some studies demonstrating that portfolios based on downside risk have outperformed traditional portfolios in emerging markets (Estrada, 2007; Ayub et al., 2020; Ang et al., 2002) as, by nature, the emerging markets are more volatile than their developed counterparts (Fama & French, 2018).

While traditional models provide valuable insights, their linear framework hinders them from capturing the nonlinear, complex, and dynamic relationships between the risk factors and returns. This situation presents an opportunity for advanced methods, such as machine learning, to improve the performance of asset pricing models.

Machine learning techniques in asset pricing

The integration of ML in asset pricing has emerged as a promising direction for empirical studies in asset pricing. This fusion of economic theory and machine learning is a natural marriage, where asset pricing explains price formation through investors' beliefs, while ML offers statistical tools capable of capturing complex patterns (Giglio et al., 2022).

Aminimehr et al. (2022) experimented with various deep learning (DL) models, an advanced area of ML, among which LSTM provided promising results owing to its memory. The memory characteristics have improved the return predictions, owing to their ability to capture long-term dependencies. The LSTM has the capability of learning from more than 1000 timesteps due to the existence of a cell state, called the LSTM cell or memory cell, in the architecture of the network, allowing for better handling of sequential data, making it particularly suitable for financial time-series forecasting (Staudemeyer & Morris, 2019). The memory cell enables the network to deal with the issue of a vanishing gradient in backpropagation in long timestep sequences through the mechanism of gating (Bhandari et al., 2022).

A growing body of literature highlights the superior predictive capabilities of LSTM over traditional and ML models. Petrozziello et al. (2022) emphasized that LSTM performs better than conventional parametric models in forecasting volatility, specifically in highly volatile market conditions. Similarly, Zhang (2022) applied various DL models to measure risk premia and found that the LSTM outperformed all the other models by achieving the highest accuracy, reinforcing its effectiveness in risk assessment. The author emphasized the underlying theory while applying a DL model for improved calculation of the asset risk premium.

A key criticism of the machine learning methods in asset pricing is the lack of theoretical grounding, as they produce better predictions but lack a clear economic interpretation. Giglio et al. (2022) suggested that the connection of the asset pricing theory with empirical models needs to be re-established. This concern has sparked debates on reconnecting ML-based asset pricing with financial theory to ensure that data-driven models remain economically meaningful.

Contribution of the study

This study contributes to the asset pricing literature by integrating the theoretical foundations of the FF6F and the predictive capabilities of LSTM to enhance return prediction while preserving economic interpretability. By leveraging the potential of LSTM for capturing long-term dependencies, the study addresses the issues of nonlinearity and complex factor interactions that are beyond the scope of traditional models.

The study also applied simple models as surrogate models to explain the interaction of the variables, which provides deeper insights into the risk–return relationship in general and under various market conditions. This approach resulted in an optimized model and analytical framework suitable for the dynamics of an emerging market, helping investors achieve risk-compensated returns.

Additionally, this study incorporates the downside beta in the construction of portfolios, providing a more suitable method of measuring systematic risk specifically for emerging markets like China. Such markets are characterized by high volatility and different investor behaviors from their developed counterparts; thus, the downside beta, which better accounts for adverse market movements, improves risk assessment and portfolio decisions.

By hybridizing DL and traditional asset pricing models, this study bridges the gap between ML and the theory of finance, offering an optimized framework that improves the predictive accuracy, theoretical insights, and practical implications for asset pricing in emerging markets.

The remainder of this paper is organized as follows: Section 2 outlines the research methodology. Section 3 presents the empirical results and analysis, and Section 4 concludes the paper with a discussion of the findings and recommendations for future research.

2. Data and methodology

2.1. Model

The study used the FF6F model as a theoretical framework for all the dependent portfolios for analysis and predictions.

2.1.1. Fama and French's Six-Factor Model (FF6F)

The choice of employing the FF6F for the study is grounded in well-documented empirical evidence in explaining stock return variations. The model extends beyond a simpler model by incorporating a large set of factors, offering a comprehensive framework for capturing risk exposures. The model has been tested across various markets, demonstrating its explanatory power and robustness in different economic environments (Goo & Wang, 2024). Given the unique characteristics of emerging markets, a model that incorporates multiple risk dimensions in its structure is crucial for examining return variations.

Furthermore, the criticism of ML models in asset pricing, highlighted by Giglio et al. (2022), emphasizes that empirical models should be aligned with financial theory. Thus, in line with these concerns, this study integrates FF6F with a data-driven technique, maintaining economic interpretability while leveraging ML's predictive power, ensuring that the results are meaningful within the asset pricing framework. Thus, the FF6F model is an appropriate and theoretically grounded benchmark for analyzing portfolio returns in this study.

The Fama–French six-factor model (FF6F) has been extensively tested in the asset pricing literature owing to its strong empirical foundation in the explanation of variations in stock returns. The model expands beyond the traditional factor models by incorporating additional risk factors such as size (SMB), value (HML), profitability (RMW), investment (CMA), and momentum (UMD) in addition to the market factor.

The following is the mathematical representation of the FF6F model:

$$R_{p,t} - R_{f,t} = \alpha_i + \beta_1(R_{mkt} - R_f)_t + \beta_2(SMB)_t + \beta_3(HML)_t + \beta_4(RMW)_t + \beta_5(CMA)_t + \beta_6(UMD)_t + \varepsilon_{p,t},$$

where

- $R_{p,t} - R_f$ represents the excess return of the portfolio calculated by deducting the risk-free rate from the return of the portfolio;
- $(R_{mkt} - R_f)_t$ symbolizes the excess market return, calculated by taking the difference between market return and the risk-free rate of return;
- $(SMB)_t$ (small minus big) captures the size effect, calculated as the difference between the returns of the small-cap firms and the large-cap firms;
- $(HML)_t$ (high minus low) represents the value premium and is calculated by deducting the return of the firms with a low Book to Market (BM) ratio from firms with a high BM ratio;
- $(RMW)_t$ (robust minus weak) accounts for profitability differences and are calculated by taking the difference between the return of the firms with robust profitability and firms with weak profitability;
- $(CMA)_t$ (conservative minus aggressive) reflects investment premium and represents the difference between the return of conservatively investing firms and aggressively investing firms;
- $(UMD)_t$ (up minus down) represents the momentum effect, which means the difference between the return of the winning firm in the last period and the losing firms;
- α_i represents the intercept, also called the abnormal return;
- $\beta_1, \beta_2, \dots, \beta_6$ represent the factor loadings;
- $\varepsilon_{p,t}$ represents the error term.

Explanation of right-hand side variables (factors)

The independent variables for the FF6F are the factors constructed upon the methodology of Fama & French (2015, 2018), also called right-hand side (RHS) variables. The first among these factors is the market factor, calculated as follows:

$$Mkt_t = Rm_t - Rf_t,$$

where Rm_t is the overall market return, proxied by a broad stock market, and Rf_t is the risk-free rate, proxied by the government bond rate.

Factors other than the market are constructed by the bivariate sorting technique (2*3), where all the available firms are first sorted by size, dividing them into two groups on the basis of the median: small and big (S, B). The subsequent sorting is based on the characteristics of firms for the corresponding factors, such as the book-to-market ratio (BM ratio), profitability, investments, and momentum. For the second sorting, the firms are divided into three groups called the low (L), neutral (N), and high (H) firms, with 30%, 40%, and 30%, respectively, in both the S and B groups. Thus, the firms are divided into six portfolios in total to form the factors. Below are the formulas for constructing the factors. The size factor in FF6F, denoted as SMB (small minus big), is calculated by subtracting the average of all big portfolios from the average of all small portfolios. The formula for calculating SMB is as follows:

$$SMB = \left(\frac{SL + SN + SH}{3} \right) - \left(\frac{BL + BN + BH}{3} \right).$$

The value factor HML (high minus low) is calculated by subtracting the average of the lower portfolios of the small and big groups from the average of the higher portfolios of the small and big groups. The formula for HML is as follows:

$$HML = \frac{(SH + BH)}{2} - \frac{(SL + BL)}{2},$$

The methodology of the HML is followed by the rest of the factors as well, like profitability, denoted by RMW (robust minus weak); investment, denoted by CMA (conservative minus aggressive); and momentum, denoted by UMD (up minus down). In all these factors, the first sorting is based on the size, forming two groups (big and small groups) at the median of the market, and the second sorting is made at 30%, 40%, and 30%, based on the respective characteristics of factors like profitability for RMW, investment for CMA, and momentum for UMD. Profitability is proxied by return on expenditure (ROE), investment is proxied by growth in assets, and momentum is proxied by the average return of the stock from $t - 12$ to $t - 2$. The formulas for RMW, CMA, and UMD are as follows.

The profitability factor, RMW (robust minus weak), is calculated as the difference between the returns of the firms with robust profitability and firms with weak profitability in both the small and big groups. The firms in the second category are divided into robust, neutral, and weak firms according to profitability from high to low, where the profitability is proxied by the ROE of the firms.

$$RMW = \frac{(SR + BR)}{2} - \frac{(SW + BW)}{2}.$$

The investment factor, CMA, is calculated by subtracting the returns of the aggressively investing firms from the conservatively investing firms in both the small and big groups. The formula for CMA is as follows:

$$CMA = \frac{(SC + BC)}{2} - \frac{(SA + BA)}{2}.$$

The momentum factor, UMD, is calculated as the difference between the return of the high-earning firms and the low-earning firms in the last 10 months

$$UMD = \frac{(SU + BU)}{2} + \frac{(SD + BD)}{2}.$$

Explanation of dependent portfolios (LHS variables)

The LHS variables are the excess return of the dependent portfolios of the model, which is constructed on bivariate dependent sorting, first by the beta and then by the downside beta. Below, we show the method of calculating the excess portfolio return of each portfolio, where $R_{p,t}$ is the return of the portfolio at time t , $R_{f,t}$ is the risk-free rate at time t , and $R_{p,t}^e$ is the excess return of the portfolio at time t .

$$R_{p,t}^e = R_{p,t} - R_{f,t}$$

The first sorting variable, beta, is a portfolio's sensitivity to the movement in the overall market, indicating the change in the return of a portfolio relative to the market. The following formula is applied to calculate the beta, where R_p is the return of the portfolio, R_m is the return of the market, $Cov(R_p, R_m)$ is the covariance between the portfolio returns and market return, and $Var(R_m)$ is the variance in the return of the market:

$$\beta = \frac{Cov(R_p, R_m)}{Var(R_m)}.$$

The variable for the second category is the downside beta (β_{down}), which is the measure of the sensitivity of the return of the portfolio to the negative market

$$\beta_{down} = \frac{Cov(R_p, R_m | R_m < rf)}{Var(R_m | R_m < rf)},$$

where R_p is the return of the portfolio, R_m is the return of the market, $Cov(R_p, R_m | R_m < rf)$ is the covariance between the portfolio returns and market returns when the market returns is negative, and $Var(R_m | R_m < rf)$ is the variance of the market return when the market return is negative.

For constructing the portfolios, the firms are first sorted on the basis of their beta (x_1), where three groups of low beta (LB), mid-beta (MB), and high beta (HB) firms were made using three quantiles [0, 0.3, 0.7, 1.0]. The second sorting was based on the downside beta (x_2) from low to high, where, within each beta group, six portfolios of firms have been formed using six quantiles [0, 0.2, 0.35, 0.5, 0.65, 0.8, 1.0]. For sorting within each group, the following formula has been used:

$$B2_{jkt} = Pctl_{p2k}(\{X2t | B1_{j-1,t} \leq X1_t \leq B1_{j,t}\}),$$

where $B2_{jkt}$ represents the k -th breakpoint for the downside beta within the j -th beta group. The values $(\{X2t | B1_{j-1,t} \leq X1_t \leq B1_{j,t}\})$ represent the values of the downside beta for all the firms in the j -th beta group. The $p2k$ percentiles are the set of downside beta quantiles, creating 18 portfolios assigned to six groups of portfolios on the basis of the downside beta within each beta group. The firms are assigned to each portfolio using the following formula:

$$P_{jkt} = \{i | B1_{j-1,t} \leq X1_t \leq B1_{j,t}\} \cap \{i | B2_{j,k-1,t} \leq X2_{i,t} \leq B2_{jkt}\}.$$

The formula indicates that the assignment of a firm to a portfolio must satisfy specific breakpoint conditions for both variables, x_1 and x_2 (beta and downside beta), resulting in $nP1 * nP2$ portfolios. For the analysis, this study initially constructed three groups of portfolios, low beta, mid-beta, and high beta (LB, MB, HB), numbered as 1,2,3 according to the beta. Within each group, six portfolios (1, 2, 3, 4, 5, 6) have been constructed on the basis of the downside beta. The first three portfolios of each group (1,1, 1,2, 1,3; 2,1, 2,2, 2,3; 3,1, 3,2, 3,3) have been termed the low-downside beta portfolios within that group, and the rest of the portfolios (1,4, 1,5, 1,6; 2,4, 2,5, 2,6; 3,4, 3,5, 3,6) in that group are termed the high downside beta portfolios.

Thus, the study formulated six groups of portfolios in each broader group of portfolios, resulting in 18 portfolios. The six sub-portfolios are termed LBLD, LBHD, MBLD, MBHD, HBLD, and HBHD.

Apart from the mean and standard deviation, the performance of the portfolios was determined using the Sharpe ratio (Sharpe, 1966) and Sortino ratio (Sortino & Van Der Meer, 1991). The Sharpe ratio measures a portfolio's risk-adjusted returns, where it is observed how much excess return is being generated per unit of total risk. The Sortino ratio measures how much return is generated per unit of downside risk.

2.1.2. Long Short-Term Memory (LSTM) framework for the FF6F

The following is the framework used to predict the portfolios' excess return $R_{p,t}^e$, based on the FF6F factors over time by applying LSTM. The key components that an LSTM cell processes at Timestep t are the input data x_t , the hidden state h_{t-1} , and the cell state C_{t-1} from the previous step. The forget gate in the LSTM structure decides which part of the previous cell state, C_{t-1} , to forget.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f).$$

The input gate decides which new information to store in the cell state, and the mathematical representation is as follows:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i);$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C).$$

The cell state update combines the outputs of the forget gate and the input gate to update the cell state

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o);$$

$$h_t = o_t * \tanh(C_t),$$

where W and b represent the weight matrices and biases learned during training, σ is the sigmoid function, and \tanh is the hyperbolic tangent function. For the return of each portfolio $R_{p,t}^e$, the factors of the FF6F at Timestep t are used as the input x_t into the LSTM network. The LSTM then predicts $R_{p,t+1}^e$ for the next timestep on the basis of the previous data. The mathematical equation for the LSTM is as follows:

$$R_{p,t+1}^e = \hat{y}_{t+1} = LSTM([x_t, x_{t-1}, x_{t-2}, \dots] + \epsilon_t),$$

where x_t represents the factors of the FF6F to be used as input features at time t , and $LSTM(\cdot)$ represents the series of LSTM operations used to predict the portfolio's excess return. The LSTM uses the previous timesteps for learning how to improve the accuracy of the prediction \hat{y}_{t+1} , and dynamically updates its hidden states by learning from the patterns and dependencies in the time series of the data by using hidden layers and recurrent connections, unlike the linear models, where the relationship remains static.

2.2. Data

The data required for the study have been retrieved from the Thomson Reuters Datastream database (TRDS) for the period of 1997 to 2022 for all nonfinancial A-share listed firms on the Shenzhen Stock Exchange (SZSE), China. The study's focus is solely on SZSE, as this market primarily hosts a large number of smaller, innovative, and growth-oriented firms which inherently carry higher volatility and information asymmetry in comparison with large-cap state-owned enterprises exhibiting different characteristics, primarily hosted in large numbers by the Shanghai Stock Exchange (SSE) (Hu et al., 2018; Carpenter & Whitelaw, 2017; Sun & Manger, 2012). The SZSE and SSE differ in their price behaviors, formation, and short-term fluctuation patterns. Despite following similar cyclical patterns, price fluctuations follow distinct behaviors (Feng et al., 2007; Liu & Cui, 2002). Individual and retail investors form a substantial portion of the SZSE. These investors often exhibit behavioral biases, which can negatively predict future returns (Tan et al., 2024), and their

preferences lead to higher volatility (Jinghan, 2010). In contrast, large investors demonstrate better information processing capabilities, contributing positively to the market's predictability (Tan et al., 2024). Since these two markets have different risk characteristics, as evidenced by the literature, combining both would dilute the unique risk profile and characteristics of the SZSE.

The data include the return of the index, the closing price of the share, number of shares outstanding, the BM ratio, return on equity, total assets, and change in total assets. Market-specific data are collected on a monthly basis. Due to the unavailability of the Chinese T-bill data in the TRDS before 2002 to be used as a risk-free rate of return, the T-bond data for China has been retrieved from the database of the Federal Reserve Bank of St. Louis, USA, where the data of the one-year T-bond returns were available, adjusted for the monthly rate. In order to avoid survivorship bias, all active and nonactive firms have been included in the data, following Agarwal and Pradhan (2018). The return of the Shenzhen Composite Index has been used as the proxy for the market return.

2.2.1. Data cleaning, splitting, and hyperparameters for the model

After the data have been cleaned for errors and missing values, the RHS and LHS variables can be constructed for the FF6F. An important decision in deploying DL/ML models is the decision about splitting the data into training, validation, and testing groups. For this study, the data were divided into three parts, namely the training, validation, and test sets, which were split 70–15–15, respectively. The data were standardized using a standard scaler, which transforms the data so that they have a mean of 0 and a standard deviation of 1. The standardization of data helps in learning the DL models, where the significant variations impact the learning process, thus making convergence faster and more effective. The standardization ensures that the RHS and LHS variables are transformed to the same scale.

The training set (70%) was used to train the model, while the validation set was used to tune the model's hyperparameters. The technique used for fine-tuning the hyperparameters was K-fold validation, in which the training set was further divided into subsets called folds. In our case, three folds were used. For this, the model was trained and tested on different subsets to ensure the model was exposed to different data in the process. The process was repeated for each fold to ensure by fine-tuning whether the model was effective, as this cross-validation ensures robustness through validation on different sets. This process ensured the model's optimization for unseen data and prevented overfitting. For the final evaluation of the models, 15% of the test set data were reserved to measure their performance and generalizability.

In the cross-validation process, a range of hyperparameters was used to train the model through a grid search cross-validation (CV). Grid searching is a technique in which the algorithm automates the process of searching exhaustively through all possible combinations of hyperparameters. The hyperparameter space of this study has five hyperparameters: the number of hidden layers and the LSTM units in it, the dropout rate, the learning rate, the batch size, and the number of epochs. The combination of the number of hidden layers and the LSTM units in it are (50,50), (100,50), (100,100), (150,100), and (150,150), where, e.g., 50 means two hidden layers which have 50 LSTM units in the first and 50 LSTM units in the second hidden layer. The dropout rate refers to the regularization techniques adopted to prevent overfitting. In this process, some of the input units are randomly set to zero during training. Different combinations are tested to strike a balance between reducing overfitting and maintaining the model's learning ability. The study applied two rates, 0.01 and 0.1.

LSTM Model Architecture for Portfolios' Return Prediction

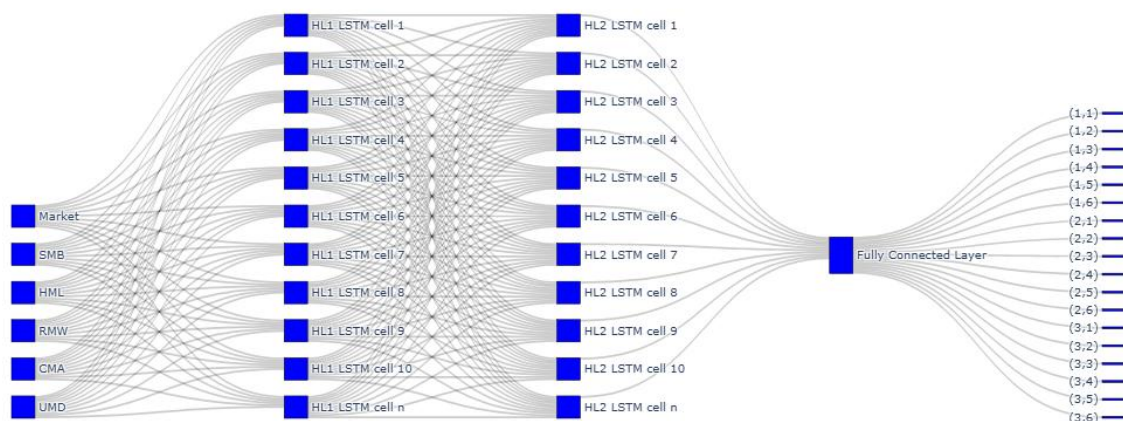


Figure 1. Architecture of the LSTM model for portfolio returns.

The next hyperparameter in the hyperparameter space is the learning rate, which refers to the step size at which the model's weights are updated after each iteration during training. The study utilizes two rates, 0.001 and 0.01, to ensure efficient convergence without undermining the optimal solution. The batch size refers to the number of samples processed by the model before updating the weights; in this study, this was set to 32. The next hyperparameter was the number of epochs, which refers to the number of times a model passes through the training dataset. The study has set these to be 25, 50, and 100.

2.3. Model performance measurement criteria

This study utilizes a total of three performance metrics for the evaluation, comparison, and selection of the best model for each portfolio. The metrics include three error metrics, namely mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and one explanatory metric, R^2 . MAE measures the average of the absolute difference between the actual and predicted values (residuals); in other words, it gives equal weight to all errors by overlooking the size. The MSE measures the average squared differences between the predicted and actual values, where, by squaring, more weight is assigned to the more significant errors and makes the MSE sensitive to outliers. The RMSE measures the root of the average squared differences between the predicted and actual values. Through the process of squaring, more weight is assigned to the more significant errors, making the RMSE sensitive to outliers and penalizing large errors more, which makes it a valuable metric for substantial deviations. The R^2 , called the coefficient of determination, is a statistical measure of the proportion of variance in the dependent variable (portfolio returns) that is predicted by the independent variables (factors).

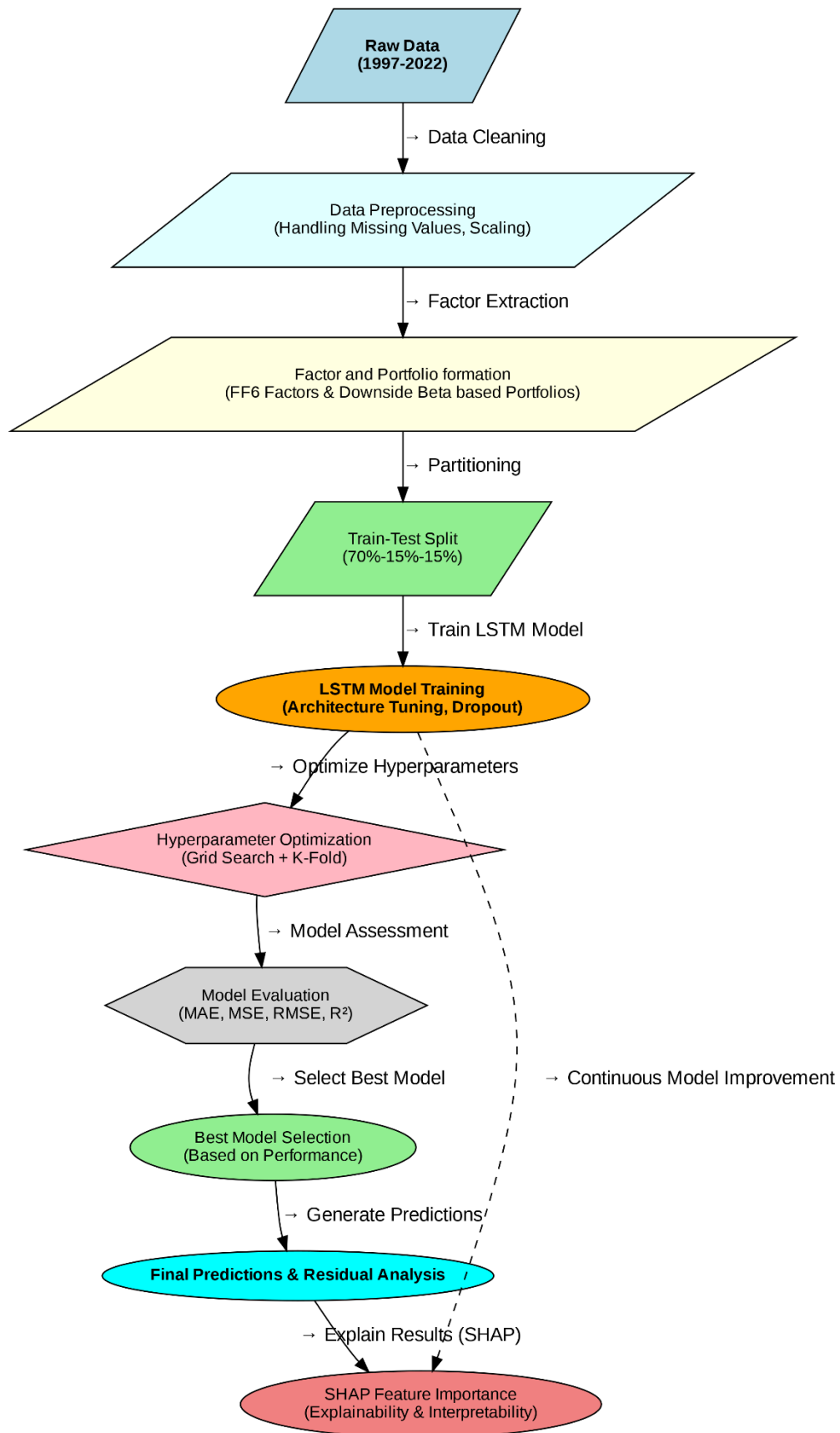


Figure 2. Workflow model for LSTM and the FF6F.

Apart from the primary model, LSTM, other simple models, such as random forest and the support vector machines (SVM) have been deployed via the SHAP (Shapley additive explanations) technique as surrogate models to show the impact of each individual feature on the output of the model for the explanation and interpretation of the main model and to check the nonlinear and conditional relationships of the factors to the returns of the portfolios.

Figure 2 represents the workflow model for the study, which includes steps like raw data collection, data preprocessing, formation of the FF6F factors and portfolios, data splitting, model training, hyperparameter optimization, model evaluation, selection of the best model, final prediction, and, finally, determining the features' importance through SHAP for explainability and interpretability.

3. Analysis and results

3.1. Analysis of fama and French's factors and dependent portfolios

Descriptive statistics of the FF6F factors

Table 2 presents the descriptive statistics for the monthly FF6F factors spanning over 25 years from 1997 to 2022 for the Shenzhen Stock Exchange, China.

Table 2. Descriptive statistics of the FF6F factors.

	Market	SMB	HML	RMW	CMA	UMD
Count	307	307	307	307	307	307
Mean	0.010	0.004	-0.028	0.016	0.003	0.006
Std. Dev.	0.098	0.044	0.050	0.062	0.038	0.115
Min	-0.284	-0.132	-0.426	-0.119	-0.065	-0.271
25%	-0.057	-0.019	-0.048	-0.010	-0.013	-0.012
50%	0.007	0.003	-0.026	0.011	-0.002	0.000
75%	0.062	0.023	-0.001	0.035	0.013	0.013
Max	0.411	0.262	0.080	0.723	0.462	1.947

The average value of the monthly mean return of the market factor is 1.01%. The figure shows that, on average, the market has rewarded investors for holding risky assets over and above the risk-free rate at the level mentioned above. The results are in line with the prediction of the traditional asset pricing theory, which advocates compensation for the risk an investor bears (Mahpudin, 2020). The SMB mean return is 0.42%, indicating that during the study's sample period, the small-cap stocks outperformed the large-cap stocks, which is in line with the literature (Hu et al., 2019). The mean value of HML (value premium) is -2.78%, indicating that, on average, the value stocks (high BM stocks) have been outperformed by growth stocks (low BM stocks) over the period of the study, which is against the expectation; however, similar exceptions has been reported where the HML has been found to be insignificant (Joshi, 2024), significantly negative (Buditomo et al., 2024), and redundant (Fama & French, 2015) for the excess returns.

The profitability factor (RMW) has a positive mean return of 1.59% over the sample period, which is aligned with the existing literature (Singh et al., 2023). The mean return of the CMA is 0.26%, which indicates that the stock of the firms with conservative investment policies has slightly outperformed the stock of the firms with aggressive investment policies, which is consistent with earlier research (Fama & French, 2015; Zheng et al., 2022). The mean value of the momentum factor

(UMD) is 0.646%, which indicates that the stocks with high momentum have performed slightly better than the stocks which have low momentum, as evident in previous works (Cederburg & O'Doherty, 2015; Daniel & Moskowitz, 2014).

Volatility analysis of the FF6F factors

Figure 3 represents the 12-month rolling volatility of the Fama and French factors over the sample period, where each line denotes one of the six factors. The following paragraph is the interpretation of the key facts in the graph.

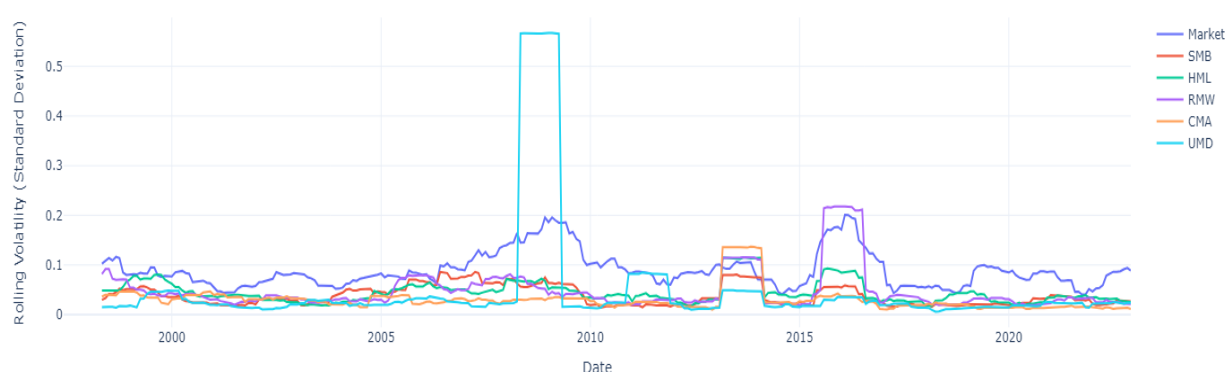


Figure 3. Volatility of the independent variables/factors.

The market factor (blue) remained relatively on the lower side but with periods of noticeable spikes around 2008 (the global financial crises) and 2016 (the Chinese stock market crises) and a minor spike around 2019–2020 (COVID-19). The SMB (red), HML (green), and RMW (purple) factors have shown stable and relatively low volatility throughout the study, apart from 2015–2016, where the SMB and RMW exhibited some spikes. The spike was during the stock market crises in China, where it is evident that the performance of the stock would have been affected by the market performance. The CMA (yellow) remained primarily stable, except for some time around 2014. The UMD factor experienced extreme spikes in volatility among all the factors. Its peak volatility was around 2008 during the global financial crisis, when it was around 0.5, which was the highest. Other spikes in UMD occurred around 2014 and 2016.

The performance of dependent portfolios

Table 3 provides overall insights into the risk and return performance of 18 portfolios and 6 groups of portfolios based on the beta and downside beta.

Table 2 reports the returns, risk, and risk-adjusted return performances of the portfolios. Portfolio 1,1, with modest positive returns and a decent Sharpe ratio and Sortino ratio, stands out in the group, suggesting good risk-adjusted returns. However, Portfolios 1,2 and 1,3 report negative returns and risk-adjusted performance, indicating underperformance for the level of risk exposure. As a group, LBLD shows a negative Sharpe ratio, and the Sortino ratio suggests overall underperformance. Portfolios 1,4, 1,5, and 1,6 as a group (LBHD) underperformed overall on all parameters. Portfolios 2,1 to 2,6 (MBLD, MBHD) reported mixed results, where Portfolio 2,1 performed very well with the best risk-adjusted

returns. The portfolio in the MBLD group reports a decent performance, while the MBHD portfolio remained short of the mark. The highest performance is reported in the high beta portfolios, specifically Portfolio 3,1, with impressive risk and downside risk-adjusted returns. Overall, a pattern of the better performance of the portfolios with lower downsides has been observed than their high downside beta counterparts in the same group. The portfolios in the higher beta group reported better returns than the portfolios in the lower beta, emphasizing the importance of the risk–return tradeoff.

Table 3. The performance of the portfolios and as a group.

Portfolio	Mean return	Std. dev.	Sharpe ratio	Sortino ratio
1,1	0.016	0.043	0.379	1.095
1,2	−0.008	0.039	−0.207	−0.242
1,3	−0.019	0.058	−0.328	−0.346
LBLD	−0.004	0.047	−0.052	0.169
1,4	−0.023	0.075	−0.300	−0.322
1,5	−0.028	0.098	−0.292	−0.313
1,6	−0.037	0.106	−0.345	−0.352
LBHD	−0.029	0.093	−0.312	−0.329
2,1	0.039	0.079	0.501	1.353
2,2	0.013	0.074	0.181	0.312
2,3	0.005	0.080	0.056	0.084
MBLD	0.019	0.078	0.246	0.583
2,4	0.000	0.090	−0.003	−0.004
2,5	−0.007	0.096	−0.078	−0.100
2,6	−0.019	0.103	−0.188	−0.220
MBHD	−0.009	0.096	−0.090	−0.108
3,1	0.107	0.297	0.362	2.029
3,2	0.081	0.233	0.347	1.482
3,3	0.047	0.149	0.319	0.808
HBLD	0.079	0.226	0.342	1.440
3,4	0.030	0.134	0.222	0.446
3,5	0.016	0.134	0.120	0.204
3,6	0.005	0.147	0.033	0.050
HBHD	0.017	0.139	0.125	0.234

The results also show that if downside risk is managed (e.g., in HBLD portfolios), the high-risk portfolios effectively offer high returns. The Sortino ratios of the LBLD, MBLD, and HBLD portfolios are better than their high downside beta counterparts, thus making them appealing in terms of the downside risk-adjusted parameter, indicating that those portfolios with a low downside beta can manage risk effectively and produce decent performances in risk-adjusted terms.

The Time-Series Performance of Portfolios' Excess Returns

Figure 4 represents the time-series return performance of the portfolios. When looking at the performance of all the portfolios, it can be inferred that in the period before 2005, the portfolios exhibited lower volatility and closely aligned returns. The spikes are there but, broadly, the excess return of all the portfolios is near zero. In the period between 2005 and 2010, a rise in volatility was observed, peaking sharply around 2006–2007. These spikes coincide with the global financial crisis, which impacted stock markets worldwide. These extreme spikes and peaks suggest high market instability in this period, where all groups of portfolios exhibited extreme swings in their returns.

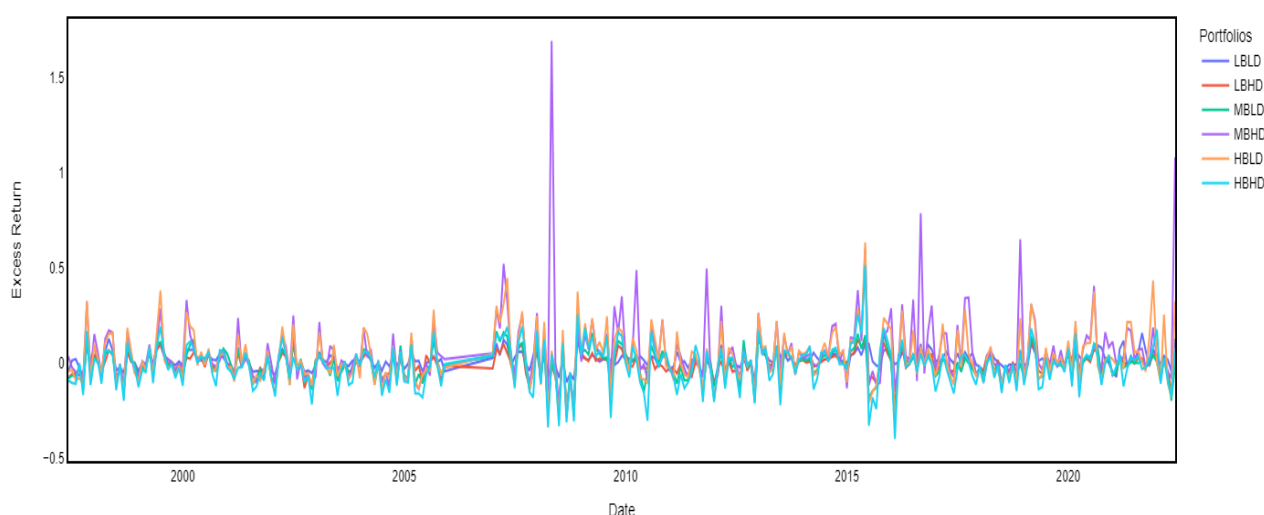


Figure 4. Time-series performance of excess portfolio returns.

In the post-2008 crisis period, an apparent reduction in volatility was observed, with few noticeable spikes, like in the period around 2015–2016, when there was turmoil in China’s stock markets, resulting in a market crash. Though fluctuation was there, the intensity was low compared with 2008. In the period after 2020, the portfolios showed some spikes, possibly due to the impacts of COVID-19, which has impacted the stock market around the world.

The LBLD portfolios delivered more stable but lower excess returns across the study period, corresponding to their low-risk profile. LBHD portfolios also exhibited a low return performance but notable spikes in periods of volatility, suggesting increased risk in a turbulent market due to their higher downside beta profile. The mid-beta portfolios (MBLD and MBHD) demonstrated higher excess returns and medium risk during volatile periods, with the MBHD portfolios exhibiting slightly higher spikes due to their higher downside risk profile.

High beta (HBLD, HBHD) portfolios exhibited the highest volatility, specifically in turbulent markets, demonstrating their high sensitivity to the market and suggesting poor performance. Around the global financial crisis (2007) and after 2020, the portfolios performed in tandem, indicating the dominance of market-wide factors. The sharp upward spikes present opportunities for higher risk-adjusted returns in HBHD portfolios.

Overall, the figure is a comprehensive representation of the behaviors of the portfolios across different market conditions, where the high beta portfolios (HBLD, HBHD) potentially offer higher returns but, at the same time, have significantly more risk. The performance of the high downside beta portfolios during 2007–2008 and 2015–2016 suggests that the vulnerability of these portfolios to the market conditions is high, especially in a downturn.

The correlation heatmap of the variables

Figure 5 is a correlation heatmap visualizing the strength and direction of the dependent (portfolios) and independent (factors) variables. The notable correlations between the factors (RHS) are moderately positive (0.51) between the market and SMD factors, indicating similar movements in the performance of the market and size factors. RMW and HML have a negative correlation (-0.46),

indicating that profitability and value move in opposite directions. The UMD exhibited a very weak correlation in both directions with all the other factors. Overall, the market factor has a strong positive correlation with all the portfolios, making it the dominant factor affecting most portfolios.

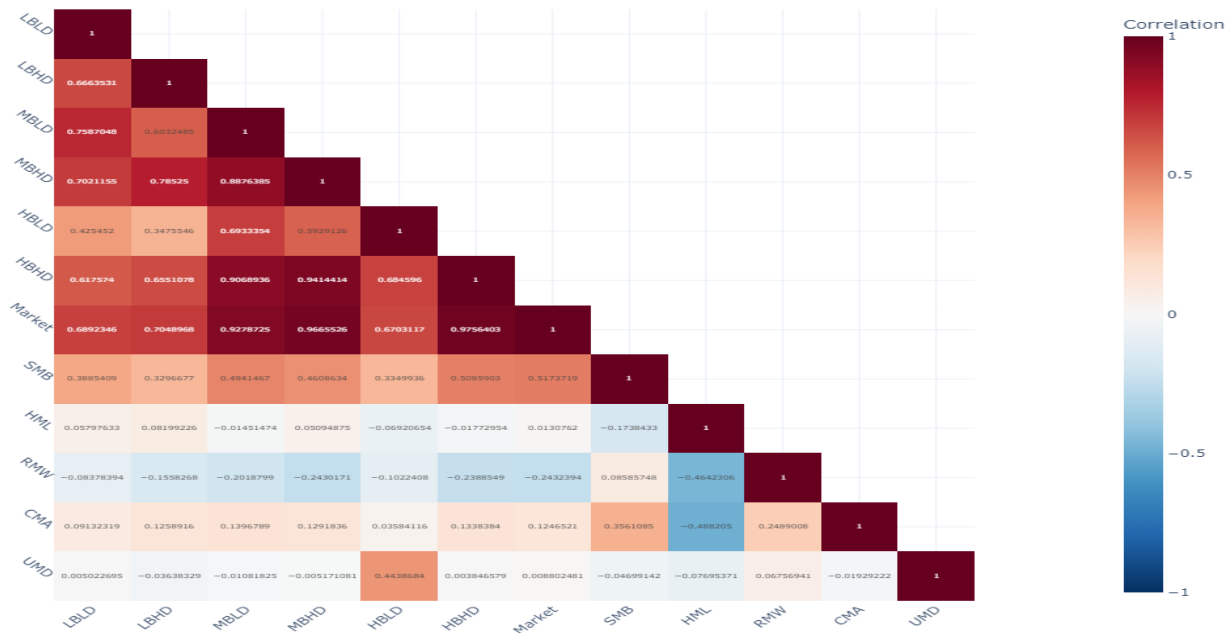


Figure 5. Correlation heatmap.

Factors like HML, SMB, and RMW exhibited weaker correlations at varying levels, suggesting they play more nuanced roles in portfolio returns. The strong interdependencies in portfolios with the same beta profile indicate that the shared characteristics of the portfolios make them move together, which is a crucial consideration for the benefits of diversification.

3.2. Factors' importance and threshold based nonlinear relationships

Factors' importance and contribution to the model's predictions

The SHAP summary plot provides insights into the individual importance and contribution of the FF6F factors (features) to predict portfolio returns, as shown in Figure 6. The SVM model has been used as a surrogate model for the more complex model (LSTM), which is in line with the methodology used by Hall and Gill (2019).

The model shows that the market factor has the most significant influence on the prediction of the model, having the most expansive spread in the SHAP values, indicating that the impact on the prediction of the model can exhibit substantial variations with both positive and negative contributions. The SMB has a somewhat moderate spread in the SHAP values, with positive and negative impacts with higher and lower values, but this is less pronounced than that of the market factor. The spread of the HML is even smaller than the SMB, where lower and higher values contribute both negatively and positively to the model's prediction.

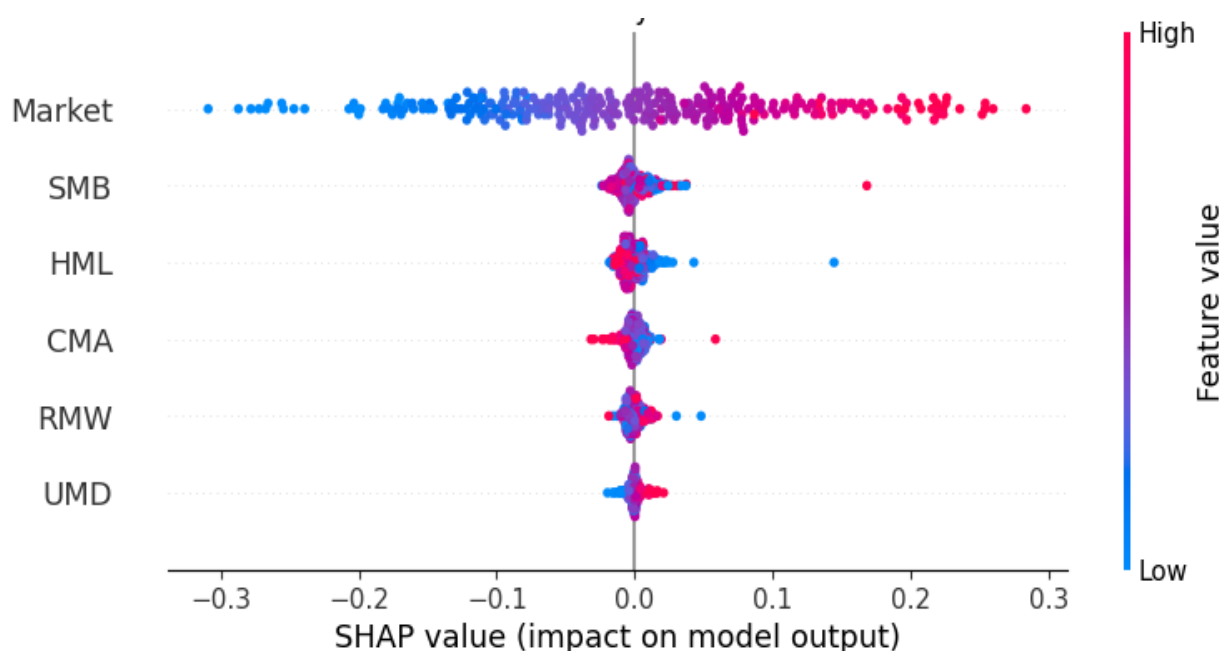


Figure 6. SHAP summary plot.

The CMA, with a narrower spread overall, exhibits low influence and indicates that the low values increase and the higher values decrease the prediction of the model. The relationship of the RMW has a narrower spread on the negative side, which behaves the same as that of CMA. The UMD has SHAP values that are very symmetrical, where the lower values decrease and the higher values increase the prediction of the model. Overall, higher values of the market and SMB, in general, lead to enhanced model prediction, while lower values decrease it. The rest of the factors have more negligible and relatively neutral effects on the model's prediction. The market factor seems to be the most influential on the model's predictions and it has significant effects.

Nonlinearity and conditional relationships of the LHS and RHS

Following the methodology of Molnar (2019), the study employed a random forest model as a surrogate model for interpreting the conditional factor–portfolio relationships presented in Figure 7. The results show that the market is a dominant factor as an essential criterion. In contrast, the other factors indicate conditional interactions, where nonlinear decision boundaries exist at various input thresholds. The results show that when the market is downturned ($\text{market} \leq -1.07$), the UMD and RMW factors significantly influence the predictions. However, in favorable market conditions ($\text{market} > -0.020$), the SMB and HML factors dominate, indicating nonlinear interactions among the factors.

Across different market conditions, the market factors exhibit asymmetric sensitivity. In unfavorable market conditions ($\text{market} \leq -0.20$), increased sensitivity has been shown to CMA and UMD, amplifying losses. Specifically, low CMA (≤ -0.39) intensified downturn losses, while high CMA predicts increased positive returns in bullish markets ($\text{market} > 1.09$), suggesting conservative investment strategies during good market conditions.

The nonlinear responses intensified in extreme market conditions, where small shifts in CMA (≤ -0.39 vs > -0.39) in a downturned market ($\text{market} \leq -2.42$) affect losses significantly, and in the

positive market (market > 1.95), they contribute disproportionately to portfolio returns, reflecting disproportionate risk–reward relationship.

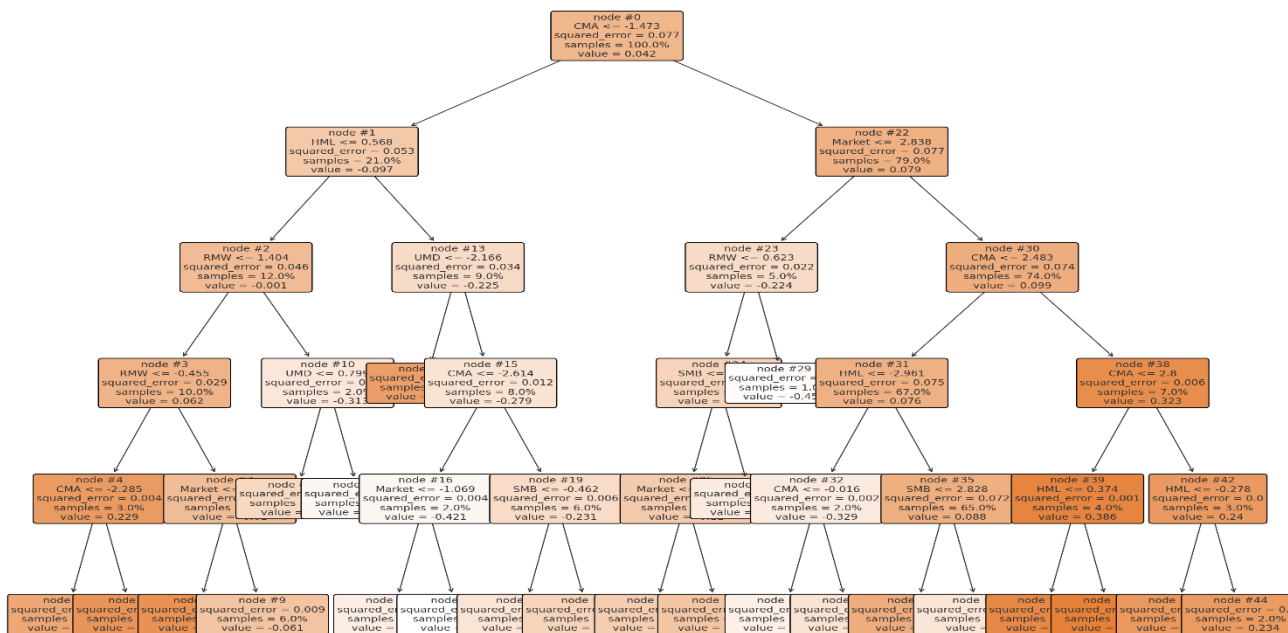


Figure 7. Surrogate model's conditional relationship tree.

The conditional tree reveals threshold-dependent, nonlinear interactions that are impossible for linear models to capture. RMW's importance emerges distinctly during severe downturns combined with negative momentum, indicating that the dynamic factors' weighting depends on the market conditions. This analysis aligns with Goo and Wang (2024), emphasizing the need for advanced nonlinear models to capture complex markets' dynamics accurately. The details of all the nodes in Figure 7 are given in Appendix II.

3.3. Performance of the model on portfolios

Table 4 provides a detailed analysis of the performance of the models for portfolios individually and then as a group for training, validation, and testing. Column 1 contains the portfolios' assignment; Column 2 has the best hyperparameters for the model; Columns 3 through 69 are the MAE, MSE, RMSE, and R^2 for the training set; Columns 7 to 10 are for the validation set of data; and Columns 11 to 14 are for test set of data. However, the analysis focuses on the test set, which is for the sample performance of the models for different portfolios.

The LBLD model group is for portfolios based on a low beta and a low downside beta. The models for the LBLD portfolio group have average test values of $R^2 = 0.4150$, MAE = 0.023, MSE = 0.0015, and RMSE = 0.0369. The error values for the LBLD group are acceptably low and signify the model's good predictive capability. However, the R^2 value is low, which suggests that the models fail to capture most of the variability in returns in the LBLD portfolios based on the input features (factors), making the models' quality average.

The performance of the LBLD portfolios is mainly driven by Model 1,2, with the lowest errors and the highest R^2 values. The second model group is for LBHD portfolios, based on a low beta and high downside beta, consisting of Portfolios 1,4, 1,5, and 1,6. The test MAE of the LBHD models is slightly higher than those of the training and validation phases, indicating a decline in prediction accuracy. The average test MSE (0.0041) and RMSE (0.0631) are similar to those for validation but lower than those for training, showing better than expected prediction ability. The average test R^2 (0.3626) is substantially lower than in previous phases and indicates the poor generalizability of models to unseen data. The LBHD models have shown moderate fitness in training and validation but struggle to maintain the same in the testing phase. The deteriorated R^2 in the test phase indicates problems with the models' robustness. These models' average errors (MAE, MSE, RMSE) show relative consistency, but the substantial drop in the test R^2 limits its ability to predict accurately. The group's performance is, however, driven by Model 1.5, which is the best among the group.

The MBLD group consists of Portfolios 2,1, 2,2, and 2,3 based on a medium and low downside beta. The average test MAE of the MBLD portfolios is 0.026, indicating good accuracy. The average test MSE (0.001) and RMSE (0.033) are slightly higher than those of the validation set, but there is no significant difference, which shows consistency in performance across all phases. The test R^2 of the MBLD portfolios is 0.811, which is higher than that of the training set and lower than that of the validation set, indicating the models' ability to explain the variance in the portfolios to a greater extent. Overall, the models have demonstrated strong performances for all parameters of all datasets, exhibiting the acceptable fitness of the models, capturing patterns, and explaining most of the variance in the data. The minimal errors across all phases indicated that the models are robust and have a strong ability to predict, which makes these models reliable and well-suited to the MBLD portfolios. In this group, the highest performer is 2,3 with the highest test R^2 and the lowest error values, followed by Model 2,2.

The models for MBHD portfolios demonstrated strong, robust, and consistent performance across all datasets, where the errors (MAE, MSE, RMSE) remained the lowest and were stable across all phases. The R^2 (0.929) remained higher, which suggests that the model can capture most of the variance in the return of the portfolios with optimal fitting. A slight increase in the MAE at the testing stage indicates room for improvement in capturing outliers or more complex variations in the data. These performances make this group the best-performing group among the others. The highest contributor to the performance in the MBHD group is Model 2,4, with the lowest errors (0.0119, 0.0002, 0.0147) and the highest R^2 (0.971), followed by Model 2,5. Model 2,4 is not only the highest performer in the group but also among all the models, with the lowest errors and highest fitness.

The HBLD group consisted of Portfolios 3,1, 3,2, and 3,3 based on a high beta and a low downside beta. The overall test performance regarding prediction errors and R^2 remained mediocre for all the models in the group. Decent performance was only displayed by Model 3,3 with a test MAE of 0.0375, an MSE of 0.0026, a RMSE of 0.051, and an R^2 of 0.858, which are significantly higher than those of the training and validation sets, indicating strong generalization and confirming the notion that the models' performance is excellent on the unseen data and predicts with minimal errors.

The HBHD consists of Portfolios 3,4, 3,5, and 3,6. The models' performances are excellent, with the average test MAE of the HBHD portfolios being 0.0246, which is almost the same as the validation MAE, reflecting consistent performance. The MSE (0.0013) and RMSE (0.0351) are very low and decent. However, the RMSE is almost identical to the training dataset, indicating minimal variation among the datasets and suggesting stable performance.

Table 4. Hyperparameters of the models and performances in the training, validation, and testing phases.

Models for the portfolios	Best parameters	Train MAE	Train MSE	Train RMSE	Train R2	Val. MAE	Val. MSE	Val. RMSE	Val. R2	Test MAE	Test MSE	Test RMSE	Test R2
1,1	((50, 50), 0.1, 0.01, 32, 100)	0.024	0.001	0.038	0.163	0.023	0.001	0.032	0.138	0.032	0.003	0.055	0.072
1,2	((100, 50), 0.01, 0.01, 32, 50)	0.018	0.001	0.025	0.555	0.019	0.001	0.029	0.557	0.018	0.001	0.023	0.630
1,3	((150, 100), 0.01, 0.01, 32, 100)	0.027	0.001	0.038	0.544	0.025	0.001	0.032	0.787	0.025	0.001	0.033	0.543
Average of the LBLD portfolios		0.023	0.001	0.034	0.421	0.022	0.001	0.031	0.494	0.025	0.002	0.037	0.415
1,4	((100, 50), 0.01, 0.01, 32, 25)	0.032	0.002	0.045	0.617	0.030	0.002	0.041	0.769	0.040	0.003	0.057	0.404
1,5	((150, 100), 0.1, 0.01, 32, 25)	0.052	0.006	0.078	0.381	0.041	0.003	0.054	0.702	0.043	0.003	0.054	0.508
1,6	((100, 50), 0.01, 0.01, 32, 25)	0.053	0.007	0.081	0.320	0.055	0.010	0.097	0.415	0.057	0.006	0.078	0.176
Average of the LBHD portfolios		0.046	0.005	0.068	0.439	0.042	0.005	0.064	0.629	0.047	0.004	0.063	0.363
2,1	((100, 100), 0.1, 0.01, 32, 25)	0.034	0.002	0.047	0.623	0.029	0.002	0.041	0.684	0.045	0.003	0.056	0.584
2,2	((150, 150), 0.01, 0.001, 32, 25)	0.020	0.001	0.027	0.855	0.017	0.001	0.028	0.880	0.019	0.001	0.023	0.909
2,3	((100, 50), 0.1, 0.01, 32, 100)	0.015	0.000	0.020	0.929	0.014	0.000	0.022	0.942	0.014	0.000	0.020	0.940
Average of the MBLD portfolios		0.023	0.001	0.031	0.802	0.020	0.001	0.030	0.835	0.026	0.001	0.033	0.811
2,4	((150, 150), 0.01, 0.001, 32, 50)	0.013	0.000	0.018	0.959	0.012	0.000	0.019	0.962	0.012	0.000	0.015	0.971
2,5	((150, 100), 0.1, 0.01, 32, 100)	0.016	0.000	0.022	0.946	0.013	0.000	0.021	0.959	0.015	0.000	0.019	0.956
2,6	((150, 150), 0.1, 0.01, 32, 50)	0.027	0.001	0.037	0.863	0.023	0.001	0.032	0.921	0.031	0.001	0.038	0.859
Average of the MBHD portfolios		0.019	0.001	0.025	0.923	0.016	0.001	0.024	0.947	0.019	0.001	0.024	0.929
3,1	((150, 150), 0.1, 0.01, 32, 50)	0.110	0.058	0.241	0.461	0.090	0.017	0.131	0.417	0.097	0.016	0.127	0.467
3,2	((100, 100), 0.01, 0.01, 32, 25)	0.073	0.019	0.139	0.698	0.060	0.009	0.094	0.736	0.064	0.010	0.102	0.571
3,3	((150, 100), 0.1, 0.01, 32, 100)	0.048	0.006	0.075	0.762	0.047	0.005	0.070	0.743	0.037	0.003	0.051	0.858
Average of the HBLD portfolios		0.077	0.028	0.152	0.640	0.065	0.010	0.098	0.632	0.066	0.010	0.094	0.632
3,4	((150, 100), 0.1, 0.01, 32, 25)	0.030	0.002	0.045	0.890	0.028	0.002	0.040	0.911	0.028	0.001	0.037	0.901
3,5	((150, 100), 0.1, 0.001, 32, 25)	0.020	0.001	0.027	0.959	0.023	0.001	0.036	0.933	0.022	0.001	0.028	0.950
3,6	((150, 150), 0.1, 0.01, 32, 100)	0.024	0.001	0.033	0.950	0.029	0.002	0.045	0.915	0.029	0.002	0.040	0.921
Average of the HBHD portfolios		0.025	0.001	0.035	0.933	0.027	0.002	0.040	0.919	0.026	0.001	0.035	0.924

The R^2 value for the testing dataset is 0.924, which is high and strong, showing that the models cover most of the variance in the return of the portfolios in the test dataset. The performance of the HBHD models is consistent and reliable across all datasets. The errors remain low, while the R^2 (0.924) values are high, suggesting a good fit and accurate predictions, with the models capturing most of the variance in the returns in these portfolios. The overall performance remained strong, highlighting the models' robustness for this group of portfolios. Model 3,5 is the highest performer in the group for all parameters, with minimum errors and a high R^2 (0.95).

Synthesis of the portfolio and predictive performance of the models

Table 3 shows that the model's predictive accuracy and fitness are generally decent across most of the portfolios but excellent in the portfolios based on a high downside beta with both medium and high betas (MBHD and HBHD). These portfolios have exhibited low errors and high R^2 values, with the maximum value of 0.97 for Portfolio 2,4, indicating excellent generalizability and strong predictive performance. The predictive accuracy of the high beta portfolios is excellent, particularly the HBHD portfolios, but they are accompanied by increased risk and a high downside risk. The high beta portfolios' (HBLD and HBHD) performance, in terms of the returns, remained excellent but comes with higher volatility, as shown by the high standard deviation reported in Table 2. The Sharpe and Sortino ratios are also high for the high beta groups, reflecting a strong risk-adjusted return. The performance of the low beta portfolios was significantly low, exhibiting negative returns and poor Sharpe and Sortino ratios. The models also struggle to capture the dynamics of these portfolios, as indicated by the low R^2 values (below 0.5); thus, the low beta seems to be an unacceptable investment strategy in the sample market.

Table 2 shows that the return performance of the low downside beta portfolios is better than the higher downside beta counterparts for both high and medium betas. Table 3 shows that the predictive accuracy and fitness of the model are excellent for the higher downside beta compared with its lower downside beta counterparts in both the medium and high beta groups.

High beta and downside beta models and portfolios remained excellent in terms of the model's predictability, fitness, and ability to deliver excess returns, making it a sound investment strategy for those willing to take the extra risk.

4. Conclusions and discussion

4.1. Discussion

The results of the study have valuable implications for portfolio construction methods, the performance of the FF6F model, and the predictive capabilities of DL models in the context of asset pricing in emerging markets like China. The study's findings highlight the relationships between risk factors and the dependent portfolios based on the beta and downside beta, and the resulting performance utilizing the DL model in a volatile market.

It is concluded that the application of DL models has proven to be highly effective for predicting portfolio returns, particularly for those with a high downside beta in the both moderate and high beta portfolios. The models demonstrated robust predictive capability for the MBHD and HBHD portfolio groups, with an R^2 value above 0.92, whereas most portfolios had R^2 values above 0.90, reaching a maximum of 0.97. These values indicate that models efficiently captured the patterns in the

relationships, both linear and nonlinear, between the risk factors and portfolio returns in these groups. These models displayed generalizability for the unseen data, emphasizing the utility of LSTM for financial modeling in emerging markets like China. Despite the complex dynamics of the stock markets, the model delivered lower errors, suggesting that they are practical tools for demystifying forecasting returns. However, the model's struggle with the low beta portfolios is evident, whether the downside risk was high or low, with a lower R^2 and higher predictive errors.

It is also concluded that the return of the portfolios reinforces the risk–return tradeoff, where the portfolios with high systematic risk delivered consistently higher returns. This relationship is specifically more pronounced in the HBLD portfolios, where the highest mean returns have been achieved. The portfolio with a low beta and a low downside beta underperforms, with minimal returns that do not justify even their low risk, suggesting that low beta portfolios in the sample market do not compensate for their risk. A distinct observation arises relating to the downside risk. While the HBLD portfolios' performance remained excellent, that performance diminished in the HBHD portfolios, highlighting that downside risk management is critical even in the presence of high systematic risk. The downside risk thus becomes a crucial factor in explaining variations in portfolios' performance.

The poor performance of the low beta portfolios, particularly those with low downside risk, coupled with the weak predictive capabilities of the model for such portfolios, suggests that the portfolios based on low risk are not able to demonstrate a predictive or strong relationship with the factors. This raises serious concerns over the effectiveness of traditional asset pricing theories for predicting returns for such portfolios in emerging markets. The situation could indicate the market's inefficiencies or the unique dynamics of the sample market in the study period, where a low beta portfolio did not compensate the investors proportionately to the risk undertaken, challenging the CAPM's assumption of a linear relationship across all levels of beta.

The strong performance of the LSTM in medium and high-risk portfolios justifies their application in asset pricing research, specifically when dealing with the portfolios with a high-risk exposure. Furthermore, it is also concluded that ML models are an effective tool for looking into complex and nonlinear associations of risk factors and portfolio returns.

The results show that the market factor is the most influential factor that contributes highly to the prediction of portfolios' return, with the broadest spread across the spectrum. Other factors like SMB, HML, CMA, RMW, and UMD also contribute but at a lower level than the market, as shown by the factors' importance and contribution analysis achieved with the SVM. The surrogate model deployed for interpretability and conditional relationships reveals that although the market factor may be the most crucial factor contributing to the prediction of returns, the importance and relationships of the other factors with returns change with changing market conditions. It is concluded from the conditional tree of the random forest model that the relationship is nonlinear and threshold-based, where a slight increase in one factor results in more significant or lesser outcomes.

It is concluded that models and portfolios with a high beta and a high downside beta remained excellent in terms of the predictability and fitness of the model and in delivering excess returns, making it a sound investment strategy for those investors who are willing to take the extra risk and gain extra returns.

4.2. Contribution and implications

This study contributes to the literature on asset pricing and portfolio management by presenting empirical evidence on the relationships among the risk factors, beta, and downside beta in explaining portfolio returns in the Shenzhen Stock Exchange (SZSE), China, over 25 years. Unlike previous studies, where the primary focus was on traditional risk factors, this study combines FF6F with LSTM, providing a novel approach to the prediction of portfolio returns in emerging markets.

By employing the downside beta in portfolio construction, the study indicates the significance of downside risk management, which is often an overlooked aspect in traditional asset pricing literature. The approach adopted in the study refines risk assessment in addition to providing in-depth insights into the nonlinear relationship between risk factors and portfolio returns, addressing the need to reconcile ML models with financial theory.

The study also has practical implications for practitioners and investors. Investors seeking higher returns should take into consideration not only the management of risk but also the downside risk for optimized portfolio performance. Furthermore, the study highlights the significance of dynamically adjusting the weights of the factors' exposure according to the market conditions, suggesting guidance for investments in emerging markets.

4.3. Limitations and future research directions

The study is limited in scope as it covers a single market in China. It can be extended to regional cooperations in countries such as the BRICS (Brazil, Russia, India, China and South Africa), SAARC (South Asian Association for Regional Cooperation), or ASEAN (Association of Southeast Asian Nations) nations. The second limitation is the application of a single DL model, LSTM, to FF6F, which can be improved by using other DL models such as GRU (Gated Recurrent Unit) and Bi-LSTM (Bidirectional Long Short-Term Memory).

Future researchers should use multiple DL models with a broad set of emerging markets to extend the generalizability of the study.

Author contributions

Tahir Afzal: Conceptualized the methodology, retrieved the data, processed and clean the data, trained the models and done the analysis; Muhammad Naveed Jan: Devise the method for factor formation and portfolios formation; Muhammad Asim Afridi: Supervision and critical review of the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflict of interest.

References

- Agarwal PK, Pradhan HK (2018) Mutual fund performance using unconditional multifactor models: Evidence from India. *J Emerg Mark Financ* 17: 157–184. <https://doi.org/10.1177/0972652718777056>
- Aharoni G, Grundy B, Zeng Q (2013). Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis. *J Financ Econ* 110: 347–357. <https://doi.org/10.1016/j.jfineco.2013.08.003>
- Aminimehr A, Raoofi A, Aminimehr A, et al. (2022) A comprehensive study of market prediction from efficient market hypothesis up to late intelligent market prediction approaches. *Comput Econ* 60: 781–815. <https://doi.org/10.1007/s10614-022-10283-1>
- Ang A, Chen JS, Xing Y (2002) Downside correlation and expected stock returns. *EFA 2002 Berlin Meetings*. <https://doi.org/10.2139/ssrn.282986>
- Ashwood AJ (2013) Portfolio selection using artificial intelligence, *Queensland University of Technology*. Available from: https://eprints.qut.edu.au/66229/1/Andrew_Ashwood_Thesis.pdf.
- Ayub U, Kausar S, Noreen U, et al. (2020) Downside risk-based six-factor capital asset pricing model (CAPM): A new paradigm in asset pricing. *Sustainability* 12: 6756. <https://doi.org/10.3390/SU12176756>
- Bhandari HN, Rimal B, Pokhrel NR, et al. (2022) Predicting stock market index using LSTM. *Mach Learn Appl* 9: 100320. <https://doi.org/10.1016/j.mlwa.2022.100320>
- Buditomo B, Candra S, Soetanto TV (2024) Fama and french five-factor study of stock market in indonesia. *Int J Organ Behav Policy* 3: 39–52. <https://doi.org/10.9744/ijobp.3.1.39-52>
- Carhart MM (1997) On persistence in mutual fund performance. *J Financ* 52: 57–82. <https://doi.org/10.1111/j.1540-6261.1997.tb03808.x>
- Carpenter JN, Whitelaw RF (2017) The development of China's stock market and stakes for the global economy. *Ann Rev Financ Econ* 9: 233–257. <https://doi.org/10.1146/annurev-financial-110716-032333>
- Cederburg S, O'Doherty MS (2015) Asset-pricing anomalies at the firm level. *J Econ* 186: 113–128. <https://doi.org/https://doi.org/10.1016/j.jeconom.2014.06.004>
- Daniel K, Moskowitz TJ (2014) Momentum crashes. *NATIONAL BUREAU OF ECONOMIC RESEARCH*. Available from: <http://www.nber.org/papers/w20439>.
- Diallo B, Bagudu A, Zhang Q (2023) Fama–French three versus five, which model is better? A machine learning approach. *J Forecast* 42: 1461–1475. <https://doi.org/https://doi.org/10.1002/for.2970>
- Dirkx P, Peter FJ (2020) The Fama-French five-factor model plus momentum: Evidence for the german market. *Schmalenbach Bus Rev* 72: 661–684. <https://doi.org/10.1007/s41464-020-00105-y>
- Estrada J (2007) Mean-semivariance behavior: Downside risk and capital asset pricing. *Int Rev Econ Financ* 16: 169–185. <https://doi.org/10.1016/j.iref.2005.03.003>
- Estrada J, Serra AP (2005) Risk and return in emerging markets : Family matters. *J Multinational Financ Manag* 15: 257–272. <https://doi.org/10.1016/j.mulfin.2004.09.002>
- Fama EF, French KR (1992) The cross-section of expected stock returns. *J Financ* 47: 427–465. <https://doi.org/10.1111/j.1540-6261.1992.tb04398.x>
- Fama EF, French KR (2006) Profitability, investment and average returns. *J Financ Econ* 82: 491–518. <https://doi.org/10.1016/j.jfineco.2005.09.009>
- Fama EF, French KR (2015) A five-factor asset pricing model. *J Financ Econ* 116: 1–22. <https://doi.org/10.1016/j.jfineco.2014.10.010>

- Fama EF, French KR (2018) Choosing factors. *J Financ Econ* 128: 234–252. <https://doi.org/10.1016/j.jfineco.2018.02.012>
- Feng G, He J, Polson NG (2018) Deep learning for predicting asset returns. *ArXiv Preprint* 1804: 1–23.
- Feng Z, Runtong Z, Lingyun H (2007) A prospective study of the price behaviors of Chinese stock markets. *Chinese Control Conference*, 558–562.
- Galagedera DU (2009) An analytical framework for explaining relative performance of CAPM beta and downside beta. *Int J Theor Appl Financ* 12: 341–358. <https://doi.org/10.1142/S0219024909005257>
- Giglio S, Kelly B, Xiu D (2022) Factor models, machine learning, and asset pricing. *Ann Rev Financ Econ* 14: 337–368. <https://doi.org/10.1146/annurev-financial-101521-104735>
- Goo Y, Wang C (2024) Do there exist nonlinear phenomena of the Fama-French six factors on stock returns? An empirical investigation on the Taiwan stock market. *Modern Econ* 15: 394–412. <https://doi.org/10.4236/me.2024.154021>
- Hall P, Gill N (2019) *An introduction to machine learning interpretability*. O'Reilly Media, Incorporated.
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Compu* 9: 1735–1780.
- Hu GX, Chen C, Shao Y, et al. (2019) Fama–French in China: Size and Value factors in Chinese stock returns. *Int Rev Financ* 19: 3–44. <https://doi.org/10.1111/irfi.12177>
- Hu GX, Pan J, Wang J (2018) Chinese capital market: An empirical overview. *NBER WORKING PAPER SERIES*. Available from: https://www.nber.org/system/files/working_papers/w24346/w24346.pdf.
- Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *J Financ* 48: 65–91. <https://doi.org/10.2307/2328882>
- Jia H (2023) Comparison on Asset Pricing Models: Application of CAPM, Fama-French 3 Factors model and Fama-French 5 factors model. *BCP Bus Manag* 40: 119–127. <https://doi.org/10.54691/bcpbm.v40i.4367>
- Jinghan C (2010) Trading preference of Chinese investors and its impact on stock volatility. *Rev Financ Stud* 3: 53–64.
- Joshi V (2024) Cross-section of expected stock returns: An application of Fama-French Five Factor Model in Nepal. *Batuk* 10: 72–89. <https://doi.org/10.3126/batuk.v10i1.62299>
- Kazmi M, Noreen U, Jadoon IA, et al. (2021) Downside beta and downside gamma: In search for a better capital asset pricing model. *Risks* 9: 223. <https://doi.org/10.3390/risks9120223>
- Kundlia S, Verma D (2021) Illiquidity premium in the Indian stock market: An empirical study. *Asian Econ Financ Rev* 11: 501–511. <https://doi.org/10.18488/JOURNAL.AEFR.2021.116.501.511>
- Lintner J (1965) Security prices, risk, and maximal gains from diversification. *J Financ* 20: 587–615. <https://doi.org/10.2307/2977249>
- Liu J (2019) A novel downside risk measure and expected returns. *The World Bank*. <https://doi.org/10.2139/ssrn.3406944>
- Liu X, Wang X (2023) Stock investment strategies with time-varying co-movements between energy market and global financial market. *Highlights Busi Econ Manag* 21: 862–866.
- Markowitz H (1952) Portfolio selection. *J Financ* 7: 77–91.
- Mahpudin E (2020) The Effect of Macroeconomics on Stock Price Index in the Republic of China. *Int J Econ Bus Adm* VIII: 228–236. <https://doi.org/10.35808/ijeba/511>
- Miller MH, Modigliani F (1961) Dividend policy, growth, and the valuation of shares. *J Bus* 34: 411. <https://doi.org/10.1086/294442>

- Miralles-Quirós JL, Miralles-Quirós MM, Nogueira JM (2020) Sustainable development goals and investment strategies: The profitability of using Five-Factor Fama-French alphas. *Sustainability* 12: 1–16. <https://doi.org/10.3390/su12051842>
- Molnar C (2020) *Interpretable machine learning*. Lulu. com.
- Mossin J (1966) Equilibrium in a capital asset market. *Econometrica* 34: 768–783. <https://doi.org/10.2307/1910098>
- Novy-Marx R (2013) The other side of value: The gross profitability premium. *J Financ Econ* 108: 1–28. <https://doi.org/10.1016/j.jfineco.2013.01.003>
- Petrozziello A, Troiano L, Serra A, et al. (2022) Deep learning for volatility forecasting in asset management. *Soft Comput* 26: 8553–8574. <https://doi.org/10.1007/s00500-022-07161-1>
- Sharma K, Bhalla R (2022) Stock market prediction techniques: A review paper. *Second International Conference on Sustainable Technologies for Computational Intelligence*, 175–188. https://doi.org/10.1007/978-981-16-4641-6_15
- Sharpe WF (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *J Financ* 19: 425–442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- Sharpe WF (1966) Mutual fund performance. *J Bus* 39: 119–138. <http://www.jstor.org/stable/2351741>
- Singh K, Singh A, Prakash P (2023) Testing factor models in an emerging market: Evidence from India. *Int J Manag Financ* 19: 203–232. <https://doi.org/10.1108/IJMF-05-2021-0245>
- Sortino FA, Van Der Meer R (1991) Downside risk. *J Portfolio Manag* 17: 27.
- Staudemeyer RC, Morris ER (2019) Understanding LSTM -- a tutorial into long short-term memory Recurrent Neural Networks. *ArXiv Preprint* 1909: 09586. <http://arxiv.org/abs/1909.09586>
- Sun J, Manger G (2012) Listing on the Shenzhen Stock Exchange: Behavioural finance implications. *ASBBS Proc* 19: 874–889. <http://search.proquest.com/docview/1445142469?accountid=79789>
- Tan L, Zhang X, Zhang X (2024) Retail and institutional investor trading behaviours: Evidence from China. *Ann Rev Financ Econ* 16: 459–483.
- Yang P, Yang L (2022) Asset pricing and nominal price illusion in China. *Humanit Soc Sci Communicat* 9: 1–9. <https://doi.org/10.1057/s41599-022-01133-4>
- Zaghoulani CA, Mezzez HD (2020). Evidence on aggregate volatility risk premium for the French Stock Market. *Managerial Financ* 46: 72–91. <https://doi.org/10.1108/MF-11-2018-0535>
- Zhang C (2022) Asset pricing and deep learning. *ArXiv Preprint* 2209: 1–18. <http://arxiv.org/abs/2209.12014>
- Zheng Y, Wen F, Deng H, et al. (2022) The relationship between carbon market attention and the EU CET market: Evidence from different market conditions. *Financ Res Lett* 50: 103140. <https://doi.org/https://doi.org/10.1016/j.frl.2022.103140>



AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)