Machine Learning Based Portfolio Analysis

Yingjie Han

School of economics and management, Nanjing University of Science and Technology, Nanjing, China

2683274103@qq.com

Abstract. This paper investigates the effectiveness of machine learning methods in predicting stock returns in China. In this paper, we use 16 liquidity, momentum, market risk and macro fundamentals indicators, as well as 112 interaction terms between indicators and macro variables with monthly stock excess returns for the lagged period from January 2000 to November 2015 as the sample set, train elasticity networks, gradient boosting, random forests, neural networks, and simple linear regression models, and based on the data of December 2015, predict the January 2016 returns, form portfolios based on the predictions and hold them for 12 months, and thereafter roll over the extended sample period to form new portfolios, and evaluate the performance of the portfolios based on out-of-sample alpha versus the estimated shap value to assess the significance of the metrics. The results show that the machine learning approach significantly outperforms the simple linear regression model and performs better in terms of weighted portfolio returns. The portfolio constructed based on the neural network model (NN1) achieves the lowest mean square error and the highest model fit, and NN3 achieves the highest out-of-sample alpha, suggesting that the model is able to achieve higher risk-adjusted returns when risk factors are taken into account. Further analysis shows that the NN3 model is able to effectively identify high-yielding stocks and control portfolio risk, making it a preferred strategy for investment portfolios in the Chinese equity market. This paper still supports Leippold's (2022) view that liquidity indicators are most important after excluding fundamental indicators.

Keywords: Machine Learning, Profit Forecast, Portfolio Analysis.

1. Introduction

China's stock market reached a record high at the end of March 2024, with a total market capitalization of 76.89 trillion yuan, making it the second largest stock market in the world after the United States. However, due to the limitations of the early capital market system, such as institutional defects and structural contradictions, the Chinese stock market has successively carried out initiatives such as the equity separation reform, the delisting system reform and the registration system reform, and gradually constructed a multi-level capital market system. Compared with mature markets, China's stock market has obvious individual investor-led characteristics. According to the Shanghai Stock Exchange 2023 Yearbook data, there are 325.2 million investors in China, of which individual investors account for the vast majority, accounting for 99.7%, a characteristic that has led to an increase in the market's trading activity, with the value of China's stock trading accounting for 208% of market capitalization in 2022. The speculative trading motives and short-term trading tendencies of individual investors may have exacerbated market volatility, leading to a decoupling from underlying economic conditions. Although Saffi and Sigurdsson (2011) point out that short selling contributes to price discovery, the Chinese stock market has strict restrictions on short selling. In addition, the Chinese stock market exhibits distinct policy market characteristics. According to Allen et al. (2005), the centralized control, bank dominance, and unique relationship-driven nature of China's financial system are among its key features. For example, China's A-shares have suspended IPOs several times, and policy adjustments inevitably affect the short- and medium-term operating trends of the market. In such a special market environment, we need to explore whether return predictability and portfolio performance are compromised, and whether technical indicators generated based on collective investment behavior are more important than solid fundamentals for asset pricing.

Asset pricing is an important topic in finance that has historically received extensive attention from scholars. From the CAPM model to Fama French's (1992) three-factor model to Carhart's (1997) four-factor model with the addition of momentum effects and Fama and French's (2015) five-factor model, these models have laid the theoretical foundation for risk premiums in asset pricing. As predictors continue to be proposed and their predictive power confirmed, we are forced to consider the plausibility and validity of extensive information sets. Green et al. (2017) find an important impact of a range of firm characteristics on average returns through Fama-Macbeth regressions. However, when confronted with high-dimensional extensive information sets, simple linear models may suffer from serious overfitting problems, and the interactions between indicators are difficult to be uncovered.

In view of the above problems, this paper draws on the study of Gu et al. (2020), which combines machine learning with asset pricing in order to explore the risk premium in the Chinese stock market for outperforming the market's favorable returns. Their results show that the expected returns are higher when using machine learning to construct investment portfolios, and the performance is much better than simple linear prediction models. Although most of the current empirical machine learning studies are keen on the US stock market, considering the characteristics of the Chinese stock market, which is characterized by a large proportion of individual investors and a policy-dominated market, this paper believes that the application of machine learning to the Chinese stock market can also achieve significant prediction results. Therefore, this paper selects four machine learning methods, namely elastic network, gradient boosting, random forest and neural network, and a simple OLS linear model to predict stock returns and compare their performance. And further research on which indicators in a large number of information sets are more important for prediction, so as to provide reference for investors.

2. Data and Methodology

This paper refers to the study of Gu et al. (2020) for indicator construction, given that their U.S.-based market study argues that fundamental indicators are unimportant for return forecasting and pricing, the paper excludes fundamental indicators, and the indicators for each variable are shown in the table below, with the source of data at the individual stock level from the CSMAR database, and the macro data from iFind, with the data period of January 2000 to December 2023, the data period is from January 2000 to December 2023.

In order to exclude the influence of contemporaneous risk factors, this paper obtains one-year treasury bill yields to calculate the excess returns, Carhart (1997) four-factor data of Chinese A-shares from the website “https://www.factorwar.com”, and Fama-French five-factor model, Daniel-Hrench five-factor model, and Daniel-Hrench five-factor model. Fama-French five-factor model, Daniel-Hirshleifer-Su three-factor model, Hou-Xue-Zhang four-factor model, Novy-Marx four-factor model, and Stambaugh-Yuan four-factor model are also obtained for the robustness test.

It is worth noting that Leippold's (2022) study for the Chinese stock market argues that the most important indicators are liquidity indicators, followed by fundamental indicators, which contrasts with Gu et al.'s (2020) view that fundamental indicators are unimportant in the U.S. market; this paper's exclusion of fundamental indicators attempts to explain that liquidity, momentum, and macro-fundamental indicators are unimportant when fundamental information about a stock is not taken into account, market risk indicators, and macro-fundamental indicators still have the same importance when stock fundamental information is not considered.

Table 1. Variable definition

Category of Variables	Category of Indicator	Acronym	Stock Characteristics	Author (s)
predicted variable	-	Mret	Monthly stock return minus monthly risk-free rate	-
Stock level predictive indicators	Liquidity	ill	Average of daily (absolute return/RMB volume) in month t	Amihud (2002)
		turn	Average monthly trading volume for month t − 3 to t − 1 scaled by number of shares outstanding in month t	Datar, Naik & Radclie (1998)
		zerotrade	Turnover weighted number of zero trading days in month t − 1	Liu (2006)
		std_dovol	Monthly standard deviation of daily RMB trading volume	Chordia, Subrahmanyam & Anshuman (2001)
		dolvol	Natural logarithm of trading volume times price per share from month t−2	Chordia, Subrahmanyam & Anshuman (2001)
		std_turn	Monthly standard deviation of daily share turnover	Chordia, Subrahmanyam & Anshuman (2001)
	Momentum	mom1m	1-month cumulative return	Jegadeesh & Titman (1993)
		mom6m	5-month cumulative returns ending one month before month end	Jegadeesh & Titman (1993)
		mom12m	11-month cumulative returns ending one month before month end	Jegadeesh (1990)
		mom36m	Cumulative returns from months t − 36 to t − 13	Jegadeesh (1990)
		maxret	Maximum daily return from returns during month t − 1	Bali, Cakici & Whitelaw (2011)
	Shareholding	top10holdrate	Percentage of common shares owned by top 10 shareholders	Gul, Kim & Qiu (2010)
		largestholdrate	Percentage of common shares owned by the largest shareholder	Gul, Kim & Qiu (2011)
	Risk	idiovol	Standard deviation of residuals of weekly returns on weekly equally-weighted market returns for three years prior to month end	Ali, Hwang & Trombley (2003)
		volatility	Standard deviation of daily	Ang, Hodrick,

			returns from month t − 1	Xing & Zhang (2006)
		beta	We estimate stock-level beta using weekly returns and value-weighted market returns for three years ending month t − 1 with at least 52 weeks of returns.	Fama & MacBeth (1973)
		svar	The stock variance is computed as sum of squared daily returns on the SSE Composite Index.	Welch (2008), Gu et al. (2020)
		ntis	The net equity expansion is the ratio of 12-month moving sums of net issues in China's A-share market divided by the total end-of-year market capitaliza-tion of A-share stocks.	Welch (2008), Gu et al. (2020)
		tms	The term spread is the difference between the yield on 10-year government bond and the 1-year government bond.	Welch (2008), Gu et al. (2020)
Macro level predictive indicators	Macro fundamental indicators	CPI	monthly consumer price index from the National Bureau of Statistics of China.	-
		m2gr	The monthly M2 growth rate (YoY) is from the National Bureau of Statistics of China.	Chen (2009)
		itgr	The international trade volume growth rate (YoY) is from the National Bureau of Statistics of China.	Rapach et al. (2013)
		mtr	The monthly turnover is the ratio of monthly trading volume measured in Chinese yuan to the average daily market value of all stocks in China's A-share market.	Baker and Stein (2004)

Machine learning methods are able to automatically learn non-linear relationships and complex patterns in data through various algorithms, which can reduce the risk of overfitting compared to simple linear models with proper regularization and cross-validation. Therefore, this paper uses the above 16 stock-level variables and 112 (7*16) interaction terms between stock-level variables and macro-level variables as predictor variables to forecast monthly excess returns. And, in this paper, to prevent forward-looking bias, 128 variables lagged one month's value with the current month's monthly excess returns are used for training to construct the model.

In this paper, four machine learning methods, namely Elastic Network (ENET), Gradient Boosting (XGB), Random Forest (RF) and Neural Network (one layer NN1 to four layers NN4), and a simple OLS linear regression model are selected to predict stock returns. For the machine learning model, the data from January 2000 to November 2015 are first selected as the training set, and 3-fold cross-validation is used, i.e., 1/3 within the training set is selected as the validation set to capture the optimal model relationship between the monthly excess returns and the variables lagged by one period, and out-of-sample return prediction is performed using December 2015 as the test set, i.e., predicting the returns for the next month, based on the model predicted returns, the top decile is selected to form a weighted and equal-weighted portfolio, and the actual returns for the twelve-month holding period are calculated. For each remaining year, an expansion window is used to expand the training set forward by one year, construct new portfolios and track their returns, and calculate an alpha assessment of out-of-sample performance excluding contemporaneous risk factors based on the factor model. For the construction of the simple OLS linear model, 128 variables with excess returns from January 2000 to November 2015 are first selected to fit the model, followed by predicting January 2016 returns using December 2015 data, constructing portfolios and holding them for twelve months, and then expanding the data window year by year to obtain a new linear fit model and re-form the portfolios.

In constructing the portfolios, this paper is long only the top decile of stocks in terms of predicted returns and does not short the bottom decile of stocks in terms of predicted returns. This is partly in reference to Leippold et al. (2022) and DeMiguel et al. (2023), and partly due to the strict shorting restrictions in the Chinese stock market.

Andrew and Jack's (2024) study on the treatment of missing values in machine learning shows that after some preprocessing of the data, the mean-value interpolation method has little effect on the returns. Therefore, this paper follows the approach of the two scholars and standardizes the data after 1% and 99% tailoring so that all predictor variables follow a normal distribution with mean 0 and standard deviation 1. After this preprocessing the missing values are filled to the mean of the variable, i.e., filled to 0. In order to mitigate the effect of outliers, reduce the bias between variables and increase the speed of convergence of the model, this paper follows Gu et al. (2020) and normalizes the predictor variables to the region of [-1, 1].

3. Portfolio Analysis

3.1. Descriptive Analysis

The correlation coefficients between the variables after normalization are shown in Figure 1. It can be found that there is a high correlation between 6 liquidity indicators, especially ill has a correlation coefficient of -0.82 with std_dovol. 5 momentum indicators have a high correlation with each other and a low correlation with the liquidity indicators. The correlation coefficient between the cumulative return of the past 1 month (mom1m) and the maximum return of the past month (maxret) reaches 0.59, and the correlation between the cumulative return of the past 12 months (mom12m) and the cumulative return of the past 6 months (mom6m) reaches 0.62. Interestingly, the volatility of the cumulative return of the past 1 month (mom1m) and the volatility of the market's return of the previous month (volatility) correlates with the volatility of the market's return of the previous month (mom1m). volatility) correlation coefficient is 0.50, and the maximum return in the past month (maxret) correlation coefficient with the volatility of last month's return (volatility) is as high as 0.91. If high volatility is regarded as a high-risk stock, following the studies of Turan G. Bali et al. (2011), (2017), whether A-share investors are more inclined to chase lottery-like stocks would be an interesting research topic.

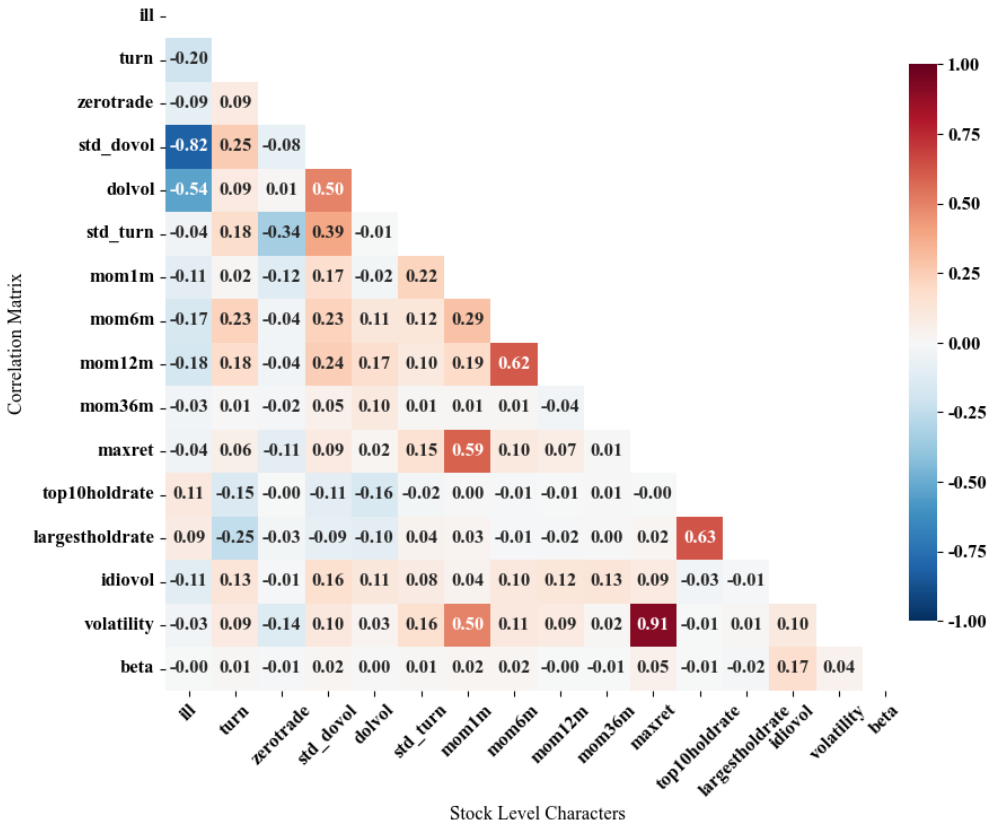


Figure 1. Matrix of correlation coefficients of variables

3.2. Portfolio Descriptive Analysis

In this paper, we consider both equal weighting and weighting approaches to constructing portfolios. For an equal-weighted portfolio, as long as the stock is in the top decile of predicted returns, it is assigned the same weight as the other top decile stocks, regardless of its predicted return. The construction of the forecast weighting refers to DeMiguel (2023), which assigns different weights to different stocks based on their forecasted returns, e.g., if the December 2016 forecast predicts that both A and B stocks will be in the top 1% in 2017 and the forecasted return is 3% for A stocks and 6% for B stocks; then A stocks are assigned a weight of 1 and B stocks are assigned a weight of 2 during the holding period. Table 2 and Table 3 report the descriptive statistics of the actual monthly excess returns, standard deviation, Sharpe ratio, and Sortino ratio of the equally weighted and weighted portfolios constructed based on the model predictions, respectively.

Table 2. Descriptive statistics for equal-weighted portfolios

	Mean	Standard Deviation	Sharpe Ratio	Sortino Ratio	Max	Max Drawdown	VaR 99%	skew	kurtosis
OLS	0.0286	0.1160	0.2462	0.9051	0.5664	-0.1676	0.5084	2.5268	8.6521
ENET	0.0113	0.0989	0.1142	0.2405	0.6287	-0.2596	0.2789	2.7024	15.9169
NN1	0.0314	0.1133	0.2776	0.9201	0.5567	-0.1586	0.4797	2.2555	7.3526
NN2	0.0130	0.0716	0.1814	0.4303	0.3387	-0.1136	0.2459	1.4676	4.4997
NN3	0.0447	0.1812	0.2467	1.1390	1.3673	-0.1748	0.7071	4.8521	31.2151
NN4	0.0131	0.0751	0.1747	0.3049	0.2588	-0.2504	0.2356	0.5983	2.6339
RF	0.0304	0.1163	0.2617	0.7580	0.5876	-0.1737	0.5652	2.4256	9.0841
XGB	0.0176	0.0963	0.1828	0.3413	0.4088	-0.2794	0.3529	1.1525	4.4534

Table 3. Descriptive statistics for value-weighted portfolios

	Mean	Standard Deviation	Sharpe Ratio	Sortino Ratio	Max	Max Drawdown	VaR 99%	skew	kurtosis
OLS	0.0296	0.1198	0.2469	0.9266	0.6235	-0.1644	0.4821	2.5713	8.9092
ENET	0.0120	0.1060	0.1136	0.2540	0.7282	-0.2629	0.2743	3.3577	21.8091
NN1	0.0381	0.1380	0.2763	1.0427	0.7332	-0.1584	0.6051	2.7118	9.6147
NN2	0.0159	0.0744	0.2142	0.5733	0.3314	-0.1034	0.2589	1.4614	3.6139
NN3	0.0522	0.2001	0.2608	1.3354	1.4277	-0.1710	0.9434	4.6446	26.9153
NN4	0.0162	0.0938	0.1728	0.3664	0.5954	-0.2582	0.2702	2.5301	15.1715
RF	0.0528	0.2144	0.2463	1.1856	1.1922	-0.1730	1.1103	3.9315	17.2093
XGB	0.0320	0.1918	0.1666	0.6101	1.5833	-0.2791	0.6204	5.9168	45.8752

From the results in the table, it can be concluded that the simple linear regression model with a variety of machine learning methods achieves positive monthly average excess returns. The model with the largest mean return in the equal weighted portfolios is the neural network model with 3 hidden layers (NN3) at 4.5% per month and it significantly outperforms the other models, while the neural network model with 1 hidden layer (NN1) and the random forest model (RF) perform second only to the NN3 model and all achieve an average return of more than 3% per month. In the weighted portfolios, the Random Forest model (RF) achieves the best performance, but the gap with the next best model, NN3, is smaller, with monthly average excess returns of 5.28% and 5.22%, respectively, and these two models are far superior to the other models. The NN3 model achieves impressive results in both the equal-weighted portfolios and the prediction-weighted portfolios, and the NN2 and NN4 models consistently perform the worst among the six machine learning models. models, even worse than the most basic OLS model, and the elastic network model (ENET) consistently performs the worst of all models. Overall, the predictive weighted combinations yielded higher returns than the equal weighted portfolios.

We consider the “value for money” of an investment from two perspectives: the Sharpe ratio and the Sortino ratio. The Sharpe ratio measures the excess return of a portfolio per unit of total risk assumed, while the Sortino ratio focuses more on the negative volatility of a portfolio, which measures the risk-adjusted return of a portfolio relative to downside risk. According to Table 1, it can be found that NN1 achieves the highest Sharpe ratio and the performance of the NN3 model, which achieves the largest monthly average return, deteriorates, with its Sharpe ratio being only the third highest. When downside risk is considered, NN3 achieves the highest Sortino ratio and NN1 achieves the second best performance, indicating that NN1 achieves a better performance when both total and downside risk are considered, with RF and OLS following closely behind. NN3 performs better in the forecast-weighted portfolio, indicating that the portfolio constructed based on the NN3 model has less volatility in general and can fully consider the downside risk, and the portfolio's loss in the downturn is relatively small, which can better protect the interests of investors.

Analyzing from the perspective of maximum negative return, XGB has the weakest ability to control the maximum retraction and NN2 is the strongest, especially under the prediction weighted portfolio, the maximum retraction of NN2 is only -10.336%, which is far better than the other models, and it is worth noting that the NN2 model always possesses the lowest return volatility.

NN3 achieves the highest VaR both in the equal-weighted and forecast-weighted portfolios, having the highest potential risk of loss; all models have positive skewness, indicating that investors are more likely to achieve above-average returns.

3.3. Model Error Measure

The mean square error MSE is obtained by calculating the average of the squares of the differences between the predicted values and the actual observed values and can be used to assess how well the model fits the given data. Smaller MSE values indicate a better fit of the model. Table 4 reports the MSE values for the different models for each year and gives the mean square error for each model under simple averaging.

Table 4. MSE values under different years for each model

	OLS	ENET	XGB	RF	NN1	NN2	NN3	NN4
2016	0.159	0.111	0.127	0.125	0.024	0.083	0.035	0.073
2017	0.017	0.018	0.019	0.013	0.016	0.065	0.043	0.119
2018	0.019	0.017	0.018	0.016	0.048	0.035	0.022	0.018
2019	0.012	0.011	0.011	0.011	0.011	0.035	0.037	0.012
2020	0.015	0.014	0.014	0.014	0.019	0.015	0.045	0.016
2021	0.024	0.02	0.021	0.021	0.021	0.031	0.024	0.019
2022	0.018	0.021	0.021	0.022	0.013	0.042	0.081	0.016
2023	0.014	0.013	0.012	0.012	0.014	0.052	0.03	0.032
Average value	0.03475	0.02813	0.03038	0.02925	0.02075	0.04475	0.03963	0.03813

In terms of means, the NN1 model has the lowest MSE, suggesting that the NN1 model may be the best fit when simply averaged. However, this does not mean that NN1 is the best choice in each individual year as the performance of the model may vary from year to year, e.g., in 2018 the model had the highest MSE. The OLS, ENET, XGB, and RF models all performed poorly in 2016, and the rest of the years were excellent, probably due to the small sample size in the early years, where the model's fit was low, and with the increase of the sample size the model fit improved. The performance of neural network models with different hidden layers varies from year to year, but in most years the NN1 model achieves lower mean square error.

3.4. Out-of-Sample Performance Based on Multifactor Modeling

To further assess the out-of-sample performance of the top decile portfolios from January 2016 to December 2023, this paper performs a time-series regression of 96 out-of-sample monthly portfolio excess returns on contemporaneous risk factors in order to assess the returns of the portfolios excluding the contemporaneous risk factors, i.e., the intercept term of the time-series regressions, alpha. This paper considers six dominant multifactor models to assess the portfolio performance: the Carhart four-factor model (Carhart4), the Fama French (2015) five-factor model (FF5), the Daniel-Hirshleifer-Sun (2020) three-factor model (DHS3), and the Hou-Xue-Zhang (2015) four-factor model (HXZ4), Novy-Marx (2013) four-factor model (NM4), and Stambaugh-Yuan (2017) four-factor model (SY4).

Table 5. Descriptive statistics for value-weighted portfolios

	FF5	Carhart4	DHS3	HXZ4	NM4	SY4
OLS	2.557**	2.807**	3.076**	3.013***	2.765**	2.722**
	(1.113)	(1.163)	(1.246)	(1.084)	(1.098)	(1.128)
ENET	1.165	1.205	2.141*	1.012	2.019*	0.822
	(0.965)	(0.955)	(1.136)	(0.730)	(1.101)	(0.622)
NN1	2.806***	3.115***	3.626***	3.217***	3.280***	3.246***
	(1.031)	(1.154)	(1.211)	(1.078)	(1.058)	(1.215)
NN2	1.156*	1.314*	2.018***	1.671**	1.725***	1.783**
	(0.616)	(0.706)	(0.716)	(0.766)	(0.629)	(0.820)
NN3	3.952**	4.500**	5.261**	4.515**	4.508**	5.246*
	(1.721)	(2.220)	(2.274)	(2.205)	(1.987)	(2.742)
NN4	1.172**	1.286**	1.690***	1.122**	1.572***	1.094**
	(0.555)	(0.537)	(0.597)	(0.530)	(0.479)	(0.469)
RF	2.641***	3.050**	3.540***	3.129**	3.159***	3.548**
	(0.961)	(1.204)	(1.205)	(1.192)	(1.077)	(1.500)
XGB	1.602**	1.753**	2.375***	1.766***	2.320***	1.683***
	(0.680)	(0.719)	(0.827)	(0.551)	(0.830)	(0.502)

Note: Standard errors in parentheses, *, **, and *** indicate significance levels of 10%, 5%, and 1%, respectively. Same as below.

Table 6. Weighted portfolio out-of-sample alpha

	FF5	Carhart4	DHS3	HXZ4	NM4	SY4
OLS	2.613**	2.896**	2.992**	3.112***	2.787***	2.840**
	(1.057)	(1.117)	(1.164)	(1.055)	(1.033)	(1.088)
ENET	1.255	1.277	2.214*	1.058	2.117*	0.844
	(1.068)	(1.051)	(1.241)	(0.799)	(1.212)	(0.669)
NN1	3.359***	3.724***	3.761***	3.912***	3.600***	3.902***
	(1.110)	(1.271)	(1.251)	(1.249)	(1.117)	(1.350)
NN2	1.413**	1.621*	2.310***	1.967**	1.967***	2.096**
	(0.701)	(0.832)	(0.866)	(0.878)	(0.746)	(0.994)
NN3	4.665**	5.217**	5.736**	5.211**	5.130**	5.865**
	(1.876)	(2.337)	(2.419)	(2.290)	(2.126)	(2.835)
NN4	1.423*	1.568**	1.641**	1.391*	1.709***	1.286**
	(0.731)	(0.750)	(0.637)	(0.770)	(0.592)	(0.643)
RF	4.507***	5.156**	4.347**	5.192**	4.435**	5.561**
	(1.585)	(2.023)	(1.944)	(2.075)	(1.728)	(2.429)
XGB	2.779*	3.096**	3.304**	3.250*	3.460**	2.963*
	(1.472)	(1.518)	(1.479)	(1.711)	(1.578)	(1.670)

When the effect of contemporaneous risk factors is excluded, it is found that the NN3 model achieves the highest out-of-sample alpha in both equally weighted portfolios and weighted portfolios. The outperformance of the NN3 model implies that the model helps the portfolios to achieve higher market excess returns, and although the model fails to achieve the smallest mean-squared error, it still shows a significant advantages. The elastic network model ENET consistently performs the worst and is not significant under most factor models, making it difficult to achieve excess returns.

The NN1 in the equal weighted portfolio takes the runner-up prize, indicating that the high fit of the model delivers significant excess returns, and the model still shines in the weighted portfolio. When shifting from the equal-weighted portfolio to the weighted portfolio, the tree model RF's performance is significantly improved, leaping to become the second best model in the weighted portfolio. Another tree model, XGB, doesn't perform as dazzlingly but also shifts to outperform the normal linear model OLS, indicating that the tree model is effective in capturing nonlinearities and interactions, and that the weighted portfolio amplifies the model's capabilities.

Overall, the weighted portfolios still achieve higher excess returns than the equal-weighted portfolios when the contemporaneous market risk factor is excluded.

3.5. Risk-Neutral Cumulative Return

To examine the persistence of the predictors and the predictive effects of the models, this paper calculates the cumulative returns of the top decile portfolios for OLS, elastic networks, random forests, gradient boosting, and the four neural network models (NN1 to NN4) for each month of the out-of-sample period from January 2016 to December 2023. The results are summarized in the following table.

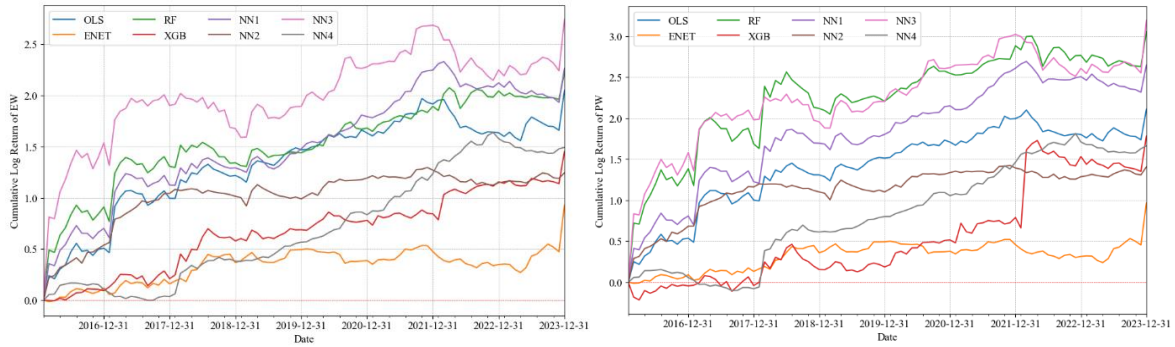


Figure 2. Cumulative return of equal-weighted and forecast-weighted portfolios

Among the equal weighted portfolios, the NN3 model achieved dazzling cumulative returns with a cumulative return of 274.05%, NN1 and RF achieved second and third place with logarithmic cumulative returns of 226.38% and 221.82%, respectively, and OLS ranked fourth with 204.83%, outperforming NN4 (149.26%), XGB (145.21%), NN2 (124.42%), while ENET was the worst performer with a cumulative gain of 92.76%.

The weighted portfolios achieved better cumulative returns than the equal-weighted portfolios, with the NN3 model achieving the highest cumulative return of 319.21%, followed by RF and NN1, with cumulative returns of 305.51% and 264.73%, respectively, and ENET remaining the worst performer, with a cumulative return of only 96.77%. The weighted portfolio gives higher weights to stocks with higher predicted returns and NN3 achieves higher returns than the equal weighted portfolio at this time and it is the best performer, suggesting that the predicted returns of NN3 can help achieve higher actual returns. When moving from the equal-weighted portfolio to the forecast-weighted portfolio, RF changes its relative performance from fourth to second place and achieves higher cumulative returns, indicating that the model's forecasted returns are effective. Interestingly, the NN4 model shows poor performance in the early stage and rapid improvement in the later stage, both in the equal-weighted and predictive-weighted portfolios, suggesting that the neural network model is able to improve its predictive ability through continuous learning, but its poor performance relative to NN1, NN2, and NN3 also indicates that too many hidden layers of the neural network model will not improve the predictive ability of the model, and even lead to a significant decrease in its ability, supporting the argument that NN1, NN2, and NN3 are effective predictors. This supports Gu et al.'s (2020) argument that the benefits of “deep” learning are limited.

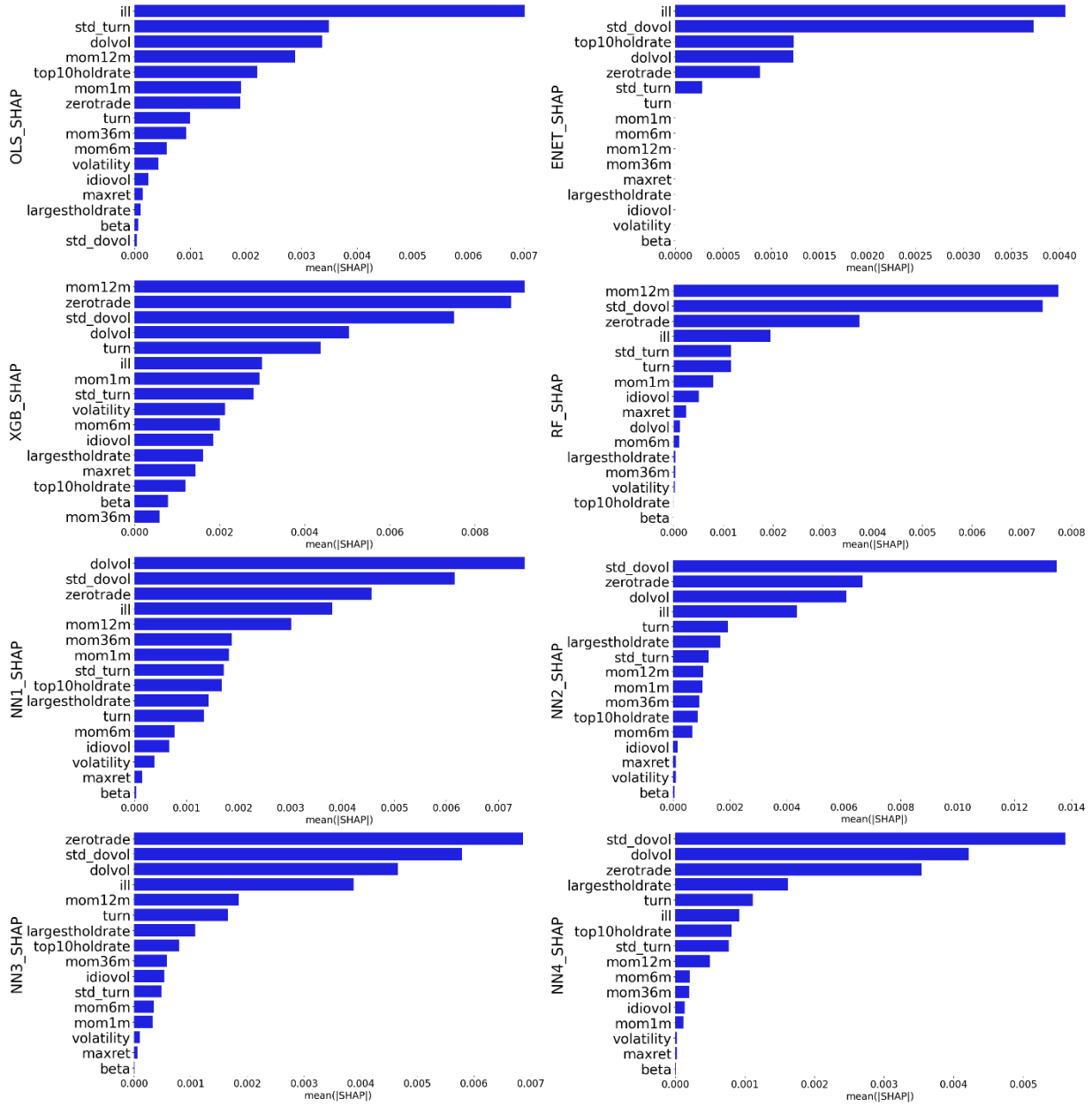
OLS performs well in both the equal-weighted and forecast-weighted portfolios, while ENET performs much worse, not only achieving the lowest cumulative returns, but also having cumulative returns of less than 1 and suffering a loss of principal. In contrast to the machine learning model, which outperforms the equal-weighted portfolio in the prediction-weighted portfolio, the linear model presents comparable cumulative returns in the prediction-weighted and equal-weighted portfolios.

4. The Importance of Variables

To investigate the importance of the predictor variables, this paper refers to DeMiguel (2023) by estimating the SHAP values (Lundberg and Lee, 2017), which is a method based on cooperative game theory that estimates the extent to which each predictor variable contributes to an individual prediction by recovering the difference between the prediction of a single observation and the average prediction of all observations. This paper assesses the importance of the predictor variables based on the average of the absolute SHAP values computed across all observations during the last estimation window (i.e., January 2016 through December 2023). In this paper, SHAP values have been calculated for each model as shown in Fig.3.

The linear models all consider liquidity indicators to be their most important variable indicators, i.e., liquidity volatility (std dolvol and std turn), zero trading days (zerotrade), and illiquidity measures (ill) are the most significant predictors, while momentum indicators and market risk indicators are not so important, and ENET even considers momentum indicators to be completely useless. This

result for linear models stems from the simplicity of their form, which manifests itself in a severe overfitting problem when dealing with high-dimensional information sets, assigning high weights to certain types of variables and ignoring relatively unimportant ones. The importance of the variables in each model shows a rapidly decreasing distribution, in contrast to the machine learning models, which are more democratic and able to utilize a wider range of metrics for prediction, supporting Gu et al.'s (2020) contention that machine learning models are able to extract predictive information from a wider range of features.



Figuree 3. SHAP values for each model

Gu et al. (2020) argue that the traditional price trend indicator, i.e., momentum, is the most influential predictor in their study for the U.S. stock market, but the importance of momentum is questionable in the case of the Chinese stock market. The NN3 model, the best performing of the machine learning models, rejects this view and argues that liquidity is the most important indicator and momentum is much less important. Only the XGB and RF models place high importance on momentum indicators and consider the long term momentum indicator mom12m to be its most important indicator, with the short term momentum indicators mom1m, mom6m and the longer term momentum indicator mom36m being less important, which is different from Gu et al. (2020) who

consider the short term indicator mom1m to be the most important of the momentum indicators, but both consider the up to 3 year price trend indicators are less important, std dolvol and zerotrade also occupy a fairly high importance. Considering the high proportion of small and medium-sized investors in the Chinese stock market and the irrational nature of their investment behavior, investors in the market may allocate most of their trades to short-term trading, which leads to the neglect of the trend.

The neural network model assigns the highest importance to the liquidity indicator, which is consistent with Leippold's (2021) view that liquidity indicators are more important in the Chinese stock market, followed by momentum, with mom12m being the most important of the momentum indicators, and the best-performing neural network models, NN1 and NN3, which both consider the long-run momentum indicator to be more important than the short-run momentum indicator, mom1m. The worst performing model, NN4, considers institutional holdings information to be the most important indicator other than the liquidity indicator, which is different from the remaining three models, and perhaps the difference in the importance of this indicator is the reason for the poor performance of this model.

In contrast to the other models that tend to place more importance on a similar set of stock level predictors, the tree-based models, XGB and RF, instead place more importance on a number of predictors, a difference that Leippold (2021) attributes to the generalization property of the tree models as they randomly select a subset of stock characteristics when constructing their decision trees. In this way, predictors like mom12m can become very influential in some decision trees and therefore more relevant to the tree model as a whole, while they play a relatively minor role in all other models.

5. Summary

Considering the special characteristics of the Chinese stock market, this paper utilizes various machine learning methods to train gradient boosting, random forest, elastic network and four hidden layer neural network models as well as the simple linear model OLS based on 128 variables with one period of lag to validate the model's forecasting ability in the Chinese stock market. Based on the empirical study, this paper finds that the three hidden-layer neural network model NN3 achieves the highest excess return but it fails to achieve the smallest mean squared error, the NN1 model achieves the smallest mean squared error with a high degree of fit, and the tree models Random Forest and Gradient Boosting also achieve a better performance, and the Elasticity Network performs the worst. For the cumulative return analysis, the predicted weighted portfolio returns are higher than the equal weighted portfolio.

Further, this paper calculates the SHAP value of each predictor variable for each model to measure the importance of each predictor variable to the predicted returns. Liquidity indicators show high importance in all models, momentum indicators show high importance only in the tree model, and market risk is relatively unimportant. This paper supports Leippold's (2022) view that liquidity is the most important indicator, market risk is relatively important, and momentum is unimportant, even after the exclusion of fundamental indicators.

References

[1] Ali A, Hwang L S, Trombley M A. Arbitrage risk and the book-to-market anomaly [J]. Journal of Financial Economics, 2003, 69 (2): 355-373.

[2] Amihud Y. Illiquidity and stock returns: cross-section and time-series effects [J]. Journal of Financial Markets, 2002, 5 (1): 31-56.

[3] Ang A, Hodrick R J, Xing Y, et al. The cross-section of volatility and expected returns [J]. The Fournal of Finance, 2006, 61 (1): 259-299.

[4] Baker M, Stein J C. Market liquidity as a sentiment indicator [J]. Journal of Financial Markets, 2004, 7 (3): 271-299.

[5] Bali T G, Brown S J, Murray S, et al. A lottery-demand-based explanation of the beta anomaly [J]. Journal of Financial and Quantitative Analysis, 2017, 52 (6): 2369-2397.

[6] Bali T G, Cakici N, Whitelaw R F. Maxing out: Stocks as lotteries and the cross-section of expected returns [J]. Journal of Financial Economics, 2011, 99 (2): 427-446.

[7] Carhart M M. On persistence in mutual fund performance [J]. The Journal of Finance, 1997, 52 (1): 57-82.

[8] Chen A Y, McCoy J. Missing values handling for machine learning portfolios [J]. Journal of Financial Economics, 2024, 155: 103815.

[9] Chordia T, Subrahmanyam A, Anshuman V R. Trading activity and expected stock returns [J]. Journal of Financial Economics, 2001, 59 (1): 3-32.

[10] Daniel K, Hirshleifer D, Sun L. Short-and long-horizon behavioral factors [J]. The Review of Financial Studies, 2020, 33 (4): 1673-1736.

[11] Datar V T, Naik N Y, Radcliffe R. Liquidity and stock returns: An alternative test [J]. Journal of Financial Markets, 1998, 1 (2): 203-219.

[12] DeMiguel V, Gil-Bazo J, Nogales F J, et al. Machine learning and fund characteristics help to select mutual funds with positive alpha [J]. Journal of Financial Economics, 2023, 150 (3): 103737.

[13] Eugene F, French K. The cross-section of expected stock returns [J]. Journal of Finance, 1992, 47 (2): 427-465.

[14] Fama E F, French K R. A five-factor asset pricing model [J]. Journal of Financial Economics, 2015, 116 (1): 1-22.

[15] Fama E F, MacBeth J D. Risk, return, and equilibrium: Empirical tests [J]. Journal of Political Economy, 1973, 81 (3): 607-636.

[16] Green J, Hand J R M, Zhang X F. The characteristics that provide independent information about average US monthly stock returns [J]. The Review of Financial Studies, 2017, 30 (12): 4389-4436.

[17] Gu S, Kelly B, Xiu D. Autoencoder asset pricing models [J]. Journal of Econometrics, 2021, 222 (1): 429-450.

[18] Gu S, Kelly B, Xiu D. Empirical asset pricing via machine learning [J]. The Review of Financial Studies, 2020, 33 (5): 2223-2273.

[19] Gul F A, Kim J B, Qiu A A. Ownership concentration, foreign shareholding, audit quality, and stock price synchronicity: Evidence from China [J]. Journal of Financial Economics, 2010, 95 (3): 425-442.

[20] Hou K, Xue C, Zhang L. Digesting anomalies: An investment approach [J]. The Review of Financial Studies, 2015, 28 (3): 650-705.

[21] Jegadeesh N, Titman S. Returns to buying winners and selling losers: Implications for stock market efficiency [J]. The Journal of Finance, 1993, 48 (1): 65-91.

[22] Leippold M, Wang Q, Zhou W. Machine learning in the Chinese stock market [J]. Journal of Financial Economics, 2022, 145 (2): 64-82.

[23] Liu W. A liquidity-augmented capital asset pricing model [J]. Journal of Financial Economics, 2006, 82 (3): 631-671.

[24] Novy-Marx R. The other side of value: The gross profitability premium [J]. Journal of Financial Economics, 2013, 108 (1): 1-28.

[25] Rapach D, Zhou G. Forecasting stock returns [M] Handbook of Economic Forecasting, 2013, 2: 328-383.

[26] Saffi P A C, Sigurdsson K. Price efficiency and short selling [J]. The Review of Financial Studies, 2011, 24 (3): 821-852.

[27] Stambaugh R F, Yuan Y. Mispricing factors [J]. The Review of Financial Studies, 2017, 30 (4): 1270-1315.

[28] Welch I. The link between fama-french time-series tests and fama-macbeth cross-sectional tests [J]. Available at SSRN 1271935, 2008.