Timing the Factor Zoo via Deep Learning: Evidence from China

Tian Ma[1], Cunfei Liao[2,*], Fuwei Jiang[3]

First Version: July 2022

Current Version: October 2022

[1] School of Economics, Minzu University of China, Beijing 100081, China, mark8938@qq.com;

[2] School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China; cfliao@njust.edu.cn

[3] School of Finance, Central University of Finance and Economics, Beijing 100081, China; jfuwei@gmail.com

[*] Corresponding Author

Timing the Factor Zoo via Deep Learning: Evidence from China

Abstract:

This paper proposes a factor timing strategy with information from 146 characteristic-based factors and a deep learning approach to capture nonlinear predictability. The deep learning-based factor timing strategy generates the highest economic value compared with the unconditional and alternative linear machine learning-based portfolios and remains robust after controlling for traditional factor models and transaction costs. With the unique market structure of the Chinese stock market, we find that mispricing-based theory helps explain the factor timing via deep learning.

JEL Classification: G11, G12, G14

Keywords: Factor timing, Deep learning, Neural network, Chinese stock market, Mispricing

1. Introduction

Factor timing, which combines factor investing from the cross section of assets and market timing from time-series predictability, has received much attention (see, e.g., Avramov et al., 2017; Arnott et al., 2021; Ehsani and Linnainmaa, 2022; Gupta and Kelly, 2019; Haddad, Kozak, and Santosh, 2020; Leippold and Yang, 2021). Relative to static factor investing, the factor timing strategy improves portfolio performance significantly in capturing the autocorrelation structure and predictability among factors.

Nowadays, deep learning methods have been widely used in asset pricing, with considerable performance in dealing with overfitting and low signal-to-noise problems (e.g., Gu, Kelly, and Xiu, 2020; Leippold, Wang, and Zhou, 2022; Nagel, 2021), and are likewise suitable for capturing the significantly time-varying components in factor investing (Stambaugh, Yu, and Yuan, 2012; Jacobs, 2015; Akbas et al., 2016; Keloharju, Linnainma, and Nyberg, 2016). We employ the neural network approach in factor timing for two main reasons. First, compared to linear models, the neural network captures the nonlinear interactions between the factors, which suits the big dataset more properly. Second, compared to tree models, the neural network has better predictability and robustness across subsamples. For example, Leippold et al. (2022) document that tree models always perform poorly in large stocks.

Inspired by Leippold et al. (2022), this paper focuses on the Chinese stock market to better understand the economic grounds behind factor timing. First, the Chinese stock market reaches the second-largest market worldwide and becomes increasingly important in the world economy, especially on these chaotic days. It is meaningful to investigate the application of the investment strategy in another crucial market, not the U.S. Second, different from the U.S. market, the factor premia are rather large and are not captured by retail investors in China (Liu, Stambaugh, and Yuan, 2019), and the market timing also performs well here (Ma, Liao, Jiang, 2021; Tang et al., 2021). Therefore, these different conditionals help study the factor timing strategy and may provide different evidence. Third, as an emerging market, the Chinese stock market has some unique characteristics, such as changeable market states and high limit-to-arbitrage, which helps us to discover the economic channels of the factor timing.

In detail, we first build a 146 characteristics-based factor set in the Chinese stock market, covering various types (Jensen, Kelly, and Pedersen, 2021), and use principal component analysis (PCA) to reduce the dimension (Kozak, Nagel, and Santosh, 2020; Haddad, Kozak, and Santosh, 2020), with the first four PCs explaining over 60% of the total variation in returns. Then, we build three factor timing strategies: the unconditional one that weights the factors simply from the 4 PCs' historical mean return; the conditional deep learning-based one that predicts the 4 PCs' returns with a three-hidden-layers feed-forward neural network (NN3) in the out-of-sample; and the conditional ordinary least squares (OLS)-based strategy that forecasts returns with the OLS.

We find that the conditional NN3-based factor timing portfolio generates the highest average returns, the lowest standard deviations, the highest Sharpe ratios and certainty equivalent return (CER) gains with the different degrees of investors' risk aversion compared with the other two strategies. For example, when the investors' risk aversion is 2, the average return of the conditional NN3-based strategy is 0.594, higher than –0.009 of the unconditional and 0.118 of the conditional OLS-based strategies. The Sharpe ratios range from 1.41 to 1.83 under the four levels of risk aversion of the NN3-based strategy. The CER gain remains economically large, suggesting that investors are willing to pay a higher annual fee to access the returns of the deep learning approach than the other two methods. Our findings show that the factor timing strategy with deep learning produces economic benefits in China.

The outperformance of the conditional NN3- relative to the conditional OLS-based strategy comes from its flexibility and effectiveness in dealing with the high-dimensional and nonlinear interaction between characteristics. As strongly predictable in the PCs of anomalies (Haddad, Kozak, and Santosh, 2020), in the investigation of out-of-sample R^2 and Diebold and Mariano (1995) tests, we find that the NN3 model significantly improves the out-of-sample R^2 in four PCs and generates smaller forecast errors than the OLS.

In addition, we find that the traditional factor models cannot absorb the returns of the conditional NN3-based factor timing approach, with the abnormal alphas of the Fama and French (2015) five-factor model (FF5), FF5 augmented with momentum factor (Carhart, 1997), and Liu, Stambaugh, and Yuan (2019) three-factor model (CH3) in China all positively significant at the 1% level. Considering the transaction costs, the returns and CER gains for the conditional NN3-based portfolio remain robust.

Notably, we exploit the economic grounds of the conditional NN3-based portfolio. First, we investigate the dominant characteristics of each PC. The results show that the quality- and profitability-related factors are important in the first PC, the valuation-, quality-, and low risk-related factors are rather crucial in PC2, the factor exposure-, seasonality-, and value-related factors play an important role in PC3, and the quality- and seasonality-related factors are essential in PC4. Then, the mispricing-based explanation helps account for the factor timing strategy. We observe that the deep learning-based factor timing strategy performs better when limits-to-arbitrage are large, i.e., smaller stocks, higher investor sentiment, and aggregate idiosyncratic volatility periods.

This paper contributes to the literature on factor timing with a large set of characteristics-based factors. As Haddad, Kozak, and Santosh (2020) mentioned, "factor timing is very valuable, above and beyond market timing and factor investing taken separately", our research improves the investment benefits of factor timing with the "factor zoo" called by Cochrane (2011). Some studies aim to discover incremental information from hundreds of potential characteristics (e.g., Harvey, Liu, and Zhu, 2016; Green, Hand, and Zhang, 2017; Freyberger, Neuhierl, and Weber 2020; Hou, Xue, and Zhang, 2020; Dong et al., 2022). By extracting the high-variance component of candidate factor returns with PCA, we aggregate the effective information from the multivariate characteristics.

Our paper is also related to the application of deep learning in finance. The benefits of deep learning have recently been documented in the literature, which has better forecasting performance (Chen, Pelger, Zhu, 2021; Fischer & Krauss, 2018) because it is suitable for fitting flexible functional forms with many covariates, especially nonlinear relationships (Feng, He, and Polson, 2018; Gu, Kelly, and Xiu, 2020, 2021). As Nagel (2021) suggests, the methods that generate the most accurate return forecasts at the individual stock level may not provide the best-performing portfolio. Our paper uses the deep learning approach in predicting portfolio-level returns, supplementing its successful application in asset pricing to provide substantial investment gains.

Our research also helps the understanding of the Chinese stock market and factor timing. Although the Chinese stock market is the second-largest market worldwide, studies on factor timing lack. As the factor premia in China is rather large (Liu, Stambaugh, and Yuan, 2019) and the market timing also performs well (Ma, Liao, Jiang, 2021; Tang et al., 2021), our paper shows the superior performance of factor timing in China, extending the application of factor timing in the emerging market. The finding that the quality- and profitability-related factors contribute most to the principal components also echoes their importance in China related to prior studies. Meanwhile, with the unique features in the Chinese stock market, for instance, changeable market states and high limit-to-arbitrage, we discover the economic channels of factor timing through the mispricing-based theory, in the spirit of Avramov, Cheng, and Metzker (2022), which contributes to the understanding of cyclical variation in investor behaviors.

The paper is organized into five sections. Section 2 describes the data and methodology. Section 3 presents the empirical results. Section 4 uncovers the economic channels of the factor timing strategy. Section 5 concludes.

2. Data and Methodology

2.1 Data

We obtain data from the China Stock Market & Accounting Research (CSMAR) from January 2004 to December 2020, including financial reports, monthly and daily stock returns, and the risk-free rate. To ensure the quality of the data, we exclude stocks with special treatment (ST) and/or particular transfer (PT) status, as they tend to be under financial distress (Carpenter, Lu, and Whitelaw, 2021; Liao, Luo, and Tang, 2021).

According to Jensen, Kelly, and Pedersen (2021), we construct 146 characteristic-based portfolios in the Chinese stock market. The 146 factors cover a wide range of factors in the literature, including accruals, debt issuance, investment, leverage, low risk, momentum, profit growth, profitability, quality, seasonality, size, skewness, and value. The usage of our input data comes from the following reasons. First, because each characteristic (Hou et al., 2020) and/or theme (Jensen et al., 2021) contributes differently to explaining stock returns, we try to explain our factor timing strategy based on the factor clusters to discover which type of factors have the higher explanatory ability in China. Second, due to the high economic and statistical similarity of factors in each theme, the marginal predictability of machine learning can be improved by introducing a factor zoo with more clusters. We employ the monthly factor return in our study, which is defined as the average return on the top decile (which earns the highest expected return) minus the bottom (which earns the lowest expected return)[1].

2.2 Factor Timing Methodology

2.2.1 Investor's Problem

We first solve the investor maximization problem under mean-variance preference. Consider a single-period investor who allocates his wealth into N risky assets and a risk-free asset and maximizes the conditional expected utility subject to a budget constraint,

$$\max_{\alpha_t \in \mathbb{R}} E[u(W_{t+1}) | E_t^K], K \in \{U, NN3, OLS\}, \tag{1}$$

$$\text{s.t. } W_{t+1} = W_t\left[R_{f,t+1} + \alpha_t\left(R_{t+1} - R_{f,t+1}\right)\right], \tag{2}$$

where α_t is the wealth proportion invested in risky assets with a gross return R_{t+1}, $1 - \alpha_t$ represents the part invested in risk-free assets with return $R_{f,t+1}$, W_t (W_{t+1}) is the investor's wealth at time t (t + 1), and $u(\cdot)$ is a utility function. $K \in \{U, NN3, OLS\}$ refers to our three-factor timing strategies, the unconditional PCs, the deep learning-, and the OLS-based conditional factor timing strategies. We assume $\alpha_t \equiv \alpha(E_t^K)$, which

[1] Table A in the Online Appendix reports the factors' annualized excess returns with their details and references.

refers to the investor's portfolio allocation or weight following Aït-Sahalia and Brandt (2001).

The mean-variance preference has the standard objective utility function, as follows:

$$u(W_{t+1}) = E[W_{t+1}] - \frac{\gamma}{2} Var[W_{t+1}], \tag{3}$$

where $\gamma \geq 0$ represents the investor's risk aversion.

Then, the conditional problem of the mean-variance investors goes,

$$E[u(W_{t+1})|E_t^K] = E[W_{t+1}|E_t^K] - \frac{\gamma}{2} Var[W_{t+1}|E_t^K], \tag{4}$$

and the investor's portfolio weight on risky assets is

$$\alpha_t = \frac{1}{\gamma W_t} \frac{E[R_{t+1} - R_{f,t+1}|E_t^K]}{Var[R_{t+1} - R_{f,t+1}|E_t^K]}. \tag{5}$$

The portfolio allocations on risky assets of the mean-variance investors are relative to the conditional expected excess return of them and reciprocal to their conditional variance. Our study aims to identify whether E_t^K influences the performance of the factor timing strategies. If one portfolio generates higher returns and/or lower variance, then the strategy is better.

Then, we compute the certainty equivalent return (CER) of the investment portfolio from the allocation perspective and the Sharpe ratios of the investment portfolio to investigate the economic value for the different factor timing strategies. CER states the following:

$$CER_p = \hat{\mu}_p - \frac{1}{2} \gamma \hat{\sigma}_p^2, \tag{6}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are the sample mean and variance, respectively, of the investor's portfolio. CER represents the risk-free return that an investor is willing to accept instead of the given risky portfolio. The CER gain is the difference between the CERs of the two strategies. We report the annualized CER gain.

Furthermore, we document the annualized Sharpe ratios of the investment portfolio to directly evaluate the economic value of the factor timing strategies.

2.2.2 Unconditional Portfolio Choice

We start our optimal portfolio strategy to extract the dominant components of the 146 characteristic-based factors through principal component analysis (PCA) to aggregate information and deal with the large dimensionality in the age of the "factor zoo" (Cochrane, 2011), following Kozak, Nagel, and Santosh (2020) and Haddad, Kozak, and Santosh (2020). To avoid look-ahead bias for the out-of-sample results, we estimate eigenvectors and eigenvalues using the first half of the data from January 2004 to December 2011. Table 1 reports the percentage of variance explained by factor principal components (PCs). The first PC explains nearly one-third of the total variation, and the first four PCs jointly account for nearly two-thirds of the variation in realized returns. Following the assumption of Haddad, Kozak, and Santosh (2020), the harmonic mean of PCs' contribution to the total variance of returns should be higher than the ratio of total R^2 to maximum squared Sharpe ratio. Considering the variance explained in Table 1, we should use the first four or five sparse PCs as the asset space.

[Insert Table 1 about here]

Then, we concentrate on the first four PCs to create three different types of factor timing strategies. The unconditional strategy, also our benchmark strategy, simply sets the expected return in equation (5) as the averaged monthly return PC_i^U from January 2004 to December 2011, as follows:

$$E_t^U = E[PC_{1,t+1}^U, PC_{2,t+1}^U, PC_{3,t+1}^U, PC_{4,t+1}^U], \tag{7}$$

where $PC_{i,t+1}^U = PC_i^U$ represents the return of PC_i at month t+1 from January 2012 to December 2020.

2.2.3 Conditional Portfolio Choice

We propose two conditional portfolio choices with different methods to predict portfolio returns. The first is our primary strategy, a conditional factor timing strategy using a deep learning approach, the feed-forward neural network (FFN). FFN is highly flexible and performs well in handling complex problems in China's stock market (Leippold, Wang, and Zhou, 2021). Specifically, we set January 2004 to December 2011 as a training period and forecast the portfolio returns from January 2012 to December 2020,

$$PC_{i,t+1} = E_t\left[PC_{i,t+1}\right] + \varepsilon_{j,t+1}. \tag{8}$$

To measure the expected returns of each PC, we separately produce individual forecasts of each factor and apply these forecasts to PCs using their loadings on the dominant components. We predict each factor return with its own characteristics by a deep learning approach,

$$E_t\left[ret_{j,t+1}\right] = f\left[Z_{j,t}\right], \tag{9}$$

where $ret_{j,t+1}$ represents the return of factor j at month t + 1; for each factor j, $Z_{j,t}$ is calculated by combining the 146 stock characteristics according to portfolio weights; and $f(\cdot)$ is our FFN method. To construct the FFN, we use an "input layer" of predictors, several "hidden layers" that transform the predictors, and an "output layer" that aggregates hidden layers into a forecast. The general equation of each hidden layer states

$$z_k^l = g(b^{l-1} + (z^{l-1})'W^{l-1}), \tag{10}$$

where $g(\cdot)$ is the nonlinear "activation function" to take the aggregated signal from the previous layer and send it to the next layer. The final output is a linear transformation of the last hidden layer output,

$$G(z, b, W) = b^{L-1} + (z^{L-1})'W^{L-1}. \tag{11}$$

We apply the rectified linear unit (ReLU) as the nonlinear active function:

$$ReLU(z_k) = \max(z_k, 0). \tag{12}$$

We employ three hidden layers for FFN (NN3), with 32, 16, and 8 neurons, respectively[2].

Thus, the deep learning-based factor timing approach is as follows:

$$E_t^{NN3} = E_t[PC_{1,t+1}^{NN3}, PC_{2,t+1}^{NN3}, PC_{3,t+1}^{NN3}, PC_{4,t+1}^{NN3}], \tag{13}$$

[2] As the previous studies show neural network with higher layers may decrease the predictability in the finance market due to the "overfitting" curse (Leippold et al.,2022), we test the performance of NN models with one to five hidden layers (shown in the appendix) and focus on the NN model with the three hidden layers.

where $PC_{i,t+1}^{NN3}$ represents the expected return of PC_i at month $t+1$. For comparison, we impose another conditional factor timing strategy using the ordinary least squares (OLS) approach instead of NN3 to predict returns. In brief, the OLS-based factor timing approach is as follows:

$$E_t^{OLS} = E_t[PC_{1,t+1}^{OLS}, PC_{2,t+1}^{OLS}, PC_{3,t+1}^{OLS}, PC_{4,t+1}^{OLS}]. \tag{14}$$

Finally, following Haddad, Kozak, and Santosh (2020), we compute an estimate of the conditional covariance matrix of the PC returns in equation (5), which we for now assume is homoscedastic to estimate the role of forecasting means. In detail, we calculate the covariance of PCs from January 2004 to December 2011 with their real returns and apply the covariance matrix to the out-of-sample period.

To summarize, we create three factor timing strategies, one unconditional with purely PCs for the full sample, one uses NN3 and one employs OLS to forecast PCs for the out-of-sample. The timing of the factors, which also refers to choosing weights of each PC, is different mainly about expected factor returns in three strategies after 2011.

3 Empirical Results

3.1 Factor Timing Performance

We present the results of three factor timing approaches, the unconditional, the conditional OLS-, and the conditional NN3-based factor timing portfolio choices, of investors with mean-variance preference and the investors' risk aversion parameter γ equals to 2, 5, 10, and 20, respectively. Table 2 reports the annualized average returns and standard deviations of all three strategies and the Sharpe ratio of the conditional NN3 approach.

[Insert Table 2 about here]

Table 2 shows that the conditional NN3-based portfolio delivers the highest mean returns and the lowest variances in the three strategies under the various investors' risk aversion parameters. For example, when $\gamma = 2$, the average annualized return of the unconditional PCs is –0.009, increases to 0.118 for the conditional OLS portfolio, and reaches 0.594 for the conditional deep learning portfolio. The standard deviation of the return is the highest in the OLS strategy because OLS might overfit the samples. The deep learning method, however, generates a more balanced trade-off between fitting and predicting and produces the lowest variance in returns. For instance, the standard deviation is only 0.042 when $\gamma = 2$, lower than 0.045 for the unconditional strategy, and 0.049 for the OLS approach.

The annualized Sharpe ratios of the conditional deep learning portfolio are remarkably large, equal to 1.41, 1.48, 1.61, and 1.83 when γ = 2, 5, 10, and 20, respectively, which are higher than the conditional OLS and the unconditional portfolios. Our results suggest large economic benefits from timing factors with the deep learning method.

Table 2 also reports the CER gain required to make the investors indifferent between the conditional deep learning-based factor timing approach and the other two portfolio choices. We observe that the return forecasts of deep learning deliver large investment profits. In particular, the CER gain is 61.07% of the conditional NN3 portfolio minus the unconditional portfolio, implying that the investor is willing to pay an annual fee of up to 6107 basis points (bps) to access the predictive regression forecasts of the deep learning approach than the unconditional method in the situation that investor's risk aversion equals 2. The results remain sizable when γ changes. In addition, the CER gains for the deep learning method minus the conditional OLS are also sizable, ranging from 5.34% (γ = 20) to 54.88% (γ = 2). Less risk-averse investors have more timing preferences, so their CER gain is higher than that of more risk-averse investors.

Figure 1 plots the cumulative returns of the deep learning portfolios and the market returns[3] from January 2012 to December 2020. The result shows that the factor timing strategies generate considerable cumulative returns during the sample period. For example, when $\gamma = 2$, the cumulative returns achieve approximately 600%, almost 12 times larger than approximately 50% of the market returns. It reaches above 200% when

[3] The market return is calculated as the averaged stock returns with market value weighted.

$\gamma = 5$. In summary, the NN3-based factor timing strategy brings large investment profits without suffering large withdrawals.

[Insert Figure 1 about here]

Overall, our result indicates that the factor timing strategy based on deep learning outperforms the buy-and-hold strategy and the OLS-based portfolio and can create economic gains.

3.2 Deep Learning vs. Ordinary Least Squares Approach

In Section 3.1, we find that the deep learning method outperforms OLS because it generates larger average returns, lower standard deviation, and positive CER gains. The two methods are only different in the out-of-sample prediction. To investigate the reason for this difference, we study the out-of-sample performance of NN3 and OLS in this section.

Following Gu, Kelly, and Xiu (2020) and Bali et al. (2020), we employ the out-of-sample R-squared to evaluate the performance of predictability,

$$R_{OS}^2 = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(Ret_{j,t+1} - \widehat{Ret}_{j,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3} Ret_{j,t+1}^2}. \tag{15}$$

[Insert Table 3 about here]

Panel A of Table 3 presents the monthly R_{OS}^2 (in percentage) of PCs returns forecasting with the 146 characteristics-based factor portfolios. We observe that R_{OS}^2 of NN3 is always larger than that of the OLS in the first four PCs. For instance, the second column presents a positive R_{OS}^2 of 0.18 of the OLS model for PC1 and a higher 0.61 R_{OS}^2 of the deep learning method, suggesting a better performance of NN3 in generating significant out-of-sample portfolio forecasting. The other columns of Panel A of Table 3 show that the deep learning model improves the R_{OS}^2 in each PC, such that it improves the R_{OS}^2 from 0.03 to 0.10 in PC2 and from a negative value to a positive value in PC3 and PC4. The results show that the nonlinear models capture the time-varying components and perform better than the linear method in forecasting the PCs in the Chinese market.

Furthermore, we use the Diebold and Mariano (1995) test to investigate the deviations in the out-of-sample predictive accuracy between NN3 and OLS (Bali et al., 2020). The modified Diebold and Mariano (1995) statistic states that

$$DM_{12} = \bar{d}_{12}/\hat{\sigma}_{\bar{d}}, \tag{16}$$

where $\bar{d}_{12}$ is the time-series mean of $d_{12,t+1}$, and $\hat{\sigma}_{\bar{d}}$ is the Newey–West standard error of $d_{12,t+1}$ in the out-of-sample period. $d_{12,t+1}$ represents the difference in the forecast error between the two models,

$$d_{12,t+1} = \frac{1}{n_{t+1}} \sum_{i=1}^{n} \left((\hat{e}_{j,t+1}^{(OLS)})^2 - (\hat{e}_{j,t+1}^{(NN3)})^2 \right), \tag{17}$$

where $\hat{e}_{j,t+1}^{(OLS)}$ and $\hat{e}_{j,t+1}^{(NN3)}$ are the forecast errors for PC j at $t + 1$ produced by model OLS and NN3, respectively, and n_{t+1} is the number of PCs in the out-of-sample period. A positive statistic implies that the NN3 model outperforms the OLS.

Panel B of Table 3 reports the Diebold and Mariano (1995) test results comparing NN3 with OLS. The results show that the deep learning model generates positive and statistically significant DM results, ranging from 2.93 to 10.88, against the OLS in all four PCs, showing the improvement of the out-of-sample forecasting performance over OLS.

Deep learning methods are suitable and more effective for handling multicollinearity, uncovering complex patterns, aggregating information (Gu, Kelly, and Xiu, 2020), and predicting PCs more accurately than OLS. Our results are also consistent with Leippold, Wang, and Zhou (2022), who find that the Chinese stock market is predictable with machine learning tools.

3.3 Abnormal Alpha of Factor Timing

Considering the economic meaning of the deep learning-based factor timing portfolio, we investigate the abnormal alphas of the strategy with factor models in this section. We regress the return of the portfolios on the Fama and French (2015) five-factor model (FF5), FF5 augmented with momentum factor (MOMe; Carhart, 1997), and Liu, Stambaugh, and Yuan (2019) three-factor model in China (CH3).

[Insert Table 4 about here]

Table 4 shows that the unconditional and conditional OLS strategies are explained in several situations: the FF5 factor model explains the conditional strategy when $\gamma = 2$; the CH3 model digests the unconditional strategies when $\gamma = 2$, 5, and 10; and the FF5+MOMe factors subsume the conditional OLS portfolio when $\gamma = 2$. However, the three types of factor models cannot explain a single portfolio of the conditional NN3 strategy, which shows the robustness significance of the deep learning-based investment. The results suggest that the deep learning-based factor timing strategy cannot be explained by the factor models and improves the economic benefits of investors.

3.4 Transaction Costs

The transaction cost is always considered in real investment. In this section, we investigate the asset allocation results after deducting a proportional transaction cost of 50 basis points. We report the turnover rate (in percentage) and Sharpe ratio for the deep learning-based portfolio and the CER gain (in percentage) of return required to make the investors indifferent between the conditional NN3- and the unconditional approach with transaction cost.

The transaction cost of portfolio choice at time t is $c_t = z * T_t$, where z is the constant transaction cost in the cross-section and T_t is the turnover at time t. We assume z = 0.5%. Following Gu, Kelly, and Xiu (2020) and Avramov, Cheng, and Metzker (2022), the portfolio choice's monthly turnover at time t states as

$$T_t = \sum_i |w_{i,t} - \frac{w_{it-1} r_{i,t}}{w_{t-1} r_t}|, \tag{18}$$

where w_t ($w_{i,t}$) refers to the weight of the wealth invested in assets (element i in w_t) with grow return r_{t+1} ($r_{i,t+1}$), $w_t = (1 - \alpha_t, \alpha_t)$, and $r_{t+1} = (R_{f,t+1}, R_{t+1})$ corresponding to Equation (5).

The average monthly turnover ratio is $TURN = \frac{1}{T}\sum_{t=1}^{T} T_t$. Table 5 reports that the turnover rate ranges from 10.29% ($\gamma = 20$) to 102.94% ($\gamma = 2$).

[Insert Table 5 about here]

The finding shows that after deducting the transaction cost, the Sharpe ratios of the conditional NN3 portfolios are still large in magnitude, valued at 1.40, 1.47, 1.59, and 1.82 when γ = 2, 5, 10, and 20, respectively. Compared to the Table 2 results, Sharpe ratios drop only 0.01 under each degree of investor risk aversion, showing the robustness of benefits that are merely influenced by transaction costs. The CER gains are also positive for the conditional deep learning-based portfolio minus the unconditional portfolio. The decrease in the CER gain between considering transaction costs and not considering them is rather small.

Our findings show that the deep learning-based factor timing portfolio still generates remarkable economic gains when including the transaction cost.

3.5 Market Crash

In 2015, the Chinese stock market experienced a tremendous rise followed by a sharp drop (Liu, Gu, and Xing (2016)). To investigate whether factor timing exists during the crash period, we report the performance of the NN3-based strategy from January 2014 to December 2017.

Table 6 shows that the conditional NN3-based factor timing strategy suits the subsample well and continues to exist and remains robust for the market crash moment. It generates an average return of 0.642 and a high Sharpe ratio of 1.07 (when $\gamma = 2$). CER gains are also positive, indicating that investors are still willing to pay higher fees for the deep learning-based investment strategy. CER gains between the deep learning- and OLS strategies are higher than in the full-time period, as the high-risk situation may enhance the overfitting issue of the OLS. Notably, compared to Table 2, the standard deviations of the returns are higher during the market crash period, resulting in lower Sharpe ratios than the full sample results.

[Insert Table 6 about here]

4 Economic Grounds

Machine learning methods, especially deep learning models, are opaque in nature and often referred to as "black boxes" (Gu, Kelly, and Xiu, 2020). It is essential to understand the economic mechanisms of the deep learning-based factor timing strategy. In this section, we provide the economic grounds of our NN3-based factor timing approach. First, we investigate the important characteristics-based factors in each PC and plot the weights of PCs in the NN3-based factor timing portfolio in the out-of-sample period. Then, we provide the mispricing-based explanation for the factor timing performance, from size, investor sentiment, and aggregate idiosyncratic volatility.

4.1 The Loadings of Factors in Principal Components

Given the large number of predictors, are certain factors more important than others? To obtain an overview of this question, we use block diagrams to illustrate the relative importance of the 146 factors in each PC. We calculate the factor weights of each PC and normalize them from 0 to 1 from January 2004 to December 2011, the same as the training period. Each line reports a factor's weight, and each column corresponds to a PC, where the color gradient indicates the degree of importance from the highest to the lowest (darkest to lightest).

[Insert Figure 2 about here]

Figure 2 plots the relative importance of the characteristics in each PC component. We find that each PC has a different set of crucial traits. The most important factors in PC1 are mostly quality-related characteristics, such as gp_atl1, mispricing_perf, and niq_at. Ebit_bev plays an important role in PC1, which stands for profitability. Additional critical factors are long-term reversal and z score in PC1. In PC2, valuation factors such as be_me and bev_mev are important. Firm quality, for example, the number of consecutive quarters with earnings increases (ni_inc8q), is one of the most important factors in PC2. Another important factor is zero_trades_21d, which represents low risk and illiquidity. The most important factors in PC3 are factor exposure-related and low-risk variables, such as beta_60 m and betabab_1260d. In addition, corr_1260d, which refers to seasonality, and sale_me, which represents value, are also important in PC3. The most important factors in PC4 are quality-related characteristics such as sale_bev and seasonality-related factors such as seas_11_15na. Overall, in the first three PCs, most of the 146 factors contribute to the PC components, whereas only a small number of factors weigh in PC4. The quality- and profitability-related variables play an important role in the first four PCs in China's stock market.

We also sort the importance of the characteristics by the 13 themes. Figure 3 shows that the quality-, profitability-, investment-, and momentum-related themes are the most important in PC1. Value plays the greatest role in PC2 and PC3. Accruals-, profit growth-, and momentum-related themes are crucial in PC4.

[Insert Figure 3 about here]

Then, as our factor timing strategy uses 4 PCs from January 2004 to December 2011 and weights each of the PCs from January 2012 to December 2020, we plot each PC's weight in it from January 2012 to December 2020, the out-of-sample period, in Figure 4, to investigate the NN3-based factor timing portfolio components.

[Insert Figure 4 about here]

We observe that the PC weights are time varying from 2012 to 2020. PC1 includes the quality-, profitability-, investment-, and leverage-related factors and contributes more in early 2019 and late 2014. PC2 includes value-, quality-, and low risk-related factors and delivers larger weights during 2013 and 2020. PC3 covers low risk-, seasonality-, and value-related factors and offers higher weights in early 2013 and 2015. PC4 contains quality- and seasonality-related factors and contributes more during 2018.

Our strategy captures the risk premium of different factors in a complicated market environment. For example, when faced with the stock market crash in 2015 and COVID-19 in 2020, PC1 dominated by profit- and quality-related factors and PC4 monopolized by quality-related characteristics have rather lower weights, while PC2 concentrated by valuation, low-risk, and illiquidity factors and PC3 ruled by beta and low-risk factors perceive higher weights. The difference in the time-varying portfolio weights for each PC highlights the effectiveness of our method of combining deep learning with large-dimensional factors.

4.2 Mispricing-based Explanation

Inspired by Avramov, Cheng, and Metzker (2022), we investigate the profitability sources of deep learning signals based on mispricing.

4.2.1. Small and Big Stocks

First, we investigate the performance of the NN3-based factor timing strategy among small and large stocks. The literature highlights that anomalies are concentrated in microcap stocks (e.g., Fama and French, 1993; Novy-Marx and Velikov, 2016; Hou, Xue, and Zhang, 2020). Moreover, small stocks may suffer more illiquidity and limits-to-arbitrage. We conjecture that the factor timing strategy performs better among small stocks. We divide stocks into small and large stocks according to the average size and construct the deep learning-based factor timing strategy among subsamples of small and large stocks.

[Insert Table 7 about here]

Panel A in Table 7 reports that the NN3-based factor timing portfolio produces higher returns among small stocks with higher Sharpe ratios but larger standard deviations. The average returns of the NN3-based factor timing strategy among small stocks are 0.629, 0.252, 0.127, and 0.065 when γ = 2, 5, 10, and 20, respectively, whereas they are 0.507, 0.204, 0.103, and 0.053 among large stocks, respectively. The Sharpe ratios of the NN3-based factor timing strategy among small stocks are 0.88, 0.89, and 0.90 when γ = 2, 5, and 10, which are 0.04, 0.05, and 0.05 higher among large stocks, respectively. The CER gains remain economically large under four levels of risk aversion. Investors are willing to pay a higher annual fee to gain the returns among small stocks compared to the large stocks. For instance, the CER gain is 1.94% of investing in the conditional NN3 portfolio among small stocks minus that among large stocks when the risk aversion is 2. Our findings suggest that the NN3-based factor timing strategy generates larger returns and higher Sharpe ratios among stocks with smaller capitalization.

4.2.2. Investor Sentiment

Furthermore, we investigate the role of investor sentiment in the factor timing performance. According to Stambaugh, Yuan, and Yu (2012), many anomalies generate higher returns during high sentiment periods because anomalies are particularly persistent in the presence of short-sale constraints due to mispricing. We suppose that the factor timing is also stronger during high investor sentiment periods. We use IPO first-day returns to measure investor sentiment and calculate the average annualized returns, standard deviation, and Sharpe ratios of the NN3-based factor timing 146 factors during high (i.e., above mean) and low (i.e., below mean) investor sentiment periods. We also calculate the CER gain between high and low investor sentiment periods.

Panel B in Table 7 reports the results of the role of investor sentiment in deep learning-based factor timing. The NN3-based factor timing strategy delivers higher returns, Sharpe ratios, and CER gains during the high-sentiment period. For instance, when $\gamma = 2$, the average underperforming strategy earns a return of 0.510 (0.649) and a Sharpe ratio of 1.22 (1.50) during the low-sentiment (high-sentiment) period; the CER gain to make the benefit indifferent from the high- and low-sentiment periods is 12.66%. The pattern remains when $\gamma = 5$, 10, and 20. Our results are consistent with Stambaugh, Yu, and Yuan (2012), who found that the deep learning-based factor timing approach is driven by investor sentiment to some extent.

4.2.3. Aggregate Idiosyncratic Volatility

In this section, we examine the role of idiosyncratic volatility in producing factor timing signals. During periods of high aggregate idiosyncratic volatility (IVOL), arbitrageurs are more likely to take advantage of high information asymmetry. In addition, a high IVOL may cause a large limits-to-arbitrage. Thus, we conjecture that the factor timing performance is stronger among high IVOL periods. Following Garcia, Mantilla-García, and Martellini (2014) and Li, Yuan, and Zhou (2022), we calculate the monthly return dispersion (RD) to estimate aggregate IVOL, which is the cross-sectional standard deviation of stock daily returns. We use the mean of RD as the index and calculate the average annualized returns, standard deviation, and Sharpe ratios of the NN3-based factor timing strategy, according to whether it is above or below the mean of RD, along with the CER gain.

Panel C in Table 7 reports the results for the factor timing strategies during high and low aggregate IVOL periods. Consistent with our conjecture, the factor timing effect is stronger in the high aggregate IVOL environment. When $\gamma = 2$, 5, 10, and 20, the NN3-based factor timing strategy delivers annualized returns of 0.753, 0.310, 0.163, and 0.089, respectively, during the high aggregate IVOL period and is stronger than the low periods of 0.534, 0.229, 0.127, and 0.076, respectively. The Sharpe ratios of the factor timing strategy reach higher values during high RD periods, of 1.80, 1.86, 1.95, and 2.13, when γ = 2, 5, 10, and 20, respectively, whereas they reach 1.22, 1.31, 1.45, and 1.74 during low RD periods, respectively. The standard deviation is also smaller in the high IVOL periods. The results of the CER gain suggest that investors are willing to pay a higher annual fee to gain the returns of investing in high IVOL periods compared to the low IVOL period in different degrees of risk aversion.

Overall, our results suggest that the deep learning method is suitable for China, as it performs well among different subsamples (Leippold, Wang, and Zhou, 2022). Moreover, the deep learning-based factor timing signal in China is influenced by mispricing (Avramov, Cheng, and Metzker, 2022).

5 Conclusion

After creating a 146 characteristics-based factor set in the Chinese stock market, we use the first four principal components of the factors and build a conditional factor timing strategy based on a three-hidden-layer feed-forward neural network. The deep learning-based factor timing portfolio generates the highest average returns, the lowest standard deviations, and the highest Sharpe ratios, compared with the unconditional and the alternative OLS-based strategies, under the mean-variance investor preference, regardless of the different degrees of investors' risk aversion. We find that our factor timing strategy delivers large economic gains for investors with large annualized CER gains. The traditional factor models cannot absorb the returns of the conditional NN3-based factor timing approach, and the performance is robust considering the transaction costs.

We investigate the economic grounds of the conditional NN3-based portfolio. First, we find that quality- and profitability-related factors are the most important ones in PCs. In addition, each PC's weight in the NN3-based out-of-sample portfolio is time-varying with economic fluctuations. In addition, mispricing-based theory helps explain the factor timing strategy because it performs better among smaller stocks, in the higher investor sentiment and higher aggregate idiosyncratic volatility periods, all are accompanied by higher mispricing. We anticipate that these economic theories will be helpful in providing direction for future research.

References

Aït-sahali, Y., & Brandt, M. W. (2001). Variable selection for portfolio choice. The Journal of Finance, 56(4), 1297-1351.

Ait-Sahalia, Y., Fan, J., & Li, Y. (2013). The leverage effect puzzle: Disentangling sources of bias at high frequency. Journal of Financial Economics, 109(1), 224-249.Haddad, V., Kozak, S., & Santosh, S. (2020). Factor timing. The Review of Financial Studies, 33(5), 1980-2018.

Akbas, F., Armstrong, W. J., Sorescu, S., & Subrahmanyam, A. (2016). Capital market efficiency and arbitrage efficacy. Journal of Financial and Quantitative Analysis, 51(2), 387-413.

Arnott, R. D., Clements, M., Kalesnik, V., & Linnainmaa, J. T. (2021). Factor momentum. Available at SSRN 3116974.

Avramov, D., Cheng, S., & Metzker, L. (2022). Machine learning vs. economic restrictions: Evidence from stock return predictability. Management Science.

Avramov, D., Cheng, S., Schreiber, A., & Shemer, K. (2017). Scaling up market anomalies. The Journal of Investing, 26(3), 89-105.

Bali, T. G., Goyal, A., Huang, D., Jiang, F., & Wen, Q. (2020). Predicting Corporate Bond Returns: Merton Meets Machine Learning. Georgetown McDonough School of Business Research Paper No. 3686164, Swiss Finance Institute Research Paper No. 20-110

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

Carpenter, J. N., Lu, F., & Whitelaw, R. F. (2021). The real value of China's stock market. Journal of Financial Economics, 139(3), 679-696.

Chen, L., Pelger, M., Zhu, J., 2020. Deep learning in asset pricing. Available at SSRN 3350138

Cochrane, J. H. (2011). Presidential address: Discount rates. The Journal of finance, 66(4), 1047-1108.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics, 13(3), 253-263.

Dong, X., Li, Y., Rapach, D. E., & Zhou, G. (2022). Anomalies and the expected market return. The Journal of Finance, 77(1), 639-681.

Ehsani, S., & Linnainmaa, J. T. (2022), Factor momentum and the momentum factor. The Journal of Finance, 77: 1877-1919.

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

Favero, C. A., Melone, A., & Tamoni, A. (2021). Macro trends and factor timing. Available at SSRN 3940452.

Feng, G., He, J., & Polson, N. G. (2018). Deep learning for predicting asset returns. arXiv preprint arXiv:1804.09314.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. The Review of Financial Studies, 33(5), 2326-2377.

Garcia, R., Mantilla-García, D., & Martellini, L. (2014). A model-free measure of aggregate idiosyncratic volatility and the prediction of market returns. Journal of Financial and Quantitative Analysis, 49(5-6), 1133-1165.

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

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

Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. Journal of Econometrics, 222(1), 429-450.

Gupta, T., & Kelly, B. (2019). Factor momentum everywhere. The Journal of Portfolio Management, 45(3), 13-36.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the cross-section of expected returns. The Review of Financial Studies, 29(1), 5-68.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. The Review of Financial Studies, 33(5), 2019-2133.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. The Review of Financial Studies, 33(5), 2019-2133.

Jacobs, H. (2015). What explains the dynamics of 100 anomalies?. Journal of Banking & Finance, 57, 65-85.

Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2021). Is there a replication crisis in finance? (No. w28432). National Bureau of Economic Research.

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics, 134(3), 501-524.

Keloharju, M., Linnainmaa, J. T., & Nyberg, P. (2016). Return seasonalities. The Journal of Finance, 71(4), 1557-1590.

Kozak, S., Nagel, S., & Santosh, S. (2018). Interpreting factor models. The Journal of Finance, 73(3), 1183-1223.

Kozak, S., Nagel, S., & Santosh, S. (2020). Shrinking the cross-section. Journal of Financial Economics, 135(2), 271-292.

Leippold, M., & Yang, H. (2021). The anatomy of factor momentum. Available at SSRN, 3517888.

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

Li, S. Z., Yuan, P., & Zhou, G. (2022). Risk-based Momentum. Available at SSRN, 4062260

Liao, C., Luo, Q., & Tang, G. (2021). Aggregate liquidity premium and cross-sectional returns: Evidence from China. Economic Modelling, 104, 105645.

Liu, J., Stambaugh, R. F., & Yuan, Y. (2019). Size and value in China. Journal of Financial Economics, 134(1), 48-69.

Ma, T., Liao, C., & Jiang, F. (2021). Factor Momentum in the Chinese Stock Market. Available at SSRN 4148445.

Nagel, S. (2021). Machine learning in asset pricing. In Machine Learning in Asset Pricing. Princeton University Press.

Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. The Review of Financial Studies, 29(1), 104-147.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. Journal of Financial Economics, 104(2), 288-302.

Tang, G., Jiang, F., Qi, X., & Huang, N. (2021). It takes two to tango: Fundamental timing in stock market. International Journal of Finance & Economics, 26(4), 5259-5277.

Table 1 Percentage of variance explained by factor PCs

This table reports the percentage of variance explained by each PC of the 146 factor strategies.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
% var. explained	34.0	13.3	7.4	5.5	4.2	3.6	3.0	2.6	2.1	2.0
Cumulative	34.0	47.3	54.7	60.2	64.4	68.0	71.0	73.6	75.7	77.7

Table 2 Unconditional and conditional portfolios for mean-variance preferences

This table reports the performance of predictive portfolio choice of mean-variance investors with various choices in timing the 146 factors, including the annualized average returns and standard deviations of all strategies and the Sharpe ratio of conditional NN3 strategy. The first column reports the investor's risk aversion parameters. The last column reports the CER (in percentage) of return required to make the investors indifferent between the conditional NN3 and the other two portfolio choices.

γ	Uncond.			Cond. OLS			Cond. NN3			CER (%)	
	Mean	Std	SR	Mean	Std	SR	Mean	Std	SR	Uncond.	Cond. OLS
2	-0.01	0.43	-0.02	0.12	0.50	0.24	0.59	0.42	1.41	61.07	54.88
5	0.01	0.17	0.06	0.06	0.20	0.30	0.25	0.17	1.48	24.71	21.84
10	0.01	0.09	0.11	0.04	0.10	0.40	0.14	0.08	1.61	12.46	10.97
20	0.02	0.05	0.40	0.03	0.07	0.43	0.08	0.04	1.83	6.36	5.34

Table 3 Out-of-sample R^2 of OLS and NN3 in each PC

This table reports the out-of-sample R-squared and Diebold-Mariano test statistics for each PC. The out-of-sample R-squared (R^2_{OS}, in percentage) is calculated with OLS and a forward-feed neural network (NN3). R^2_{OS} states that

$$R^2_{OS} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}(Ret_{j,t+1} - \widehat{Ret}_{j,t+1})^2}{\sum_{(i,t)\in\mathcal{T}_3} Ret^2_{j,t+1}}.$$

The Diebold-Mariano test statistics compare the out-of-sample prediction performance (R^2_{OS}) among NN3 and OLS. Positive DM statistics indicate that the NN3 model outperforms the OLS. The full sample covers the periods from January 2004 to December 2020, the training period from January 2004 to December 2011, and the test period from January 2012 to December 2020.

	PC1	PC2	PC3	PC4
Panel A: R^2_{OS}				
OLS	0.18	0.03	-0.31	-0.29
NN3	0.61	0.10	0.03	0.05
Panel B: DM-test				
DM-test	2.76	3.57	10.88	2.93

Table 4 Alpha in unconditional and conditional portfolios

This table reports the abnormal alphas, including FF5 alpha, FF5 augmented with momentum (MOMe) alpha, and CH3 alpha, of the unconditional portfolio, conditional OLS portfolio and conditional NN3 portfolio. The first column reports the investor's risk aversion parameters. Newey–West *t*-statistics are reported in square brackets.

Panel A: Uncond.	FF5	FF5+MOMe	CH3
γ=2	0.056 [1.57]	0.057 [1.71]	0.010 [0.58]
γ=5	0.011 [1.77]	0.011 [1.87]	0.009 [0.69]
γ=10	0.004 [1.94]	0.004 [1.91]	0.005 [0.98]
γ=20	0.002 [2.11]	0.002 [2.23]	0.006 [2.13]
Panel B: OLS	**FF5**	**FF5+MOMe**	**CH3**
γ=2	0.024 [1.69]	0.025 [1.46]	0.015 [1.65]
γ=5	0.011 [1.91]	0.011 [1.68]	0.009 [1.88]
γ=10	0.006 [1.97]	0.006 [1.81]	0.008 [1.98]
γ=20	0.004 [1.99]	0.004 [2.18]	0.005 [2.43]
Panel C: NN3	**FF5**	**FF5+MOMe**	**CH3**
γ=2	0.052 [3.94]	0.052 [3.86]	0.055 [4.23]
γ=5	0.022 [4.13]	0.022 [4.06]	0.023 [4.43]
γ=10	0.012 [4.46]	0.012 [4.39]	0.012 [4.77]
γ=20	0.007 [5.12]	0.007 [5.03]	0.007 [5.44]

Table 5 Portfolio performance with transaction cost

The table reports the results of the conditional NN3-based predictive portfolio choice with the constant transaction cost. The proportional transaction cost is 0.5%, constant across stocks and over time. The first column reports the investor's risk aversion parameters. We report the turnover rate (in percentage) and Sharpe ratio for the conditional NN3 strategy. The last column reports the CER (in percentage) of return required to make the investors indifferent between the conditional NN3 and the unconditional strategy.

γ	Conditional NN3		CER (%)
	Turnover (%)	SR	
2	102.94	1.40	60.55
5	41.17	1.47	24.51
10	20.58	1.59	12.35
20	10.29	1.82	6.31

Table 6 Portfolio performance in the market crash period

This table reports the performance of unconditional and conditional portfolios in the market crash period from January 2014 to December 2017. The first column reports the investor's risk aversion parameters. The last column reports the CER (in percentage) of return required to make the investors indifferent between the conditional NN3 and the other two portfolio choices.

γ	Uncond.			Cond. OLS			Cond. NN3			CER (%)	
	Mean	Std	SR	Mean	Std	SR	Mean	Std	SR	Uncond.	Cond. OLS
2	0.04	0.56	0.07	0.12	0.66	0.17	0.64	0.60	1.07	55.68	60.65
5	0.03	0.22	0.13	0.06	0.26	0.23	0.27	0.24	1.13	22.48	24.12
10	0.03	0.11	0.22	0.04	0.13	0.32	0.15	0.12	1.23	11.40	12.08
20	0.02	0.06	0.41	0.03	0.07	0.50	0.09	0.06	1.43	5.88	6.06

Table 7 Portfolio performance in subsamples

The table reports the results of the conditional NN3-based predictive portfolio choice among subsamples. Panel A reports results in small and large stocks, Panel B reports results in low and high sentiment periods and Panel C reports results in low aggregate idiosyncratic volatility (IVOL) and high IVOL periods. Investor sentiment is defined as the first-day return of an IPO. We calculate the monthly return dispersion (RD) to estimate aggregate IVOL, which is the cross-sectional standard deviation of stock daily returns. We use the mean of RD to represent the aggregate IVOL. The first column reports the investor's risk aversion parameters. The last column reports the CER (in percentage) of return required to make the investors indifferent between investing in the large and the small stocks.

Panel A	Big			Small			CER (%)
	Mean	Std	SR	Mean	Std	SR	Small-Big
γ=2	0.507	0.606	0.84	0.629	0.713	0.88	1.94%
γ=5	0.204	0.242	0.84	0.252	0.284	0.89	0.66%
γ=10	0.103	0.121	0.85	0.127	0.142	0.90	0.37%
γ=20	0.053	0.059	0.90	0.065	0.073	0.90	0.62%
Panel B	**Low Sentiment**			**High Sentiment**			**CER (%)**
	Mean	Std	SR	Mean	Std	SR	High-Low
γ=2	0.510	0.418	1.22	0.649	0.433	1.50	12.66%
γ=5	0.220	0.167	1.32	0.269	0.172	1.56	4.41%
γ=10	0.124	0.084	1.48	0.142	0.086	1.66	1.65%
γ=20	0.075	0.042	1.80	0.080	0.043	1.86	0.36%
Panel C	**Low IVOL**			**High IVOL**			**CER (%)**
	Mean	Std	SR	Mean	Std	SR	High-Low
γ=2	0.534	0.438	1.22	0.753	0.418	1.80	23.56%
γ=5	0.229	0.175	1.31	0.310	0.167	1.86	8.80%
γ=10	0.127	0.088	1.45	0.163	0.084	1.95	3.94%
γ=20	0.076	0.044	1.74	0.089	0.042	2.13	1.47%

Figure 1 Accumulative return in conditional portfolios

This figure plots the accumulative returns of the deep learning portfolios with the market returns. The sample period is from January 2012 to December 2020.

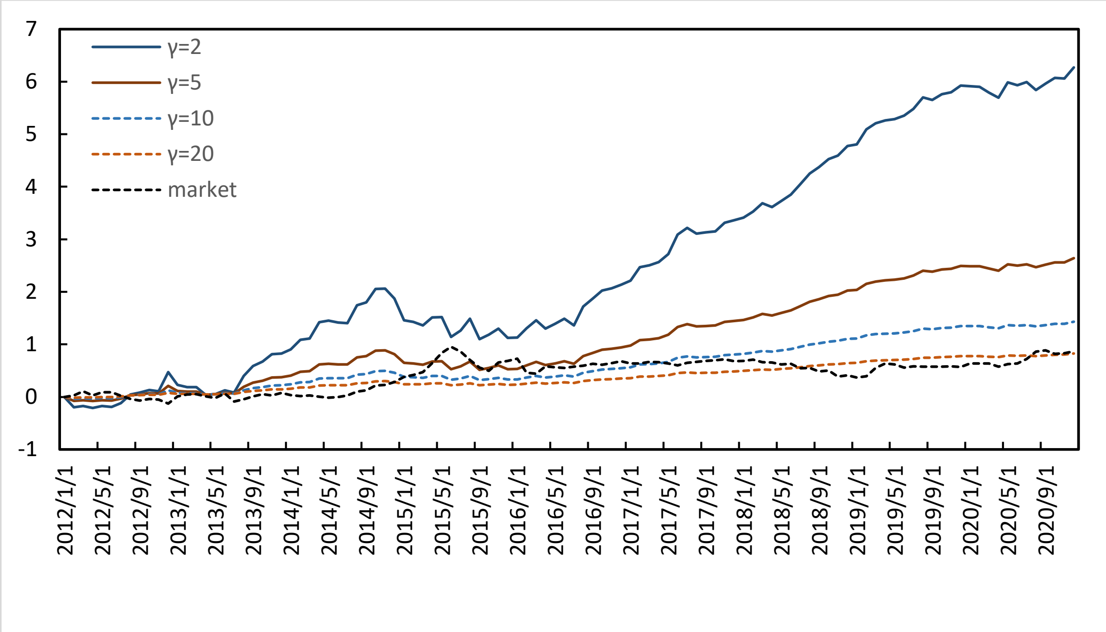


Figure 2 Characteristic importance of each PC

This figure plots the contribution of 146 factors to the first four PCs. The vertical axis reports the characteristics. Columns correspond to the individual PCs. The color gradients within each column indicate the most influential (dark blue) to the least influential (white) variables. The sample period is from January 2004 to December 2011.

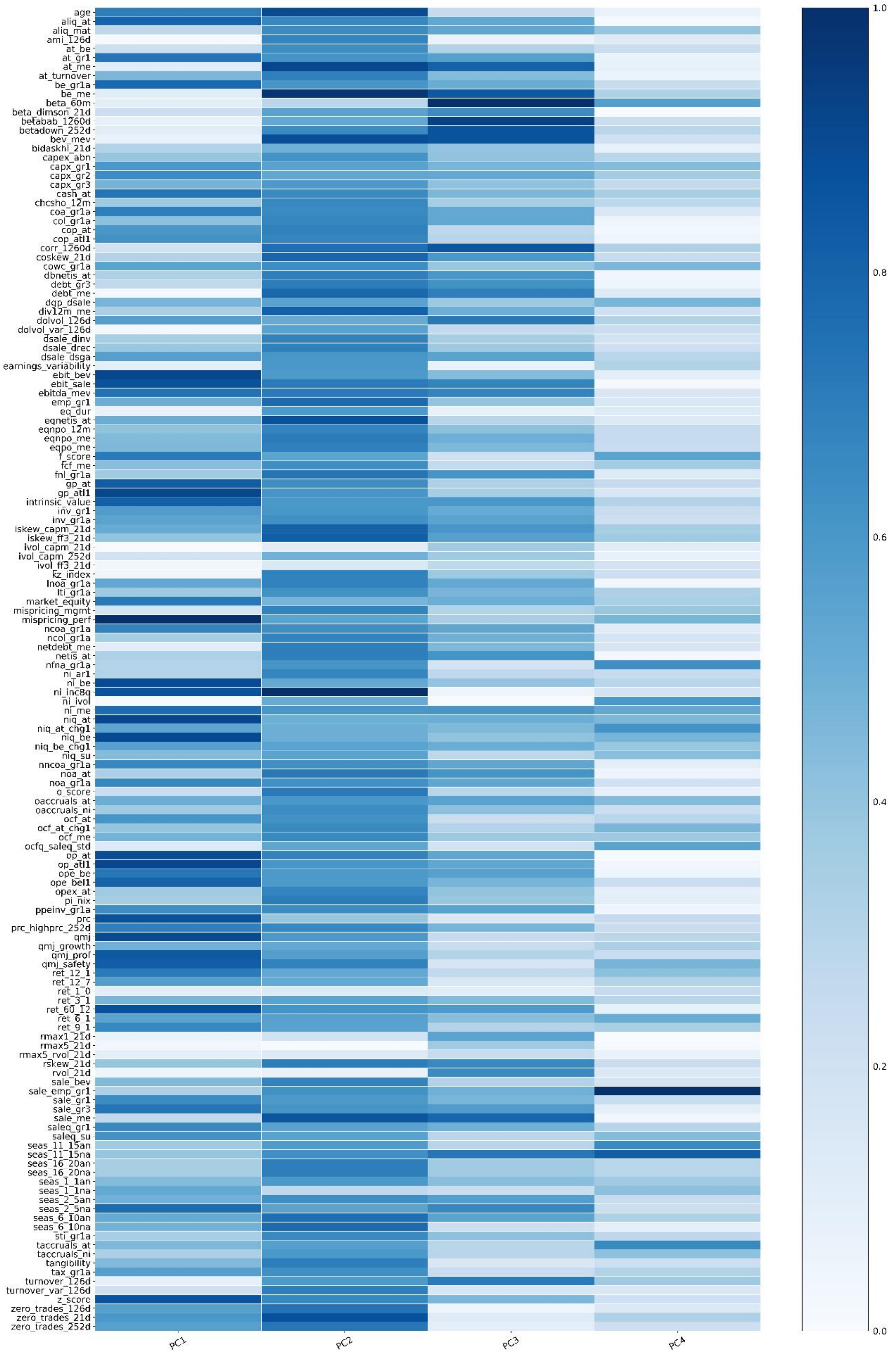


Figure 3 Characteristic theme importance of each PC

This figure plots the contribution of 13 factor themes to the first four PCs. The vertical axis reports the characteristics. Columns correspond to the individual PCs. The color gradients within each column indicate the most influential (dark blue) to the least influential (white) variables. The sample period is from January 2004 to December 2011.

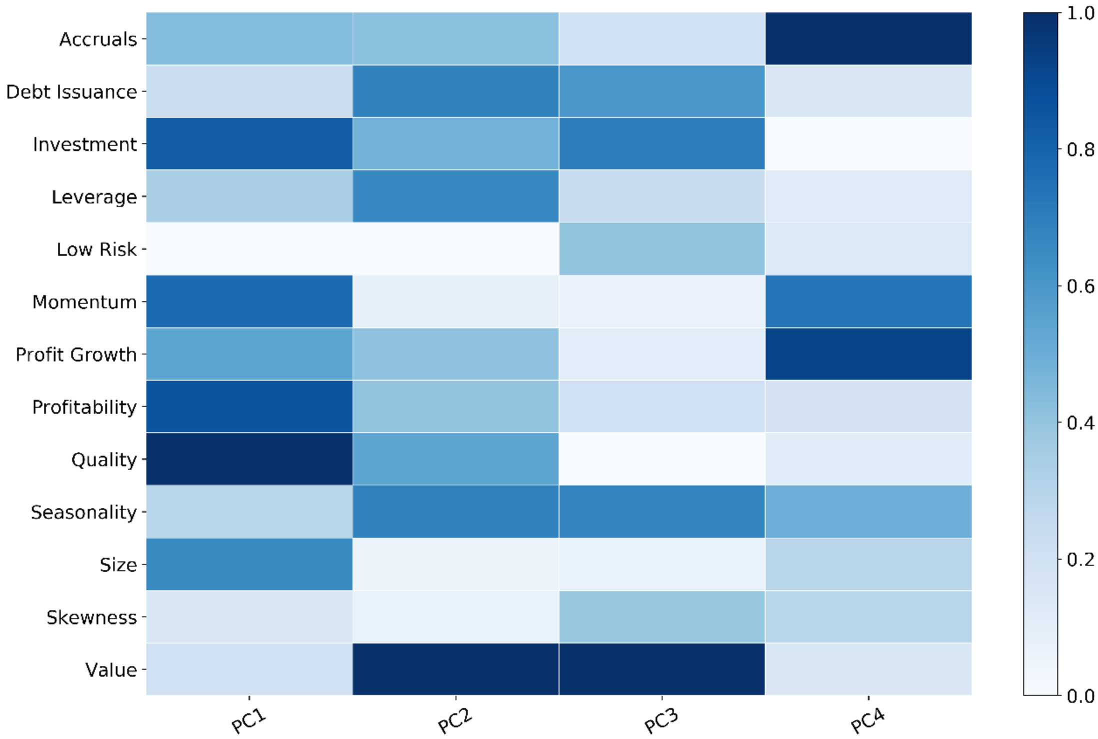
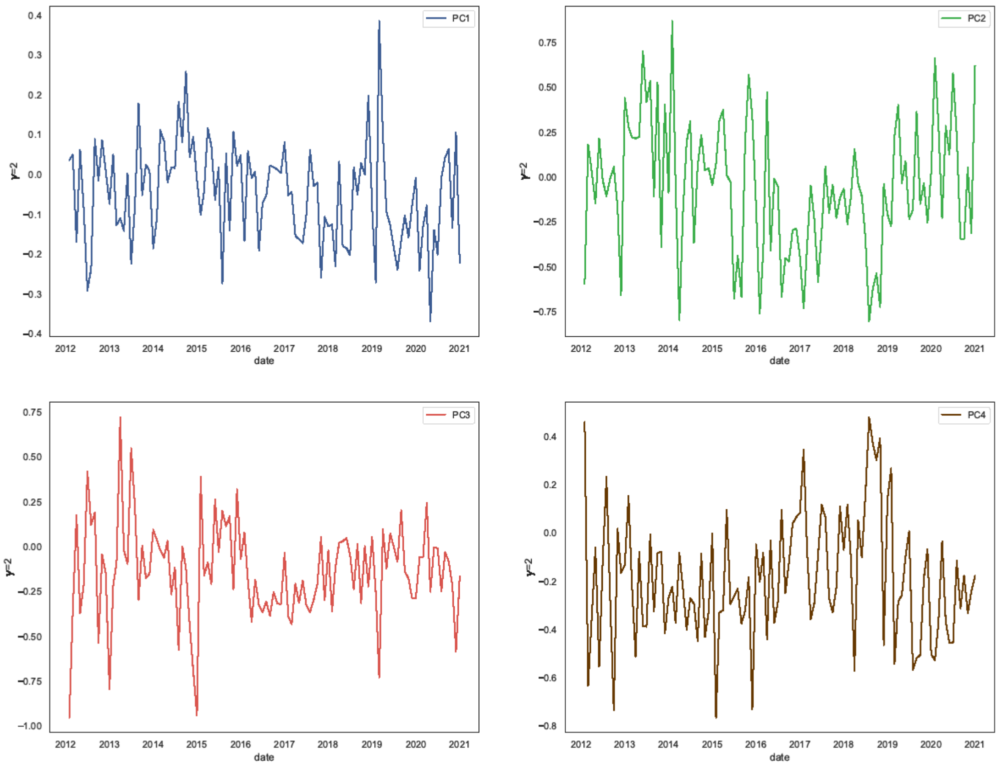


Figure 4 Each PC's weight

This figure plots the weight of each PC of the first four PCs in the deep learning-based factor timing portfolio from January 2012 to December 2020. The vertical axis reports the weight. characteristics. The horizontal axis reports time.



Online Appendix

Table A Factor return and construction

This table reports the average annualized excess return (in %), which is the annualized returns minus the risk-free rate, of each characteristic-based factor from January 2004 to December 2020, along with the details and references of each factor. We construct the decile portfolios based on the monthly characteristics. The columns “Low”, “High”, and “H-L” represent decile 1 (which earns the lowest expected return), decile 10 (which earns the highest expected return), and the long-minus-short strategy, respectively.

Factors	Low	High	H-L	Detail	Reference
Panel A Accruals					
cowc_gr1a	1.20	1.50	0.30	Change in current operating working capital	Richardson et al. (2005)
oaccruals_at	1.15	1.55	0.40	Operating accruals	Sloan (1996)
oaccruals_ni	0.87	1.40	0.53	Percent operating accruals	Hafzalla Lundholm and Van Winkle (2011)
taccruals_at	1.28	1.48	0.20	Total accruals	Richardson et al. (2005)
taccruals_ni	1.00	1.20	0.20	Percent total accruals	Hafzalla Lundholm and Van Winkle (2011)
Panel B Debt Issuance					
capex_abn	0.95	1.24	0.29	Abnormal corporate investment	Titman Wei and Xie (2004)
debt_gr3	1.07	1.78	0.71	Growth in book debt (3 years)	Lyandres Sun and Zhang (2008)
fnl_gr1a	1.04	1.64	0.61	Change in financial liabilities	Richardson et al. (2005)
ncol_gr1a	1.12	1.26	0.14	Change in noncurrent operating liabilities	Richardson et al. (2005)
nfna_gr1a	1.14	1.57	0.44	Change in net financial assets	Richardson et al. (2005)
noa_at	1.05	1.85	0.80	Net operating assets	Hirshleifer et al. (2004)
Panel C Investment					
aliq_at	1.38	1.41	0.04	Liquidity of book assets	Ortiz-Molina and Phillips (2014)
at_gr1	1.49	1.68	0.19	Asset Growth	Cooper Gulen and Schill (2008)
be_gr1a	0.99	1.33	0.34	Change in common equity	Richardson et al. (2005)

capx_gr1	1.16	1.15	0.00	CAPEX growth (1 year)	Xie (2001)
capx_gr2	1.05	1.25	0.20	CAPEX growth (2 years)	Anderson and Garcia-Feijoo (2006)
capx_gr3	1.19	1.27	0.08	CAPEX growth (3 years)	Anderson and Garcia-Feijoo (2006)
coa_gr1a	1.33	1.37	0.04	Change in current operating assets	Richardson et al. (2005)
col_gr1a	1.18	1.57	0.39	Change in current operating liabilities	Richardson et al. (2005)
emp_gr1	1.03	1.62	0.59	Hiring rate	Belo Lin and Bazdresch (2014)
inv_gr1	1.20	1.25	0.05	Inventory growth	Belo and Lin (2011)
inv_gr1a	1.25	1.29	0.04	Inventory change	Thomas and Zhang (2002)
lnoa_gr1a	1.22	1.57	0.35	Change in long-term net operating assets	Fairfield Whisenant and Yohn (2003)
mispricing_mgmt	1.15	1.60	0.45	Mispricing factor: Management	Stambaugh and Yuan (2016)
ncoa_gr1a	1.27	1.71	0.44	Change in noncurrent operating assets	Richardson et al. (2005)
nncoa_gr1a	1.26	1.66	0.40	Change in net noncurrent operating assets	Richardson et al. (2005)
noa_gr1a	1.28	1.65	0.37	Change in net operating assets	Hirshleifer et al. (2004)
ppeinv_gr1a	1.27	1.50	0.23	Change PPE and Inventory	Lyandres Sun and Zhang (2008)
ret_60_12	0.68	2.48	1.81	Long-term reversal	De Bondt and Thaler (1985)
sale_gr1	1.12	1.40	0.28	Sales Growth (1 year)	Lakonishok Shleifer and Vishny (1994)
sale_gr3	1.24	1.46	0.22	Sales Growth (3 years)	Lakonishok Shleifer and Vishny (1994)
saleq_gr1	0.85	1.86	1.01	Sales growth (1 quarter)	
seas_2_5na	0.82	1.93	1.11	Years 2-5 lagged returns, nonannual	Heston and Sadka (2008)
Panel D Leverage					
age	1.19	2.45	1.25	Firm age	Jiang Lee and Zhang (2005)
aliq_mat	0.75	1.83	1.08	Liquidity of market assets	Ortiz-Molina and Phillips (2014)
at_be	1.23	1.39	0.16	Book leverage	Fama and French (1992)
bidaskhl_21d	-0.67	6.01	6.68	The high-low bid-ask spread	Corwin and Schultz (2012)
cash_at	1.02	1.34	0.32	Cash-to-assets	Palazzo (2012)
netdebt_me	0.18	0.25	0.07	Net debt-to-price	Penman Richardson and Tuna

					(2007)
ni_ivol	1.14	2.12	0.98	Earnings volatility	Francis et al. (2004)
tangibility	0.95	1.45	0.50	Asset tangibility	Hahn and Lee (2009)
z_score	0.96	1.34	0.37	Altman Z score	Dichev (1998)
Panel E Low Risk					
beta_60 m	1.12	1.83	0.71	Market Beta	Fama and MacBeth (1973)
beta_dimson_21d	2.65	5.00	2.34	Dimson beta	Dimson (1979)
betabab_1260d	1.27	1.49	0.21	Frazzini-Pedersen market beta	Frazzini and Pedersen (2014)
betadown_252d	0.94	1.65	0.70	Downside beta	Ang Chen and Xing (2006)
chcsho_12 m	1.18	1.29	0.10	Net stock issues	Pontiff and Woodgate (2008)
earnings_variability	1.20	1.72	0.52	Earnings variability	Francis et al. (2004)
eqnetis_at	-0.04	1.16	1.20	Net equity issuance	Bradshaw Richardson and Sloan (2006)
fcf_me	-0.08	0.65	0.73	Free cash flow-to-price	Lakonishok Shleifer and Vishny (1994)
ivol_capm_21d	-2.19	1.17	3.37	Idiosyncratic volatility from the CAPM (21 days)	
ivol_capm_252d	0.27	1.44	1.17	Idiosyncratic volatility from the CAPM (252 days)	Ali Hwang and Trombley (2003)
ivol_ff3_21d	-2.53	1.07	3.59	Idiosyncratic volatility from the Fama-French 3-factor model	Ang et al. (2006)
netis_at	0.97	1.64	0.67	Net total issuance	Bradshaw Richardson and Sloan (2006)
ocfq_saleq_std	1.11	1.69	0.58	Cash flow volatility	Huang (2009)
rmax1_21d	-4.28	1.65	5.93	Maximum daily return	Bali Cakici and Whitelaw (2011)
rmax5_21d	-5.37	1.22	6.59	Highest 5 days of return	Bali, Brown, Murray and Tang (2017)
rvol_21d	-0.61	4.08	4.68	Return volatility	Ang et al. (2006)
turnover_126d	0.51	1.37	0.87	Share turnover	Datar Naik and Radcliffe (1998)
zero_trades_126d	0.52	1.36	0.84	Number of zero trades with	Liu (2006)

				turnover as tiebreaker (6 months)	
zero_trades_21d	-0.62	6.32	6.94	Number of zero trades with turnover as tiebreaker (1 month)	Liu (2006)
zero_trades_252d	0.64	1.14	0.50	Number of zero trades with turnover as tiebreaker (12 months)	Liu (2006)
Panel F Momentum					
prc_highprc_252d	-5.18	1.49	6.67	Current price to high price over last year	George and Hwang (2004)
ret_12_1	0.97	1.27	0.30	Price momentum t-12 to t-1	Jegedeesh and Titman (1993)
ret_3_1	0.32	1.62	1.31	Price momentum t-3 to t-1	Jegedeesh and Titman (1993)
ret_6_1	0.68	1.35	0.67	Price momentum t-6 to t-1	Jegedeesh and Titman (1993)
ret_9_1	0.73	1.09	0.35	Price momentum t-9 to t-1	Jegedeesh and Titman (1993)
seas_1_1na	-3.46	2.07	5.54	Year 1-lagged return, nonannual	Heston and Sadka (2008)
Panel G Profit Growth					
dsale_dinv	1.23	1.28	0.05	Change sales minus change Inventory	Abarbanell and Bushee (1998)
dsale_drec	1.16	1.24	0.08	Change sales minus change receivables	Abarbanell and Bushee (1998)
dsale_dsga	1.15	1.29	0.15	Change sales minus change SG&A	Abarbanell and Bushee (1998)
niq_at_chg1	0.81	1.98	1.16	Change in quarterly return on assets	
niq_be_chg1	0.75	1.67	0.92	Change in quarterly return on equity	
niq_su	0.50	1.76	1.26	Standardized earnings surprise	Foster Olsen and Shevlin (1984)
ocf_at_chg1	1.21	1.40	0.19	Change in operating cash flow to assets	Bouchard, Krueger, Landier and Thesmar (2019)
qmj_safety	1.26	1.30	0.04	Quality minus Junk: Safety	Assness, Frazzini and Pedersen (2018)
ret_12_7	0.86	1.52	0.66	Price momentum t-12 to t-7	Novy-Marx (2012)
sale_emp_gr1	0.42	0.85	0.43	Labor force efficiency	Abarbanell and Bushee (1998)
saleq_su	0.52	2.10	1.58	Standardized Revenue surprise	Jegadeesh and Livnat (2006)
seas_1_1an	0.95	1.24	0.29	Year 1-lagged return, annual	Heston and Sadka (2008)
sti_gr1a	0.87	1.29	0.41	Change in short-term investments	Richardson et al. (2005)

Panel H Profitability					
dolvol_var_126d	0.38	4.01	3.63	Coefficient of variation for dollar trading volume	Chordia Subrahmanyam and Anshuman (2001)
ebit_bev	1.52	1.56	0.04	Return on net operating assets	Soliman (2008)
ebit_sale	1.18	1.80	0.62	Profit margin	Soliman (2008)
f_score	0.84	1.87	1.03	Pitroski F-score	Piotroski (2000)
intrinsic_value	1.18	1.51	0.33	Intrinsic value-to-market	Frankel and Lee (1998)
ni_be	1.26	1.58	0.32	Return on equity	Haugen and Baker (1996)
niq_be	0.98	1.84	0.86	Quarterly return on equity	Hou Xue and Zhang (2015)
o_score	1.23	1.52	0.29	Ohlson O-score	Dichev (1998)
ocf_at	1.06	1.60	0.54	Operating cash flow to assets	Bouchard, Krueger, Landier and Thesmar (2019)
ope_be	1.48	1.58	0.10	Operating profits-to-book equity	Fama and French (2015)
ope_bel1	1.20	1.33	0.13	Operating profits-to-lagged book equity	
turnover_var_126d	0.77	2.86	2.09	Coefficient of variation for share turnover	Chordia Subrahmanyam and Anshuman (2001)
Panel I Quality					
at_turnover	1.25	1.26	0.01	Capital turnover	Haugen and Baker (1996)
cop_at	1.16	1.51	0.34	Cash-based operating profits-to-book assets	
cop_atl1	1.18	1.46	0.28	Cash-based operating profits-to-lagged book assets	Ball et al. (2016)
dgp_dsale	1.05	1.47	0.42	Change gross margin minus change sales	Abarbanell and Bushee (1998)
gp_at	1.05	1.57	0.52	Gross profits-to-assets	Novy-Marx (2013)
gp_atl1	1.12	1.51	0.39	Gross profits-to-lagged assets	
mispricing_perf	1.54	1.64	0.10	Mispricing factor: Performance	Stambaugh and Yuan (2016)
ni_inc8q	-0.43	0.94	1.37	Number of consecutive quarters with earnings increases	Barth Elliott and Finn (1999)
niq_at	1.09	1.99	0.90	Quarterly return on assets	Balakrishnan Bartov and Faurel

					(2010)
op_at	1.28	1.82	0.54	Operating profits-to-book assets	
op_atl1	1.40	1.73	0.33	Operating profits-to-lagged book assets	Ball et al. (2016)
opex_at	0.98	1.27	0.28	Operating leverage	Novy-Marx (2011)
qmj	1.03	1.65	0.63	Quality minus Junk: Composite	Assness, Frazzini and Pedersen (2018)
qmj_growth	1.11	1.72	0.61	Quality minus Junk: Growth	Assness, Frazzini and Pedersen (2018)
qmj_prof	1.18	1.70	0.52	Quality minus Junk: Profitability	Assness, Frazzini and Pedersen (2018)
sale_bev	0.97	1.32	0.35	Assets turnover	Soliman (2008)
tax_gr1a	0.91	1.44	0.53	Tax expense surprise	Thomas and Zhang (2011)
Panel J Seasonality					
corr_1260d	0.75	1.99	1.25	Market correlation	Assness, Frazzini, Gormsen, Pedersen (2020)
coskew_21d	0.41	1.48	1.07	Coskewness	Harvey and Siddique (2000)
dbnetis_at	0.95	1.59	0.64	Net debt issuance	Bradshaw Richardson and Sloan (2006)
kz_index	0.87	1.16	0.29	Kaplan-Zingales index	Lamont Polk and Saa-Requejo (2001)
lti_gr1a	1.09	1.23	0.14	Change in long-term investments	Richardson et al. (2005)
ni_ar1	1.04	1.24	0.20	Earnings persistence	Francis et al. (2004)
pi_nix	1.25	1.42	0.17	Taxable income-to-book income	Lev and Nissim (2004)
seas_11_15an	1.05	1.14	0.09	Years 11-15 lagged returns, annua	Heston and Sadka (2008)
seas_11_15na	1.31	1.36	0.05	Years 11-15 lagged returns, nonannual	Heston and Sadka (2008)
seas_16_20an	0.85	1.75	0.90	Years 16-20 lagged returns, annual	Heston and Sadka (2008)
seas_16_20na	0.47	1.40	0.93	Years 16-20 lagged returns, nonannual	Heston and Sadka (2008)
seas_2_5an	0.98	1.92	0.94	Years 2-5 lagged returns, annual	Heston and Sadka (2008)
seas_2_5na	0.82	1.93	1.11	Years 2-5 lagged returns, nonannual	Heston and Sadka (2008)

seas_6_10an	0.79	1.07	0.28	Years 6-10 lagged returns, annual	Heston and Sadka (2008)
seas_6_10na	0.70	0.86	0.16	Years 6-10 lagged returns, nonannual	
Panel K Size					
ami_126d	0.96	2.11	1.15	Amihud Measure	Amihud (2002)
dolvol_126d	1.13	1.14	0.01	Dollar trading volume	Brennan Chordia and Subrahmanyam (1998)
market_equity	-0.57	2.47	3.04	Market Equity	Banz (1981)
prc	-1.00	3.91	4.91	Price per share	Miller and Scholes (1982)
Panel L Skewness					
iskew_capm_21d	-2.99	2.21	5.21	Idiosyncratic skewness from the CAPM	
iskew_ff3_21d	-2.39	1.59	3.98	Idiosyncratic skewness from the Fama-French 3-factor model	Bali Engle and Murray (2016)
ret_1_0	-4.11	4.64	8.75	Short-term reversal	Jegadeesh (1990)
rmax5_rvol_21d	-2.29	5.56	7.84	Highest 5 days of return scaled by volatility	Assness, Frazzini, Gormsen, Pedersen (2020)
rskew_21d	-2.88	3.93	6.80	Total skewness	Bali Engle and Murray (2016)
Panel M Value					
at_me	-0.02	4.12	4.14	Assets-to-market	Fama and French (1992)
be_me	-0.51	4.40	4.92	Book-to-market equity	Rosenberg Reid and Lanstein (1985)
bev_mev	-0.34	3.55	3.88	Book-to-market enterprise value	Penman Richardson and Tuna (2007)
debt_me	0.10	1.65	1.55	Debt-to-market	Bhandari (1988)
div12m_me	0.42	3.34	2.92	Dividend yield	Litzenberger and Ramaswamy (1979)
ebitda_mev	0.17	1.43	1.25	Ebitda-to-market enterprise value	Loughran and Wellman (2011)
eq_dur	-0.21	2.80	3.01	Equity duration	Dechow Sloan and Soliman (2004)
eqnpo_12 m	1.17	1.63	0.46	Equity net payout	Daniel and Titman (2006)

eqnpo_me	0.71	0.96	0.25	Net payout yield	Boudoukh et al. (2007)
eqpo_me	0.64	1.82	1.18	Payout yield	Boudoukh et al. (2007)
ni_me	0.16	1.34	1.18	Earnings-to-price	Basu (1983)
ocf_me	0.42	0.53	0.11	Operating cash flow-to-market	Desai Rajgopal and Venkatachalam (2004)
sale_me	0.15	3.50	3.35	Sales-to-market	Barbee Mukherji and Raines (1996)

Table B The performance of conditional NN-based predictive portfolios with different hidden layers

The table reports the results of the conditional NN-based predictive portfolio with different hidden layers in investor risk aversion $\gamma = 2$. We report the annualized average returns, standard deviations, and Sharpe ratios.

γ=2	NN1	NN2	NN3	NN4	NN5
Mean	0.451	0.504	0.594	0.572	0.565
STD	0.396	0.400	0.421	0.440	0.431
SR	1.14	1.26	1.41	1.30	1.31