International Review of Financial Analysis 87 (2023) 102627

Contents lists available at ScienceDirect

International Review of Financial Analysis

journal homepage: www.elsevier.com/locate/irfa

Nonlinear asset pricing in Chinese stock market: A deep learning approach ☆

Shuiyang Pan [a,e], Suwan(Cheng) Long [b,c,*], Yiming Wang [e], Ying Xie [d]

[a] China Southern Asset Management Co., Ltd, Shenzhen, China
[b] Judge Business School, University of Cambridge, Trumpington St, Cambridge CB2 1AG, United Kingdom
[c] Trinity Business School, Trinity College Dublin, Dublin 2, Ireland
[d] Cranfield School of Management, Cranfield University, College Road, Cranfield, MK43 0AL, United Kingdom
[e] School of Economics, Peking University, Beijing, China

ARTICLE INFO

Keywords:
Nonlinear asset pricing
Long–short-term memory neural network
Deep learning

ABSTRACT

The redesign of asset pricing models failed to integrate the frequent financial phenomenon that stock markets exhibit a non-linear long- and short-term memory structure. The difficulty lies in developing a nonlinear pricing structure capable of depicting the memory influence of the pricing variable. This paper presents a Long- and Short-Term Memory Neural Network Model (LSTM) to capture the non-linear pricing structure among five elements in the Chinese stock market, including market portfolio return, market capitalisation, book-to-market ratio, earnings factor, and investment factor. The long–short-term memory structure implies that the autocorrelation function of the stock return series decays slowly and has a long-term characteristic. The LSTM model surpasses the standard Fama–French five-factor model in terms of out-of-sample goodness-of-fit and long–short strategy performance. The empirical findings indicate that the LSTM nonlinear model properly represents the nonlinear relationships between the five components.

1. Introduction

Empirical evidence reveals that traditional linear asset pricing models, when viewed through an econometric lens, suffer from several problems. These problems include inadequate explanatory power of the model, regression intercept terms that are significantly greater than zero, and insignificant beta coefficients. The root causes of these problems may be the omission of explanatory variables, the time-varying nature of beta coefficients, or the time-varying nature of the variance of the random terms, such as conditional heteroskedasticity. As a result, revisions to linear asset pricing models have emerged in three different directions.

Firstly, researchers have attributed the failure of linear asset pricing models to the omission of explanatory variables and have proposed multi-factor asset pricing models accordingly (Fama & French, 1992, 2015; Hou et al., 2015; Jegadeesh & Titman, 1993). Secondly, researchers have attributed the failure to the time-varying nature of beta coefficients and have proposed a variable parameter (beta) asset pricing model (Adrian & Franzoni, 2009; Lettau & Ludvigson, 2001). Finally, researchers have attributed the failure to the time-varying nature of the variance of the stochastic term and have proposed GARCH-like asset pricing models, among others (Engel & Rodrigues, 1989; French et al., 1987). However, the revised models given in the above three directions do not fundamentally address the weaknesses of linear asset pricing models in terms of their explanatory power.

Increasing the model's explanatory power simply by adding explanatory variables is an inexhaustible process. Lewellen and Shanken (Lewellen & Nagel, 2006) show that the covariance between the beta coefficients estimated by the time-varying beta-parameter pricing model and the excess returns of the market portfolio has no significant differences, which implies that the returns obtained by the time-varying beta-parameter pricing model are similar to those calculated through the CAPM model.

Numerous studies have identified the presence of long- and short-term memory non-linearity in financial markets (Andrew, 1991, Ding et al., 1993, Panas, 2001, Davidson & Teräsvirta, 2002, Sadique & Silvapulle, 2001, Cao et al., 2010, Wang et al., 2019, Ma et al., 2006). The long–short-term memory structure of financial data refers to the slow decay of the autocorrelation function of the stock return series and its long-term autocorrelation characteristic. The non-linear nature of financial markets with long- and short-term memory effects challenges the random walk and efficient market assumptions, which modern capital market theory and other financial econometric models based

☆ This paper is part of the project "Nonlinear Asset Pricing Model Based on Deep Learning Theory in the Context of Big Data (18CJY057)" funded by the National Social Science Fund of China. We express our gratitude for this support.

* Corresponding author at: Trinity Business School, Trinity College Dublin, Dublin 2, Ireland.
E-mail address: longch@tcd.ie (S. Long).

https://doi.org/10.1016/j.irfa.2023.102627
Received 9 January 2023; Received in revised form 1 March 2023; Accepted 21 March 2023
Available online 24 March 2023

on normal distributions or finite variance properties rely on. Despite the widespread occurrence of the financial anomaly of non-linear long- and short-term memory structure in stock markets, it has been ignored in the revision of asset pricing models and not incorporated into the modelling framework. This poses a challenge because the non-linear pricing structure that characterises the factor memory effect cannot be formulated.

To address this issue, this study proposes a long and short-term memory neural network model (LSTM) to learn the non-linear pricing structure between the five factors, namely market portfolio return, market capitalisation, book-to-market ratio, earnings, and investment in the Chinese stock market. LSTM can learn long and short-term dependencies between time steps of data, capturing the long-term autocorrelation of financial data. The intuition behind using the LSTM model is to retain effective information through memory units and filter out "noise" information through forgetting units, resulting in more accurate predictions of stock returns.

This study's contributions are twofold: First, it departs from the traditional linear regression framework for constructing asset pricing models. Instead, it improves explanatory power by increasing the number of explanatory variables and building asset pricing models with better predictive power. Second, this study uses LSTM to automatically learn the non-linear pricing structure between the five factors, namely market portfolio return, market capitalisation, book-to-market ratio, earnings, and investment, providing a new research methodology for exploring asset pricing models.

2. Literature review

The question of what determines stock returns is an enduring research topic in academic and investment communities. Previous studies have used a linear regression econometric modelling empirical research framework, which implicitly presupposes a linear relationship between individual pricing factors and stock returns. Meanwhile, much literature (Andrew, 1991, Ding et al., 1993, Panas, 2001, Davidson & Teräsvirta, 2002, Sadique & Silvapulle, 2001,Yang & Chen, 2014, Ma et al., 2006) finds that there are non-linear characteristics of financial markets, such as long memory, spikes and thick tails, and fractals; the stock return process obeys the typical peak, thick-tailed non-normal distribution; The stock market price change is not an independent and identical distribution process; the return has a significant serial autocorrelation structure and shows long-term correlation. The above phenomena indicate that financial time series may have non-linear dynamical systems.

According to the existing literature, the memory process is a common non-linear model that characterises the aforementioned distribution (Granger & Ding, 1996, Panas, 2001). In the case of the Chinese stock market, Yang and Chen (2014) observed that Chinese log returns deviate from a normal distribution and are not homogeneously distributed independently, due to non-linear effects. Meanwhile, Wang et al. (2019) conducted an empirical study on the long-term correlation between Chinese stock market returns and their volatility using a modified R/S analysis and an ARFIMA model, and found that the Chinese stock market has significant non-linear characteristics. Although the autocorrelation of the return series is weak, the volatility series exhibits significant long-run memory effects.

Similarly, Ma et al. (2006) employed non-linear dynamics analysis to examine the essential characteristics and formation mechanism of Chinese stock market volatility, finding that it has significant fractal dynamics characteristics and long-term memory effects. Using the R/S method and the ARFIMA model, Cao et al. (2010) also found evidence of a long memory effect in the Chinese stock market, which is more robust in the Shenzhen market than in the Shanghai market. These findings are consistent with those of Yong (2008) and Zhang (2017).

The observed memory effect in the stock market suggests that modern capital market theory, which is based on the theoretical assumptions of random wandering and efficient market hypotheses, and other financial econometric models that rely on the normal distribution or finite variance properties, will face significant challenges. Despite this, the widespread financial anomaly of the stock market having a non-linear memory structure has been selectively ignored in the revision of asset pricing models. This paper argues that this is mainly due to the inability to formulate the non-linear pricing structure that characterises the factor memory effect. To address this issue, we propose using a Long–Short-Term Memory (LSTM) model to automatically capture the non-linear pricing structure between factors. As long as a non-linear pricing structure exists between the factor dataset and the stock returns, the deep learning model can learn the pricing structure hidden in the data from the historical data. Deep learning is a powerful tool for identifying non-linear pricing structures between factors by building models with a data-driven core.

Current applications of machine learning (ML) and deep learning (DL) in the field of economics are limited, but have recently gained significant attention from researchers (Varian, 2014, Mullainathan & Spiess, 2017). Doudchenko & Imbens, 2016 demonstrate how ML can be combined with double difference DID for causal identification. Moreover, Varian (2016) provides insights into future directions for combining ML with breakpoint regression. In another study, Athey (2017) describes the application of DL for public resource allocation and causal inference. Additionally, Chalfin et al. (2016) highlights the significant social welfare gains that can be achieved by using DL to predict employee productivity. Non-linear methods such as neural networks in DL have been used by Hartford et al. (2017) to estimate the first stage of instrumental variable selection. Furthermore, ML has been employed to detect Chinese stock market manipulation by Liu et al. (2021). In the area of asset pricing, Rapach et al. (2013) uses LASSO regression in DL to predict global stock returns. Neural networks have been used to predict derivative prices by Hutchinson et al. (1994), and regression tree models have been used by Khandani et al. (2010) and Butaru et al. (2016) to predict credit card default probabilities. In a different approach, Harvey et al. (2016) tested the validity of multiple asset pricing factors using a self-help approach in DL. Kelly and Pruitt (2013) and Kelly and Pruitt (2015) achieved better prediction results by using a factor dimensionality reduction approach in DL and constructing a combination of multiple prediction models to predict stock returns. Another study by Lu et al. (2022) uses ML models for oil futures volatility forecasting. Other relevant literature includes Gu et al. (2020) and Leippold et al. (2022). The above literature demonstrates the unique advantages of ML and DL models over conventional linear forecasting models in predicting stock returns. Conventional time series forecasting models such as ARAM, GARCH, and Kalman filter are linear models and are effective only when the financial and economic system is linear. However, non-linear forecasts cannot be accurately captured by conventional linear models, and DL models can capture the non-linear patterns in financial data without human intervention.

3. Data and methodology

3.1. Data

This paper utilises a sample of monthly return, closing price, total equity, total owners' equity, operating profit, and total assets data extracted from consolidated financial statements of A-share stocks listed on the Shanghai and Shenzhen Stock Exchange during the period spanning from January 2000 to June 2022. Stocks that have been delisted, *ST stocks, and stocks that have been listed for less than one year are excluded from the sample. To derive relevant measures, the closing price is multiplied by the total equity to obtain the total market capitalisation of the stock, while the total owners' equity is divided by the total market capitalisation of the stock to obtain the book-to-market ratio. Additionally, the profitability factor is derived by dividing the operating profit by the owners' equity, whereas the investment factor is obtained by dividing the new total assets by the total assets. The

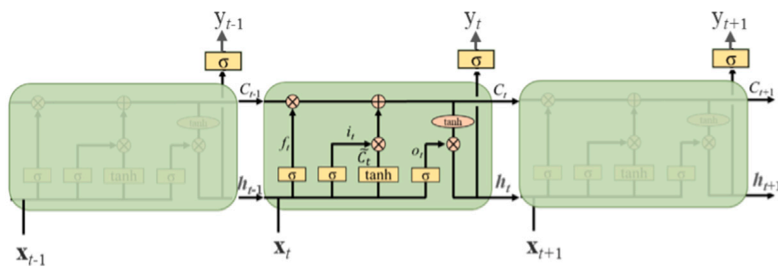


Fig. 1. LSTM structure.

market portfolio return is based on the CSI All A Index return. Data for this study is sourced from the WIND information database.[1]

3.2. Methodology

The general expression for the stochastic discount factor asset pricing model proposed by academics is

$$E_t(M_{t+1}R_{t+1}) = 1 \quad (1)$$

In Eq. (1), $\boldsymbol{M}_{t+1}$ denotes the stochastic discount factor. The stochastic discount factor can be expressed as a function of the risk factor. The relationship between the stochastic discount factor and the risk factor may be linear or non-linear. When the stochastic discount factor is a linear combination of risk factors, the return on the asset can be portrayed by a linear factor model (i.e., beta model) (Campbell, 2000). These results do not hold when the stochastic discount factor is not a linear combination of risk factors. In order to better characterise the asset pricing model and achieve a perfect combination of theory and practice, the stochastic discount factor should satisfy the following conditions: $\boldsymbol{M}_{t+1}$ is non-linear, and the pricing structure of $\boldsymbol{M}_{t+1}$ can reflect the memory property. This paper implements the memory structure and non-linear mapping of $\boldsymbol{M}_{t+1}$ by introducing LSTM, a long and short-term memory neural network model in the field of artificial intelligence.

3.2.1. Long–short term memory neural network model

Long short-term memory(LSTM) neural networks were first proposed by Hochreiter and Schmidhuber (1997). LSTM was explicitly designed to solve long-term dependency problems, and remembering information over long periods is the model's most important feature. The simplest LSTM structure is shown in Fig. 1.

In Fig. 1, $\mathbf{X}_t = \left[x_1^t, x_2^t, \ldots, x_N^t\right]$, is the external input variable at time t. y_t is the output at time t. C_t is the cell state, h_t is the implied state, and $f_t, i_t, and\, o_t$ represent the forgetting gate, input gate, and output gate, respectively. tanh and σ denote the hyperbolic tangent and sigmoid functions, respectively. The non-linear mapping between $\boldsymbol{X}_t$ and y_t can be obtained by various operations on the internal structure of the LSTM. The mathematical description of the LSTM structure is given in the following equation:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \quad (2)$$

$$f_t = \sigma\left(\sum_{n=1}^{N} w_{fn}x_n^t + w_{fh}h_{t-1} + b_f\right) \quad (3)$$

$$i_t = \sigma\left(\sum_{n=1}^{N} w_{in}x_n^t + w_{ih}h_{t-1} + b_i\right) \quad (4)$$

$$\widetilde{C}_t = \tanh\left(\sum_{n=1}^{N} w_{cn}x_n^t + W_{ch}h_{t-1} + b_c\right) \quad (5)$$

$$o_t = \sigma\left(\sum_{n=1}^{N} w_{on}x_n^t + W_{oh}h_{t-1} + b_o\right) \quad (6)$$

$$h_t = o_t * \tanh\left(C_t\right) \quad (7)$$

$$y_t = \sigma\left(wh_t + b\right) \quad (8)$$

Eq. (2) is an expression for updating the cell state C_t, with the historical state information C_{t-1} and the new arrival information $\widetilde{C}_t$ controlled by the forgetting gate f_t and the input gate i_t, respectively.

Eq. (3) shows how the forgetting gate f_t is implemented. The forgetting gate selectively forgets the information in the cell state at the previous moment. The inputs to the forgetting gate are h_{t-1} and $\mathbf{x}_{\mathrm{m}}$, which are then mapped to the [0-1] interval by the σ function. A value of 1 for f_t means that the historical information is fully preserved C_{t-1}, and a value of 0 means that the historical information is completely forgotten.

Eq. (4) is the implementation of the input gate i_t. The input gate selectively records the new arrival information into the cell state. The input gate inputs h_{t-1} and x_t are also mapped to the [0-1] interval by the σ function. i_t takes a value of 1 to absorb the new arrival information $\widetilde{C}_t$ completely and a value of 0 to completely discard the new arrival information $\widetilde{C}_t$.

Eq. (5) is the new arrival information $\widetilde{C}_t$ implementation. The new arrival information is taken as input from the external input variable x_t at the current moment and the implicit state h_{t-1} of the neural network at the previous moment, which is then mapped to the interval $[-1, 1]$ by the tanh function.

Eq. (6) is the output gate o_t implementation. The output gate determines the prediction result of the model. The output gate inputs h_{t-1} and x_t are then mapped to the [0-1] interval by the σ function, which determines which parts of the new cell states are selected as outputs.

Eq. (7) is the cell state output. The cell states are mapped to the interval $[-1, 1]$ by the tanh function, which is then multiplied by the output gate o_t.

Eq. (8) is the final output of the LSTM. The implicit states are mapped nonlinearly by the σ function to obtain the final output result of the model.

The w_{fn}, w_{in}, w_{cn}, w_{on}, b_f, b_i, b_c, b_o, n=1,2...N in Eqs. (2)–(8) are the parameters to be estimated. The way the LSTM models the input, forgetting and output gates from the mathematical expressions is essentially a linear weighting of the input at the current moment and the output of the implicit state at the previous moment, followed by a non-linear transformation through the σ function.

3.2.2. LSTM non-linear pricing model

An important reason for applying the LSTM model to asset pricing in this paper is that the model is able to characterise the long- and short-term memory non-linearity in the stock market that has been derived from much of the empirical literature. Assuming that there are M stocks

[1] For more information about the database, please refer to: https://www.wind.com.cn/en/edb.html.

in the market and N stock pricing factors, at time t we have a dataset of stock factors $\boldsymbol{X}_t$ and stock returns $\boldsymbol{y}_t$

$$\boldsymbol{X}_t = \begin{bmatrix} x_{11}^t & \cdots & x_{1N}^t \\ \vdots & \ddots & \vdots \\ x_{M1}^t & \cdots & x_{MN}^t \end{bmatrix}$$
$$\boldsymbol{y}_t = \begin{bmatrix} y_1^t & \cdots & y_M^t \end{bmatrix}$$

The essence of asset pricing is to find the mapping (F) between the factor dataset $\boldsymbol{X}_t$ and the stock return y_t:

$$\boldsymbol{y}_t = F\left(\boldsymbol{X}_t\right) \tag{9}$$

The linear asset pricing model assumes a linear relationship between the factor dataset $\boldsymbol{X}_t$ and the stock return $\boldsymbol{y}_t$, where Eq. (9) becomes:

$$\boldsymbol{y}_t = \boldsymbol{X}_t \boldsymbol{\beta}_t \tag{10}$$

The LSTM model assumes a non-linear relationship between the factor dataset $X_{\leftarrow}$ and the stock return y_t:

$$\boldsymbol{y}_i = F\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots \mid \boldsymbol{W}, \boldsymbol{b}\right) \tag{11}$$

W, b in Eq. (11) are the parameters to be estimated in the LSTM model, and F is the LSTM non-linear mapping described in Eqs. (2)–(8). Comparing Eqs. (10) and (11) and combining with the LSTM structure in Fig. 1, it can be found that the LSTM asset pricing model has a specific long- and short-term memory structure compared to the traditional linear asset pricing model, not only the historical stock returns, the historical values of the factor data can have an impact on future returns and achieve a specific memory structure, but also the non-linear mapping of the tanh function and the σ function can capture to the non-linear mapping relationship between the factor data set X_t and the stock returns y_t. In fact if the tanh function and σ function are replaced with linear functions, it can be found from Eqs. (2)–(8) that the LSTM will degenerate into a linear model.

The paper by George (1989) demonstrates that multilayer neural network structures can fit arbitrary continuous functions very well. Thus as long as there is a non-linear pricing structure between the factor dataset X_t and the stock return y_t, an LSTM neural network structure containing multiple implicit layers is theoretically able to identify the non-linear pricing structure hidden behind the data. In the latter part of this paper, both the Fama–French five-factor linear pricing model and the LSTM non-linear asset pricing model are used to empirically test the effectiveness of the LSTM pricing model for the Chinese A-share market by comparing the out-of-sample goodness of fit and the performance of the long–short strategies of the two models.

For parameter estimation in the LSTM asset pricing model, we use the widely used stochastic gradient descent method. Assuming that the LSTM model output value is y and the actual value is y, the error sum of squares function is defined as

$$E = \frac{1}{2}\sum_{t=1}^{T}\left(y_t - \hat{y}_t\right)^2 = L\left(F\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots; \boldsymbol{W}, \boldsymbol{b}\right), \boldsymbol{y}_t\right) \tag{12}$$

The parameters to be estimated for the LSTM asset pricing model can be obtained by minimising Eq. (12). A brief description of the solution algorithm is shown below, and a detailed derivation of the procedure can be found in the literature (Duchi et al., 2011; Kingma & Ba, 2014; Zhang et al., 2015).

(1) Determine the learning rate ϵ

(2) Initialise the parameters $\boldsymbol{W}_0, \boldsymbol{b}_0$

(3) Randomly select m subsamples from the training set $\left\{\boldsymbol{X}_t^i, \boldsymbol{X}_{t-1}^i, \ldots\right\}_{i=1,2\cdots m}$, where the ith sample corresponds to the target $\boldsymbol{y}_t^i$, the error back-propagation algorithm is used to obtain the gradient estimate:

$$\hat{\boldsymbol{g}} \leftarrow \frac{1}{m}\nabla_{W,\boldsymbol{b}} L\left(F\left(\boldsymbol{X}_t^i, \boldsymbol{X}_{t-1}^i \ldots; \boldsymbol{W}, \boldsymbol{b}\right), \boldsymbol{y}_t^i\right)$$

(4) Update the argument $\left[W_0 b_0\right] \leftarrow \left[W_0 b_0\right] - \epsilon\hat{\boldsymbol{g}}$

(5) If the iteration stopping criterion is satisfied, stop the computation, otherwise go to step (3).

4. Result

The Fama–French five-factor asset pricing model suggests that stock returns can be explained by five factors, namely market portfolio return, market capitalisation, book-to-market ratio, earnings and investment. Based on these findings, we can construct a five-factor linear pricing forecasting model:

$$\begin{aligned} r_{it} = \beta_0 + \beta_1 R_{m,t-1} + \beta_2 SMB_{i,t-1} + \beta_3 HML_{i,t-1} \\ + \beta_4 RMW_{i,t-1} + \beta_5 CMA_{i,t-1} + e_{it} \end{aligned} \tag{13}$$

where r_{it} denotes the return of stock i at time t; $R_{m,t-1}$ denotes the market portfolio return weighted by market capitalisation at time $t-1$; $SMB_{i,t-1}$ denotes the market capitalisation factor return of stock i at time $t-1$; $HML_{i,t-1}$ denotes the book-to-market return of stock i at time $t-1$; $RMW_{i,t-1}$ denotes the earnings factor return; $CMA_{i,t-1}$ denotes the investment factor return; e_{it} denotes the prediction error of the five-factor pricing linear forecasting model; β_0, β_1, β_2, β_3, β_4, β_5 are the parameters of the linear pricing model.

In order to test the validity of the LSTM asset pricing model, we use the out-of-sample model predictions of the goodness-of-fit R2 and the performance of the constructed long–short strategy to test the validity of the model. The benchmark model is the fama–french five-factor linear pricing model. The criteria for judging the superiority of the LSTM asset pricing model over the fama–french five-factor linear pricing model are:

1. the out-of-sample fit R2 of the LSTM asset pricing model is greater than the out-of-sample fit R2 of the fama–french five-factor linear pricing model.

2. The LSTM asset pricing model outperforms the fama–french five-factor linear pricing model in constructing long–short investment strategies based on the LSTM asset pricing model and the fama–french five-factor linear pricing model, respectively.

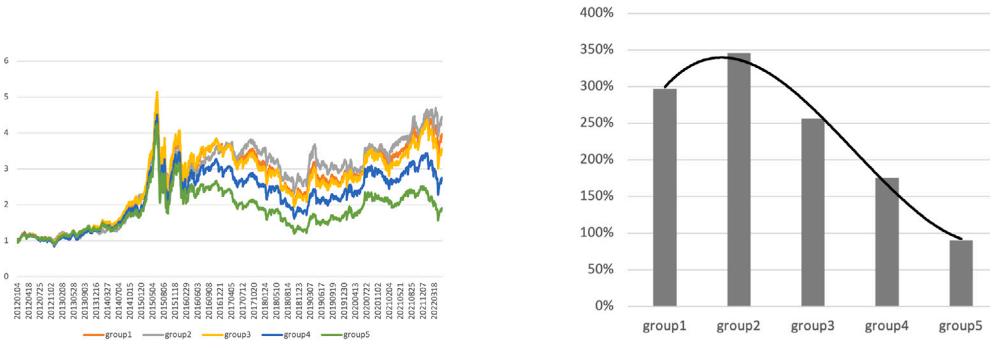
Suppose the LSTM asset pricing model satisfies the above two conditions. In that case, we have good reasons to believe that the LSTM asset pricing model can identify the non-linear relationships hidden behind the pricing factor data and that the non-linear pricing model outperforms the linear pricing model, provided that the model input variables are consistent.

4.1. Non-linear phenomenon in the Chinese A-share stock market

Fig. 2 depicts a strategy for constructing a book-to-market ratio based on empirical data collected from January 2012 to June 2022. Specifically, individual stocks' book-to-market ratios are computed at the beginning of each month and subsequently partitioned into five equal groups based on descending order, with the largest group assigned to the first quintile and the smallest group assigned to the fifth. Based on the Fama–French linear pricing model, the book-to-market ratio positively correlates with the expected return on the portfolio. Hence, the first quintile should have the highest expected return, while the fifth quintile should have the lowest. However, such a linear relationship does not hold in the Chinese A-share market, as Fig. 2 illustrates. Instead, the second quintile has the highest returns, and the relationship between the book-to-market ratio and stock returns is non-linear.

4.2. Comparison of the out-of-sample fit of the model

The study's data collection period extended from January 2000 to June 2022, spanning a duration of 22 years. The sample size consisted of nearly 6,000 stocks, with an average monthly stock count exceeding 2,600, resulting in approximately 216,000 observations. Simple networks with limited layers and nodes have typically yielded the best outcomes in small-scale datasets. However, selecting an appropriate network architecture through cross-validation poses a challenging task. Recent advancements in training and regularising neural networks have



(a) Strategy Net Value (b) Yield rate

Fig. 2. Strategy net value and yield rate.

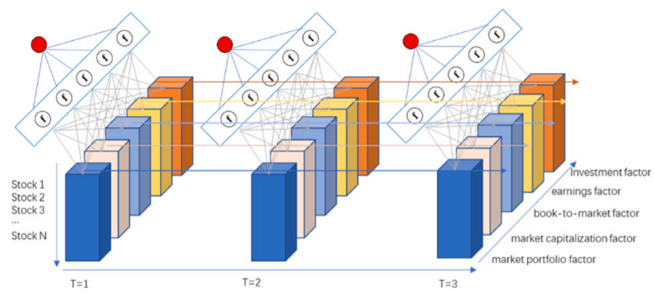


Fig. 3. Model parameter estimation and prediction process.

simplified this process. It is now only necessary to determine the maximum number of neurons per layer and the total number of layers required. In this study, the shallowest neural network was employed, consisting of a single hidden layer with five neurons and a total of 36 parameters (six parameters required to reach each neuron and six weights to consolidate the neurons into a single output). Fig. 3 displays the LSTM structure utilised, which has a total of 108 parameters (36 x 3).

This paper employs data from January 2000 to December 2011 as model training data to estimate the parameters of the Fama–French five-factor linear pricing model and the parameters of the LSTM non-linear asset pricing model. Data from January 2012 to June 2022 are used as out-of-sample validation data to test the forecasting ability of the model. In consideration of the differences in magnitude between various variables, this paper standardises the monthly cross-sectional data of each variable through a z-score transformation, where $Z_{\text{Score}} = \frac{z-\bar{z}}{\sigma_z}$. Here, $\bar{z}$ and σ_z represent the mean and standard deviation of the monthly data, respectively. After standardisation, each variable approximates a normal distribution before being input into the model for training and prediction.

The model parameter estimation and out-of-sample forecasting process are shown in Fig. 3.

After estimating the model parameters, at the beginning of each month starting from January 2012, the latest market capitalisation factor data, book-to-market ratio factor, earnings factor, investment factor and CSI All A Index monthly return data of the stocks were input into the Fama–French five-factor linear pricing forecasting model and LSTM non-linear asset pricing forecasting model respectively, to obtain each model's return for the next month of the stocks forecasts, and then calculate the model's out-of-sample goodness-of-fit R2.

$$R^2 = 1 - \frac{\sum_{(i,t)} \left(r_{i,t} - \widehat{r_{l,t}}\right)^2}{\sum \left((i,t) r_{i,t}^2\right.} \quad (14)$$

In Eq. (14), $r_{l,t}$ denotes the actual return of the ith stock at moment t, and $\widehat{r_{l,t}}$ denotes the predicted return of the ith stock at moment t.

Table 1
Out-of-sample model goodness-of-fit.

Date	Fama_French	LSTM
2012	0.57	0.81
2013	0.61	0.86
2014	0.51	0.79
2015	0.62	0.83
2016	0.55	0.77
2017	0.59	0.85
2018	0.47	0.75
2019	0.53	0.82
2020	0.64	0.85
2021	0.56	0.76
Overall out-of-sample R2	0.61	0.83
Out-of-sample R2 mean	0.56	0.81

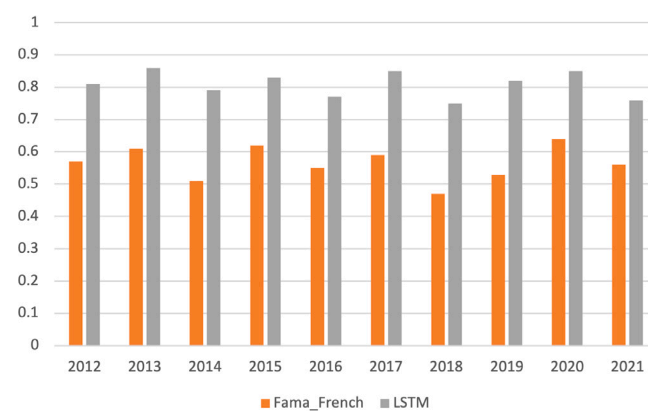


Fig. 4. Out-of-sample model goodness-of-fit per year.

Table 1 and Fig. 4 show the fit superiority of the fama–french five-factor model and the LSTM model in predicting stock returns. As can be seen from the table, the overall out-of-sample goodness-of-fit of the Fama–French model and the LSTM model are 0.61 and 0.83 respectively, and the mean out-of-sample goodness-of-fit is 0.56 and 0.81 respectively, with the LSTM model significantly outperforming the Fama–French model in terms of out-of-sample prediction.

4.3. Comparison of the performance of long–short strategies

To further assess the efficacy of the LSTM model, this study constructs two combinations of long–short strategies based on the Fama–French model and the LSTM model, namely the Fama–French five-factors linear combination forecasting strategy and the LSTM factor non-linear combination forecasting strategy.

Table 2
Fama–French Five-factors linear portfolio model subgroup forecasting performance.

Group	Annualised return	Annualised volatility	Sharpe ratio	Maximum drawdown
Group 1	19.61%	27.95%	0.704135868	49.92%
Group 2	14.96%	27.73%	0.541770529	57.51%
Group 3	14.16%	27.98%	0.507923678	59.38%
Group 4	9.89%	28.34%	0.350291651	66.58%
Group 5	2.09%	29.34%	0.071460505	75.17%

Table 3
LSTM non-linear portfolio model subgroup forecasting performance.

Group	Annualised return	Annualised volatility	Sharpe ratio	Maximum drawdown
Group 1	29.32%	28.23%	1.04263439	48.14%
Group 2	20.00%	28.03%	0.716201252	51.58%
Group 3	12.86%	28.05%	0.46007926	61.84%
Group 4	6.91%	28.17%	0.246272852	69.23%
Group 5	−6.11%	28.71%	−0.213512686	82.98%

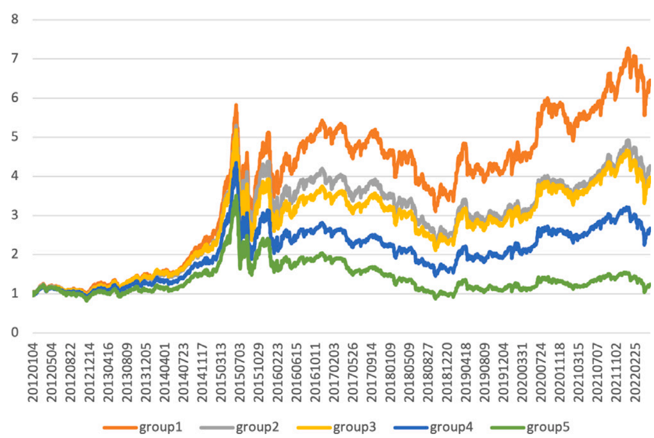


Fig. 5. Fama–french five-factor linear portfolio subgroup net worth curve.

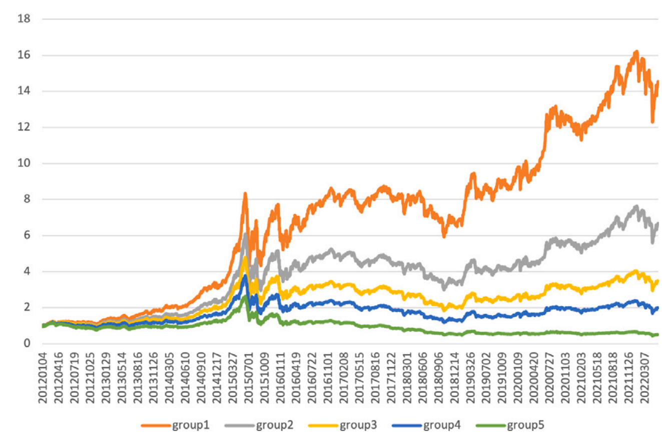


Fig. 6. LSTM non-linear portfolio subgroup net worth curve.

The Fama–French factor linear portfolio forecasting strategy is established by employing the following steps: at the beginning of each month, the Fama–French model anticipates the stock returns for the next period based on the latest factor cross-section data, and then sorts the stock return predictions into five groups in decreasing order. Subsequently, the strategy goes long on the first group and short on the fifth group.

The LSTM factor non-linear portfolio forecasting strategy, on the other hand, is established by performing the following steps: at the outset of each month, the LSTM model forecasts the next period's stock return based on the most recent factor cross-section data and then distributes the stock return forecasts into five equal groups in decreasing order. This strategy then takes long position on group 1 and short position on group 5.

Table 2 illustrates the grouping performance of the Fama–French factor linear portfolio forecasting strategy, while Table 3 presents the grouping performance of the LSTM factor non-linear portfolio forecasting strategy.

Fig. 5 shows the net value curve of the Fama–French factor linear portfolio forecasting strategy grouping, while Fig. 6 shows the net value curve of the LSTM factor non-linear portfolio forecasting strategy grouping.

From the performance of the strategy grouping, the Fama–French grouping strategy fails to capture the non-linear pricing structure among the factors and has a mediocre forecasting effect. In contrast, the LSTM grouping strategy is able to capture the non-linear pricing structure among the factors and therefore has better forecasting performance than the Fama–French grouping strategy.

Table 4 shows the performance and significance tests of the two models' long–short strategies.

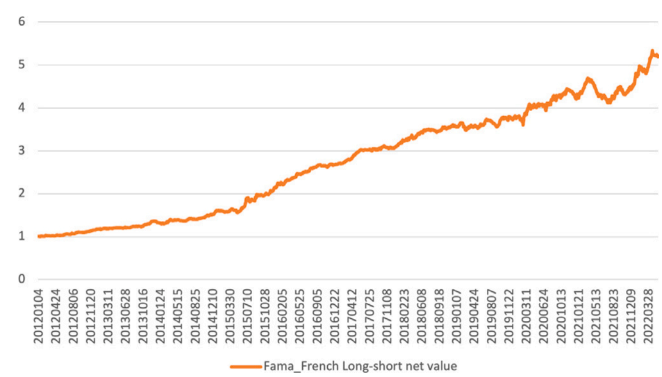


Fig. 7. Fama–french five-factors long–short strategy net value curve.

Fig. 7 shows the net value performance of the Fama–French five-factors model long–short strategy and Fig. 8 shows the net value performance of the LSTM model long–short strategy.

Based on the results presented above, it is evident that the LSTM model long–short strategy outperforms the Fama–French model long–short strategy in terms of annualised return, annualised volatility, Sharpe ratio, and maximum retracement. This finding suggests that the LSTM model effectively captures the non-linear relationship between the factors, leading to a superior investment strategy compared to the Fama–French model. Additionally, the present study calculates the significant difference between the two models using the DM test (Diebold & Mariano, 1995), which tests the null hypothesis that the two models have the same predictive outcome and are not significantly different.

Table 4
Long–short strategy performance and Significance test.

Strategy	long–short portfolio annualised return	long–short portfolio annualised volatility	Sharpe ratio	Maximum drawdown
Fama_French	17.15%	7.02%	2.36	12.34%
LSTM	37.72%	6.53%	5.08	6.44%
	Standard deviation test (DM test) statistic	Standard deviation test (DM test) *P*-value	Sharpe ratio test statistic	Sharpe ratio test *p*-value
	−3.408**	0.035	4.38***	<0.01

Note: The original hypothesis of the DM test is that the volatility of the LSTM model strategy is equal to that of the Fama–French model strategy; the original hypothesis of the Sharpe ratio test is that the Sharpe ratio of the LSTM model strategy is equal to that of the Fama–French model strategy. *, ** , *** denote rejection of the original hypothesis at the 10%, 5%, and 1% levels, respectively.

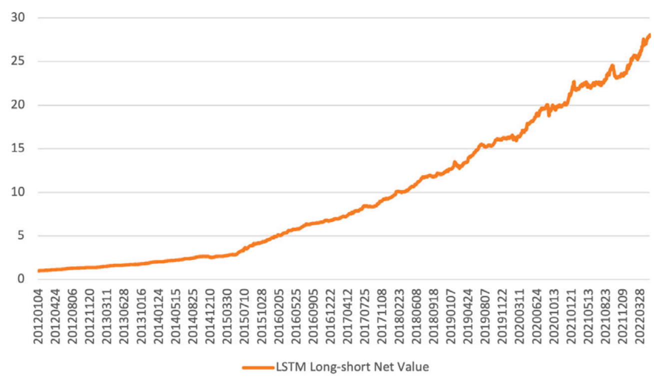


Fig. 8. LSTM long–short strategy net value curve.

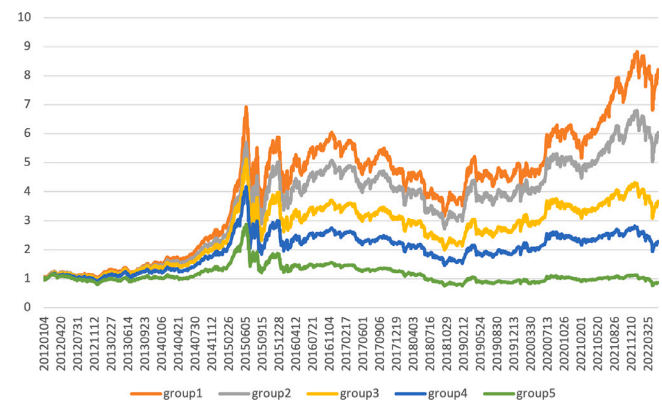


Fig. 9. Residual series non-linear effects grouped into net value curves.

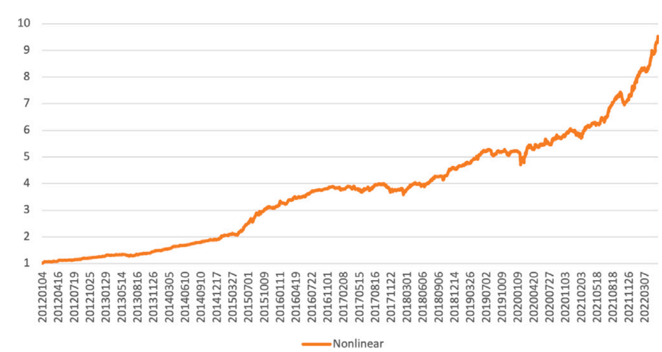


Fig. 10. Residual series non-linear effects long–short portfolio strategy net curve.

Moreover, the test for the difference in the Sharpe ratios of the two models employs the test provided by Ledoit and Wolf (2008), which tests the null hypothesis that the Sharpe ratios of the two strategies are equal.

Specifically, Table 4 demonstrates that the annualised return of the LSTM model strategy is 20.57% higher than that of the Fama–French model strategy, with a DM statistic of −3.408, indicating that the difference is statistically significant at the 10% level. Furthermore, the Sharpe ratio of the LSTM model strategy is 2.72 higher than that of the Fama–French model strategy, and the test statistic is 4.02, indicating that the difference is statistically significant at the 1% level. Thus, the results suggest that the LSTM model strategy outperforms the Fama–French model strategy regarding investment performance.

4.4. Testing for non-linear effects

In this paper, we try to strip out the linearity effect from the LSTM non-linear asset pricing model to verify the non-linear effect further. This paper decomposes the problem based on the following simple model.

Total LSTM non-linear asset pricing effect = linear effect + non-linear effect

This paper utilises the cross-sectional forecasts of individual stock returns in the LSTM non-linear asset pricing model as the total effect, and the cross-sectional forecasts of individual stock returns in the fama–french five-factor linear pricing model as the linear effect component. The linear utility is extracted through linear regression by taking the residuals. Specifically, the following equation is used:

$$\text{Predit}_{LSTM} = \alpha + \beta \cdot \text{Predict}_{\text{Fama–French}} + e \tag{15}$$

The residual obtained from the regression of Eq. (15) is regarded as the non-linear effect factor. A higher residual value indicates a higher non-linear predictive effect and a lower linear predictive effect for an individual stock. Given that the LSTM model identifies the non-linear pricing structure, the residual e, which is separated from the total effect, captures information on the non-linear pricing structure of the factor and can predict the future return of the stock.

At the beginning of each month, both the fama–french five-factor model and the LSTM model predict the next stock return based on the latest factor cross-sectional data. The Fama–french five-factor model predictions are then regressed on the LSTM model predictions to derive the residual series. Finally, the residual series is divided into ten groups in descending order, and the long group 1 stock and short group 10 stocks are utilised to construct a long–short strategy portfolio.

Table 5 shows the performance of the residual series non-linear effect grouping.

Fig. 9 shows the net value curve of the residual series non-linear effect grouping.

The grouped returns show a strictly decreasing relationship from the results, and the non-linear effect has good predictive power.

Table 6 shows the performance of the residual series non-linear effects long–short portfolio strategy.

Fig. 10 shows the net value curve of the residual series non-linear effect long–short portfolio strategy.

Table 5
Grouped performance of residual series non-linear effects.

Group	Annualised return	Annualised volatility	Sharpe ratio	Maximum drawdown
Group 1	22.40%	28.89%	0.778441162	54.33%
Group 2	18.85%	28.89%	0.654868022	52.33%
Group 3	13.29%	28.65%	0.46575066	61.23%
Group 4	8.27%	28.07%	0.295842984	65.28%
Group 5	−1.42%	27.01%	−0.052630254	74.02%

Table 6
Residual series non-linear effects long–short portfolio strategy.

Strategy	Long–short portfolio annualised return	Long–short portfolio annualised volatility	Sharpe ratio	Maximum drawdown
Non-linear effects strategy	24.15%	8.19%	2.76	6.01%

This further validates that the LSTM model can automatically learn the linear and non-linear relationships between the pricing factors. After stripping out the linear effects, the residual series carries information on the non-linear effects, which can provide additional useful information for individual stock return forecasting.

5. Conclusions

This paper proposes an innovative approach to addressing the non-linear feature of long and short-term memory structure in the stock market that is often ignored in asset pricing models. Specifically, we introduce a long and short-term memory (LSTM) neural network model, a machine learning technique, to automatically learn the non-linear pricing structure among the five factors of market portfolio return, market capitalisation, book-to-market ratio, earnings, and investment pattern. We rigorously compare our new model with the classical Fama–French five-factor model in terms of out-of-sample goodness-of-fit and performance of long–short strategies.

Our empirical results demonstrate that the LSTM non-linear model outperforms the traditional five-factor linear pricing model in terms of out-of-sample fit and performance of long–short strategies. Our findings suggest two key points: First, the Chinese A-share market has typical non-linear characteristics, and the non-linear relationships between pricing factors can be learned automatically with the help of machine learning techniques. Second, the LSTM non-linear model demonstrates superior performance in comparison to the traditional five-factor linear pricing model.

It is important to note that the improved predictions obtained from these models merely serve as measurements and do not elucidate economic mechanisms or equilibria. Furthermore, machine learning methods, when used alone, cannot identify deep underlying associations between asset prices and conditioning variables. Nonetheless, machine learning techniques can be beneficial in understanding economic mechanisms, provided that economists incorporate structure into the estimation problem and determine how to introduce a machine learning algorithm within that structure. A nascent literature has begun to make progress marrying machine learning with equilibrium asset pricing, and this remains an exciting direction for future research.

CRediT authorship contribution statement

Shuiyang Pan: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Funding acquisition. **Suwan(Cheng) Long:** Conceptualization, Formal analysis, Data curation, Software, Writing – original draft, Writing – review & editing. **Yiming Wang:** Conceptualization, Investigation, Resources, Supervision, Project administration, Writing – original draft, Funding acquisition. **Ying Xie:** Conceptualization, Formal analysis, Data curation, Software, Writing – review & editing.

Data availability

Data will be made available on request.

References

Adrian, T., & Franzoni, F. (2009). Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *Journal of Empirical Finance*, *16*(4), 537–556.

Andrew, W. L. (1991). Long-term memory in stock market prices. *Econometrica*, *59*(5), 1279–1313.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, *355*(6324), 483–485.

Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A. W., & Siddique, A. (2016). Risk and risk management in the credit card industry. *Journal of Banking & Finance*, *72*, 218–239.

Campbell, J. Y. (2000). Asset pricing at the millennium. *The Journal of Finance*, *55*(4), 1515–1567, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00260.

Cao, S.-n., Li, H.-d., & Wang, Y. (2010). Long-term memory in realized volatility: Evidence from Chinese stock market. In *2010 third international conference on business intelligence and financial engineering* (pp. 323–327). http://dx.doi.org/10.1109/BIFE.2010.82.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., & Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. *American Economic Review*, *106*(5), 124–127.

Davidson, J., & Teräsvirta, T. (2002). Long memory and nonlinear time series. *Journal of Econometrics*, *110*(2), 105–112.

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106.

Doudchenko, N., & Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. In *Working paper series*, National Bureau of Economic Research.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*(61), 2121–2159.

Engel, C., & Rodrigues, A. P. (1989). Tests of international CAPM with time-varying covariances. *Journal of Applied Econometrics*, *4*(2), 119–138, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.3950040203.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1992.tb04398.x.

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

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, *19*(1), 3–29.

Granger, C. W. J., & Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, *73*(1), 61–77.

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

Hartford, J., Lewis, G., Leyton-Brown, K., & Taddy, M. (2017). *Vol. 3*, *Deep IV: A flexible approach for counterfactual prediction* (pp. 2257–2268).

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

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780, Conference Name: Neural Computation.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, *28*(3), 650–705.

Hutchinson, J. M., Lo, A. W., & Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance*, *49*(3), 851–889, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1994.tb00081.x.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *The Journal of Finance*, *48*(1), 65–91, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1993.tb04702.x.

Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance*, *68*(5), 1721–1756, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12060.

Kelly, B., & Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, *186*(2), 294–316.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767–2787.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859.

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

Lettau, M., & Ludvigson, S. (2001). Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, *109*(6), 1238–1287.

Lewellen, J., & Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, *82*(2), 289–314.

Liu, Q., Wang, C., Zhang, P., & Zheng, K. (2021). Detecting stock market manipulation via machine learning: Evidence from China securities regulatory commission punishment cases. *International Review of Financial Analysis*, *78*, Article 101887.

Lu, X., Ma, F., Xu, J., & Zhang, Z. (2022). Oil futures volatility predictability: New evidence based on machine learning models. *International Review of Financial Analysis*, *83*, Article 102299.

Ma, C., Li, H., Zou, L., & Wu, Z. (2006). Long-term memory in emerging markets: evidence from the chinese stock market. *International Journal of Information Technology and Decision Making*, *05*(03), 495–501.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

Panas, E. (2001). Estimating fractal dimension using stable distributions and exploring long memory through ARFIMA models in athens stock exchange. *Applied Financial Economics*, *11*(4), 395–402.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International stock return predictability: What is the role of the United States? *The Journal of Finance*, *68*(4), 1633–1662, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12041.

Sadique, S., & Silvapulle, P. (2001). Long-term memory in stock market returns: international evidence. *International Journal of Finance & Economics*, *6*(1), 59–67, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijfe.143.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28.

Varian, H. R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences*, *113*(27), 7310–7315.

Wang, X., Wang, X., Li, B., & Bai, Z. (2019). The nonlinear characteristics of Chinese stock index futures yield volatility: Based on the high frequency data of CSI300 stock index futures. *China Finance Review International*, *10*(2), 175–196.

Yang, K., & Chen, L. (2014). Realized volatility forecast: Structural breaks, long memory, asymmetry, and day-of-the-week effect. *International Review of Finance*, *14*(3), 345–392.

Yong, L. (2008). The long-term memory analysis of industrial indices of the Chinese stock market. *Journal of Physics: Conference Series*, *96*, Article 012120.

Zhang, L. (2017). The investment CAPM. *European Financial Management*, *23*(4), 545–603, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/eufm.12129.

Zhang, S., Choromanska, A., & LeCun, Y. (2015). Deep learning with elastic averaging SGD. arXiv:1412.6651, [cs, stat].