Predicting Stock Returns based on Convolutional Neural Networks with Feature Operators

Jing-zhi Huang Peilin Liu Jun Ni Jingyi Zhou

Abstract

We propose a Convolutional Neural Network (CNN) model consisting of six feature operators, which are designed to extract valuable features from the stock trading data. We apply this method to the Chinese stock market and construct a model with significant out-of-sample predictive power for stock returns, termed as AlphaNet. To verify the model's effectiveness, its forecast for the 5-day return is back-tested as a trading signal, using the portfolio sort method. Furthermore, it is controlled by the multi-factor risk model and previous China anomalies and compared with other common machine learning methods. The portfolio position optimization is also employed to evaluate and improve its performance on real market investment. These results indicate that AlphaNet shows promise as a skillful method to predict stock returns in the cross-section.

Smeal College of Business, The Pennsylvania State University, University Park, PA 16802, USA. Email: jxh56@psu.edu.
College of Engineering, Peking University, Beijing, 100871, China. Email: 1800011081@pku.edu.cn
Changzhou Institute of Technology, Changzhou, 213032, China. Email: nij@czu.cn.
Halıcıoğlu Data Science Institute, University of California, San Diego, CA 92093, USA. Email: jiz238@ucsd.edu.

1 Introduction

The multi-factor model is one of the most important models in the field of empirical asset pricing. The model believes that the expected return of a stock is determined by the factor exposure and factor return of several efficient risk factors. The factors can generally be divided into two categories, which are accounting-based fundamental factors and trading-based technical factors. The data source of the fundamental factors is the financial indicators of the listed companies, while the technical factors are derived from the trading information of the stocks.

In the Chinese stock market, there are numerous stocks trading data, including trading price and turnover volume, which contain information about trading activity, market microstructure, and investor sentiment (Chordia et al., 2001). The update frequency of these data can reach the second level, but the noise of them is pretty large and their signal-to-noise ratio is low. As a result, it is of great significance to apply some efficient data analytics methods on them to extract valuable information. The empirical studies, such as Chan et al., 1996, Bremer and Sweeney, 1991, and Shimizu et al., 2019, indicate that specific trading-based signals, such as the momentum factor, the reversal factor, and the low-volatility factor, have the ability to predict stock returns. However, as the signals mentioned above are continuously applied to the stock market, their effectiveness and excess return from them have been declining, which is called factor crowding. Meanwhile, the correlation between the new signals and the old ones is also on the rise, so quantitative portfolio managers are starting to design more complex signals with more sophisticated logic and calculations.

Considering that the end-to-end deep learning method can directly establish a relationship between future stock trends and the historical information (Chen and He, 2018) and the trading-based signals are typically more effective in the Chinese stock market than the accounting-based signals (Hou et al., 2023), a well-designed deep learning model has the potential to be directly applied on the stock trading data to build an effective model for stock returns predicting. The forecast of the established model can be used as a trading signal, which can avoid the complicated process of manually designing factors. Recent studies, such as Gu et al., 2020 and Leippold et al., 2022, adopt machine learning techniques to combine anomaly signals. In contrast, our study focuses on designing unique deep learning models to directly extract information from raw trading data to construct anomaly signals, bypassing the traditional process of anomaly design and discovery. In particular, we design a CNN model with six feature operators to perform feature engineering on the raw trading data and improve the prediction ability and training efficiency of the model and finally we construct the model, termed as AlphaNet, through the systematic training process according to Gu et al., 2020.

The evaluation metrics show that AlphaNet has a significant prediction ability for the 5-day return. In particular, we use the model's forecast as a trading signal and analyze the performance of the factor mimicking portfolio. Specially, considering that the Chinese market has a limited history of short sales even after 2010 (Gao and Ding, 2019), while most of the studies on factor investing in US and European markets relies on long-short hedge strategies (1-5 Hedge portfolios), we also analyze long-only portfolios (Top1 portfolios). The results show that both in-sample and out-of-sample, the Top 1 portfolio achieves good annualized returns, which are 23.30% and 38.73%, respectively. The 1-5 Hedge portfolio also performs well both in-sample and out-of-sample, with annualized returns of 26.30% and 22.34% and the annualized Sharpe Ratio reaching up to 6.35 and 4.86, respectively. What's more, we also analyze the excess return against the CSI 500 Index of the five long-only portfolios and the Top 1 portfolio achieve an annualized excess return of 27.84% in-sample and 15.45% out-of-sample.

To prove AlphaNet's significance, we take the China 4-factor (CH-4) model proposed by Liu et al., 2019 as a control and the regression results indicate that return of AlphaNet is an anomaly of the CH-4 model. What's more, we replicate 22 trading-based signals according to Hou et al., 2023 and take them as a control. The regression results further indicate that the return of AlphaNet is an anomaly of previous China anomalies. Additionally, we apply several machine learning methods on the same input data, including Ordinary Least Squares (OLS), Random Forest (RF), and Artificial Neural Network (ANN) and the results indicate that AlphaNet has attained significantly better performance than the other models.

Finally, to evaluate the performance of AlphaNet on real market investment, on each trading day, we construct a long-only portfolio on the five hundred stocks with largest factor exposure and optimize the portfolio position by minimizing its tracking error with the CSI 500 Index subject to the individual weight constraint and the turnover rate constraint. Then the optimized portfolio achieves an excess annualized return of 22.88% and an excess annualized Sharpe Ratio of 3.24 against the CSI 500 Index.

2 CNN with Feature Operators

2.1 Introduction to ANN

2.1.1 Structure

ANN is developed on the basis of neuroscience and has been a hot research topic in the field of machine learning and artificial intelligence (AI) since the 1980s. From the perspective of information processing, scientists hope to simulate the behavior of real neurons, including the activation of neurons and neurotransmitter transmission, and thus the process of brain thinking by using computer technology.

ANN is a graphical model that consists of nodes (neurons) connected by some connection weights, and the set of these weights is the parameters of the model. The structure of ANN includes input layer, hidden layer, and output layer: each neuron in the input layer corresponds to a variable dimension, which is used to receive the input raw data; each neuron in the hidden layer represents the feature learned by the model in that layer. The deep ANN can be composed of multiple hidden layers and the number of hidden layers determines the depth of the model, while the output layer requires problem-specific analysis: for classification problems the outputs are a finite number of discrete variables while continuous variables are for regression problems. Each neuron node in an ANN has a specific output function called the activation function. The artificial neuron imitates the behavior of real neuron cells during information transmission through the activation function such as RELU and Sigmoid, which is defined as:

$$\mathrm{ReLU}(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{1}$$

$$\mathrm{Sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{2}$$

The neuronal cells secrete neurotransmitters when they receive a strong enough electrical stimulus, while the artificial neurons act only when the input reaches a certain level. Actually, the activation function introduces nonlinearity to the model and enables it to learn nonlinear features and approximate highly nonlinear functions (Hornik et al., 1989), which equips deep learning models with the potential to solve the highly nonlinear problems in the financial markets.

2.2.2 Back Propagation Algorithm

Since the structure of ANN is complicated and gradient of the its loss function with respect to its connection weights is always not available, how to effectively train the parameters has been the focus of AI research and scientists have proposed some effective algorithms to solve this problem, among which the most famous one is the Back Propagation (BP) algorithm proposed by Rumelhart et al., 1986. Along the computational graph, this algorithm firstly calculates the model's loss in the forward propagation process and then calculates the model's gradient in the back propagation process, according to the Chain Rule of the composite function derivative. To be more specific, it calculates the partial derivative of the output layer neurons to the hidden layer neurons and the partial derivative of the hidden layer neurons to the input layer neurons. Then the parameters of the model (including connection weights and biases) are optimized using some optimization algorithms and finally, the optimal solution is found through continuous iterations. Figure 1 is the sketch map of a simple ANN and ω_{ij}^{k} is the weight of the j^{th} neuron in the $(k-1)^{th}$ layer connected to the i^{th} neuron in the k^{th} layer and b_i^k is the bias of the i^{th} neuron of the k^{th} layer. The computation process of the BP algorithm is shown as follows,

z_i^k is the input of the i^{th} neuron of the k^{th} layer, which is calculated as:

$$z_i^k = \sum_j \omega_{ij}^k a_j^{k-1} + b_i^k \tag{3}$$

where a_i^k is the output of the i^{th} neuron in the k^{th} layer, which is calculated as:

$$a_i^k = \sigma(\sum_j \omega_{ij}^k a_j^{k-1} + b_i^k) = \sigma(z_i^k) \tag{4}$$

where σ is the activation function. Defining δ_j^k as the gradient of the loss function E with respect to z_j^k, we have:

$$\delta_j^k = \frac{\partial E}{\partial z_j^k} = \sum_i \frac{\partial E}{\partial z_j^{k+1}} \frac{\partial z_i^{k+1}}{\partial z_j^k} = \sum_i \frac{\partial E}{\partial z_j^{k+1}} \frac{\partial z_i^{k+1}}{\partial \alpha_j^k} \frac{\partial a_j^k}{\partial z_j^k} = \frac{\partial \sigma(z_j^k)}{\partial z_j^k} \sum_i \delta_j^{k+1} \omega_{ij}^{k+1} \tag{5}$$

Hence, the gradient of E with respect to ω_{ij}^k is:

$$\frac{\partial E}{\partial \omega_{ij}^k} = \frac{\partial E}{\partial z_i^k} \frac{\partial z_i^k}{\omega_{ij}^k} = \delta_j^k a_j^{k-1} \tag{6}$$

Similarly, the gradient of E with respect to b_i^k is：

$$\frac{\partial E}{\partial b_i^k} = \frac{\partial E}{\partial z_i^k} \frac{\partial z_i^k}{b_i^k} = \delta_j^k \tag{7}$$

[Insert Figure 1 Here]

Based on this algorithm, we can transform the process of computing the gradient of the loss function with respect to the connection weight and the bias into the process of computing the gradient of the loss function with respect to the input and the output of the neuron, which in turn allows us to continuously compute the gradient and update the parameters.

2.2 Architecture of CNN

CNN has been widely used in the field of Computer Vision (CV), mainly for solving image processing and recognition problems. The basic structure and principle of the CNN are similar to those of the ANN, but the difference is that the input of the CNN is usually an image or a matrix and we can apply convolution operation on it to extract spatial features.

2.2.1 Convolution Operation

Assume there are two functions f(x) and g(x), which are integrable, then the convolution of them is defined as:

$$f(x) * g(x) = \int_{-\infty}^{+\infty} f(\tau)g(x-\tau)d\tau \tag{8}$$

Assume there are two sequences $\{x\}$ and $\{y\}$, which are infinite, then the convolution of x(n) and y(n) is defined as:

$$x(n) * y(n) = \sum_{i \in R} x(i)y(n-i) \tag{9}$$

To be concluded, convolution operation implements a mapping from a function (or a sequence) to another function (or a sequence).

2.2.2 Convolution Layer

The function of the convolutional layer is to extract features from the input data by convolution operation, which is equivalent to systematic feature engineering and its core is the convolution kernel, also named a filter. The filter is usually a matrix of small size with optimizable weights and bias terms, and its function is to convolve a small region of the input image (or matrix) called the receptive field, i.e., the pixels (or values) in this small region are weighted and averaged according to the parameters of the filter. Figure 2 is an example of how the filter works.

[Insert Figure 2 Here]

The convolution operation has the properties of localization and translation invariance, and the feature map obtained after it will also undergo a corresponding translation transformation, which is suitable for the extraction of spatial features of images.

2.2.3 Weight Sharing

Weight sharing is an important mechanism of the CNN, which refers that the filter takes the same weight on each receptive field each time it moves over the input image. The mechanism is based on the basic assumption that if a filter used to extract a certain spatial feature of an image (or matrix) is useful when acting on a certain spatial location (x_1, x_1+a)×(y_1, y_1+a), then it is also expected to be useful when acting on another spatial location (x_2, x_2+a)×(y_2, y_2+a).

2.3 Data Description

As is shown in Table 1, on each trading day, 9 kinds of daily frequency stock trading data of the whole Chinese market stocks in the past 30 days are used as the independent variable, while the returns from holding these stocks for the next 5 days are used as the dependent variable. The time range of the data is from Feb. 25, 2015, to Nov. 19, 2020, with a total of 1400 trading days. At the first trading day, the amount of stock is 2590, while it is 4034 at the last trading day.

[Insert Table 1 Here]

Then the price and volume data are combined into a 9×30 matrix and used as the input of AlphaNet, while the 5-day return is used as the output. At each trading day t, the formulae for the return, turnover, free_turnover, and 5-day return are as follows,

$$\mathrm{return}_t = \frac{\mathrm{close}_t}{\mathrm{close}_{t-1}} - 1 \tag{10}$$

$$\mathrm{turnover}_t = \frac{\mathrm{close}_t \times \mathrm{volume}_t}{\mathrm{total\ share}_t} \tag{11}$$

$$\mathrm{free\_turnover}_t = \frac{\mathrm{close}_t \times \mathrm{volume}_t}{\mathrm{free\ float\ share}_t} \tag{12}$$

$$5 - \mathrm{day\ return}_t = \frac{\mathrm{close}_{t+5}}{\mathrm{close}_t} - 1 \tag{13}$$

2.4 Architecture of AlphaNet

Considering that the covariance matrix (Sharpe, 1964), correlation matrix, moving average (Huang and Huang, 2020), historical volatility (Shimizu et al., 2019), and rate of change (Chan et al., 1996) are all important time-series characteristic of stock and other securities, six kinds of feature operators, including Cov(X, Y, d), Corr(X, Y, d), MA(X, d), Std(X, d), Zscore(X, d), and RoC(X, d), are designed to extract features from the input matrix and their definition is shown in Table 2 and d is an adjustable hyperparameter. Figure 3 shows the working mechanism of the feature operators. Every feature operator takes a permutation approach to traverse the feature dimension of the input matrix, while the time dimension is traversed according to the time step (an adjustable parameter). In our model, we take d and the time step both equal to 10. When all the operators are completed, the matrices obtained by each operator are stitched together vertically to obtain the complete feature map.

[Insert Table 2 Here]

[Insert Figure 3 Here]

The complete architecture of AlphaNet is shown in Figure 4. To avoid the overfitting and increase the convergence speed of the model, the Batch Normalization (BN) layer (Ioffe and Szegedy, 2015) is added after the feature extraction layer consisting of 6 feature operators. After the BN layer is the convolutional layer and its role mainly includes three points: (1) continuing to extract valuable features (Chen and He, 2018); (2) up-dimensioning the data through the channels of the convolution kernel, i.e., transforming large-size low-dimensional features into small-size high-dimensional features (Szegedy et al., 2014); (3) adjusting the judgment of the model on which part of the time series is more valuable for predicting the future stock returns through the weights of the convolution kernel and its weight sharing mechanism. The convolutional layer is followed by three fully-connected hidden layers and finally the output layer.

[Insert Figure 4 Here]

2.4 Model Training

Our procedure from training, to model tuning, and finally to prediction follows the basic procedure outlined by Gu et al., 2020. Each training step selects 1000 trading days as the sample space, dividing the train set, validation set, and test set according to the ratio of 8:1:1, and the model is iterated every 100 trading days. The training step minimizes the Mean Square Error (MSE), which is the standard objective function for regression problems. It is defined as:

$$\mathrm{MSE}(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{14}$$

where $\hat{y}$ is the output from the final step in the model and N is the amount of sample. Then according to the BP algorithm, we can get the gradient of MSE with respect to the parameters of the model and train the parameters.

We apply the Xavier initializer for weights in each layer (Glorot and Bengio, 2010). Loss function optimization uses the Adam algorithm (Kingma and J Ba, 2015) with an initial learning rate of 1×10^{-6} and batch size of 1000. We apply 50% dropout (Srivastava et al., 2014) to all the hidden layers to enhance the generalization of the model and avoid overfitting. Finally, we use early stopping to stop the training process once the loss function of the validation sample fails to improve for two consecutive epochs, which can be generally described as the patience is 2. Because the optimization procedure is stochastic, we independently re-train the model five times and average their forecasts (following Gu et al., 2020).

3 Model Validation

3.1 Factor Analysis Theory

3.1.1 Evaluation Metrics

In the field of quantitative investment, Information Coefficient (IC), Rank Information Coefficient (Rank IC), Information Rate (IR), and win rate are all important evaluation metrics to evaluate a single factor's effectiveness.

IC_t is the Pearson correlation coefficient between the vector of factor values (f_t) and the vector of future stock returns (R_{t+1}) in the cross-section, which is defined as:

$$IC_t = \frac{\sum_{i=1}^{n}(f_{t,i}-\bar{f_t})(R_{t+1,i}-\overline{R_{t+1}})}{\sqrt{\sum_{i=1}^{n}(f_{t,i}-\bar{f_t})^2}\sqrt{\sum_{i=1}^{n}(R_{t+1,i}-\overline{R_{t+1}})^2}} \tag{15}$$

where n is the amount of stock at time t. Rank IC_t is the Spearman's rank correlation coefficient between f_t and R_{t+1}. Basically, we obtain rank_f_t and rank_R_{t+1} by arranging f_t and R_{t+1} in ascending (or descending) order and then calculate the Pearson's correlation coefficient between them. IR is the ratio of the mean of the stock return to its standard deviation, which can be approximated by the ratio of the multi-period mean of IC to its standard deviation. The formula for IR is shown as follows:

$$IR = \frac{\frac{1}{T}\sum_{t=1}^{T} IC_t}{\sqrt{\frac{\sum_{t=1}^{T}(IC_i-\overline{IC})^2}{T-1}}} \tag{16}$$

where T is the length of the IC time series. The win rate generally refers to the ratio of the terms greater than zero in the time series of IC to the length of the time series.

A larger absolute value of IC means that the factor has a better ability on stock returns forecast. In general, a factor with an average IC greater than 0.04 can be considered to be effective. Rank IC effectively avoids the problem of possible distortion of IC, which happens when the difference between the factor values is too large. IR takes into account the stability of a factor's stock picking ability and stock selection ability and represents the factor's ability to obtain stable returns and the win rate can also measure the stability of the factor.

3.1.2 Portfolio Sort Method

Basically, the evaluation metrics in above are effective but do not fully represent the performance of the factors, portfolio sort method, as a result, are adopted to construct simulated factor mimicking portfolios to observe the real stock selection ability of the factors. In general, we sort the stocks into equal-weighted quintile portfolios. The specific steps are as follows: (1) in each period, divide all the stocks into five groups according to the sequence of the factor values; (2) construct five long-only portfolios, termed as Top 1 to Top 5, and a long-short hedge portfolio by longing Top 1 and shorting Top 5 stocks at the same time, termed as 1-5 Hedge; (3) obtain the net value curve and related evaluation metrics of the simulated portfolios by adjusting the positions at each rebalance day according to the latest factor values.

3.2 Forecast of Stock returns

The forecast of AlphaNet for the 5-day return is used as a trading signal and the evaluation metrics and the results of the z-test are shown in Table 3. These results indicate that AlphaNet performs well both in-sample and out-of-sample, which indicates that the model has excellent generalization.

[Insert Table 3 Here]

In the following contents, we will be committed to analyzing the performance of AlphaNet. Both of the in-sample period (from Feb. 25, 2015, to Oct. 30, 2018) and the out-of-sample period (from Oct. 31, 2018, to Nov. 19, 2020) are used in the back-testing process in order to evaluate its in-sample and the out-of-sample performance. The portfolios are equal-weighted, the rebalance period is one day and the net value is calculated at the adjusted closing price using the compounded interest pattern, without considering transaction costs. Figure 5 plots the net value curves of the Top 1 to Top 5 and the 1-5 Hedge portfolios. It can be seen that the difference between the tendency of the net value curves of the five long-only portfolios is significant and the net value curve of the 1-5 Hedge portfolio grows steadily, which indicates that the structure of the AlphaNet is valid and indeed possible to forecast future stock returns.

[Insert Figure 5 Here]

The relevant metrics of the portfolios, including return, annualized return, annualized volatility, annualized Sharpe Ratio, accumulated turnover rate, and maximum drawdown during the in-sample period and the out-of-sample period are shown in Table 4. And we calculate the accumulated turnover rate as:

$$\text{Accumulated turnover} = \sum_{i=1}^{T}\left(\sum_{i}\left|w_{i,t+1} - \frac{w_{i,t}(1+r_{i,t+1})}{1+\sum_{j} w_{j,t} r_{j,t+1}}\right|\right) \tag{17}$$

where T is the number of rebalance days, $r_{i,t+1}$ is the return of stock i at time t+1, and $w_{i,t}$ is the portfolio weight of stock i at time t.

[Insert Table 4 Here]

Table 5 shows the performance of the Top 1 and the 1-5 Hedge Portfolio each year. As can be seen, the Top 1 portfolio suffers losses in 2017 and 2018, but the 1-5 Hedge portfolio still shows significant annualized returns (15.12%, 20.12%) and annualized Sharpe Ratio (4.70, 6.24) at the same time, indicating that the broader market declined during this period, but the factor can still distinguish the trend of the relative stocks for better or worse. In addition, the annualized return and the annualized Sharpe Ratio of the 1-5 Hedge portfolio differ significantly across years, suggesting that the effectiveness of the model varies over time.

[Insert Table 5 Here]

Since the Chinese stock market cannot be shorted, we give the excess net value curves (Figure 6) and relevant metrics (Table 6) of the five long-only portfolios against the CSI 500 index. It can be seen that the Top 1 and the Top 2 portfolio can steadily outperform the broader market and both show good excess returns, while the Top 4 and Top 5 portfolios lose to the broader market.

[Insert Figure 6 Here]

[Insert Table 6 Here]

The metrics of the excess net value of Top 1 in each year are shown in Table 7. The Top 1 portfolio can earn an excess return most of the time except in 2017, which indicates again that AlphaNet has the ability to outperform the broader market but its effectiveness varies over time.

[Insert Table 7 Here]

3.3 Control of Risk Model

To test and verify the contribution of AlphaNet to risk models, we take the CH-4 model as a control, which is defined as:

$$E(R_i) - R_f = \beta_{MKT,i} MKT + \beta_{SMB,i} SMB + \beta_{VMG,i} VMG + \beta_{PMO,i} PMO \tag{18}$$

where R_i is the return of asset i, R_f is the risk-free return, MKT is the factor return of the market factor (the excess return on the market), SMB is the factor return of the size factor (small-minus-big), VMG is the factor return of the value factor (value-minus-growth), PMO is the factor return of the turnover factor (pessimistic-minus-Optimistic), and $\beta_{MKT,i}$, $\beta_{SMB,i}$, $\beta_{VMG,i}$, and $\beta_{PMO,i}$ are the factor exposure of asset i to the above four factors, respectively. If there exists an anomaly which cannot be explained by MKT, SMB, VMG, and PMO, then the above equation can be rewritten as:

$$R = \alpha + \beta_{MKT} MKT + \beta_{SMB} SMB + \beta_{VMG} VMG + \beta_{PMO} PMO + \epsilon \tag{19}$$

where R is the anomaly's long-short return.

To test whether the return of AlphaNet is still significant after controlling the above four risk factors, we regress the return of AlphaNet on the return of the four risk factors by performing OLS regression and the standard errors are under Newey-West adjustment (Newey and West, 1987) and White adjustment (White, 1980). The results are shown in Table 8, respectively.

[Insert Table 8 Here]

It can be seen that the intercept is significant with a significance level equal to 0.001, so it can be proved that AlphaNet is indeed an anomaly which cannot be explained by the CH-4 model. In Figure 7, we decompose the return of AlphaNet according to the results of the OLS regression and it can be seen that the anomaly is significant.

[Insert Figure 7 Here]

3.4 Control of China Anomalies

To further test and verify the contribution of AlphaNet to the Chinese stock market, we replicate the China anomalies by Hou et al., 2023. More specifically, we select 22 trading-based signals that can be constructed directly based on the raw data used by AlphaNet to ensure the comparability and the details on the construction of the 22 signals are provided in Online Appendix A. Then we regress the return of AlphaNet on the return of the China anomalies by performing OLS regression and the standard errors are under Newey-West adjustment and White adjustment. The results are shown in Table 9, respectively.

[Insert Table 9 Here]

It can be seen that the intercept is significant with a significance level equal to 0.001, indicating that AlphaNet is an anomaly which cannot be explained by the previous China anomalies. In Figure 8, we decompose the return of AlphaNet according to the results of the OLS regression and it can be seen that the anomaly is significant.

[Insert Figure 8 Here]

What's more, considering that there exists correlation among the China anomalies, which might lead to strong multicollinearity and inaccurate t-statistics estimation, we calculate the average cross-sectional correlation among AlphaNet and the 22 China anomalies and the time-series correlation among the return of them, as shown in Figure 9, respectively.

[Insert Figure 9 Here]

We select 11 anomalies with the smallest correlation, including size1, turn1, cvturn1, dtv1, isc1, isch3_1, ts1, cs1, betaDM1, R1, and pps1, and regress the return of AlphaNet on the return of the selected anomalies. The results are shown in Table 10. It can be seen that the intercept is still significant with a significance level equal to 0.001, which further proved the significance of AlphaNet and its contribution to the Chinese stock market.

[Insert Table 10 Here]

3.4 Comparison with Other Machine Learning Methods

To verify the effectiveness of AlphaNet's unique design, we apply several other machine learning methods on the same input data, including OLS, RF, and ANN. In all these models, the former 900 trading days are used as the train set and latter 500 trading days are used as the test set, the 9×30 input matrix is transformed into a 1×270 vector. We standardize the in-sample data as a whole while standardizing the daily data in the out-of-sample period, which is in line with the principle of prediction. In the RF model, the last fifth of the train set is selected as the validation set and the Grid Search method is used to choose the best hyperparameters. The ANN model contains two hidden layers with 128 and 30 neurons, and the activation function is ReLU and Sigmoid, respectively. The training method of the three machine learning models are the same as that of AlphaNet to ensure the comparability. More specifically, we divide the train set, validation set, and test set by 8:1:1, and the model is iterated every 100 trading days. The training step minimizes the MSE and the loss function optimization uses the Adam algorithm with an initial learning rate of 1×10^{-5} and batch size of 1000. We apply 50% dropout and the patience of early stopping is 2.

Each model's forecast is used as a trading signal and their evaluation metrics are compared with AlphaNet in Table 11. It can be preliminarily seen from the evaluation metrics that the OLS model performs poorly both in-sample and out-of-sample, and the RF model performs well in-sample but poorly out-of-sample, i.e., it cannot achieve any generalization. Among the three benchmarks, the ANN model shows the best predictive power and generalization, which is consistent with Leippold et al., 2022. What's more, AlphaNet is even significantly better than the ANN model, indicating that its unique design is effective. It can be concluded that among the four models, not only does AlphaNet have the best out-of-sample predictive power, but it also attains the best generalization.

[Insert Table 11 Here]

To further compare the model's performance, for each model we simulate a Top 1 portfolio and a 1-5 Hedge portfolio according to the portfolio sort method. The portfolios are equal-weighted, the rebalance period is one trading day, and the net value is calculated with the adjusted closing price using the compounded interest pattern, without considering transaction costs. Their net value curves and relevant evaluation metrics are shown in Figure 10 and Table 12, respectively.

[Insert Figure 10 Here]

[Insert Table 12 Here]

It can be seen that both AlphaNet's long-only return and its long-short hedge return significantly outperform other machine learning methods and it also achieves the best out-of-sample annualized Sharpe Ratio. What's more, its turnover rate is limited, which will help reduce transaction costs in real market investment. These results indicate that out-of-sample predictive power of AlphaNet is significantly better than the other machine learning models and its unique design is pretty effective.

4 Performance on Real Market Investment

In the above process of simulating the portfolios, we adopt the equal-weighted strategy, however, in the real market investment, investors usually use some methods to optimize the position of the assets. Therefore, we choose the out-of-sample period to simulate a long-only portfolio, apply the linear rank method on the factor, and perform the portfolio position optimization process by minimizing the exponential decay tracking error (Gaivoronski et al., 2005). The optimization procedure is subject to several constraints, including the individual weight constraint, the long-only constraint, the total weight constraint, and the turnover rate constraint. Then the mathematic formulation is as follows:

$$\begin{aligned}
\text{minimize} \quad & \text{Tracking Error}(\mathrm{w_t}) \\
\text{subject to} \quad & \text{Turnover}(\mathrm{w_t}, \mathrm{w_{t-1}}) \leq 0.1 \\
& 0 \leq \mathrm{w_{t,i}} \leq 0.05 \quad \text{for } \mathrm{i} = 1,2,\ldots,\mathrm{N} \\
& \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{w_{t,i}} = 1
\end{aligned}$$

where w_t is the vector of stock weights in the cross-section at time t, and N is the amount of stock in the portfolio each day. We select the parameters of this optimization process based on the in-sample data and finally let N to be 300, i.e., on each trading day we construct the portfolio with the 300 largest factor exposure stocks, the estimation window of the tracking error to be 50 trading days, and the exponential decay coefficient of the tracking error to be 0.99. Then the curves of the net value and the excess net value of the long-only portfolio against the CSI 500 Index are shown in Figure 11, while the evaluation metrics are shown in Table 13. These results show that the position optimization process can bring more stable excess return to the portfolio, which is valuable for real market investment.

[Insert Figure 11 Here]

[Insert Table 13 Here]

5 Conclusion

In this article, we use the historical price and volume data as the independent variable and the future 5-day return as the dependent variable to train a CNN model consisting of six feature operators, termed as AlphaNet. To evaluate the model, we use the forecast of the model as a trading signal to simulate several factor mimicking portfolios and analyze their performance. What’s more, we analyze the contribution of AlphaNet to previous risk models and the Chinese stock market by performing OLS regression on the CH-4 model and previous China anomalies, respectively. Then we compare AlphaNet with three other machine learning methods based on the same data and the same training process. The results show that AlphaNet has obtained excellent effectiveness and robustness by properly mining the raw trading data. Additionally, to evaluate and improve the performance of AlphaNet on real market investment, we perform the portfolio position optimization on the out-of-sample period and the results indicate that the model can be applied to real market investment and achieve significantly positive return and Sharpe Ratio, which apparently outperform the market index.

The model we propose has been shown effective in generating trading signals from daily trading data and it can also be used in a higher-frequency situation and other technical data, for example, using the price and volume data in the past 30 minutes to forecast the 5-minute return or generating trading signals by modelling the original order book data. In light of this, our findings highlight feature engineering and deep learning as a future research direction with great potential to forecast stock returns and conduct investment strategies. However, since the data and methods for building models are still in their infancy, there is still a lot of space for improvement. With the development of big data and deep learning technology, the model will be continuously tested and improved through real market investment and will become more and more effective.

References

Amihud, Yakov. 2002. Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets 5(1):31–56.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam. 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. Journal of Financial Economics 49(3):345–373.

Chan, Louis K.C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies. The Journal of Finance 51(5):1681–1713.

Chordia, Tarun, Avanidhar Subrahmanyam, and V. Ravi Anshuman, 2001, Trading activity and expected stock returns, Journal of Financial Economics 59(1):3–32.

Datar, Vinay T., Narayan Y. Naik, and Robert Radcliffe. 1998. Liquidity and stock returns: An alternative test. Journal of Financial Markets 1(2):203–219.

Diederik P. Kingma, Jimmy Ba 2015. Adam: method for stochastic optimization. International Conference for Learning Representations (ICLR) 5.

Gaivoronski Alexei A, Krylov S, and Nico van der Wijst. 2005. Optimal portfolio selection and dynamic benchmark tracking. European Journal of Operational Research 163(1):115–131.

Gao, K., Ding, M., 2019. Short-sale refinancing and price adjustment speed to bad news: evidence from a quasi-natural experiment in China. China Journal of Accounting Research 12(4), 379–394.

Glorot, Xavier, and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the thirteenth international conference on artificial intelligence and statistics, 249–256.

Ioffe, Sergey, and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

Jianan Liu, Robert F. Stambaugh, and Yu Yuan. 2019. Size and value in China. Journal of Financial Economics 134(1):48–69.

Jing-Zhi Huang, Zhijian (James) Huang. 2020. Testing moving average trading strategies on ETFs. Journal of Empirical Finance 57:16–32.

Kewei Hou, Fang Qiao, and Xiaoyan Zhang. 2023. Finding Anomalies in China.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2(5):359–366.

Marc Bremer and Richard J. Sweeney. 1991. The Reversal of Large Stock-Price Decreases. The Journal of Finance 46(2): 747-754.

Markus Leippolda, Qian Wanga, and Wenyu Zhou. 2022. Machine learning in the Chinese stock market. Journal of Financial Economics 145(2):64–82.

Newey, W. K., West, K. D. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and

Autocorrelation Consistent Covariance Matrix. Econometrica 55(3):703–708.

Rumelhart, D., Hinton, G., and Williams, R. 1986. Learning representations by back-propagating errors. Nature 323(6088):533–536.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. The journal of finance 19(3):425–442.

Sheng Chen and Hongxiang He. 2018. Stock Prediction Using Convolutional Neural Network. IOP Conf. Series: Materials Science and Engineering 435 012026.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. 2020. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies 33(5):2223–2273.

Shimizu, H., Shiohama, T. 2019. Multifactor Portfolio Construction by Factor Risk Parity Strategies: An Empirical Comparison of Global Stock Markets. Asia-Pac Finance Markets 26(4):453–477.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15(56):1929–1958.

Szegedy C, Liu W, Jia Y, et al. 2015. Going deeper with convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 1–9.

Weimin Liu. A liquidity-augmented capital asset pricing model. 2006. Journal of Financial Economics 82(3):631–671

White, H. 1980. A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. Econometrica 48(4):817–838.

Table 1: Data Description

The 9 kinds of input data contain the basic market conditions of stocks and are easy to obtain, which imply the microstructure of the stock market and the information about future stock returns. The output data the 5-day stock returns.

Data	Abbreviation
Adjusted opening price	open
Adjusted closing price	close
Adjusted high price	high
Adjusted low price	low
Adjusted average price (volume weighted)	vwap
Adjusted trading volume	volume
Daily return	return
Total share capital turnover rate	turnover
Free float share capital turnover rate	free_turnover
Return in the next 5 days	5-day return

Table 2: Definition of The Feature Operators

The 6 kinds of feature operators can extract time series features from the input matrix. For example, let X to be the time series of adjusted closing price and Y to be the time series of adjusted volume, then we can get the price-volume correlation, which is a classic signal in the field of quantitative investment, by using the Corr operator. What's more, if we apply the Std operator on X, we will get a kind of volatility-family signal.

Name	Definition
Cov(X, Y, d)	The covariance between time series X and time series Y for the past d days
Corr(X, Y, d)	The Pearson correlation coefficient between time series X and time series Y for the past d days
MA(X, d)	The mean of time series X for the past d days
Std(X, d)	The standard deviation of time series X for the past d days
Zscore(X, d)	The mean of time series X divided by the standard deviation for the past d days
RoC(X, d)	The rate of change of time series X for the past d days

Table 3: Evaluation Metrics of The Trading Signal from AlphaNet

Generally, a signal is considered valid if its IC exceeds 0.04, therefore, we apply the z-test on the series of IC and to show whether its average is more than 0.04. The null and alternative hypotheses are as follows:

$$H_0: \mu \leq \mu_0$$

$$H_1: \mu > \mu_0$$

where μ_0 is 0.04. Assume that the population conforms to the normal distribution assumption, the z-statistic and its distribution are as follows:

$$z = \frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma}$$

$$z \sim N\left(\frac{\sqrt{n}(\mu - \mu_0)}{\sigma}, 1\right)$$

where n is the amount of sample. Then according to the results of the z-test, whether in-sample or out-of-sample, we can reject the null hypothesis with 99% confidence, i.e., the mean of IC is greater than 0.04

	In-sample	Out-of-sample
Mean IC	0.064	0.055
(z-statistic, p-value)	(11.242, 0.000)	(4.724, 0.000)
Mean Rank IC	0.084	0.079
IR	0.80	0.56
Win Rate	78.4%	73.4%

Table 4: Performance of The Top 1 to Top 5 and 1-5 Hedge Portfolios

This table shows that from Top1 portfolio to Top 5 portfolio, there exists a clear downward trend in the annualized return and annualized Sharpe Ratio both in-sample and out-of-sample and the 1-5 Hedge portfolio has a in-sample annualized Sharpe Ratio of 6.35 and an out-of-sample annualized Sharpe Ratio of 4.86, indicating the significant predictive power and the monotonicity of the signal. Incidentally, the five long-only portfolios perform better out-of-sample than in-sample, which is attributed to the fact that the Chinese stock market is better during the out-of-sample period.

Panel A. In-sample performance of the Top 1 to Top 5 and 1-5 hedge portfolios

Portfolio	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
Top 1	93.77%	23.30%	32.33%	0.72	165.39	−48.33%
Top 2	31.36%	12.50%	31.67%	0.39	302.04	−49.17%
Top 3	−2.98%	4.03%	30.94%	0.13	315.80	−54.95%
Top 4	−43.61%	−10.32%	32.21%	−0.32	281.80	−71.96%
Top 5	−72.43%	−29.23%	33.88%	−0.86	144.12	−84.93%
1-5 Hedge	161.77%	26.30%	4.14%	6.35	152.83	−3.29%

Panel B. Out-of-sample performance of the Top 1 to Top 5 and 1-5 hedge portfolios

Portfolio	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
Top 1	108.79%	38.73%	22.90%	1.69	102.76	−16.39%
Top 2	105.59%	38.05%	23.20%	1.64	166.23	−15.16%
Top 3	82.53%	32.39%	23.95%	1.35	175.09	−16.52%
Top 4	47.44%	22.22%	25.16%	0.88	151.11	−21.83%
Top 5	−18.69%	−6.27%	27.71%	−0.23	74.18	−38.88%
1-5 Hedge	57.40%	22.34%	4.60%	4.86	87.76	−7.03%

Table 5: Performance of The Top 1 and 1-5 Hedge Portfolios Each Year

This table shows that the signal also performs well in each year. In general, the performance of the 1-5 Hedge portfolio, including the annualized return and the annualized Sharpe Ratio, has decreased over time, which may be due to the increasing effectiveness of the Chinese stock market and the information decay of the underlying trading data we use.

Panel A. Performance of the Top 1 portfolio each year

In-sample						
Year	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
2015	144.64%	115.29%	48.16%	2.39	51.98	−48.33%
2016	10.68%	15.28%	31.65%	0.48	44.92	−31.14%
2017	−4.82%	−3.63%	16.24%	−0.22	37.60	−17.93%
2018	−24.81%	−31.55%	25.80%	-1.22	30.89	−31.59%
Out-of-sample						
Year	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
2018	7.79%	45.20%	22.19%	2.04	8.11	−5.58%
2019	48.49%	42.13%	21.96%	1.92	49.92	−16.39%
2020	30.44%	33.53%	24.16%	1.39	44.74	−13.75%

Panel B. Performance of the 1-5 Hedge portfolio each year

In-sample						
Year	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
2015	50.28%	47.31%	6.24%	7.58	49.23	−3.29%
2016	27.27%	24.27%	2.92%	8.29	41.49	−0.96%
2017	16.19%	15.12%	3.22%	4.70	33.99	−1.48%
2018	17.80%	20.12%	3.22%	6.24	28.13	−3.29%
Out-of-sample						
Year	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
2018	6.66%	36.85%	3.70%	9.96	6.93	−0.59%
2019	29.74%	26.23%	3.78%	6.94	42.69	−1.78%
2020	13.74%	14.96%	5.49%	2.73	38.14	−2.33%

Table 6: Excess Performance of The Top 1 to Top5 Portfolios

This table shows that the Top 1, Top 2, and Top 3 portfolios can outperform the market and earn a positive annualized excess return and a positive annualized Sharpe Ratio, while the Top 4 and Top 5 cannot.

Panel A. In-sample excess performance of the Top 1 to Top5 portfolios

Portfolio	Ret	Ann ret	Ann vol	Ann SR	Max dd
Top 1	122.64%	27.84%	42.61%	0.65	−31.97%
Top 2	60.23%	17.04%	42.25%	0.40	−28.24%
Top 3	25.89%	8.56%	41.87%	0.20	−25.91%
Top 4	−14.74%	−5.79%	42.79%	−0.14	−30.09%
Top 5	−43.56%	−24.69%	44.29%	−0.56	−59.02%

Panel B. Out-of-sample excess performance of the Top 1 to Top5 portfolios

Portfolio	Ret	Ann ret	Ann vol	Ann SR	Max dd
Top 1	57.51%	15.45%	33.31%	0.46	−14.90%
Top 2	54.31%	14.76%	33.65%	0.44	−15.74%
Top 3	31.25%	9.10%	34.05%	0.27	−16.31%
Top 4	−3.84%	−1.07%	34.80%	−0.03	−20.82%
Top 5	−69.97%	−29.56%	36.65%	−0.81	−76.27%

Table 7: Excess Performance of The Top 1 Portfolio Each Year

Panel A. In-sample excess performance of the Top 1 portfolio each year

Year	Ret	Ann ret	Ann vol	Ann SR	Max dd
2015	115.77%	75.10%	59.37%	1.26	−31.97%
2016	28.45%	30.39%	45.34%	0.67	−21.25%
2017	−4.61%	−4.49%	21.94%	-0.20	−14.21%
2018	7.92%	14.06%	36.77%	0.38	−19.87%

Panel B. Out-of-sample excess performance of the Top 1 portfolio each year

Year	Ret	Ann ret	Ann vol	Ann SR	Max dd
2018	8.66%	47.63%	31.42%	1.52	−6.10%
2019	22.10%	15.96%	32.13%	0.50	−9.44%
2020	9.70%	8.36%	35.11%	-0.24	−14.90%

Table 8: Results of The OLS Regression of AlphaNet on CH-4 Model

Whether applying Newey-West adjustment or White adjustment on the covariance matrix, the absolute t-statistics of the intercept is large and the corresponding p-value is less than 0.001, showing that the intercept of the OLS regression is significant with a significance level equal to 0.001, which indicate that a large portion of the return of AlphaNet cannot be explained by the return of the CH-4 model.

Panel A. OLS Regression with Newey-West adjusted standard errors

	In-sample (lag=7)		Out-of-sample (lag=6)	
	t-statistics	p-value	t-statistics	p-value
Intercept	9.740	0.000	6.797	0.000
MKT	0.926	0.355	1.074	0.284
SMB	3.132	0.002	−1.894	0.059
VMG	2.936	0.003	1.020	0.308
PMO	−0.550	0.582	0.176	0.860

Panel B. OLS Regression with White adjusted standard errors

	In-sample		Out-of-sample	
	t-statistics	p-value	t-statistics	p-value
Intercept	11.607	0.000	6.953	0.000
MKT	0.734	0.463	0.958	0.339
SMB	3.270	0.001	−1.590	0.112
VMG	3.402	0.001	0.887	0.375
PMO	−0.535	0.593	0.198	0.843

Table 9: Results of The OLS Regression of AlphaNet on China Anomalies

Whether applying Newey-West adjustment or White adjustment on the covariance matrix, the absolute t-statistics of the intercept is large and the corresponding p-value is less than 0.001, showing that the intercept of the OLS regression is significant with a significance level equal to 0.001, which indicate that a large portion of the return of AlphaNet cannot be explained by the return of the 22 China anomalies. Specially, betaDM1 is the only signal that is both significant in the in-sample period and the out-of-sample period, indicating that the return of AlphaNet is partly from the market beta.

	Newey-West adjusted standard errors					White adjusted standard errors			
	In-sample (lag=7)		Out-of-sample (lag=6)			In-sample		Out-of-sample	
	t-statistics	p-value	t-statistics	p-value		t-statistics	p-value	t-statistics	p-value
Intercept	8.466	0.000	6.257	0.000	Intercept	9.045	0.000	5.678	0.000
size1	0.672	0.502	−0.773	0.440	size1	0.579	0.563	−0.748	0.455
turn1	−0.851	0.395	−0.237	0.812	turn1	−0.939	0.348	−0.304	0.762
vturn1	0.551	0.582	0.082	0.935	vturn1	0.628	0.530	0.082	0.935
cvturn1	−0.596	0.551	0.147	0.883	cvturn1	−0.616	0.538	0.134	0.894
dtv1	0.786	0.432	2.018	0.044	dtv1	0.684	0.494	1.914	0.056
vdtv1	−1.630	0.103	−1.853	0.064	vdtv1	−1.679	0.093	−1.854	0.064
cvdtv1	0.403	0.687	0.963	0.336	cvdtv1	0.429	0.668	0.848	0.397
Ami1	−0.162	0.872	0.257	0.797	Ami1	−0.139	0.890	0.237	0.813
Lm1	−0.685	0.494	−0.580	0.562	Lm1	−0.618	0.536	−0.686	0.493
ivc1	−0.600	0.549	−2.355	0.019	ivc1	−0.537	0.591	−2.305	0.022
ivch3_1	−0.118	0.906	2.548	0.011	ivch3_1	−0.111	0.912	2.361	0.019
ivch4_1	0.364	0.716	−2.059	0.040	ivch4_1	0.296	0.767	−1.969	0.050
tv1	0.836	0.403	0.795	0.427	tv1	0.763	0.446	0.702	0.483
isc1	0.731	0.465	0.847	0.397	isc1	0.737	0.461	0.809	0.419
isch3_1	−0.017	0.986	−1.825	0.069	isch3_1	−0.017	0.987	−1.867	0.063
isch4_1	−0.403	0.687	1.374	0.170	isch4_1	−0.488	0.626	1.422	0.156
ts1	0.854	0.394	−0.523	0.601	ts1	0.881	0.379	−0.563	0.574
cs1	0.023	0.981	−1.376	0.169	cs1	0.020	0.984	−1.059	0.290
betaDM1	1.965	0.050	1.901	0.058	betaDM1	1.975	0.049	1.976	0.049
R1	−0.709	0.478	1.335	0.182	R1	−0.620	0.536	1.184	0.237
mdr1	0.277	0.782	−1.085	0.279	mdr1	0.277	0.782	−0.941	0.347
pps1	−0.523	0.601	3.226	0.001	pps1	−0.403	0.687	3.571	0.000

Table 10: Results of The OLS Regression of AlphaNet on selected China Anomalies with Smallest Correlation

Whether applying Newey-West adjustment or White adjustment on the covariance matrix, the absolute t-statistics of the intercept is large and the corresponding p-value is less than 0.001, showing that the intercept of the OLS regression is significant with a significance level equal to 0.001, which indicate that a large portion of the return of AlphaNet cannot be explained by the return of the selected 11 China anomalies. Specially, dtv1 and betaDM1 are the only two signals that are both significant in the in-sample period and the out-of-sample period, indicating that the return of AlphaNet is partly from the liquidity and the market beta.

	Newey-West adjusted standard errors					White adjusted standard errors			
	In-sample (lag=7)		Out-of-sample (lag=6)			In-sample		Out-of-sample	
	t-statistics	p-value	t-statistics	p-value		t-statistics	p-value	t-statistics	p-value
Intercept	7.805	0.000	6.189	0.000	Intercept	8.577	0.000	6.022	0.000
size1	0.271	0.787	1.500	0.134	size1	0.221	0.826	1.403	0.161
turn1	−0.425	0.671	0.630	0.529	turn1	−0.370	0.712	0.643	0.521
cvturn1	−3.001	0.003	0.831	0.406	cvturn1	−2.992	0.003	0.890	0.374
dtv1	−2.415	0.016	−2.468	0.014	dtv1	−2.537	0.011	−2.419	0.016
isc1	0.709	0.478	2.338	0.020	isc1	0.802	0.423	2.068	0.039
isch3_1	−0.740	0.460	−1.805	0.072	isch3_1	−0.805	0.421	−1.349	0.178
ts1	1.759	0.079	−0.796	0.427	ts1	1.466	0.143	−0.859	0.391
cs1	0.851	0.395	−1.015	0.311	cs1	0.680	0.497	−0.715	0.475
betaDM1	3.689	0.000	2.612	0.009	betaDM1	3.334	0.001	2.787	0.006
R1	−0.501	0.616	−0.155	0.877	R1	−0.492	0.623	−0.133	0.894
pps1	−0.607	0.544	2.064	0.040	pps1	−0.465	0.642	2.417	0.016

Table 11: Evaluation Metrics of The Trading Signals from Different Machine Learning Methods

The OLS model doesn't have any in-sample or out-of-sample predictive power, which shows that simple linear models are indeed insufficient to predict the complex stock market and further illustrates the necessity of constructing effective deep learning models. The RF model performs well in-sample, but its out-of-sample performance is poor, showing the characteristics of overfitting and indicating that it cannot extract valuable features from the train set. The ANN model outperforms the OLS model and the RF model, but its performance and generalization are still worse than those of AlphaNet, which indicates that a deep learning model is necessary for trading data modeling and the unique design of AlphaNet is also significantly effective.

Panel A. In-sample evaluation metrics of different machine learning methods

Model	Mean IC	Mean Rank IC	IR	Win rate
OLS	0.001	−0.005	0.025	51.7%
RF	0.087	0.045	0.80	81.6%
ANN	0.034	0.047	0.36	65.7%
AlphaNet	0.064	0.084	0.80	78.4%

Panel B. Out-of-sample evaluation metrics of different machine learning methods

Model	Mean IC	Mean Rank IC	IR	Win rate
OLS	−0.001	0.003	−0.025	50.2%
RF	0.004	−0.010	0.088	51.4%
ANN	0.014	0.032	0.17	58.2%
AlphaNet	0.055	0.079	0.56	73.4%

Table 12: Performance of The Top 1 Portfolios and 1-5 Hedge Portfolios Based on Different Machine Learning Methods

Looking at the four Top 1 portfolios, the OLS model's out-of-sample performance is significantly better than its in-sample performance, due to the fact that the Chinese market is better in the out-of-sample period rather than the model actually learning some useful information, and the same is true for the RF model. And both on the in-sample period and the out-of-sample period, AlphaNet's annualized return is the best and AlphaNet's annualized Sharpe Ratio is worse than the ANN model's on the in-sample period, but better on the out-of-sample period, indicating that AlphaNet has better out-of-sample predictive power. Looking at the four 1-5 Hedge portfolios, the OLS model performs weakly both in-sample and out-of-sample. And the RF model achieves the best in-sample annualized Sharpe Ratio (7.79), but its performance is pretty weak on the out-of-sample period with an annualized Sharpe Ratio equal to −0.20, indicating again that the RF model has little generalization. The ANN model exhibits some out-of-sample predictive power but its annualized return and annualized Sharpe Ratio (7.66%, 1.46) are still significantly inferior to that of AlphaNet (22.34%, 4.86), which indicates that AlphaNet achieves better results than general deep learning models through its special design.

Panel A: Performance of the Top 1 portfolio from different machine learning models

In-sample						
Model	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
OLS	−34.25%	−5.59%	33.87%	−0.17	315.06	−70.26%
RF	63.94%	16.70%	25.35%	0.66	375.53	−33.96%
ANN	97.28%	22.30%	27.47%	0.81	482.34	−42.28%
AlphaNet	93.77%	23.30%	32.33%	0.72	165.39	−48.33%
Out-of-sample						
Model	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
OLS	34.03%	17.62%	25.48%	0.69	169.59	−24.85%
RF	61.31%	26.15%	23.22%	1.13	313.24	−18.94%
ANN	62.61%	26.50%	23.02%	1.15	294.59	−19.71%
AlphaNet	108.79%	38.73%	22.90%	1.69	102.76	−16.39%

Panel B: Performance of the 1-5 Hedge portfolio from different machine learning models

In-sample						
Model	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
OLS	4.75%	1.34%	3.82%	0.35	282.93	−7.88%
RF	58.93%	12.63%	1.62%	7.79	391.62	−0.71%
ANN	133.88%	23.46%	8.04%	2.92	466.70	−4.53%
AlphaNet	161.77%	26.30%	4.14%	6.35	152.83	−3.29%
Out-of-sample						
Model	Ret	Ann ret	Ann vol	Ann SR	Acc tvr	Max dd
OLS	1.86%	0.95%	3.03%	0.31	153.87	−3.17%
RF	−0.58%	−0.28%	1.40%	−0.20	320.56	−2.17%
ANN	16.60%	7.66%	5.25%	1.46	267.44	−4.81%
AlphaNet	57.40%	22.34%	4.60%	4.86	87.76	−7.03%

Table 13: Performance and Excess Performance of The Position-optimized Portfolio

Comparing Table 11 and Table 6, it can be seen that the excess performance of the long-only portfolio is apparently improved after the portfolio position optimization process, as shown by the excess annualized return improving from 15.45% to 22.88% and the excess annualized Sharpe Ratio improving from 0.46 to 3.24.

Panel A: Performance of the position-optimized portfolio

Year	Ret	Ann ret	Ann vol	Ann SR	Max dd
2018	5.76%	38.50%	22.04%	1.62	−6.89%
2019	55.71%	57.13%	21.98%	2.18	−17.93%
2020	39.50%	47.55%	25.78%	1.65	−15.05%
All	129.75%	51.57%	23.64%	1.88	−17.93%

Panel B: Excess performance of the position-optimized portfolio

Year	Ret	Ann ret	Ann vol	Ann SR	Max dd
2018	6.80%	46.57%	3.82%	10.28	−0.32%
2019	23.06%	23.59%	6.00%	3.57	−2.23%
2020	14.88%	17.59%	7.29%	2.27	−4.14%
All	50.98%	22.88%	6.44%	3.24	−4.14%

Figure 1: Structure of ANN

This figure shows the structure of a simple ANN model, which has one input layer with three neurons, one hidden layer with four neurons, and one output layer with two neurons. Typically, this model used to deal with binary classification problems according to its output layer. All these neurons are connected by connection weights (in blue).

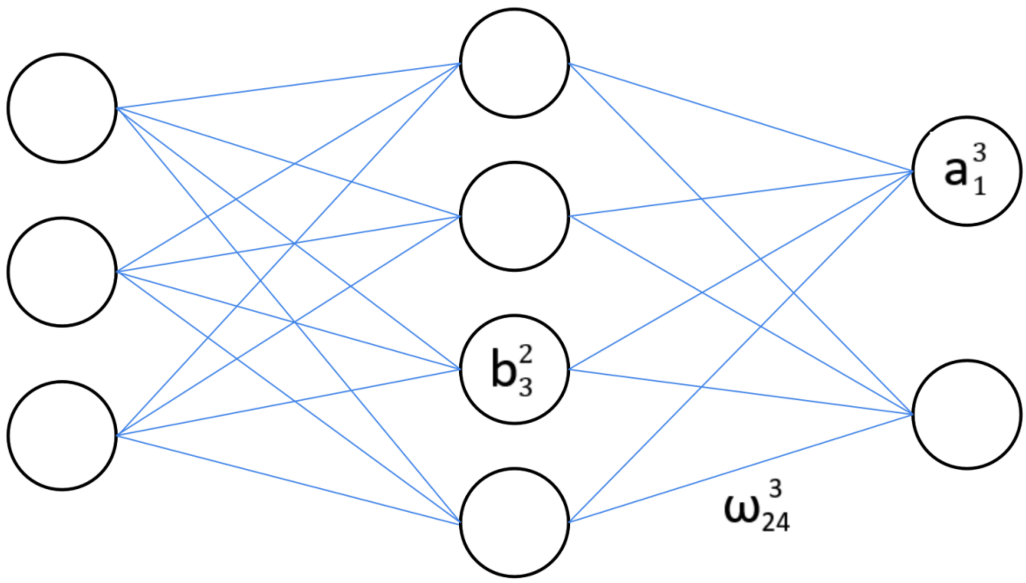


Figure 2: How Convolution Operation Works on The Input Matrix

This figure is an illustrative example with a 6×6 input matrix and a 3×3 filter. For each element (i, j) of the input matrix, the convolution operation sums the element-wise product of the filter and the 3×3 matrix contents centered on (i, j). The result is stored in the corresponding element of the 4×4 output matrix. The stride is an important parameter in the convolution operation, which refers to the distance of each move of the filter on the input matrix, and the stride in Figure 2 is 1. The shape of the matrix obtained after the convolution operation is usually determined by the shape of the input matrix, the shape of the filter, and the value of the stride.

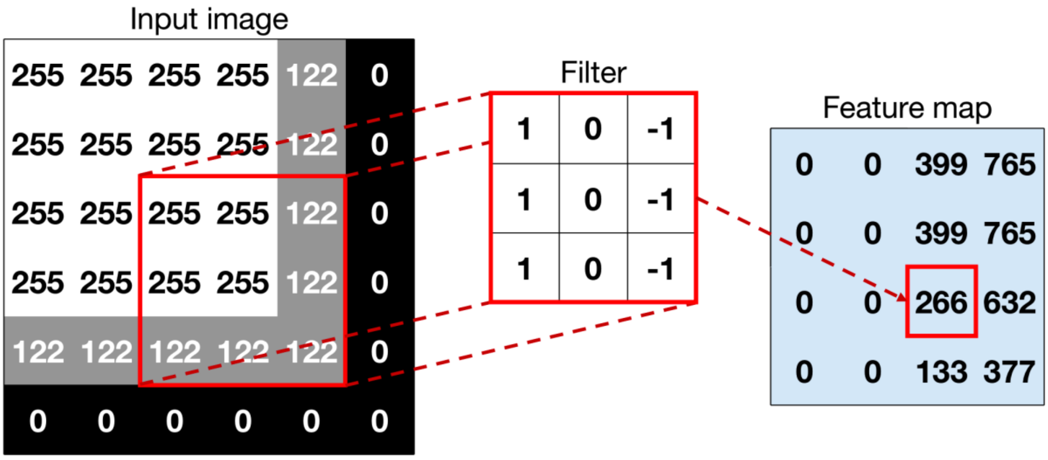


Figure 3: Working Mechanism of The Feature Operators

This figure is an example of how the operator works with the Corr(X, Y, 3) and the time step (similar to the stride of a convolution kernel) is 1. In AlphaNet, for the feature operators acting on two time series, the 9×30 input matrix will be converted into a 36×3 feature matrix and for the feature operators acting on one time series, the 9×30 input matrix will be converted into a 9×3 feature matrix. Then after the merging process, the dimension of the feature map will be 108×3.

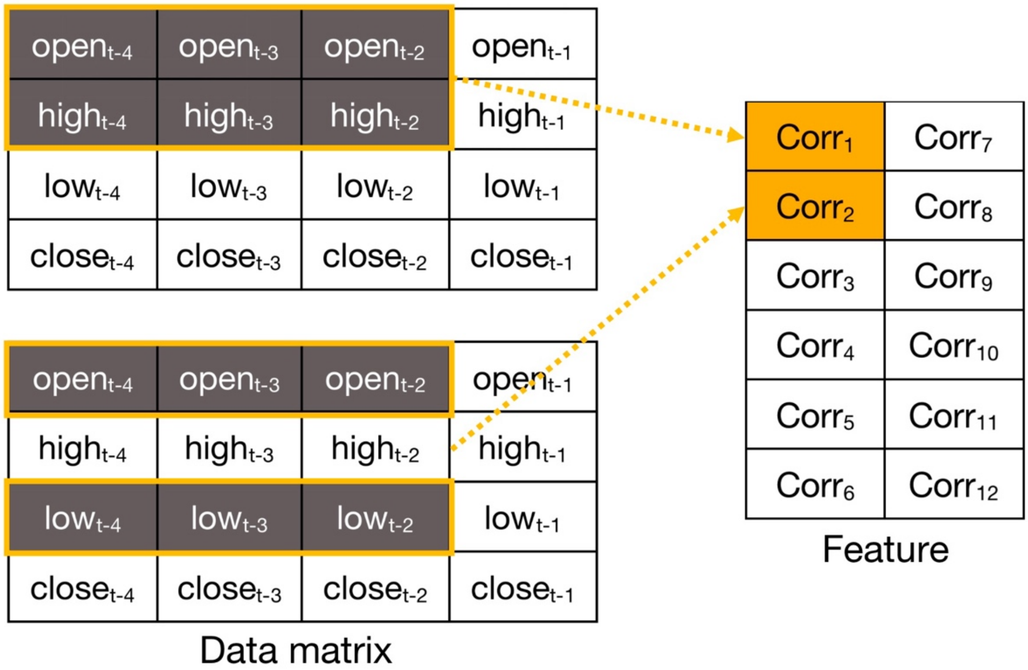


Figure 4: Architecture of AlphaNet

The shape of the convolutional filter is 1×3, the number of the channels is 16, and the stride is 1. Since the feature map's shape is 108×3, then after the convolutional layer, the data's shape becomes 16×108×1. After that, the data will be flattened into a 1728×1 vector and sent into the following layers. The amounts of neuron in the three fully-connected hidden layers are 1728, 256, and 30, respectively, and the activation functions are chosen as ReLU, ReLU, and Sigmoid in sequence. Finally, there is the output layer with one neuron since it is a regression model.

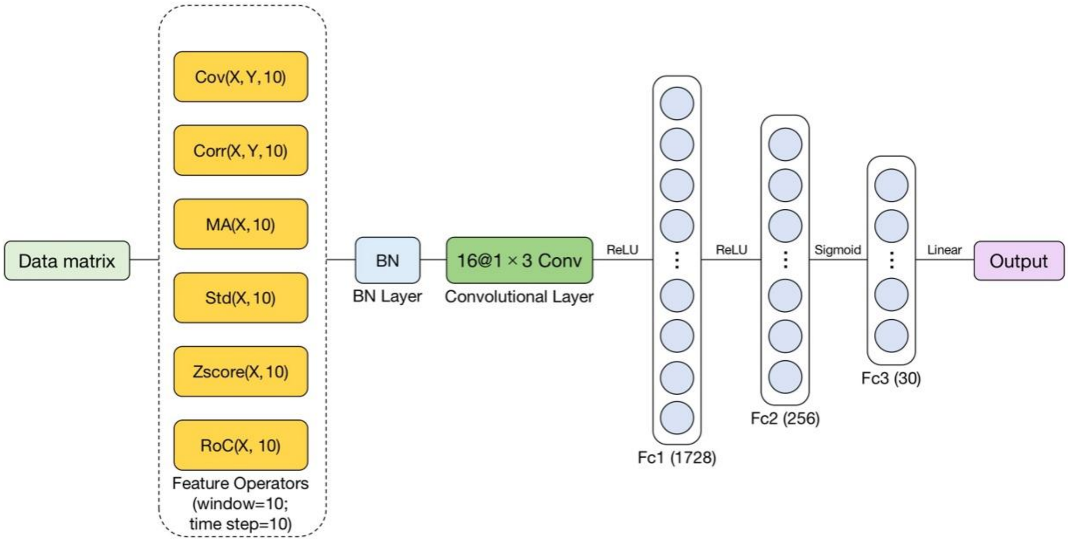


Figure 5: Net Value of The Top 1 to Top 5 Portfolios

This figure shows apparent gaps between the five long-only portfolios, which indicates that the model has excellent monotonicity, i.e., the model is able to distinguish between stocks that are more likely to perform well and stocks that are more likely to perform poorly. And the 1-5 Hedge portfolio (in purple) also performs well and stably both in-sample and out-of-sample, which indicates that the model has excellent generalization and significant in-sample and out-of-sample predictive power.

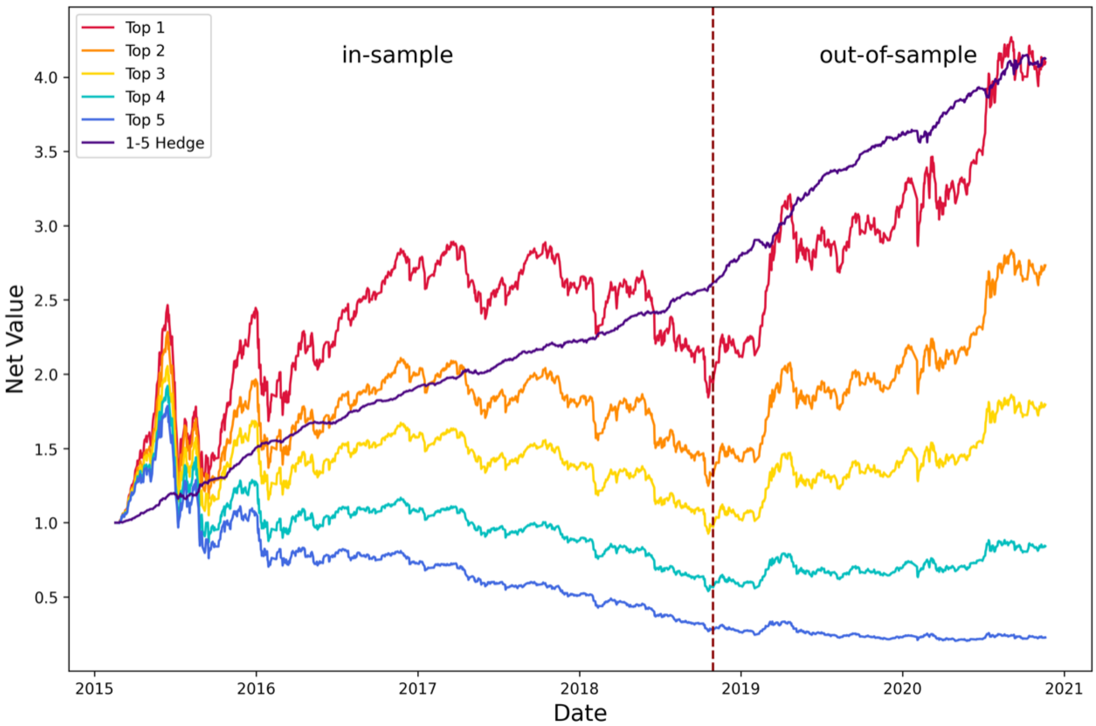


Figure 6: Excess Net Value of The Top 1 to Top 5 Portfolios

Most of the time, the Top 1 portfolio (in red) and the Top 2 (in orange) portfolio outperforms the Chinese stock market and obtain stable excess return, while the Top 5 portfolio (in blue) underperforms the market.

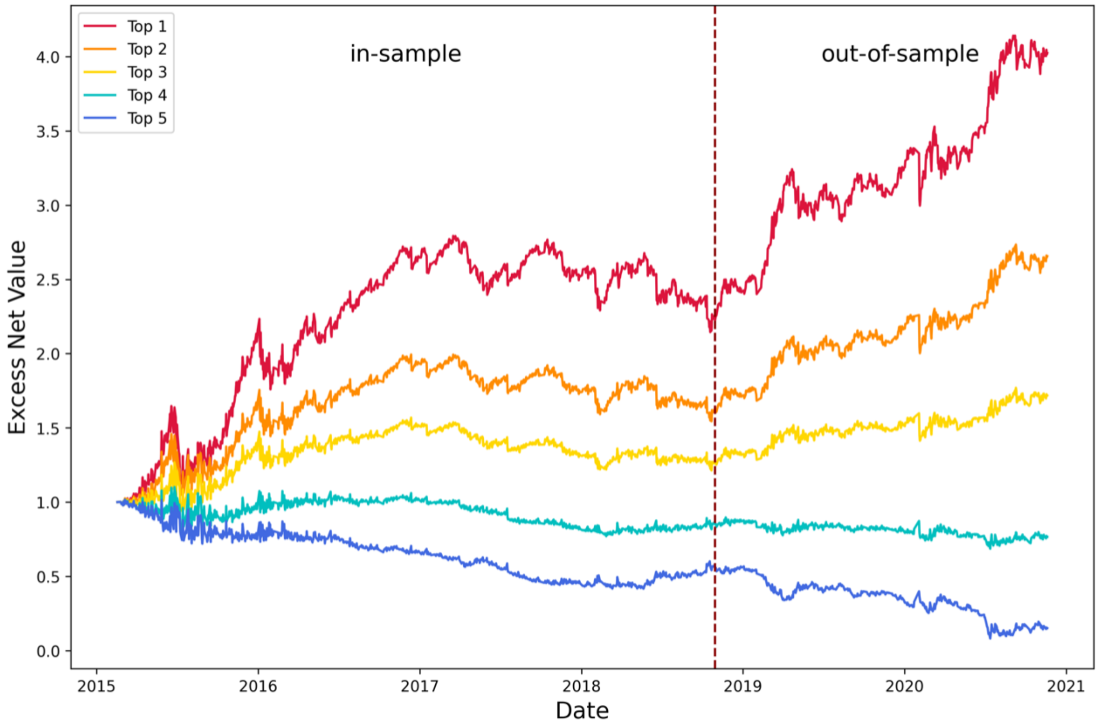


Figure 7: Decomposition of The Return of AlphaNet on CH-4 model

In this figure, according to the results of the OLS regression, the accumulated return of AlphaNet (in red) is decomposed into the four risk factors' accumulated return and the anomaly (in blue). It can be seen that only a small fraction of the return of AlphaNet can be explained by the return of CH-4 model and most of it is the unexplained anomaly, which further illustrates the significance of AlphaNet.

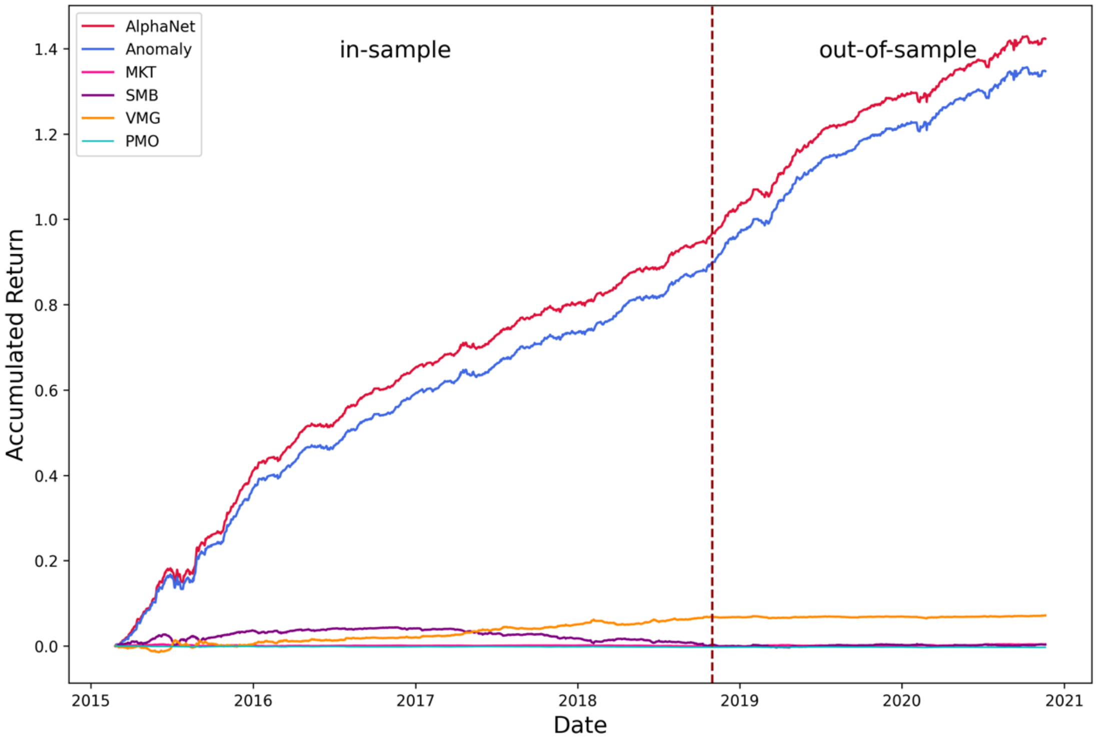


Figure 8: Decomposition of The Return of AlphaNet on China Anomalies

In this figure, according to the results of the OLS regression, the accumulated return of AlphaNet (in red) is decomposed into the 22 China anomalies' accumulated return (in purple) and the anomaly (in blue). It can be seen that only a small fraction of the return of AlphaNet can be explained by the return of the 22 China anomalies and most of it is the unexplained anomaly, which further illustrates the significance of AlphaNet.

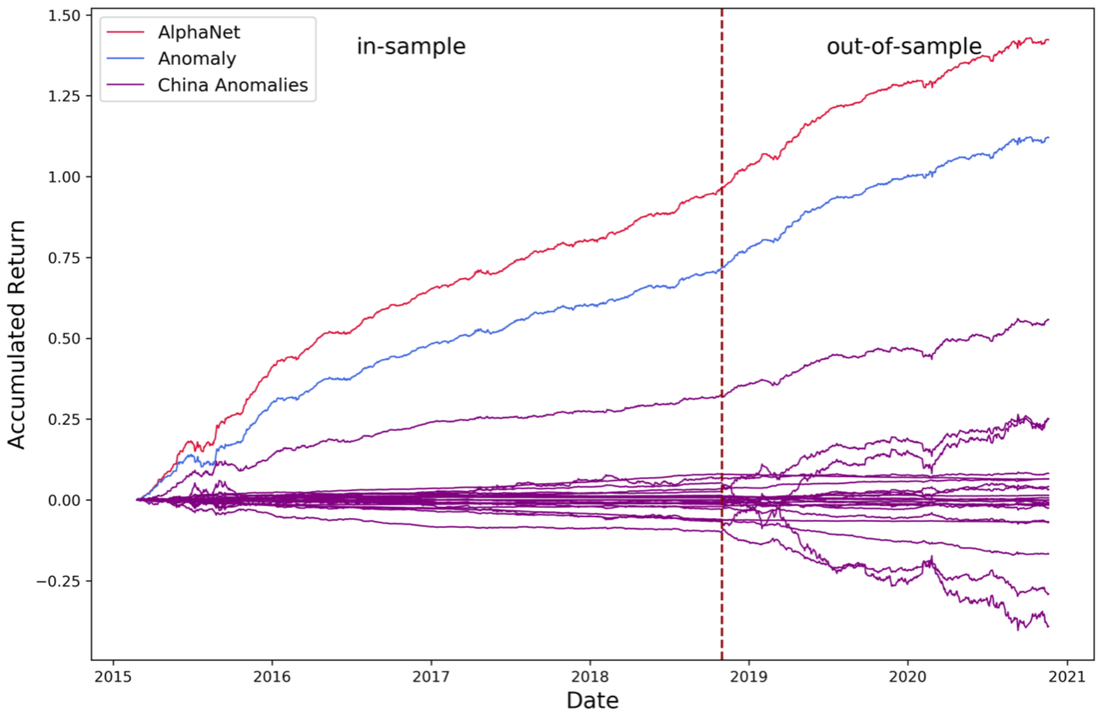
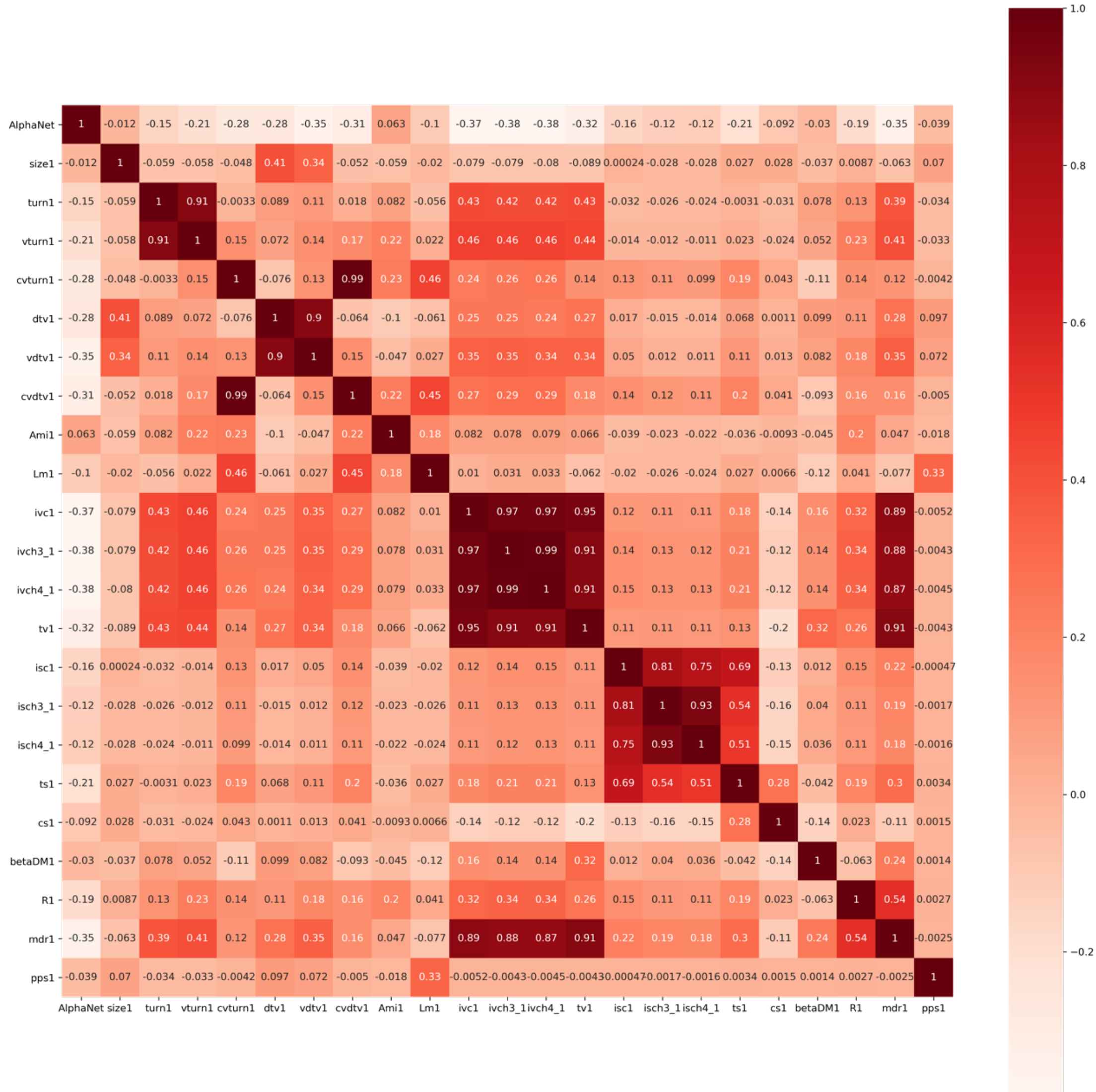


Figure 9: Correlation among AlphaNet and The China Anomalies

In this figure, according to the average cross-sectional correlation among AlphaNet and the China Anomalies (in red) and the time-series correlation among the return of them (in blue), there are several anomaly groups with high correlation, including {turn1, vturn1, ivc1, ivch3_1, ivch4_1, tv1, mdr1}, {dtv1, vdtv1}, {isch3_1, isch4_1}, etc. Then we select 11 anomalies with the smallest correlation, including size1, turn1, cvturn1, dtv1, isc1, isch3_1, ts1, cs1, betaDM1, R1, and pps1. In particular, for two anomalies whose correlation exceeds 0.80, we only keep one of them.



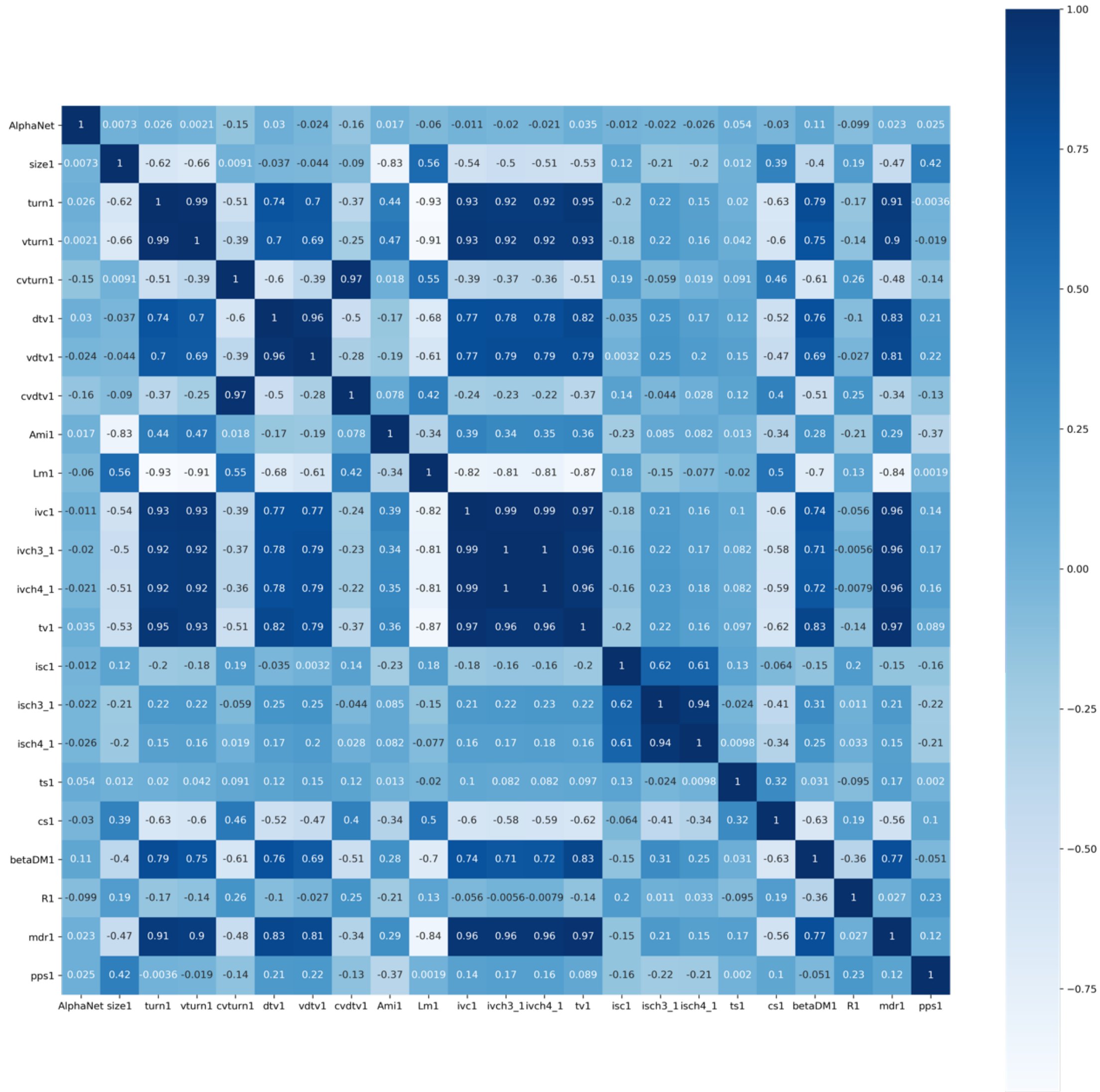
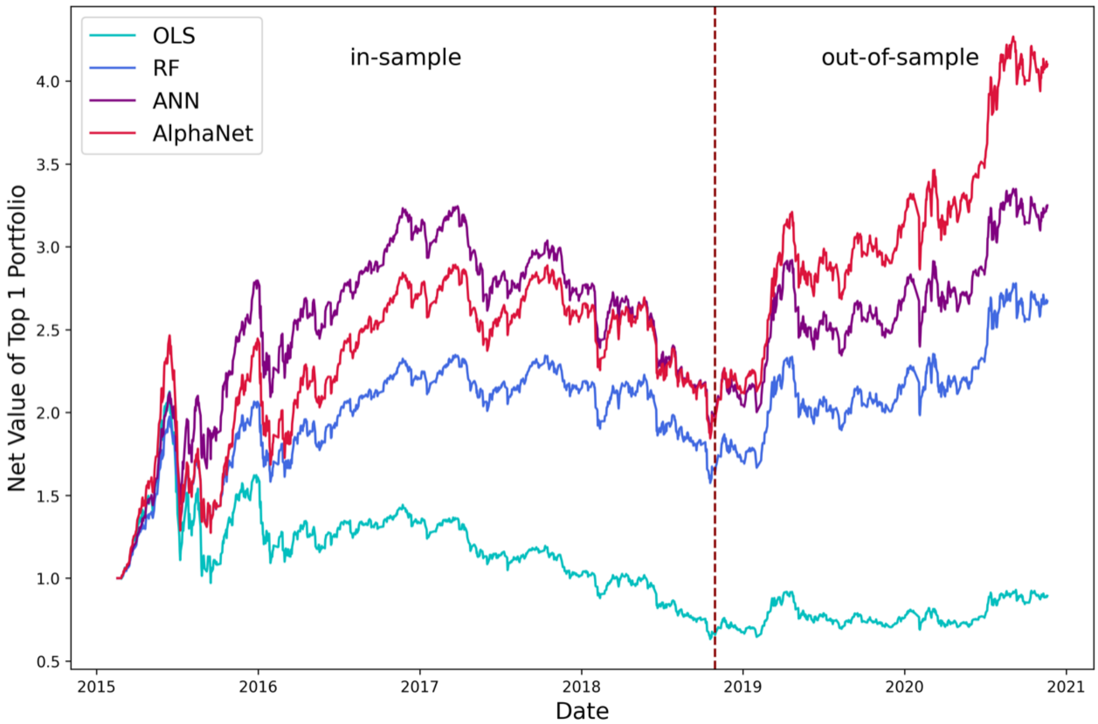

AlphaNet
size1
turn1
vturn1
cvturn1
dtv1
vdtv1
cvdtv1
Ami1
Lm1
ivc1
ivch3_1
ivch4_1
tv1
isc1
isch3_1
isch4_1
ts1
cs1
betaDM1
R1
mdr1
pps1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

Figure 10: Net Value of The Top 1 Portfolios and 1-5 Hedge Portfolios Based on Different Machine Learning Methods

Looking at the net value curves of the four Top 1 portfolios (the left one), the in-sample performances of AlphaNet (in red) and the ANN model (in purple) are better the other two models and the out-of-sample performance of AlphaNet is better than the other three models. Looking at the net value curves of the four 1-5 Hedge portfolios (the right one), both the in-sample performance and the out-of-sample performance of AlphaNet (in red) are better the other three models. These results show that AlphaNet outperforms the other common machine learning methods due to its unique design.



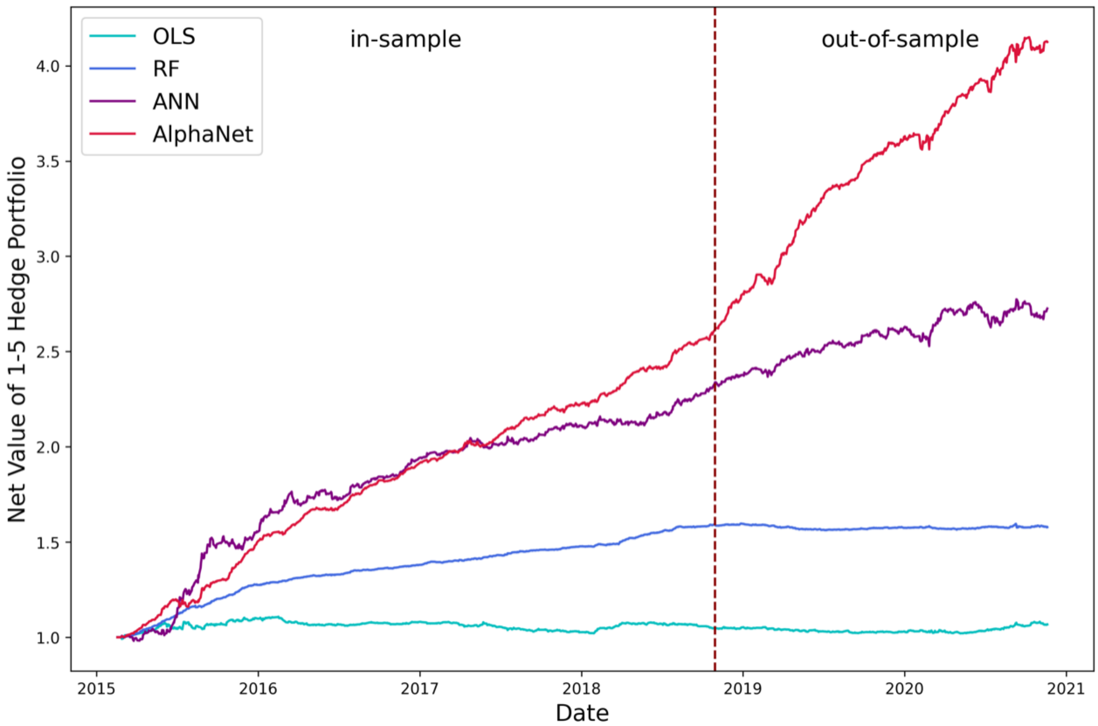
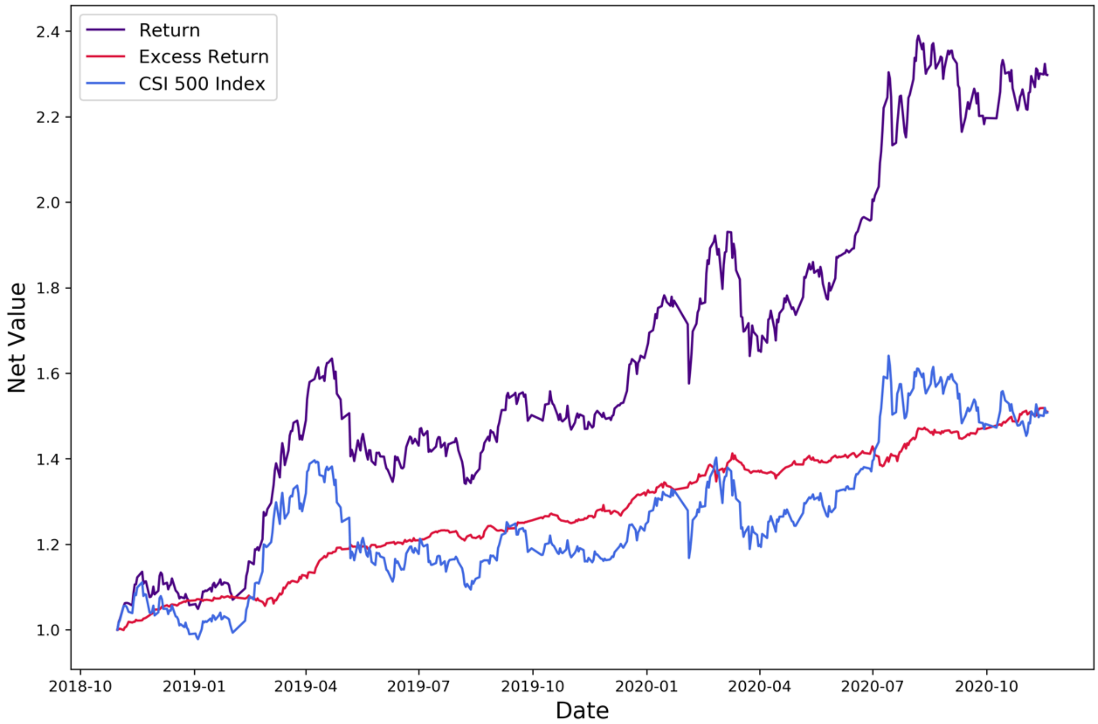


Figure 11: Net Value and Excess Net Value of The Position-optimized Portfolio

This figure shows that after the portfolio position optimization process, the excess return of the long-only portfolio (in red) is much more stable than before.



Predicting Stock Returns based on Convolutional Neural Networks with Feature Operators

---Online Appendix

A. Definitions of Signals

A.1 Liquidity

A.1.1 Firm Size (size1)

According to Liu et al., 2019, daily firm size is unadjusted closing price multiplied by the number of total A-shares outstanding. At each trading day t, we sort stocks into quintiles based on $size1_t$, which is the average value of the daily firm size estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.1.2 Share Turnover (turn1)

According to Datar et al., 1998, daily share turnover is adjusted trading volume divided by the number of free-float A-shares outstanding. At each trading day t, we sort stocks into quintiles based on $turn1_t$, which is the average value of the daily share turnover estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.1.3 Variation of Share Turnover (vturn1)

According to Chordia et al., 2001, variation of share turnover is the standard deviation of daily share turnover. At each trading day t, we sort stocks into quintiles based on $vturn1_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.1.4 Coefficient of Variation of Share Turnover (cvturn1)

According to Chordia et al., 2001, coefficient of variation of share turnover is the ratio of the standard deviation to the mean for daily share turnover. At each trading day t, we sort stocks into quintiles based on $cvturn1_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.1.5 RMB Trading Volume (dtv1)

According to Brennan et al., 1998, RMB trading volume is the average value of daily RMB trading volume. At each trading day t, we sort stocks into quintiles based on $dtv1_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.1.6 Variation of RMB Trading Volume (vdtv1)

According to Chordia et al., 2001, variation of RMB trading volume is the standard deviation of daily RMB trading volume. At each trading day t, we sort stocks into quintiles based on $vdtv1_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.1.7 Coefficient of Variation of RMB Trading Volume (cvdtv1)

According to Chordia et al., 2001, coefficient of variation of RMB trading volume is the ratio of the standard deviation to the mean for daily RMB trading volume. At each trading day t, we sort stocks into quintiles based on $cvdtv1_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.1.8 Amihud Illiquidity (Ami1)

According to Amihud, 2002, Amihud illiquidity measure is the average value of the ratio of absolute daily stock return to daily RMB trading volume. At each trading day t, we sort stocks into quintiles based on $\mathrm{Ami1}_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.1.9 Turnover-adjusted Number of Zero Daily Trading Volume (Lm1)

According to Liu, 2006, we calculate turnover-adjusted number of zero daily trading volume over the prior one month as follows,

$$\mathrm{Lm1} = \text{Number of days with volumes} < 150{,}000 \text{ in prior one month} + \frac{\frac{1}{1 - \text{month turnover}}}{\text{Deflator}}$$

$$\text{Deflator} = \max\left\{\frac{1}{1 - \text{month turnover}}\right\} + 1$$

where 1-month turnover is the sum of daily turnover over the prior one month. At each trading day t, we sort stocks into quintiles based on $\mathrm{Lm1}_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2 Risk

A.2.1 Idiosyncratic Volatility per the CAPM (ivc1)

According to Hou et al., 2023, idiosyncratic volatility per the CAPM is the standard deviation of the residuals from regressing a stock's daily excess returns on the CAPM over the prior one month. At each trading day t, we sort stocks into quintiles based on $\mathrm{ivc1}_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2.2 Idiosyncratic Volatility per the China 3-Factor Model (ivch3_1)

According to Liu et al., 2019, idiosyncratic volatility per the China 3-factor (CH-3) model is the standard deviation of the residuals from regressing a stock's daily excess returns on the CH-3 factors over the prior one month. At each trading day t, we sort stocks into quintiles based on $\mathrm{ivch3\_1}_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2.3 Idiosyncratic Volatility per the China 4-Factor Model (ivch4_1)

According to Liu et al., 2019, idiosyncratic volatility per the China 4-factor (CH-4) model is the standard deviation of the residuals from regressing a stock's daily excess returns on the CH-4 factors over the prior one month. At each trading day t, we sort stocks into quintiles based on $\mathrm{ivch4\_1}_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2.4 Total Volatility (tv1)

According to Hou et al., 2023, total volatility is the standard deviation of a stock's daily returns over the prior one month. At each trading day t, we sort stocks into quintiles based on $\mathrm{tv1}_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2.5 Idiosyncratic Skewness per the CAPM (isc1)

According to Hou et al., 2023, idiosyncratic skewness per the CAPM is the skewness of the residuals from regressing a stock's daily excess returns on the CAPM over the prior one month. At each trading day t, we sort stocks into quintiles based on $\mathrm{isc1}_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2.6 Idiosyncratic Skewness per the China 3-Factor Model (isch3_1)

According to Hou et al., 2023, idiosyncratic skewness per the China 3-factor (CH-3) model is the skewness of the residuals from regressing a stock's daily excess returns on the CH-3 factors over the prior one month. At each trading day t, we sort stocks into quintiles based on isch3_1$_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2.7 Idiosyncratic Skewness per the China 4-Factor Model (isch4_1)

According to Hou et al., 2023, idiosyncratic skewness per the China 4-factor (CH-4) model is the skewness of the residuals from regressing a stock's daily excess returns on the CH-4 factors over the prior one month. At each trading day t, we sort stocks into quintiles based on isch4_1$_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2.8 Total Skewness (ts1)

According to Hou et al., 2023, total skewness is the skewness of a stock's daily returns over the prior one month. At each trading day t, we sort stocks into quintiles based on ts1$_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2.9 Co-skewness (cs1)

According to Harvey and Siddique, 2000, We measure co-skewness as follow,

$$\mathrm{cs1} = \frac{\mathrm{E}[\epsilon_i \epsilon_m^2]}{\sqrt{\mathrm{E}[\epsilon_i^2]\mathrm{E}[\epsilon_m^2]}}$$

where ϵ_i is the residuals from regressing a stock's excess returns on the CAPM and ϵ_m is the demeaned return of CAPM. At each trading day t, we sort stocks into quintiles based on cs1$_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.2.10 The Dimson Beta (betaDM1)

According to Dimson, 1979, we measure the market beta by using the current as well as the lead and lag of market returns.

$$r_{i,d} - r_f = \alpha_i + \beta_{i,1}(r_{m,d-1} - r_f) + \beta_{i,2}(r_{m,d} - r_f) + \beta_{i,3}(r_{m,d+1} - r_f) + \epsilon_{i,d}$$

where $r_{i,d}$ is the stock i's return on day d, $r_{m,d}$ is the CAPM return on day d, and r_f is the risk-free rate. The Dimson beta for stock i is calculated as $\widehat{\beta_1} + \widehat{\beta_2} + \widehat{\beta_3}$. At each trading day t, we sort stocks into quintiles based on betaDM1$_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.3 Past Returns

A.3.1 Prior One-month Return (R1)

According to Hou et al., 2023, at each trading day t, we sort stocks into quintiles based on R1$_t$. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.3.2 Maximum Daily Return (mdr1)

According to Bali et al., 2017, maximum daily return is the average of the 5 highest daily returns of a given stock from the prior month. At each trading day t, we sort stocks into quintiles based on mdr1$_t$ estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).

A.3.3 Share Price (pps1)

According to Hou et al., 2023, at each trading day t, we sort stocks into quintiles based on $pps1_t$, which is the average value of the daily share price estimated with daily data over the prior one month. We calculate the daily decile portfolio returns over one day (the rebalance period is one trading day).