

RESEARCH

Open Access



Stock return prediction with multiple measures using neural network models

Cong Wang^{1*}

*Correspondence:
cong.wang@uniroma1.it

¹ Department of Economics
and Law, Sapienza University
of Rome, Rome, Italy

Abstract

In the field of empirical asset pricing, the challenges of high dimensionality, non-linear relationships, and interaction effects have led to the increasing popularity of machine learning (ML) methods. This study investigates the performance of ML methods when predicting different measures of stock returns from various factor models and investigates the feature importance and interaction effects among firm-specific variables and macroeconomic factors in this context. Our findings reveal that neural network models exhibit consistent performance across different stock return measures when they rely solely on firm-specific characteristic variables. However, the inclusion of macroeconomic factors from the financial market, real economic activities, and investor sentiment leads to substantial improvements in the model performance. Notably, the degree of improvement varies with the specific measures of stock returns under consideration. Furthermore, our analysis indicates that, after the inclusion of macroeconomic factors, there is a dissimilarity in model performance, variable importance, and interaction effects among macroeconomic and firm-specific variables, particularly concerning abnormal returns derived from the Fama–French three- and five-factor models compared with excess returns. This divergence is primarily attributed to the extent to which these factor models remove the variance associated with the macroeconomic variables. These findings collectively offer valuable insights into the efficacy of neural network models for stock return predictions and contribute to a deeper understanding of the intricate relationship between factor models, stock returns, and macroeconomic conditions in the domain of empirical asset pricing.

Keywords: Neural network model, Stock return, Macroeconomic conditions, Factor model

JEL Classification: G1, C8

Introduction

Factor models play a crucial role in empirical asset pricing by providing a framework for explaining the cross-sectional variance in stock returns and identifying systematic risk factors (Fama and French 2020). Within this framework, abnormal returns derived from these factor models represent the proportion of stock returns that cannot be explained by systematic risk factors included in factor models. These abnormal returns can be influenced by a myriad of factors, including firms' idiosyncratic characteristics, market efficiencies, macroeconomic conditions, financial market dynamics, and investor

sentiment. Understanding abnormal returns holds significant value for researchers, as it provides insights into the efficacy of factor models, the presence of market efficiency, and various other facets of the financial market and the real economy. Meanwhile, excess returns measure the returns of stocks above a risk-free rate, such as the return on Treasury Bills. They represent the additional compensation that investors receive to shoulder the risks associated with investing in the stock market.

In the empirical asset pricing literature, to resolve the challenges posed by high dimensionality, non-linear relationships, and interaction effects, Cochrane (2011) and Giglio et al. (2022) have prompted researchers to explore alternative methodologies that can improve the accuracy of stock return predictions. In recent years, machine learning (ML) methods have gained popularity in empirical asset pricing research, referring to Israel et al. (2020). These methods offer powerful capabilities for analyzing complex financial data and capturing non-linear relationships and interaction effects that may be overlooked by traditional statistical approaches. Although ML methods have predominantly been applied to predict excess stock returns, it is crucial to acknowledge the economic significance of abnormal stock returns derived from various factor models. Therefore, this study concentrates on the use of neural network models for stock return prediction, assesses their performance in predicting different measures of stock returns, investigates variations in feature importance, and explores the interplay between firm-specific variables and macroeconomic factors when using neural network models to predict these different measures of stock returns.

This empirical study focuses exclusively on the US stock market and covers the period from 1971 to 2021. This study employs 49 stock characteristic variables to capture stock fundamentals and includes 14 distinct macroeconomic variables from three dimensions: financial market, real economic activities, and investor sentiment. Given the dataset spans over five decades, the factor models considered in this study have evolved from the capital asset pricing model (CAPM) to the Fama–French five-factor model. To thoroughly compare and analyze abnormal returns, this study extensively examines abnormal returns derived from various factor models, including the CAPM model, the Fama–French three-factor model (FF3), and the Fama–French five-factor model (FF5). By considering different factor models and their corresponding abnormal return estimates, this study elucidates the impact of factor selection on abnormal return analysis and its implications for empirical asset pricing research.

Prior to the main analysis using the neural network model, a univariate long–short portfolio analysis was conducted. Stocks are sorted into deciles based on 45 firm-specific characteristic feature variables (four variables cannot be cut into deciles), and long–short portfolios are constructed by a long top decile and a short bottom decile. Most stock characteristic feature variables exhibiting predictive power, with portfolios achieving higher-than-average mean returns and Sharpe ratios. Next, the full sample is divided into three subsamples based on the Chicago Fed National Activity Index (CFNAI): recession, normal, and expansion periods. Interaction effects have been observed between stock characteristic variables and macroeconomic conditions when predicting stock returns. Portfolios realized higher mean return and Sharpe ratio in the recession and expansion periods, and lower mean return and Sharpe ratio in the normal period when the market was less volatile regardless of different measures of stock returns. It is

important to note that the analysis of univariate long–short portfolios represents only the first step in exploring the relationship between our candidate predictors and stock returns. This approach does not account for the non-linear relationship between predictors and stock returns or the interaction effects among firm-specific characteristic variables.

In the main analysis, only 49 firm-specific characteristic feature variables are used to predict stock returns as a first step. The empirical analysis reveals that the performance of neural network models in predicting abnormal returns is almost equal to their performance in predicting excess returns. The out-of-sample R-squared value in the testing dataset for predicting abnormal returns based on the CAPM model is the highest at 0.825%, while the lowest R-squared value comes from predicting abnormal returns based on the FF5 model, which is 0.664%. However, there is no significant difference in prediction accuracy in terms of the R-squared values for predicting various measures of stock returns, which are around 0.7% and in a range of less than 0.18%.

Next, 14 macroeconomic variables were incorporated into the model. These variables provide a summary of the macroeconomic conditions from different perspectives. The results reveal a significant improvement in the models' performance for predicting excess stock returns in terms of R-squared values. The addition of these three macroeconomic variables from the financial market increases the out-of-sample R-squared value for excess stock returns by 89.6%, from 0.78 to 1.479%. Furthermore, after adding five macroeconomic variables from real economic activities to the model, the out-of-sample R-squared value for excess stock returns more than doubled to 2.458%. Finally, after adding the remaining six macroeconomic variables from the investor sentiment dataset, the R-squared value reached 5.474%. Similarly, the model's performance in predicting abnormal stock returns derived from various factor models also shows an improvement in prediction accuracy, although the magnitude of the improvement is not as pronounced as that observed for predicting excess stock returns, as reflected in the increased R-squared values.

The performance of value-weighted portfolios based on model predictions vary significantly. The distribution of the portfolios' CARs is nearly symmetrical, with the spectrum evenly allocated on both sides of the zero line. Investing in the top decile could yield up to eight times the abnormal returns for the entire period, whereas investing in the bottom decile could result in a loss of the same magnitude. In contrast to the spectrum of abnormal stock returns, the portfolios' cumulative excess returns exhibit an upward trend for most deciles. Investing in the top decile can yield nearly 12 times the excess return, whereas investing in the bottom decile can result in a loss close to double the excess return. After incorporating the macroeconomic variables, there are slight improvements in distinguishing the performances of the different portfolios. This result is consistent with the improvements in R-squared values observed after including these additional variables.

The relative importance of firm characteristics and macroeconomic variables is then investigated to predict various stock return measures using neural network models. To quantify their contributions, we employ SHAP values developed by Lundberg et al. (2018). This analysis reveals that the relative importance rankings of these variables differ across the different stock return measures. When predicting abnormal stock returns

derived from the FF3 and FF5 models, the firm characteristic variables from the Momentum and Trading Fractions groups emerge as the most important contributors, whereas the role of macroeconomic variables is less significant. Conversely, when predicting excess stock returns, macroeconomic variables are the most important contributors, particularly those from the investor sentiment dimension. However, when predicting abnormal stock returns derived from the CAPM, feature importance presents a mixed pattern, with macroeconomic variables from investor sentiment data and firm characteristic variables from the Trading Fractions group playing prominent roles. A potential explanation for these findings is that factor models, such as the FF3 and FF5 models, account for and eliminate the variance associated with macroeconomic conditions. By contrast, the CAPM model only includes market returns as a factor, thereby failing to remove the variance linked to macroeconomic conditions.

To gather additional evidence, the interaction effects between the firm characteristic variable short-term reversal (STreversal) and the macroeconomic variable CFNAI are explored to predict different measures of stock returns. These variables are chosen because of their significant roles as feature variables in neural network models for stock return predictions and their economic significance in empirical asset pricing. The results demonstrate a strong interaction effect between short-term reversal and CFNAI when predicting excess and abnormal stock returns derived from the CAPM model. However, this interaction effect is less pronounced when predicting the abnormal returns derived from the FF3 and FF5 models, suggesting that these factor models have already accounted for and removed the periodic variance associated with macroeconomic conditions. Similarly, the interaction effects between short-term reversals and investor sentiment are more evident in predicting excess stock returns and CAPM-derived abnormal returns, but less prominent in predicting abnormal returns derived from the FF3 and FF5 models. These consistent patterns across different macroeconomic variables highlight the importance of considering the information already incorporated and removed by the factor models.

Finally, we investigate whether the portfolios predicted by the neural network model can achieve higher mean returns and Sharpe ratios under different macroeconomic conditions. To do so, we divide the full sample into three subsamples based on the CFNAI index, similar to the univariate long–short portfolio analysis, defining them as recession, normal, and expansion. The results reveal that the long–short portfolios, constructed by buying stocks in the top decile and selling stocks in the bottom decile, consistently achieve the highest mean returns and Sharpe ratios during recession and expansion periods when the market is more volatile. Furthermore, a clear progression of the mean return and Sharpe ratio is observed from the bottom to the top decile across different economic periods for each long-only portfolio. These findings emphasize the importance of market timing, based on macroeconomic conditions, when constructing prediction-based portfolios.

Literature review

In recent years, ML methods have been increasingly used in the field of empirical asset pricing. The goal is to develop and implement models that can predict future returns or pricing of assets; exhaustive reviews in this area have been undertaken by Giglio et al.

(2022) and Bagnara (2022). Freyberger et al. (2020) use adaptive grouped LASSO to predict excess stock returns non-parametrically and explore which firm characteristics provide independent information for cross-sectional stock return prediction. One drawback of linear models, such as LASSO and Elastic-net, is that the interaction effects among the feature variables are not accounted for. Bryzgalova et al. (2020) use decision trees to group stocks endogenously and select optimal portfolio splits to span the stochastic discount factor. Non-linear tree models can accommodate interaction effects among feature variables and predict more interpretable results.

As Cochrane (2011) mentions in his AFA (American Finance Association) Presidential Address, the current challenge in asset pricing research is the “curse of high-dimensionality,” and hence, another research direction is to reduce the dimensionality of predictors. Kelly et al. (2019) propose instrumented principal component analysis to predict cross-sectional stock returns. They use observable characteristics that instrument for unobserved dynamic factor loadings such that the formulated latent factors have lower dimensions and process time-varying characteristics. Kozak et al. (2020) impose a prior on SDF coefficients that shrink the low-variance principal components of the firm characteristic factors, and an SDF built by a smaller number of reduced components performs even better than the four- or five-factor models used in the recent literature. There are further studies in this direction, such as Feng et al. (2023), who use the deep factor model to generate latent factors.

Gu et al. (2020) conduct an empirical study that compared several prediction models. They find that ML models, particularly neural network models, can outperform traditional statistical models in terms of accuracy and precision, particularly when dealing with high-dimensional and complex financial data. Further research was conducted using deep neural network models. Gu et al. (2019) propose a new latent conditional asset pricing model, following the seminal research of Kelly et al. (2019). Cong et al. (2021) use deep reinforcement learning to achieve high prediction accuracy and model interpretation. Chen et al. (2019) use deep neural networks, including feed-forward networks, long–short-term memory, and adversarial networks to estimate a pricing kernel model that predicts cross-sectional stock returns based on a large amount of information. The deep neural network learning method outperforms the benchmark approaches in terms of the out-of-sample Sharpe ratio.

While most ML studies on empirical asset pricing focus on predicting excess stock returns, Kaniel et al. (2022) examine the predictive power of neural network models in predicting abnormal fund returns. Their findings suggest novel and substantial interaction effects among market sentiment, fund flows, and momentum. This raises the question of whether abnormal stock returns are more predictable than excess returns, and whether there are significant interaction effects between macroeconomic conditions and stock characteristics.

Stock return prediction models have evolved significantly over the last century, predating the incorporation of ML methods. The CAPM introduced by Sharpe (1964) describes the concept of beta as a measure of an asset’s risk relative to the market. Later, Fama and French (1993) expand the model to include two additional factors, size and value, which explain the cross-section of expected returns. The size factor captures the tendency of smaller firms to have higher expected returns, whereas the value factor

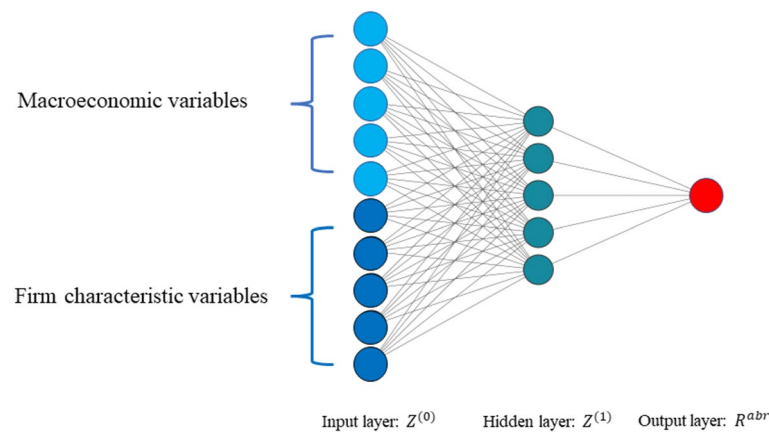


Fig. 1 Demonstration of feed-forward neural network. This figure showcases a simple neural network model structure with one hidden layer

captures the tendency of value stocks (i.e., stocks with low price-to-book ratios) to have higher expected returns. More recently, Fama and French (2015) added two factors to create a five-factor model. The profitability factor captures the risk associated with a firm's profitability, the investment factor captures the risk associated with a firm's investment policy. Recently, researchers have added that the momentum factor captures the risk associated with a stock's recent price momentum. These factors help to capture additional sources of risk and returns that are not covered by the original three factors.

Exploring the similarities and differences between stocks' abnormal returns (i.e., the return that deviates from the expected return based on factor models) and excess returns (i.e., the return that deviates from the risk-free rate) is intriguing. Investigating how macroeconomic variables predict stock returns provides valuable insights for both theoretical and empirical studies. These questions are worth further dissecting and exploring to provide empirical evidence.

Methodology

Neural network model

The artificial neural network is a potent ML modeling method. Our empirical analysis employs a simple feed-forward neural network architecture to capture the complex relationship between stock feature variables and returns. The input layer comprised a set of firm characteristics and macroeconomic variables. One or more hidden layers capture the interactive effects between different variables, and perform non-linear transformations through activation functions on the input variables. The output layer aggregates all information from the last hidden layer to generate the ultimate prediction output. This is illustrated in Fig. 1.

As posited by Gu et al. (2020), stock returns can be effectively described using an additive prediction error model represented by Eq. 1. In this equation, individual stocks are indexed as $i = 1, 2, \dots, N$ and the time periods are indexed as $t = 1, 2, \dots, T$. The abnormal return of stock i at time $t + 1$, denoted by $R_{i,t+1}^{abr}$, can be understood as the expected stock return $E_t(R_{i,t+1}^{abr})$ based on the conditional information available at time t , along with an error term $\epsilon_{i,t+1}$.

$$R_{i,t+1}^{abr} = E_t(R_{i,t+1}^{abr}) + \epsilon_{i,t+1} \quad (1)$$

In most ML studies, the primary objective is to estimate the expected return at time $t + 1$ based on the information available to practitioners at time t . This conditional expectation of returns can be denoted as $E_t(R_{i,t+1}^{abr}) = g(z; \theta)$. Here, the function $g(z, \theta)$ represents a mapping function that incorporates a set of predictors z and a set of parameters θ , which correspond to the weights and biases in each layer of the neural network model. By utilizing a feed-forward neural network, the model learns to establish a mapping between the inputs (predictors) and outputs (expected returns) by adjusting the weights of the nodes in each layer. This adjustment process occurs iteratively in response to errors caused by the model on the training dataset. A neural network model aims to minimize these errors and improve its performance in predicting stock returns. This process can be expressed mathematically as follows:

$$g(z; \theta) = \theta_0 + \sum_{j=1}^n x_j \theta_j \quad (2)$$

Objective function

The objective function (also known as the loss function) quantifies the differences between the predicted and actual outcomes in a neural network and guides the training process to determine the optimal weights. Because of the vast space available for potential weight combinations, achieving a perfect fit is virtually impossible. Instead, the training process was reformulated as an optimization problem. The goal was to search for the best possible weight set that yielded accurate predictions. Neural networks typically employ a stochastic gradient descent (SGD) optimization algorithm. The weights of the models were adjusted using the backpropagation of errors method. Predictions were made using initial weights. Subsequently, the errors between the predicted and actual values were calculated. Based on this error, the SGD algorithm modifies the weights to reduce the errors in future predictions. Commonly used objective functions to measure this error include the mean squared error (MSE), mean absolute error, and Huber loss. The MSE is often preferred for stock return predictions, where large outliers are of particular concern. The MSE loss function is given by

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3)$$

Activation function

Activation functions play an important role in neural networks. The choice of an activation function in the hidden layers dictates the transformation of the weighted sum of the inputs into output from one layer to the next. Meanwhile, the choice of the output layer determines the nature of the model's predictions. Activation functions, such as softmax, sigmoid, and ReLU, offer a range of choices. The architecture of a neural network ensures the capture of the interaction effects between variables. Simultaneously, the correct activation function can account for the non-linearity of the individual predictors

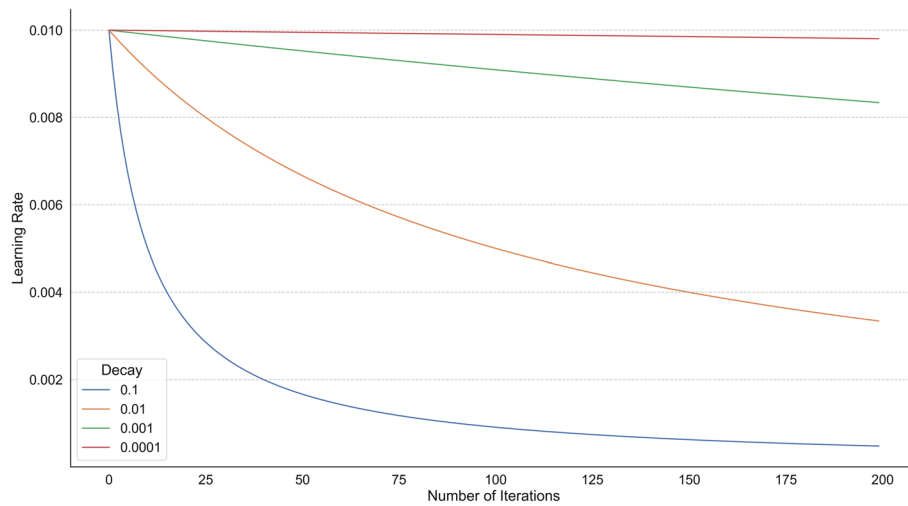


Fig. 2 Learning rate schedule. This figure shows how the learning rate changes in a learning rate schedule after each iteration. The initial learning rate is set to be 0.01, with decay rates of 0.1, 0.001, 0.001, and 0.0001 respectively

of the expected returns. In this study, a linear function is chosen for the output activation, because the returns we aim to predict have random characteristics and extreme values. For the hidden layers, the rectified linear unit (ReLU) function was adopted as the activation function because of its computational simplicity and alignment with existing literature, as discussed by He et al. (2015). The ReLU function is piecewise linear, delivering the input value if it is positive and zero otherwise, as shown in Eq. 4:

$$ReLU(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{otherwise.} \end{cases} \quad (4)$$

Learning rate

Learning rate is a vital hyperparameter in neural networks that influences the extent to which a model adjusts during weight updates. In feed-forward neural networks, the SGD algorithm aims to find the best weights to minimize loss. This adjustment is scaled by the learning rate, which controls the speed of learning; a higher rate means faster learning but potential over-adjustments, whereas a lower rate ensures careful and more precise adjustments, but can be slower. Values typically lie between 0 and 1; a very high rate may cause erratic performance, whereas a very low rate may lead to slow or stagnant optimization. To optimize the training, this study incorporates momentum to adjust the learning rate over time. Momentum considers past updates to influence future updates and, when combined with a dynamic learning rate, can lead to faster and more efficient model training. A common adaptive learning rate strategy is represented in formula 5:

$$Learning\ Rate = Initial\ Rate \times \frac{1}{1 + decay \times iteration} \quad (5)$$

This strategy, shown in Fig. 2, often outperforms static rates by better adapting to the training process.

Fix overfitting with regularization

Overfitting is a common challenge in neural networks, where models may perform exceptionally well on training data but underperform on unseen test data. The primary goal of ML is to generalize training data to new, unseen instances, and overfitting compromises this ability. There are two fundamental approaches for mitigating overfitting: enlarging the training dataset and adjusting the complexity of the model. While expanding a dataset can enhance a neural network's performance to its maximum learning capacity, adjusting the complexity involves either tweaking the network's structure, such as the number of nodes or layers, or altering its parameters, primarily the weight values. Numerous regularization techniques such as activity regularization, weight constraints, dropout, and noise addition have been developed to address overfitting. Among these, weight regularization is prominent. If left unchecked during prolonged training, the weights may adapt overly to the training data, causing overfitting. Large weights can make the model overly reactive to minor changes in test data, leading to inconsistent predictions. Weight regularization combines this by imposing penalties based on the magnitude of the weights, ensuring that they remain controlled. Various penalties like $L1$ (sum of absolute weight values) and $L2$ (sum of squared weight values) exist. This study specifically utilizes $L1$ penalties on network weights, following the LASSO approach in linear regression, a method commonly adopted in many studies.

Enhancing predictions with ensemble learning

Although adept at generalizing complex mapping functions from inputs to predicted outputs, neural networks exhibit high flexibility. This implies that they may yield different weights each time they are trained even for identical datasets. This characteristic, referred to as their low bias and high variance nature, is due to their sensitivity to initial conditions, such as initial random weights and specific nuances in the training dataset. Consequently, different initializations can lead neural networks to converge to different weights, causing variations in the predictions of new data. This variability poses a challenge for the development of a consistent predictive model. A widely adopted solution to this challenge is ensemble learning, in which multiple models are trained, and their predictions are aggregated to produce a more stable and potentially more accurate output. Although ensembling may introduce bias, it counteracts the high variance often observed in stand-alone models. Thus, ensembled models, by virtue of being less reliant on training data specifics and initial conditions, not only mitigate the variance of individual models but can also outperform single models, especially if the aggregated models excel in diverse ways and their prediction errors are uncorrelated.

Finally, the "He initializer" is used to avoid the problem of "dying ReLU." The maximum number of epochs during each training process was set to 100, and early stopping was implemented based on the R-squared value of the validation dataset. To enhance the stability of the training process, we used batch normalization, and to improve the robustness of our predictions, we employed ensemble learning by combining the predictions of the 10 models. The detailed model specifications are

Table 1 Model summary

| Parameters | Candidates | Final model | Objective function | MSE |
|--------------------------|--------------|-------------|---------------------|----------------|
| <i>Tuning parameters</i> | | | | |
| Hidden layers | 1, 2, 3 | 1 | Activation function | ReLU |
| Hidden nodes | 32, 16, 4 | 32 | Kernel initializer | He initializer |
| Learning rate | 0.01, 0.001 | 0.01 | Ensemble models | 10 |
| Decay rate | 0.01, 0.001 | 0.001 | Maximum Epochs | 100 |
| Momentum | 0.95, 0.99 | 0.99 | Early stop monitor | Validation R2 |
| L1 weight penalty | 0.001, 0.005 | 0.005 | Early stop patience | 10 |
| Batch size | 3200, 32000 | 3200 | Batch normalization | Applied |

This table presents the candidate hyperparameters and the finalized hyperparameters after tuning in the left panel. Other information related to the neural network model is presented in the right panel

presented in Table 1, where the optimal network architecture comprised one hidden layer with 32 hidden nodes. The other hyperparameters were tuned using the validation dataset and are listed in the left panel of the table.

Stock abnormal return

The objective of this study is to investigate how neural network models behave differently when predicting various stock return measures. Specifically, we concentrate on predicting excess stock returns, which are computed by deducting the one-month treasury bill rate from actual stock returns. Furthermore, we generate abnormal stock returns from various factor models using a method inspired by Kaniel et al. (2022). For each stock, the abnormal return is defined as the difference between the stock's excess return and the expected return based on various factor models, which consist of the alpha of the factor models and residuals.

To ensure a robust analysis, a criterion is set such that each stock has a minimum of 37 observations. Then, the β s are estimated for each factor using data from the past 36 months by employing Eq. 6 as follows:

$$R_{i,t-36:t} = \alpha_{i,t-36:t} + F_{t-36:t} \hat{\beta}_{i,t} + \epsilon_{i,t-36:t} \quad (6)$$

in which $R_{i,t}$ is the excess return of stock i in month t , F_t is a vector of factors; taking the FF5 model as an example, F_t would be $[SMB_t + HML_t + RMW_t + CMA_t + (R_m - R_f)_t]$, and $\hat{\beta}_{i,t}$ is a vector of estimated coefficients accordingly. We then use the estimated $\hat{\beta}_{i,t}$ from the past 36 months to predict the stock returns for the 37th month, $R_{i,t+1}^E = F_{t+1} \hat{\beta}_{i,t}$. Finally, the abnormal return for the stock is obtained as the difference between the excess return and the Fama–French five-factor model's expected return, using equation 7.

$$R_{i,t+1}^{abn} = R_{i,t+1} - R_{i,t+1}^E \quad (7)$$

Data

Firm characteristic data

This study deviates from the traditional approach of constructing all variables from raw data typically obtained from databases, such as Compustat or Thomson Reuters. Instead, we use firm characteristics data sourced from open-source asset pricing¹, a dataset created and maintained by Chen and Zimmermann (2021). There are several reasons for this observation. First, ML studies often involve the analysis of a large number of feature variables. For instance, Kozak et al. (2020) use 50 anomaly characteristics and 80 financial ratios from WRDS, while Freyberger et al. (2020) construct 62 firm characteristic variables based on raw data from Compustat. The process of constructing these variables is time-consuming. Second, the construction of these feature variables based on descriptions in original papers requires meticulous attention. Even slight, subjective variations can result in different datasets, making it challenging to reproduce some studies, let alone achieve results identical to those of the authors. Third, the use of a self-constructed dataset can introduce complexity into replication efforts and hinder comparability with other ML methods.

This study focuses on all publicly traded stocks in the U.S. stock market. The full dataset from Open Source Asset Pricing contains 319 distinct firm characteristic variables based on previous asset pricing studies. Among these, 161 characteristic variables exhibited significant t-statistics in the original papers, 44 showed mixed evidence, and the remaining 114 were found to be insignificant. Not all variables from the dataset were included in this analysis; rather, a specific subset was selected for this study. These selections were drawn from a subset of 161 characteristic variables that exhibited noteworthy t-statistics in the original research. The statistical significance of these variables underscores their relevance for predicting stock returns. Within a pool of 161 characteristic variables, the selection criteria considered the completeness of each variable. The time period ranges from 1971 to 2021, and variables with a significant number of missing observations, specifically those with more than 40% missing data, are excluded. This is related to the approach used for handling missing values in this study, thereby ensuring a maximally sized dataset in terms of observations. Gu et al. (2020), Leippold et al. (2022), and Drobetz and Otto (2021) replace all missing values with the cross-sectional median of each firm-characteristic variable for that month. Although this approach allows us to include as many firm characteristic variables as desired, it requires strong assumptions about the structure of the data and creates artificial time-series fluctuations in the firm characteristic variables. Given the non-random nature of missing data in firm characteristics, as noted by Bryzgalova et al. (2022), we chose not to use artificial imputation methods instead of dropping all missing values to obtain a balanced dataset. This was performed to avoid additional sources of error and to maintain the dependency structure in the dataset. Consequently, the final selection consists of 49 firm characteristic variables, neatly categorized into seven groups: Value, Investment, Trading Fractions, Profitability, Momentum, Intangible, and Risk. These categories are similar to those employed

¹ The dataset consists of 319 characteristics that are based on previous asset pricing studies. The database is maintained and updated annually by the authors, and open-source code is provided for the purpose of enhancing reproducibility and credibility.

Table 2 Firm characteristics by category

| Value | | |
|--------------------------|-----------------|---|
| 1 | Asset growth | Asset growth |
| 2 | Size | Size |
| 3 | SP | Sales-to-price |
| 4 | NOA | Net operating assets |
| 5 | AM | Total assets to market |
| 6 | ChEQ | Growth in book equity |
| 7 | DelEqu | Change in equity to assets |
| 8 | ChNWC | Change in Net Working Capital |
| 9 | BMdec | Book to market using December ME |
| Investment | | |
| 10 | ShareRepurchase | Share repurchases |
| 11 | ShareIss1Y | Share issuance (1 year) |
| 12 | InvestPPEInv | change in ppe and inv/assets |
| 13 | dNoa | change in net operating assets |
| 14 | DelLTI | Change in long-term investment |
| 15 | DelCOA | Change in current operating assets |
| 16 | GrLTNOA | Growth in long term operating assets |
| Trading frictions | | |
| 17 | Price | Price |
| 18 | Beta | CAPM beta |
| 19 | High52 | 52 week high |
| 20 | BidAskSpread | Bid-ask spread |
| 21 | VolSD | Volume Variance |
| 22 | DoIVol | Past trading volume |
| 23 | Illiquidity | Amihud's illiquidity |
| 24 | BetaFP | Frazzini-Pedersen Beta |
| 25 | VolMkt | Volume to market equity |
| 26 | MaxRet | Maximum return over month |
| Momentum | | |
| 27 | LRreversal | Long-run reversal |
| 28 | STreversal | Short term reversal |
| 29 | Mom12m | Momentum (12 month) |
| 30 | Mom6m | Momentum (6 month) |
| 31 | MRreversal | Medium-run reversal |
| 32 | IntMom | Intermediate Momentum |
| 33 | ResidualMom | Momentum based on FF3 residuals |
| Risk | | |
| 34 | CF | Cash flow to market |
| 35 | IdioRisk | Idiosyncratic risk |
| 36 | Leverage | Market leverage |
| 37 | BookLeverage | Book leverage (annual) |
| 38 | IdioVol3F | Idiosyncratic risk (3 factor) |
| 39 | DelFINL | Change in financial liabilities |
| 40 | cfp | Operating cash flows to price |
| 41 | DelCOL | Change in current operating liabilities |
| Profitability | | |
| 42 | CashProd | Cash productivity |

Table 2 (continued)

| Profitability | | |
|----------------------|------------|----------------------------|
| 43 | roaq | Return on assets (qtrly) |
| 44 | GP | gross profits/total assets |
| 45 | RoE | net income/book equity |
| Intangible | | |
| 46 | Accruals | Accruals |
| 47 | OPLeverage | Operating leverage |
| 48 | AbnormalAc | Abnormal accruals |
| 49 | PctAcc | Percent operating accruals |

This table presents all the firm characteristic variables used in this paper, these variables are categorized into 7 groups aligning with Chen et al. (2019)

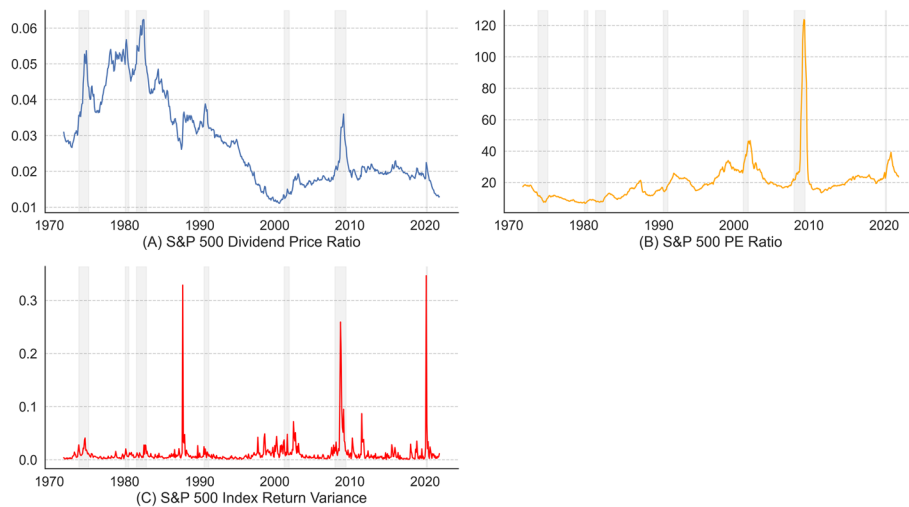


Fig. 3 Macroeconomic variables from the financial market. This graphic plots the historical trends of macroeconomic variables from the financial market dimension in conjunction with the National Bureau of Economic Research (NBER) suggested recession periods. The data for S&P 500 index return variance is computed as the variance of daily index returns within a given month

by Chen et al. (2019), as Table 2 presents. It is noteworthy that the Open Source Asset Pricing database lacks such variables as “short-term reversal,” “size,” and “return,” which are obtained from the Center for Research in Security Prices (CRSP) and merged using the unique stock identifier PERMNO and date.

Macroeconomic data

Together with the incorporation of firms’ characteristic variables, an extensive array of macroeconomic variables was employed to capture the broader economic landscape in favor of stock return prediction. The macroeconomic data used in this study are sourced from three principal dimensions: the financial market, real economic activities, and investor sentiment. Within the financial market dimension, this study coincides with that of He et al. (2021, 2023), who demonstrate the significance of macroeconomic variables in predicting bond returns. Specifically, such metrics as the S&P

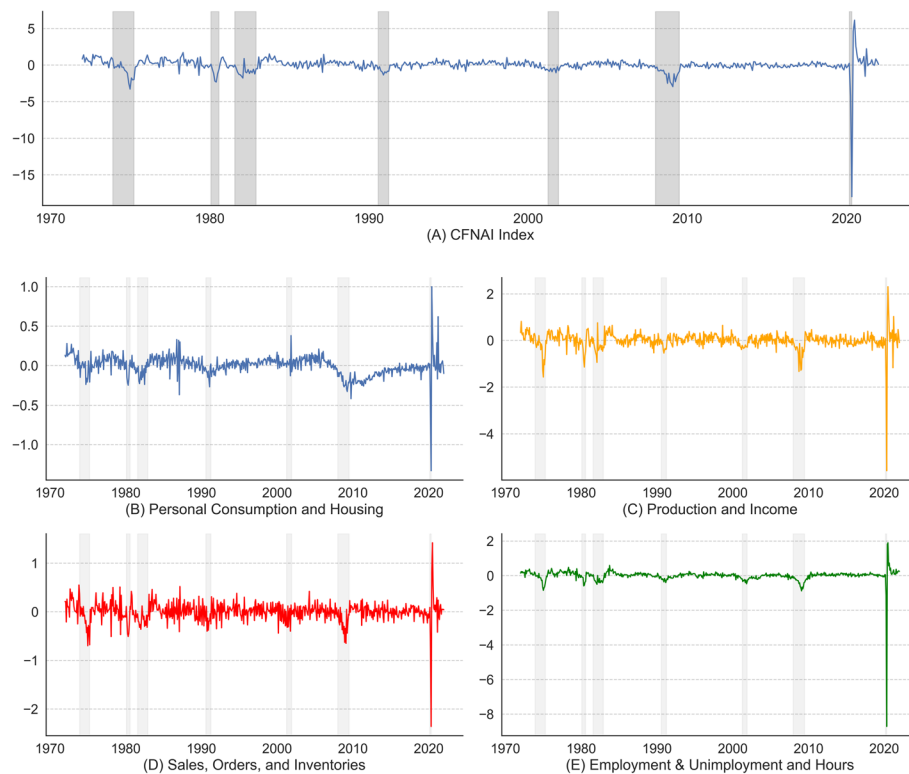


Fig. 4 Macroeconomic variables from real economic activities. This graphic plots the historical CFNAI index from 1971 to 2022 together with NBER suggested recession periods, in 2020 the CFNAI index together with the other four subsectional indices fell sharply due to the economic impact of the COVID-19 pandemic

500 dividend-to-price ratio ($S\&P500_divprc$), S&P 500 PE ratio ($S\&P500_pe$), and S&P 500 index return variance ($S\&P500_var$) are selected to quantify the state of the financial market, these variables initially used by Welch and Goyal (2008), are important in assessing market conditions. The monthly trend of each indicator is plotted in Fig. 3.

From the real economic activities dimension, the selection of macroeconomic indicators includes the CFNAI and its subsectional macroeconomic indexes, including production and income (P_I), employment, unemployment, hours (EU_H), personal consumption and housing (C_H), and sales, orders, and inventories (SO_I). These variables serve as proxies for real economic activities and are commonly used to quantify the overall macroeconomic conditions in the United States, as discussed by Evans et al. (2002) and Brave et al. (2009). The CFNAI, reported monthly by the Federal Reserve Bank of Chicago, holds significance as a leading indicator of economic growth or recession and is closely monitored in both academic and policy circles. Figure 4 illustrates the comprehensive CFNAI index's historical trends, and provides a detailed representation of the subsectional macroeconomic indexes, both in conjunction with the NBER's suggested recession indicators.

For the investor sentiment dimension, the selection of macroeconomic data originates from Baker and Wurgler (2007), who find that market sentiment, reflecting investors' collective disposition towards the market, significantly influences stock

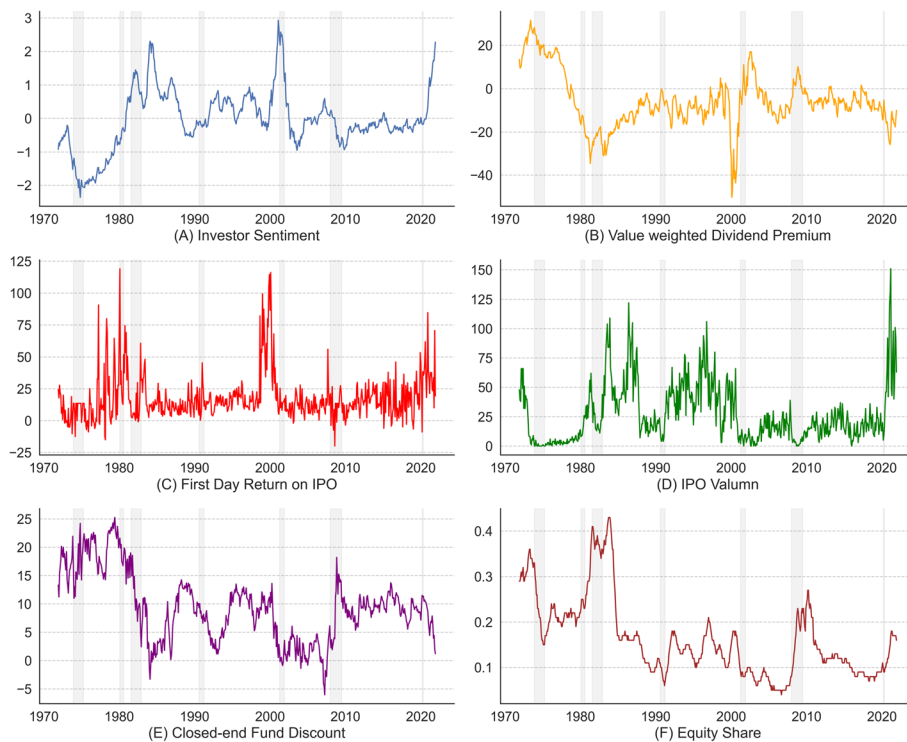


Fig. 5 Macroeconomic variables from investor sentiment. This plot showcases the historical trends in investor sentiment and its five key indicators that encompass various facets of investor sentiment, drawing from data primarily referenced Baker and Wurgler (2007) and Baker and Wurgler (2006)

Table 3 Factors from Fama–French five-factor model

| Acronym | Defination |
|---------|---|
| SMB | SMB (Small Minus Big) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios |
| HML | HML (High Minus Low) is the average return on the two value portfolios minus the average return on the two growth portfolios |
| RMW | RMW (Robust Minus Weak) is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios |
| CMA | CMA (Conservative Minus Aggressive) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios |
| Mkt-RF | $R_m - R_f$, the excess return on the market, value-weight return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t , good shares and price data at the beginning of t , and good return data for t minus the one-month Treasury bill rate (from Ibbotson Associates) |
| RF | One-month Treasury bill rate (from Ibbotson Associates) |

The Fama/French 5 factors (2×3) are constructed using the 6 value-weight portfolios formed on size and book-to-market, the 6 value-weight portfolios formed on size and operating profitability, and the 6 value-weight portfolios formed on size and investment

returns. To quantify market sentiment, they compiled a sentiment index based on five key indicators that encompass various facets of investor sentiment. Figure 5 presents the historical trend chart of investor sentiment and the components used to construct the investor sentiment index.

Table 4 Summary statistics of stock returns

| | CAPM | FF3 | FF5 | Excess return |
|-------|-----------|-----------|-----------|---------------|
| Count | 1,397,937 | 1,397,937 | 1,397,937 | 1,397,937 |
| Mean | 0.001 | − 0.000 | 0.001 | 0.009 |
| Std | 0.163 | 0.165 | 0.173 | 0.170 |
| Min | − 2.306 | − 2.748 | − 4.295 | − 0.985 |
| 25% | − 0.070 | − 0.072 | − 0.074 | − 0.067 |
| 50% | − 0.007 | − 0.008 | − 0.007 | − 0.002 |
| 75% | 0.057 | 0.058 | 0.062 | 0.069 |
| Max | 18.920 | 18.928 | 19.043 | 18.998 |

This table presents the summary statistics of different measures of stock returns. Moving from left to right, the table showcases abnormal returns derived from the CAPM model, followed by the Fama-French 3-factor and 5-factor models, and finally, the last column shows stock excess return

Fama-French 5 factor data

In this study, the Fama–French five-factor dataset and risk-free rates were incorporated from the Fama–French data library and merged with the primary dataset. A summary of the variables can be found in Table 3, whereas an in-depth discussion of their construction and economic rationale can be found in Fama and French (2015) work. The collected data span from July 1963 to the present, with a monthly frequency.

Stock abnormal return

Over the years, the field of finance has witnessed significant changes in the prevalence of factor models used for asset pricing. Starting from the classic CAPM (1964), the evolution of asset pricing models led to the intertemporal capital asset pricing model (ICAPM) by Merton (1973). The FF3 proposed by Fama and French (1992) has gained popularity, because it can better explain a cross-section of stock returns than the CAPM. However, it has been replaced by the more recent FF5 introduced by the same authors, Fama and French (2015). Moreover, researchers have proposed unofficial extensions to the FF5 model, such as the Fama–French five-factor model with momentum (FF5F+M) and the Fama–French six-factor model (FF6), which includes a volatility factor, as suggested by Harvey et al. (2016). To ensure that the study covers historical trends without drifting too much from the main academic consensus, we aim to compare the abnormal returns obtained using the CAPM, FF3, and FF5 models, given the long time horizon of our analysis spanning over five decades.

Figure 6 graphically represents the key features of excess stocks and abnormal returns. The left panel shows the cumulative monthly average excess and abnormal returns, together with the shaded NBER suggested recession periods. The cumulative excess return displays a clear upward slope, indicating a considerable increase of over four times from the first month to the last month. Several drawdowns occurred during or near the NBER suggested recession period. By contrast, the cumulative abnormal returns from the CAPM model oscillate within the range of [0,1], whereas those from the three- and five-factor models fluctuate alongside the zero bar. Among the abnormal stock returns derived from different factor models, the abnormal return generated by the five-factor model deviates the least from the zero line, suggesting that it is the best

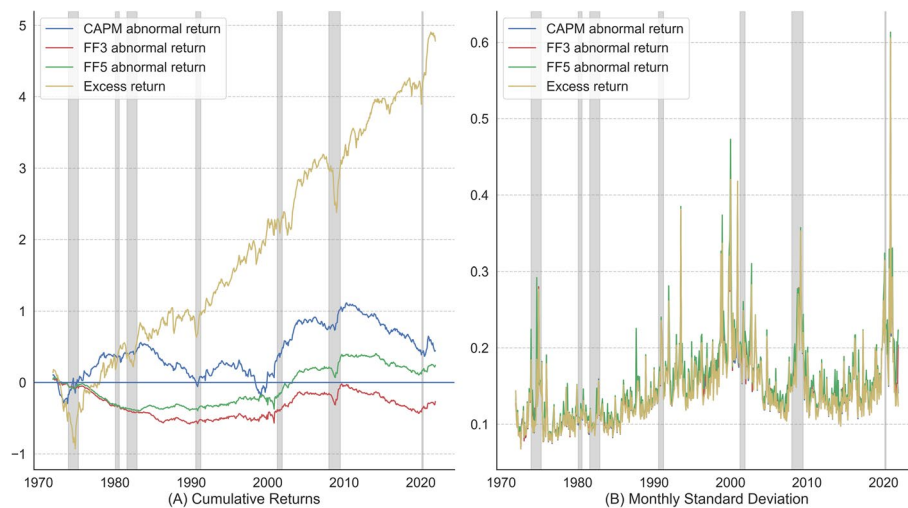


Fig. 6 Excess and abnormal returns. The left panel of this figure shows the monthly average excess and abnormal return accumulated for the whole time period in our dataset. The right panel presents the standard deviation of stock excess and abnormal returns in each month

factor model for predicting stock returns among other factor models. The right panel of Fig. 6 displays the monthly standard deviations of excess stocks and abnormal returns. Generally, the monthly standard deviations of excess and abnormal stock returns overlap. However, the standard deviation of the abnormal stock returns derived from the FF5 model was the most dispersed. Table 4 presents a summary of the statistics for various measurements of stock returns.

Result

Univariate long-short Portfolios' return

Before conducting the main analysis, we examine the predictive power of the firm characteristic variables with respect to different stock return measures. To do so, we construct univariate long-short portfolios based on each of the 45 firm characteristic feature variables. Four firm characteristic variables (ShareIss1Y, ShareRepurchase, DelFINL, and DelLTI) cannot be sorted into deciles and thus, are excluded from this analysis. This approach involves sorting stocks into deciles based on the remaining 45 firm characteristic variables. We then purchase stocks in the top decile and sell them in the bottom deciles to build the long–short portfolios. To ensure comparability, all firm feature variables are signed to ensure that the top decile exhibits higher returns than the bottom deciles, on average. Portfolios are subject to periodic rebalancing contingent on the update frequency of the respective firm characteristic variables. For instance, variables such as investment in property, plants, & equipment (InvestPPEInv) and the change in net operating assets (dNoa) undergo annual updates, whereas others such as 12-month momentum (Mom12m) and long-run reversal (LRreversal) receive monthly updates. Transaction costs are an important consideration, referring to Avramov et al. (2023) and Jensen et al. (2022), particularly when it comes to rebalancing the portfolio based on the update frequency of each specific firm characteristic variable. However, the primary objective of this analysis is to demonstrate the predictive power of each firm's

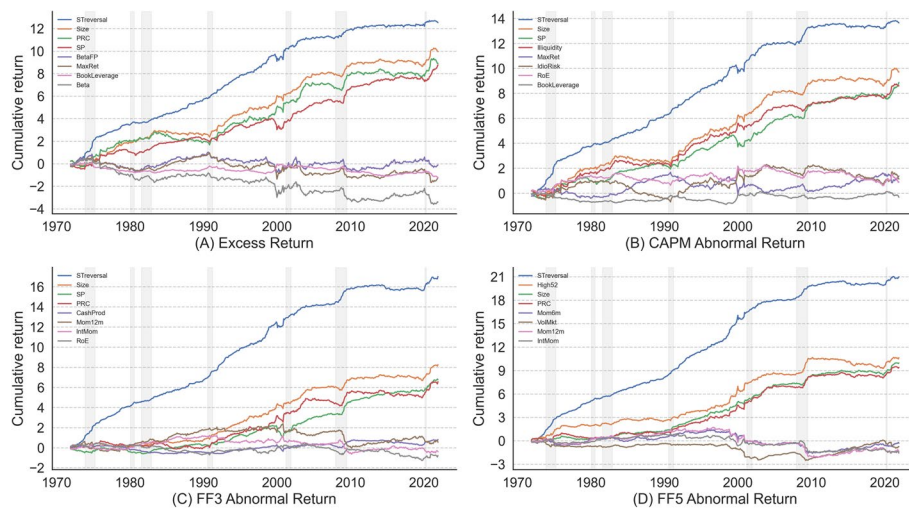


Fig. 7 Univariate long–short Portfolios’ Cumulative Return. This graphic showcases the cumulative returns of univariate long–short portfolios sorted based on firm characteristics over the entire sample period. Each panel represents different measures of stock returns, and within each panel, we highlight the four best-performing and four worst-performing portfolios in terms of cumulative returns throughout the entire sample period

characteristic variables. Therefore, the exact cumulative returns achieved by each portfolio were less critical in the context of this analysis. Additionally, the varying update frequencies of these characteristic variables, ranging from monthly to annual, would complicate the consideration of transaction costs and potentially obscure the comparison of their predictive powers of these characteristic variables. Hence, the consideration of transaction costs was relaxed in this study.² It is worth mentioning that the univariate long–short portfolio strategy solely concentrates on the predictive ability of individual firm characteristic variables. It disregards the time-varying patterns in the panel data, does not consider potential interaction effects among different variables, and overlooks the non-linear relationships between firm characteristic variables and returns.

Figure 7 shows the top and bottom three long–short univariate portfolios in terms of cumulative returns over the entire sample period using various measures of stock returns. The main finding is that regardless of the different measures of stock returns employed, the most successful univariate long–short portfolios are those based on short-term reversal (STreversal). This finding aligns with the conclusions of Welch and Goyal (2008), who demonstrate that historical average excess stock returns provide more accurate forecasts of future excess stock returns than do regressions of excess returns on predictor variables. By contrast, the bottom three long–short portfolios exhibit cumulative returns that oscillate around the zero line, indicating a lack of predictive power in the associated firm characteristic variables for stock returns. The cumulative returns of portfolios based on other firm characteristics are not shown in this figure, although most achieve significant cumulative returns at the end of the sample period. It is important to highlight that the worst performing portfolios

² To maintain the consistency of the analysis and out the similar reason, in the following analysis, we also relax the influence of the transaction fee when conducting the comparison of portfolio performance based on ML methods.

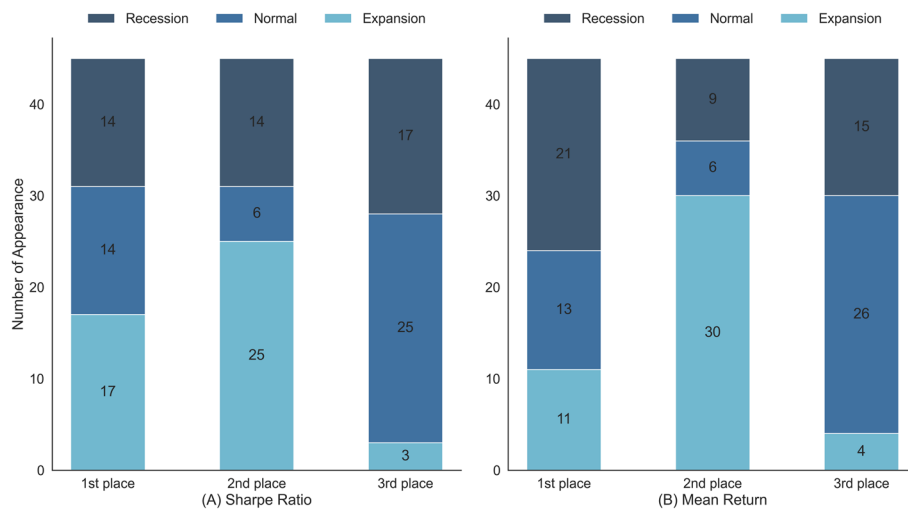


Fig. 8 Portfolio performance and macroeconomic conditions. This graphic showcases firm characteristics sorted long–short univariate portfolios’ performance across different macroeconomic conditions. We compare univariate long–short portfolios’ Sharpe ratio and mean return in recession, normal, and expansion periods, which is the horizontal comparison among the last 6 columns in table 5

differ depending on the specific measures of stock returns used. Nonetheless, it is essential to acknowledge that this analysis does not consider potential non-linear relationships and interaction effects among the different characteristic variables.

Subsequently, an analysis of univariate long–short portfolio performance across various macroeconomic conditions was conducted. In this study, the analysis involves 14 macroeconomic factors from three dimensions, collectively capturing the overall economic conditions. Similar to firm characteristic variables, these macroeconomic factors display varying predictive power with respect to stock returns. In the initial exploration of the interaction between the macroeconomic and firm characteristic variables, we focused solely on the interaction effect of the CFNAI. The CFNAI indicates that a positive index reading corresponds to growth above the trend, whereas a negative index reading corresponds to growth below the trend. Using the CFNAI index, data were sorted into tertiles, defined as expansion, normal, and recession periods from top to bottom, based on their CFNAI values. In Table 5, the first two columns present the Sharpe ratio and mean return for each univariate long–short portfolio across the entire sample period, based on abnormal stock returns calculated using the FF5 model. Notably, approximately half of these long–short portfolios display an average monthly return exceeding 0.1%, with the highest Sharpe ratio reaching 0.035. This observation aligns with the findings in Fig. 7, where certain firm characteristic variables demonstrate higher predictive power for stock returns, whereas others do not.

In Table 5, columns 3–8 provide information on the Sharpe ratio and mean return of univariate long–short portfolios under different macroeconomic conditions. A comparison of these numbers across three distinct macroeconomic scenarios is presented in Fig. 8, that is, a horizontal comparison of the portfolios formulated by the same firm characteristic variable but under different macroeconomic conditions.

Table 5 FF5 abnormal return: univariate long–short Portfolios

| | Full sample | | Recession | | Normal | | Expansion | |
|------------------|-------------|--------|-----------|--------|--------|--------|-----------|--------|
| | SR | Mean | SR | Mean | SR | Mean | SR | Mean |
| STreversal | 0.035 | 0.484 | 0.041 | 0.499 | 0.029 | 0.403 | 0.035 | 0.559 |
| Size | 0.017 | 0.284 | 0.017 | 0.260 | 0.015 | 0.283 | 0.018 | 0.313 |
| Illiquidity | 0.012 | 0.237 | 0.010 | 0.175 | 0.012 | 0.281 | 0.014 | 0.284 |
| DolVol | 0.010 | 0.224 | 0.008 | 0.164 | 0.010 | 0.245 | 0.012 | 0.280 |
| High52 | 0.018 | 0.215 | 0.024 | 0.249 | 0.011 | 0.136 | 0.017 | 0.257 |
| PRC | 0.016 | 0.213 | 0.022 | 0.253 | 0.010 | 0.158 | 0.014 | 0.216 |
| SP | 0.009 | 0.174 | 0.011 | 0.165 | 0.008 | 0.195 | 0.008 | 0.185 |
| BidAskSpread | 0.011 | 0.164 | 0.013 | 0.175 | 0.006 | 0.103 | 0.013 | 0.203 |
| NOA | 0.005 | 0.159 | 0.010 | 0.255 | 0.002 | 0.070 | 0.004 | 0.120 |
| VolSD | 0.006 | 0.159 | 0.002 | 0.038 | 0.010 | 0.361 | 0.006 | 0.175 |
| dNoa | 0.005 | 0.158 | 0.007 | 0.217 | 0.002 | 0.057 | 0.006 | 0.183 |
| InvestPPEInv | 0.005 | 0.150 | 0.006 | 0.157 | 0.004 | 0.132 | 0.005 | 0.158 |
| DelCOA | 0.005 | 0.144 | 0.005 | 0.146 | 0.001 | 0.028 | 0.008 | 0.229 |
| AssetGrowth | 0.006 | 0.137 | 0.007 | 0.155 | 0.004 | 0.115 | 0.005 | 0.137 |
| AM | 0.008 | 0.134 | 0.011 | 0.144 | 0.003 | 0.067 | 0.009 | 0.195 |
| OPLEverage | 0.005 | 0.127 | 0.007 | 0.179 | 0.005 | 0.170 | 0.002 | 0.040 |
| Accruals | 0.003 | 0.118 | 0.005 | 0.162 | −0.000 | −0.006 | 0.005 | 0.172 |
| GP | 0.004 | 0.115 | 0.003 | 0.078 | 0.008 | 0.213 | 0.002 | 0.063 |
| BookLeverage | 0.003 | 0.112 | 0.002 | 0.064 | 0.007 | 0.289 | 0.001 | 0.042 |
| DelEqu | 0.004 | 0.104 | 0.008 | 0.171 | 0.001 | 0.015 | 0.004 | 0.105 |
| ChEQ | 0.004 | 0.100 | 0.008 | 0.166 | −0.001 | −0.031 | 0.005 | 0.128 |
| Leverage | 0.005 | 0.093 | 0.007 | 0.095 | 0.000 | 0.012 | 0.007 | 0.180 |
| ChNWC | 0.002 | 0.092 | 0.004 | 0.185 | 0.000 | 0.013 | 0.001 | 0.060 |
| roaq | 0.004 | 0.083 | −0.001 | −0.023 | 0.011 | 0.223 | 0.004 | 0.084 |
| MRreversal | 0.004 | 0.080 | 0.007 | 0.116 | 0.000 | 0.002 | 0.005 | 0.114 |
| AbnormalAccruals | 0.002 | 0.074 | 0.005 | 0.166 | −0.002 | −0.104 | 0.003 | 0.110 |
| GrLTNOA | 0.002 | 0.073 | 0.002 | 0.096 | −0.001 | −0.037 | 0.004 | 0.140 |
| IdioVol3F | 0.005 | 0.072 | 0.008 | 0.105 | 0.001 | 0.017 | 0.004 | 0.079 |
| BetaFP | 0.003 | 0.069 | −0.003 | −0.049 | 0.009 | 0.230 | 0.004 | 0.112 |
| IdioRisk | 0.004 | 0.065 | 0.008 | 0.107 | 0.000 | 0.003 | 0.004 | 0.065 |
| ResidualMomentum | 0.003 | 0.051 | 0.002 | 0.030 | 0.001 | 0.017 | 0.006 | 0.104 |
| MaxRet | 0.003 | 0.045 | 0.000 | 0.006 | 0.005 | 0.091 | 0.003 | 0.055 |
| CashProd | 0.001 | 0.039 | −0.000 | −0.011 | 0.000 | 0.008 | 0.004 | 0.145 |
| DelCOL | 0.001 | 0.034 | 0.000 | 0.006 | 0.000 | 0.004 | 0.003 | 0.091 |
| BMdec | 0.001 | 0.029 | −0.004 | −0.082 | 0.001 | 0.018 | 0.007 | 0.198 |
| Beta | 0.002 | 0.029 | −0.002 | −0.031 | 0.010 | 0.139 | −0.000 | −0.004 |
| PctAcc | 0.001 | 0.027 | 0.000 | 0.003 | −0.001 | −0.041 | 0.002 | 0.098 |
| RoE | 0.001 | 0.017 | 0.007 | 0.127 | −0.005 | −0.107 | −0.000 | −0.000 |
| cfp | 0.000 | 0.006 | −0.002 | −0.052 | 0.002 | 0.050 | 0.002 | 0.038 |
| LReversal | −0.000 | −0.001 | 0.006 | 0.100 | −0.007 | −0.111 | −0.001 | −0.009 |
| CF | −0.000 | −0.005 | −0.004 | −0.074 | 0.002 | 0.061 | 0.001 | 0.022 |
| Mom6m | −0.000 | −0.007 | −0.002 | −0.029 | 0.001 | 0.015 | 0.000 | 0.001 |
| VolMkt | −0.001 | −0.023 | −0.008 | −0.118 | 0.005 | 0.117 | 0.000 | 0.009 |
| Mom12m | −0.002 | −0.030 | −0.008 | −0.086 | 0.003 | 0.043 | −0.002 | −0.025 |
| IntMom | −0.003 | −0.041 | −0.007 | −0.090 | 0.001 | 0.022 | −0.002 | −0.037 |

This table presents firm characteristics sorted univariate long–short portfolios' Sharpe ratios and mean returns. The dataset has further split into three different subsamples based on the macroeconomic condition indicator—CFNAI index, named by the recession, normal, and expansion period. The return in this table refers to the abnormal stock return derived from the FF5 model

First, we compare each long–short portfolio’s Sharpe ratio under these three macroeconomic conditions and subsequently rank the macroeconomic conditions based on the portfolios’ Sharpe ratios. Panel A of Fig. 8 presents the analysis results. Among the 45 long–short portfolios, 17 achieved their highest Sharpe ratios during the expansion period, whereas 25 portfolios exhibited their lowest Sharpe ratios during normal periods. With the same methodology, Panel B presents the comparison of different macroeconomic conditions based on the mean returns of the portfolios. Among the 45 long–short portfolios, 21 achieved their highest mean returns during recession periods, whereas 26 portfolios exhibit their lowest mean returns during normal periods. This initial exploration demonstrates that univariate long–short portfolios tend to achieve higher Sharpe ratios and mean returns during recession and expansion periods when the market is more volatile. Conversely, they tended to achieve lower Sharpe ratios and mean returns when the market was relatively stable during normal periods.³ These findings lead to the conclusion that an interaction effect exists between firm characteristics and macroeconomic conditions when predicting stock returns.

Predicting with neural network model

In the main analysis, neural network models are used for stock return prediction, with the aim of capturing non-linear relationships and interaction effects. Using the 49 firm characteristics listed in Table 2, we specify the methodology employed in this study as follows. First, we randomly split the sample into three equal-sized subsamples: training, validation, and test datasets. The neural network model is then trained using the training dataset. Subsequently, we conduct a meticulous tuning process to optimize all relevant hyperparameters, with the primary objective of maximizing the R-squared value in the validation dataset. A summary of the hyperparameters and final model settings can be found in Table 1. In the final stage, the fitted model, along with different sets of feature variables, is used to predict stock returns for the following month in the testing dataset. This ensured that all the predictions were based on out-of-sample data. In the context of using ML methods for stock return prediction in finance, “out-of-sample” refers to the data that have not been used during the model training and parameter tuning phases and are instead reserved to test the model’s performance. Predicting “unseen data” can provide a more realistic evaluation of the potential predictive accuracy in real-world scenarios.

Predicting performance

We consistently adopt this methodology to predict different measures of stock returns, aiming for a comparative analysis. The results are summarized in Table 6. In the first row of the table, the R-squared value in the testing dataset is presented, using the 49 firm characteristic variables exclusively. It is important to emphasize that the performance of the model displays minimal variation across different stock return measures. This is shown by the R-squared values for the different target variables, which remain

³ The similar patterns have also been observed when using stock abnormal returns from FF3 and CAPM model as well as stock excess returns, but results are not presented here.

Table 6 Full sample R-squared values

| Variable added | Excess | CAPM | FF3 | FF5 |
|--------------------------------|--------|-------|-------|-------|
| <i>Firm characteristics</i> | | | | |
| | 0.780 | 0.825 | 0.705 | 0.664 |
| <i>Financial market</i> | | | | |
| S &P 500 DivPrc | 0.752 | 0.854 | 0.663 | 0.676 |
| S &P 500 PE | 1.089 | 0.936 | 0.780 | 0.756 |
| S &P 500 Var | 1.479 | 0.955 | 0.754 | 0.808 |
| <i>Real economy activities</i> | | | | |
| P_I | 2.157 | 0.965 | 0.777 | 0.796 |
| EU_H | 2.160 | 0.961 | 0.803 | 0.792 |
| C_H | 2.242 | 0.958 | 0.814 | 0.808 |
| SO_I | 2.482 | 0.979 | 0.797 | 0.832 |
| CFNAI | 2.458 | 1.021 | 0.800 | 0.786 |
| <i>Investors sentiment</i> | | | | |
| PDND | 3.157 | 1.072 | 0.802 | 0.848 |
| RIPO | 3.994 | 1.102 | 0.905 | 0.907 |
| NIPO | 4.498 | 1.229 | 0.944 | 0.923 |
| CEFD | 4.859 | 1.299 | 0.930 | 0.935 |
| S | 5.010 | 1.367 | 0.938 | 0.933 |
| SENT | 5.474 | 1.411 | 0.993 | 0.946 |

The table presents R-squared values (expressed as percentages) for the prediction of stock abnormal and excess returns in the testing dataset using neural network models. Each successive row, starting from the first one, adds one more macroeconomic factor in addition

near 0.7%. Among these measures, the model demonstrates the highest accuracy when predicting CAPM abnormal returns, with an R-squared value of 0.825% in the testing dataset. Conversely, the model's predictive accuracy was lowest when forecasting FF5 abnormal returns, yielding an R-squared value of 0.664% for the testing dataset.

As extensively addressed in the literature, including studies by Cochrane et al. (2005), Welch and Goyal (2008), and as examined in “Univariate long–short Portfolios’ return” section, macroeconomic variables exhibit predictive capabilities for stock returns. Moreover, interaction effects exist between firm characteristic feature variables and macroeconomic conditions in the context of stock return predictions. In the following analysis, to encompass a comprehensive view of the economic landscape in stock return prediction, we gradually incorporate 14 macroeconomic factors from three distinct dimensions. These 14 distinct macroeconomic variables are from the financial market, real economic activities, and investor sentiment, with the expectation of capturing macroeconomic conditions as comprehensively as possible. In Table 6, each subsequent row reports the R-squared values with all the firm characteristic variables, along with the addition of one more macroeconomic variable. The rows progress from the top, which includes only the firm characteristic variables, to the bottom, where all 49 firm characteristic variables are combined with all 14 macroeconomic factors for prediction. As shown in the table, there is a consistent increase in the predictive power of the neural network models, as indicated by the rising R-squared values for predicting stock returns. The R-squared values for predicting ‘excess stock returns increased from 0.78 to 5.47%, while those for predicting abnormal stock returns based on FF5 models increased from 0.66 to 0.95% after incorporating all macroeconomic factors.

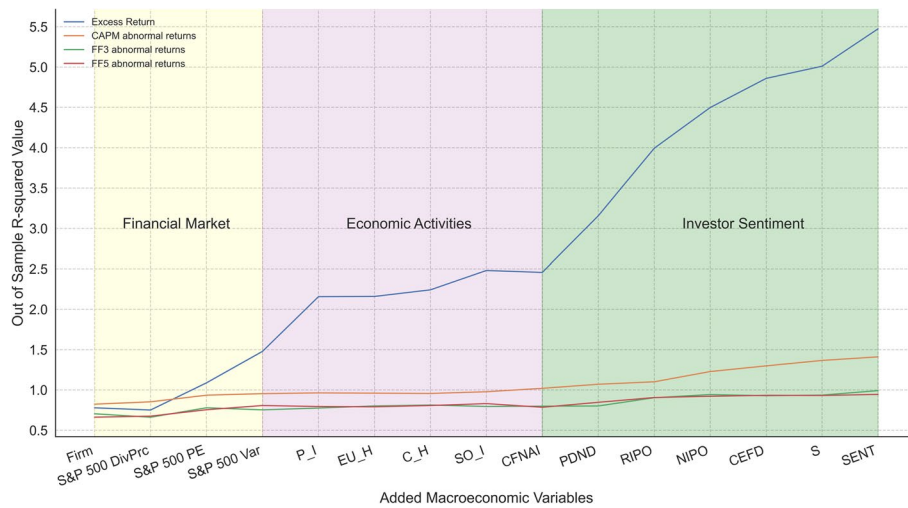


Fig. 9 Full sample R-squared values with macroeconomic variables. This graphic shows the R-squared values in stock return prediction through neural network models, with the impact of an increasing number of macroeconomic factors. Moving from the far left to the far right, it demonstrates the R-squared values, starting with only firm characteristic variables and gradually incorporating all 14 macroeconomic factors together

Although the addition of macroeconomic variables led to a significant increase in the predictive power of the neural network models, the extent of improvement varied when predicting different stock return measures. To compare the performances of the models in predicting different stock return measures, we visualize the results in Table 6 and Fig. 9. As the graph shows, there was a substantial enhancement in the prediction of excess stock returns. The model’s predictive power increases more than seven times when all firm characteristics and macroeconomic variables are incorporated, elevating the R-squared values from 0.78 to 5.47%. By contrast, improvements in predicting abnormal stock returns are relatively moderate. Notably, the neural network models behave most similarly in predicting abnormal stock returns to FF3 and FF5 models, as the increase in R-squared values follows a consistent pattern as more macroeconomic variables are added. Even though the CAPM model is the simplest factor model, the abnormal returns derived from it still exhibit significant differences from excess stock returns, but behave more similarly to abnormal returns from the FF3 and FF5 models.

Neural network Portfolios

Next, we build value-weighted portfolios based on model predictions. Each month we sort stocks into deciles based on their predicted returns, in each decile we assign weights to each component stock based on their market capitalization. It is important to note that this portfolio formulation process involves rebalancing portfolios every month and does not consider transaction costs for the same reason as mentioned before. Figure 10 shows the portfolios’ cumulative returns with respect to various measures of stock returns throughout the sample period using only firm characteristic variables for the prediction. The prediction-based portfolios can be differentiated from good to bad regardless of the kind of measures used for stock returns. By following this investing

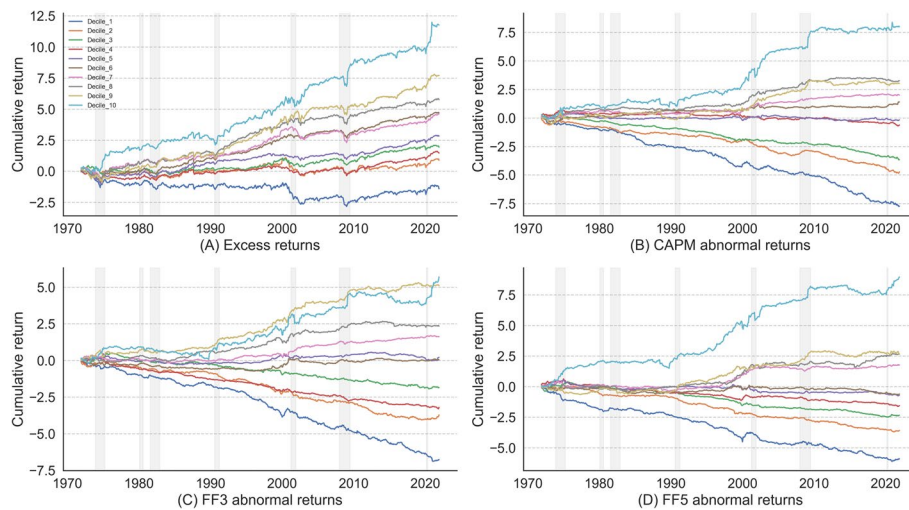


Fig. 10 Cumulative return with only firm features. This graphic presents the cumulative return of value-weighted portfolios based on predictions made solely with firm characteristic variables using neural network models. The portfolios are re-balanced each month based on model predictions without considering transaction costs

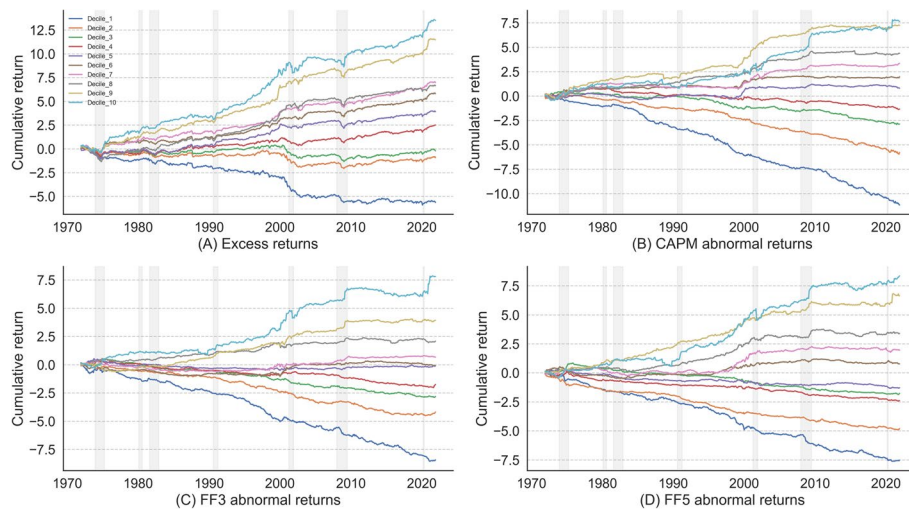


Fig. 11 Cumulative return with all features. This graphic presents the cumulative return of value-weighted portfolios based on predictions made with firm characteristic variables together with all 14 macroeconomic variables using neural network models. The portfolios are re-balanced each month based on model predictions without considering transaction costs

strategy, we could achieve approximately seven times the cumulative abnormal returns across the sample period while avoiding potential losses at the same level. The portfolio spectrum is almost symmetrically allocated on both sides of the zero line for abnormal stock returns derived from the different factor models. Panel (A) of this figure depicts the cumulative value-weighted portfolios based on the prediction of excess stock returns. By contrast to abnormal returns, most portfolios present upward slopes across time instead of being nearly symmetrically distributed on both sides of the zero line. The model can better differentiate between good and bad portfolios in terms of cumulative

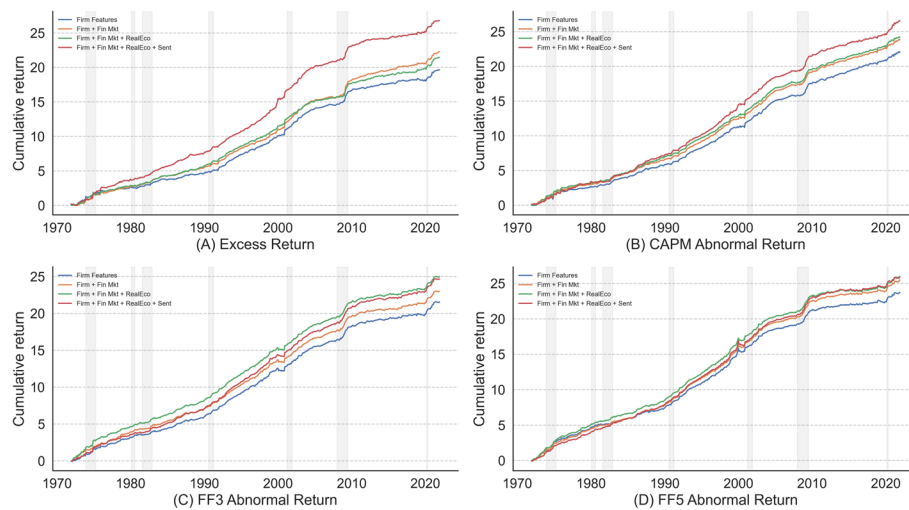


Fig. 12 Cumulative return of long-short Portfolios based on prediction. This graphic showcases the cumulative returns of equal-weighted long-short portfolios formulated using neural network models' predictions based on various predictor combinations. The graphic consists of four panels, each illustrating cumulative portfolio returns based on different measures of stock returns. Within each panel, we present cumulative long-short portfolio returns generated by neural network models' predictions using different sets of predictor variables

returns; investing in the top decile can achieve nearly 12 times the excess return, while investing in the bottom decile would result in a negative return across the entire sample period.

We employ a similar strategy to construct value-weighted portfolios, using predictions based on firm characteristic variables in conjunction with all 14 macroeconomic variables. The cumulative returns of these portfolios are shown in Fig. 11. As previously explored, the highest R-squared value in the testing dataset, listed in Table 6, is achieved when predicting using a combination of firm characteristic features and all 14 macroeconomic variables, indicating the highest level of prediction accuracy. A comparison between the cumulative returns of value-weighted portfolios based on all the variables and those based solely on firm features corroborates our previous findings. In Fig. 11, we can better distinguish the performance of different portfolios sorted based on the model predictions, especially for excess stock returns. The portfolio formulated using the top decile of the predictions achieves 13 cumulative returns across the entire sample period, whereas the portfolio formulated using the bottom decile of the predictions results in a loss of more than five times. It is important to note that these improvements extend beyond predicting excess stock returns; there are also enhancements in predicting abnormal stock returns. While linear factor models capture the linear relationship of macroeconomic conditions, they may overlook the interaction effects and non-linear relationships, which are effectively captured by neural network models.

Long-short Portfolios based on prediction

In the univariate long-short portfolio analysis, the most successful portfolio was created using the short-term reversal (STreversal) factor. Over the entire sample period, this portfolio demonstrates outstanding performance, achieving more than 12 cumulative

excess returns, nearly 14 abnormal returns based on the CAPM factor model, and over 17 and 20 abnormal returns derived from the FF3 and FF5 models, respectively. To compare the performance of long–short portfolios based on neural network model predictions with those constructed from univariate sorted portfolios, we initially sorted model-predicted returns into deciles. Subsequently, we select the top decile for a long position and the bottom decile for a short position, thereby forming equally weighted long–short portfolios based on predictions. We present the cumulative returns of these long–short portfolios for various measures with different information sets in Fig. 12. Cumulative returns can exceed 25 times for portfolios formulated using neural network model predictions that incorporate both firm characteristics and macroeconomic variables, regardless of the specific measure used for stock returns.

It is important to note that even in Fig. 7, the best-performing univariate long–short portfolio based solely on short-term reversal (STreversal) achieved only 20 abnormal returns derived from the FF5 model. Most other univariate long–short portfolios do not generate cumulative returns that exceed 10 times. When we examine each panel in Fig. 12, we observe that the red line, representing prediction-based long–short portfolios with both firm features and all 14 macroeconomic variables, outperforms most portfolios based on predictions with only firm features or a limited set of macroeconomic variables. This is also evident from the blue lines, which represent model prediction-based long–short portfolios with only firm characteristic variables that have achieved the lowest cumulative returns, regardless of which measure of stock returns is considered. Again, when comparing the different measures of stock returns, the addition of macroeconomic variables has the most significant impact on the cumulative returns of long–short portfolios, particularly excess stock returns. As previously discussed, linear factor models already capture macroeconomic conditions, and their addition does not lead to significant improvements in abnormal returns compared to predicting excess stock returns.

Feature importance

As we show in the univariate long–short portfolio analysis, some predictors exhibit stronger predictive power, whereas others are less influential in predicting stock returns. In the subsequent analysis, we assessed the relative importance of firm characteristics and macroeconomic variables in predicting various measures of stock returns. To quantify the significance of features in the neural network model, we employ SHAP (SHapley Additive exPlanation) values, a novel approach introduced by Lundberg and Lee (2017). SHAP values track how the predicted outputs change when conditioning specific variables. Figure 13 illustrates the relative importance rankings of firm characteristics and macroeconomic variables for predicting various measures of stock returns. This includes the importance of each specific variable and the cumulative feature importance within each group. For a comprehensive list of these groups, please refer to Table 2.

The importance of feature variables varies in predictions when different measures of stock returns are compared. For instance, when predicting abnormal stock returns using the FF3 and FF5 models, the most critical feature variables belong to a group of Momentum and Trading Fractions, such as short-term reversals and past trading volumes (DoIVol). Although macroeconomic variables contribute to the prediction, their

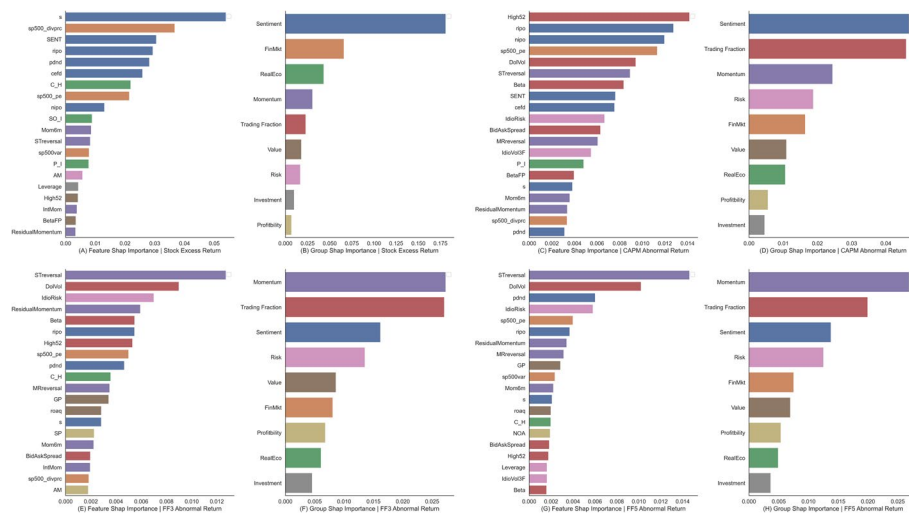


Fig. 13 Shap feature importance. This figure shows the importance ranking for predictor variables and variable groups in predicting stock abnormal returns derived from factor models and stock excess return. The ranking is the average of the feature variables’ Shap force. The variable importance measures are evaluated on the testing dataset. Group Shap importance is the feature importance’s sum value within each group

relative importance is low. Conversely, macroeconomic variables play a dominant role in predicting excess stock returns. Among the macroeconomic variables, those from the investor sentiment and financial market groups rank highest in relative importance within the neural network. Similarly, when predicting abnormal stock returns derived from the CAPM, macroeconomic variables from investor sentiment take the lead, which is consistent with the prediction of excess stock returns. The most influential variables are from the Trading Fractions and Momentum groups, aligned with predicting abnormal stock returns using the FF3 and FF5 models. Abnormal returns derived from the simplest factor model, the CAPM, exhibit similarities with both excess stock returns and abnormal returns derived from FF3 and FF5 models.

As depicted in Fig. 6, abnormal stock returns are adjusted using factor models to remove periodic fluctuations related to the macroeconomic conditions inherent in stock returns. Only non-linear relationships and interactive effects from macroeconomic variables contribute to the prediction of abnormal stock returns because linear relationships have already been captured by linear factor models. Consequently, the relative importance of the macroeconomic variables in predicting abnormal stock returns is low. However, macroeconomic variables contain valuable information that contributes the most to predicting excess stock returns, which has not been touched upon by factor models. Hence, the macroeconomic variables rank highest in terms of relative feature importance for predicting excess stock returns. Finally, when predicting abnormal stock returns using the CAPM, the relative feature importance shows mixed evidence. This is because the original CAPM solely employs market returns to capture macroeconomic conditions, which fail to capture a substantial portion of macroeconomic conditions and neglect the periodic fluctuations associated with macroeconomic conditions. In summary, the importance of macroeconomic variables

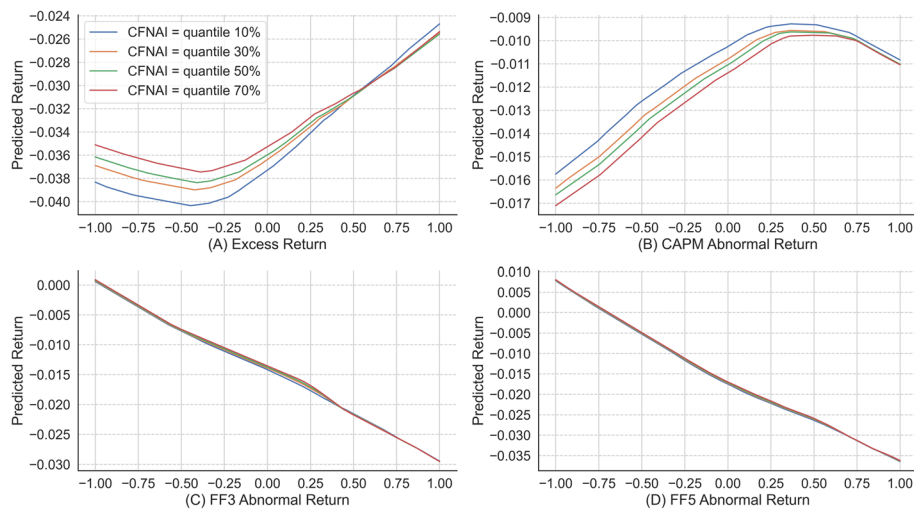


Fig. 14 Interaction effect between STreveral and CFNAI. This figure shows the sensitivity of expected monthly returns (vertical axis) to interaction effects for short-term reversal and macroeconomic variable the CFNAI index in the Neural network model (holding all other covariates fixed at their median values, while the macroeconomic variable the CNFAI index is taken at 10%, 30%, 50%, and 70% level)

in predicting different measures of stock returns depends on the specific measure being considered and the adjustment made to the returns.

Interaction effects

In the analysis of interaction effects, we explore how a short-term reversal (STreveral) interacts with two macroeconomic indicators: the CFNAI and the investor sentiment index. We specifically chose these variables because of the pronounced influence of short-term reversals and macroeconomic indicators, which typically act as feature variables in neural network models dedicated to stock return predictions. In addition, the CFNAI and investor sentiment index are comprehensive indicators of the overall health of real economic activities and investors’ enthusiasm levels, respectively. However, it is important to note that the interaction effects observed in the neural network model do not imply a causal relationship between the predictor variable and outcomes, nor do they offer actionable investment insights into practical scenarios. Instead, these findings are specific to the context of stock return predictions within specific modeling frameworks.

Figure 14 illustrates the interaction effect between short-term reversals and the macroeconomic variable CFNAI. Notably, we observe a strong interplay between these variables in predicting both excess and abnormal stock returns derived from the CAPM. However, when predicting abnormal stock returns using the FF3 and FF5 models, the interaction effects were less pronounced. This finding reinforces our previous results, indicating that abnormal stock returns derived from the CAPM exhibit similar properties to excess stock returns when fitted to neural network models. Meanwhile, abnormal stock returns derived from the FF3 and FF5 models lack the periodic variance associated with macroeconomic conditions, as they have

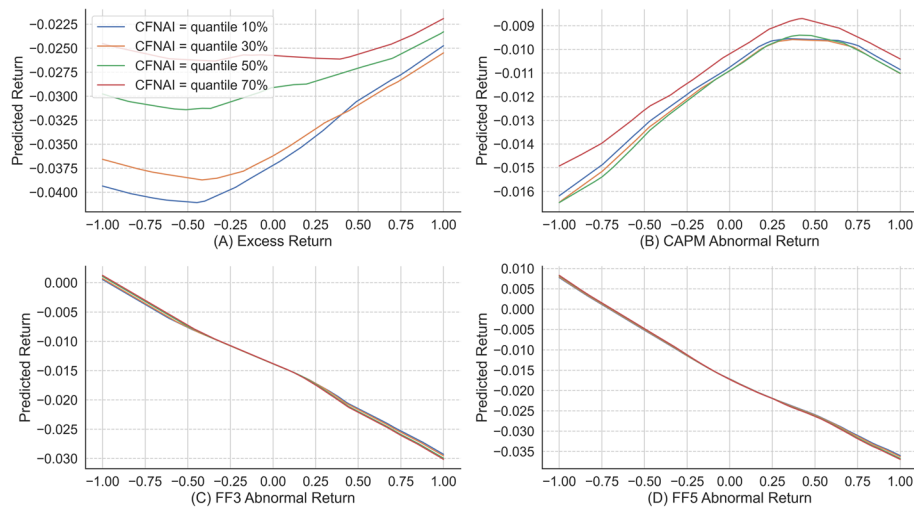


Fig. 15 Interact effect between *STreveral* and investor sentiment. This figure shows the sensitivity of expected monthly returns (vertical axis) to interaction effects for short-term reversal and macroeconomic variable the investor sentiment index in the Neural network model (holding all other covariates fixed at their median values, while the macroeconomic variable the investor sentiment index is taken at 10%, 30%, 50%, and 70% level

been accounted for and removed by the factor models. Consequently, the interaction effects of the macroeconomic variables are not as prominent in these cases.

Figure 15 presents the interaction effects between short-term reversals and the macroeconomic variable of the investor sentiment index. Consistent with our previous findings, we observe that the interaction effects are more pronounced when predicting excess and abnormal stock returns derived from the CAPM. Conversely, the interaction effects are less evident when predicting abnormal stock returns derived from the FF3 and FF5. These results further reinforce our earlier conclusions, highlighting the consistent patterns observed across different macroeconomic variables. Stronger interaction effects in predicting excess stock returns and CAPM-derived abnormal returns suggest that short-term reversals and investor sentiment synergistically contribute to these stock return measures. However, the diminished interaction effects in predicting abnormal returns derived from the FF3 and FF5 models indicate that these factor models have already accounted for and incorporated the influence of investor sentiment, thus diminishing its additional impact.

Market timing based on macroeconomic conditions

In the previous sections, we find a clear interaction effect between the macroeconomic variable CFNAI and the firm characteristic variable short-term reversal (*STreveral*). We also confirm that the inclusion of macroeconomic variables substantially improves the predictive accuracy of the neural network models. This prompts us to consider whether we can identify the optimal market timing to enhance the performance of our ML-driven portfolios. In this section, we extend the approach outlined in “[Neural network Portfolios](#)” section to construct prediction-based portfolios by using a neural network model. However, instead of constructing portfolios using the entire sample, we follow

Table 7 Market timing based on CFNAI

| | Recession | | Normal | | Expansion | |
|-----------------------------|-----------|------------|---------|------------|-----------|------------|
| | Mean | Sharpe rat | Mean | Sharpe Rat | Mean | Sharpe rat |
| <i>Excess return</i> | | | | | | |
| Bottom | - 0.856 | - 0.108 | - 1.089 | - 0.180 | - 1.592 | - 0.228 |
| 2 | - 0.087 | - 0.013 | 0.006 | 0.001 | - 0.711 | - 0.121 |
| 3 | 0.422 | 0.064 | 0.099 | 0.021 | - 0.249 | - 0.044 |
| 4 | 0.503 | 0.080 | 0.524 | 0.119 | 0.232 | 0.042 |
| 5 | 0.845 | 0.130 | 0.822 | 0.198 | 0.263 | 0.049 |
| 6 | 1.408 | 0.195 | 0.705 | 0.157 | 0.635 | 0.110 |
| 7 | 1.403 | 0.199 | 1.126 | 0.235 | 0.829 | 0.140 |
| 8 | 1.771 | 0.234 | 1.297 | 0.261 | 0.970 | 0.152 |
| 9 | 2.134 | 0.238 | 1.474 | 0.247 | 1.504 | 0.210 |
| Top | 4.398 | 0.323 | 2.476 | 0.313 | 2.879 | 0.311 |
| Long-short | 5.253 | 0.488 | 3.565 | 0.539 | 4.471 | 0.622 |
| <i>CAPM abnormal return</i> | | | | | | |
| Bottom | - 1.724 | - 0.352 | - 1.999 | - 0.427 | - 2.791 | - 0.558 |
| 2 | - 0.547 | - 0.169 | - 0.910 | - 0.277 | - 1.373 | - 0.379 |
| 3 | - 0.599 | - 0.218 | - 0.439 | - 0.155 | - 0.694 | - 0.242 |
| 4 | - 0.132 | - 0.046 | - 0.146 | - 0.057 | - 0.424 | - 0.135 |
| 5 | - 0.001 | - 0.001 | 0.208 | 0.098 | - 0.077 | - 0.028 |
| 6 | 0.232 | 0.059 | 0.299 | 0.122 | 0.292 | 0.086 |
| 7 | 0.704 | 0.182 | 0.513 | 0.187 | 0.228 | 0.060 |
| 8 | 0.955 | 0.190 | 0.719 | 0.255 | 0.541 | 0.140 |
| 9 | 1.209 | 0.210 | 0.946 | 0.250 | 0.801 | 0.175 |
| Top | 3.076 | 0.299 | 1.615 | 0.285 | 1.947 | 0.272 |
| Long-short | 4.801 | 0.492 | 3.614 | 0.586 | 4.738 | 0.646 |
| <i>FF3 abnormal return</i> | | | | | | |
| Bottom | - 1.914 | - 0.402 | - 1.758 | - 0.454 | - 2.129 | - 0.540 |
| 2 | - 1.156 | - 0.331 | - 0.878 | - 0.328 | - 0.796 | - 0.301 |
| 3 | - 0.611 | - 0.261 | - 0.505 | - 0.270 | - 0.474 | - 0.220 |
| 4 | - 0.378 | - 0.152 | - 0.161 | - 0.104 | - 0.161 | - 0.071 |
| 5 | - 0.515 | - 0.270 | 0.006 | 0.004 | - 0.176 | - 0.089 |
| 6 | 0.012 | 0.005 | 0.072 | 0.044 | - 0.163 | - 0.086 |
| 7 | 0.151 | 0.061 | 0.352 | 0.194 | - 0.133 | - 0.073 |
| 8 | 0.474 | 0.150 | 0.278 | 0.129 | 0.316 | 0.138 |
| 9 | 1.062 | 0.222 | 0.803 | 0.309 | 0.569 | 0.210 |
| Top | 3.038 | 0.375 | 1.512 | 0.305 | 1.897 | 0.362 |
| Long- Short | 4.952 | 0.589 | 3.270 | 0.573 | 4.026 | 0.689 |
| <i>FF5 abnormal return</i> | | | | | | |
| Bottom | - 1.798 | - 0.349 | - 1.736 | - 0.415 | - 1.914 | - 0.469 |
| 2 | - 1.113 | - 0.340 | - 0.749 | - 0.339 | - 0.811 | - 0.351 |
| 3 | - 0.471 | - 0.173 | - 0.531 | - 0.294 | - 0.282 | - 0.120 |
| 4 | - 0.321 | - 0.121 | - 0.435 | - 0.215 | - 0.458 | - 0.186 |
| 5 | - 0.197 | - 0.079 | 0.061 | 0.037 | - 0.404 | - 0.186 |
| 6 | - 0.206 | - 0.065 | - 0.003 | - 0.002 | - 0.097 | - 0.047 |
| 7 | 0.302 | 0.110 | 0.354 | 0.195 | 0.084 | 0.041 |
| 8 | 0.594 | 0.176 | 0.692 | 0.268 | 0.274 | 0.117 |
| 9 | 1.094 | 0.231 | 1.247 | 0.383 | 1.004 | 0.273 |
| Top | 3.243 | 0.401 | 2.051 | 0.366 | 2.154 | 0.406 |
| Long-short | 5.041 | 0.577 | 3.787 | 0.532 | 4.067 | 0.596 |

Table 7 (continued)

This table presents monthly portfolio mean returns and Sharpe ratios based on predictions generated by neural network models. Each panel in the table corresponds to different measures of stock returns, with portfolios listed from the bottom deciles to the top. The last row in each panel provides the returns and Sharpe ratios for the long-short portfolios. All the figures are expressed as percentages

the strategy used in univariate long–short portfolio analysis and divide the entire sample into tertiles based on the macroeconomic variable CFNAI. Following the methodology described in “[Univariate long–short Portfolios’ return](#)” section, we define each tertile as representing the expansion, normal, and recession periods. Within each economic period, we sort the stocks into deciles on a monthly basis. Table 7 presents the monthly mean returns and Sharpe ratios for each portfolio. The final row of each panel reports the mean return and Sharpe ratio of the long–short portfolio. The long–short portfolios are constructed by buying stocks in the top decile and selling stocks in the bottom decile.

By segmenting the sample data based on CFNAI, our objective is to capture the varying performance of prediction-based portfolios across different economic conditions. Interestingly, we consistently observe that, more often than not, both the long–short portfolios and long-only portfolios yield higher mean returns and Sharpe ratios during recession and expansion periods. It is notable that only a few long-only portfolios managed to achieve higher mean returns or Sharpe ratios during the “normal” period, characterized by CFNAI index values within the middle range, which signifies a more stable market environment with lower volatility. Moreover, for each long-only portfolio, we observe a distinct progression in the mean return and Sharpe ratio from the top to the bottom decile across different economic periods, regardless of the specific measure of stock return being examined. In conclusion, ML-based portfolios tend to exhibit improved performance, with higher mean returns and Sharpe ratios during recession and expansion periods, when the CFNAI reflects extreme values characterized by higher market volatility.

Conclusion

This study offers an in-depth exploration of the network models across various stock return measures. When relying solely on firm characteristic variables, neural network models exhibit consistent performances across different stock return measures. However, including macroeconomic variables significantly boosts the predictive accuracy for excess stock returns, as evidenced by the substantial increase in out-of-sample R-squared values. By contrast, the enhancement in predicting abnormal stock returns derived from different factor models is modest.

The analysis of feature importance in predicting various stock return measures reveals distinct patterns. Macroeconomic variables emerged as the primary drivers of excess stock returns. However, when predicting stock abnormal returns derived from the FF3 and FF5 models, the variables from firm characteristic groups Momentum and Trading Fractions dominate. In the context of predicting abnormal returns based on the CAPM model, both macroeconomic variables from the investor sentiment dataset and firm characteristic variables from the Trading Fractions group take precedence. This divergence can be attributed to the nuances of the factor models, in which advanced factor models effectively adjust for and remove the periodic fluctuations associated with

macroeconomic conditions. By contrast, the primitive CAPM relies solely on market returns to account for macroeconomic variations, missing a significant amount of macroeconomic information and disregarding associated periodic fluctuations. This makes its predictions similar to those of excess stock returns.

Furthermore, this study uncovers strong interaction effects between macroeconomic variables and firm-specific characteristics in predicting excess and abnormal stock returns derived from the CAPM. This interplay highlights the significance of considering macroeconomic conditions and their interactions with firm-specific factors to understand stock performance. These findings also offer valuable guidance for market timing strategies, suggesting that portfolios informed by neural network models tend to outperform during both recession and expansion periods, particularly when market volatility is heightened.

In summary, this study contributes to the advancement of empirical asset pricing research by exploring the capabilities of neural network models in predicting different measures of stock returns. The insights gained from this study can inform future research and assist market participants in making informed investment decisions by comprehensively considering the firm-specific characteristics, macroeconomic factors, and interaction effects. Despite its contributions, this study had several limitations. First, we chose a 36-month data window to estimate β s to compute abnormal returns; different choices of rolling windows may introduce potential bias. Additionally, neural network models are highly dynamic, and variations in the model structure or hyperparameters may yield slightly different results, especially concerning the interaction effects highlighted in this study. Second, although our methodology can be readily applied to other stock markets or different periods, we did not conduct a comprehensive robustness analysis. Thirdly, the factor models examined are not exhaustive, omitting emerging factors like Momentum (Carhart 1997), Liquidity (Acharya and Pedersen 2005), and Green-Brown (Pástor et al. 2022), etc. from recent asset pricing literature for example. Finally, although we identified the underlying reasons for the similarities and variances across different measures of stock returns, formal empirical validation of these hypotheses remains lacking. Addressing these nuances and embarking on pragmatic hypothesis testing will undoubtedly be valuable for future studies.

Abbreviations

| | |
|---------------|---|
| ML | Machine learning |
| CAPM | Capital asset pricing model |
| FF3 | Fama and French 3 factors |
| FF5 | Fama and French 5 factors |
| CFNAI | Chicago fed national activity index |
| STreversal | Short-term reversal |
| LASSO | Least absolute shrinkage and selection operator |
| AFA | American Finance Association |
| SDF | Stochastic discount factor |
| SDG | Stochastic gradient descent |
| MSE | Mean squared error |
| ReLU | Rectified linear unit |
| CRSP | Center for research in security prices |
| S&P500_divprc | S&P 500 index dividend-to-price ratio |
| S&P500_pe | S&P 500 index PE ratio |
| S&P500_var | S&P 500 index return varianc |
| P_I | Production and income |
| EU_H | Employment, unemployment, and hours |
| C_H | Personal consumption and housing |

| | |
|--------------|--|
| SO_I | Sales, orders, and inventories |
| NBER | National Bureau of Economic Research |
| SMB | Small minus big |
| HML | High minus low |
| CMA | Conservative minus aggressive |
| Mkt-RF | Market return minus risk free rate |
| InvestPPEInv | Investment in property, plant, and equipment |
| dNoa | Change in net operating assets |
| Mom12m | 12-Month momentum |
| LReversal | Long-run reversal |
| DolVol | Past trading volume |
| SHAP | SHapley additive explanation |

Acknowledgements

This paper was initially conceptualized during a conference at the Barcelona School of Economics. A significant portion of it was developed during my visit at the University of California, Los Angeles, and I am thankful for the fellowship and support of my colleagues and students throughout this period. I received invaluable suggestions from my supervisor, Professor Giacomo Morelli, at Sapienza University of Rome; my host, Professor Saki Bigio, at the University of California, Los Angeles; and Professor Andre B.M. Souza at the Barcelona School of Economics. Special thanks to the journal's associate editor and anonymous referees for their extremely useful suggestions to improve the quality of the article.

Author contributions

I am the sole author of this paper.

Funding

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The author declares that there are no competing interests.

Received: 1 August 2023 Accepted: 29 December 2023

Published online: 01 June 2024

References

- Acharya VV, Pedersen LH (2005) Asset pricing with liquidity risk. *J Financ Econ* 77(2):375–410
- Avramov D, Cheng S, Metzker L (2023) Machine learning vs economic restrictions: Evidence from stock return predictability. *Manag Sci* 69(5):2587–2619
- Bagnara M (2022) Asset pricing and machine learning: a critical review. *J Econ Surv*
- Baker M, Wurgler J (2006) Investor sentiment and the cross-section of stock returns. *J Financ* 61(4):1645–1680
- Baker M, Wurgler J (2007) Investor sentiment in the stock market. *J Econ Perspect* 21(2):129–151
- Brave S et al (2009) The Chicago fed national activity index and business cycles. *Chicago Fed Letter*
- Bryzgalova S, Lerner S, Lettau M, Pelger M (2022) Missing financial data. Available at SSRN 4106794
- Bryzgalova S, Pelger M, Zhu J (2020) Forest through the trees: building cross-sections of stock returns. Available at SSRN 3493458
- Carhart MM (1997) On persistence in mutual fund performance. *J Financ* 52(1):57–82
- Chen AY, Zimmermann T (2021) Open source cross-sectional asset pricing. *Crit Finance Rev* (Forthcoming)
- Chen L, Pelger M, Zhu J (2019) Deep learning in asset pricing. arXiv preprint [arXiv:1904.00745](https://arxiv.org/abs/1904.00745)
- Cochrane JH et al (2005) Financial markets and the real economy. *Found Trends Finance* 1(1):1–101
- Cochrane JH (2011) Presidential address: discount rates. *J Financ* 66(4):1047–1108
- Cong LW, Tang K, Wang J, Zhang Y (2021) Alphaportfolio: direct construction through deep reinforcement learning and interpretable AI. Available at SSRN 3554486
- Drobtz W, Otto T (2021) Empirical asset pricing via machine learning: evidence from the European stock market. *J Asset Manag* 22:507–538
- Evans CL, Liu CT, Pham-Kanter G (2002) The 2001 recession and the Chicago fed national activity index: identifying business cycle turning points. *Econ Perspect Federal Reserve Bank Chicago* 26(3):26–43
- Fama EF, French KR (1992) The cross-section of expected stock returns. *J Finance* 47(2):427–465
- Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. *J Financ Econ* 33(1):3–56

- Fama EF, French KR (2015) A five-factor asset pricing model. *J Financ Econ* 116(1):1–22
- Fama EF, French KR (2020) Comparing cross-section and time-series factor models. *Rev Financ Stud* 33(5):1891–1926
- Feng G, He J, Polson NG, Xu J (2023) Deep learning in characteristics-sorted factor models. *J Financ Quant Anal*. <https://doi.org/10.1017/S0022109023000893>
- Freyberger J, Neuhierl A, Weber M (2020) Dissecting characteristics nonparametrically. *Rev Financ Stud* 33(5):2326–2377
- Giglio S, Kelly B, Xiu D (2022) Factor models, machine learning, and asset pricing. *Annu Rev Financ Econ* 14:337–368
- Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. *Rev Financ Stud* 33(5):2223–2273
- Gu S, Kelly BT, Xiu D (2019) Autoencoder asset pricing models
- Harvey CR, Liu Y, Zhu H (2016) The cross-section of expected returns. *Rev Financ Stud* 29(1):5–68
- He X, Feng G, Wang J, Wu C (2021) Predicting individual corporate bond returns. Available at SSRN 4374213
- He X, Feng G, Wang J, Wu C (2023) Corporate bond pricing via benchmark combination model. In: *Corporate bond pricing via benchmark combination model: He, Xin–uFeng, Guanhao–uWang, Junbo–uWu, Chunchi*. [SI]: SSRN
- He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE international conference on computer vision*, pp 1026–1034
- Israel R, Kelly BT, Moskowitz TJ (2020) Can machines ‘learn’ finance? *J Invest Manag*
- Jensen TI, Kelly BT, Malamud S, Pedersen LH (2022) Machine learning and the implementable efficient frontier. Available at SSRN 4187217
- Kaniel R, Lin Z, Pelger M, Van Nieuwerburgh S (2022) Machine-learning the skill of mutual fund managers. Technical report. National Bureau of Economic Research
- Kelly BT, Pruitt S, Su Y (2019) Characteristics are covariances: a unified model of risk and return. *J Financ Econ* 134(3):501–524
- Kozak S, Nagel S, Santosh S (2020) Shrinking the cross-section. *J Financ Econ* 135(2):271–292
- Leippold M, Wang Q, Zhou W (2022) Machine learning in the Chinese stock market. *J Financ Econ* 145(2):64–82
- Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds), *Advances in neural information processing systems*, vol 30. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>
- Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, Liston DE, Low DK-W, Newman S-F, Kim J et al (2018) Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2(10):749–760
- Merton RC (1973) An intertemporal capital asset pricing model. *Econo J Econ Soc* 8:867–887
- Pástor L, Stambaugh RF, Taylor LA (2022) Dissecting green returns. *J Financ Econ* 146(2):403–424
- Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. *J Financ* 19(3):425–442
- Welch I, Goyal A (2008) A comprehensive look at the empirical performance of equity premium prediction. *Rev Financ Stud* 21(4):1455–1508

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
