CENTER FOR ADVANCED RESEARCH IN FINANCE
GRADUATE SCHOOL OF ECONOMICS, THE UNIVERSITY OF TOKYO

CARF Working Paper

CARF-F-588

A New Equity Investment Strategy with Artificial Intelligence, Multi Factors, and Technical Indicators

Daiya Mita
Nomura Asset Management Co, ltd.,
Graduate School of Economics, The University of Tokyo

Akihiko Takahashi
Graduate School of Economics, The University of Tokyo

July 8, 2024

CARF is presently supported by Nomura Holdings, Inc., Mitsubishi UFJ Financial Group, Inc., Sumitomo Mitsui Banking Corporation., Sumitomo Mitsui Trust Bank, Limited, The University of Tokyo Edge Capital Partners Co., Ltd., Brevan Howard Asset Management LLP, Ernst & Young ShinNihon LLC, SUMITOMO LIFE INSURANCE COMPANY, and All Nippon Asset Management Co., Ltd.. This financial support enables us to issue CARF Working Papers.

CARF Working Papers can be downloaded without charge from:
https://www.carf.e.u-tokyo.ac.jp/research/

Working Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Working Papers may not be reproduced or distributed without the written consent of the author.

A New Equity Investment Strategy with Artificial Intelligence, Multi Factors, and Technical Indicators

Daiya Mita[a b], Akihiko Takahashi[b]

[a] Nomura Asset Management Co, ltd., 12F Toyosu Bayside Cross Tower, 2-2-1 Toyosu, Koto-ku, Tokyo 135-0061, Japan

[b] Graduate School of Economics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Abstract

This study proposes a novel equity investment strategy that effectively integrates artificial intelligence (AI) techniques, multi factor models and financial technical indicators. To be practically useful as an investment fund, the strategy is designed to achieve high investment performance without losing interpretability, which is not always the case especially for complex models based on artificial intelligence.

Specifically, as an equity long (buying) strategy, this paper extends a five factor model in Fama & French [1], a well-known finance model for its explainability to predict future returns by using a gradient boosting machine (GBM) and a state space model. In addition, an index futures short (selling) strategy for downside hedging is developed with IF-THEN rules and three technical indicators. Combining individual equity long and index futures short models, the strategy invests in high expected return equities when the expected return of the portfolio is positive and also the market is expected to rise, otherwise it shorts (sells) index futures. To the best of our knowledge, the current study is the first attempt to develop an equity investment strategy based on a new predictable five factor model, which becomes successful with effective use of AI techniques and technical indicators.

Finally, empirical analysis shows that the proposed strategy outperforms not only the baseline buy-and-hold strategy, but also typical mutual funds for the Japanese equities.

Disclaimer: The views expressed in this paper are our own and do not reflect the institutions the authors are affiliated with. The authors are not responsible or liable in any manner for any losses and/or damages caused by the use of any contents in this research.

1 Introduction

1.1 Introduction

In recent years, artificial intelligence (AI) has proven to be useful across various industries. One of the most challenging applications is finance, where there exists a serious problem that many active equity mutual funds can not outperform a simple index buy-and-hold strategy. For instance, Tables 1 and 2 compare the performance of typical equity mutual funds for Japanese stocks and a buy-and-hold strategy of standard stock index (TOPIX[1]) futures. Table 1 shows that only three funds outperform the index futures buy-and-hold strategy in terms of Sharpe Ratio (SR), one of the most popular metrics for evaluating return/risk of investment strategies. Table 2 reveals that after taking management fee and fund distribution into account, only one fund outperforms the strategy in terms of SR, where distributions in mutual funds are payments to shareholders from the fund's income and capital gains. Under such circumstances, many researchers have been trying to develop new investment strategies using AI techniques to outperform the simple index buy-and-hold strategy.

Table 1: Performance comparison (01/04/2012-12/29/2023)

	AR	SD	SR
Fund 1	13.0%	19.4%	0.67
Fund 2	12.6%	19.0%	0.66
Fund 3	12.7%	20.6%	0.62
Index futures buy-and-hold	12.0%	19.5%	0.62
Fund 4	11.3%	19.6%	0.58
Fund 5	11.3%	19.7%	0.57
Fund 6	11.3%	20.7%	0.55
Fund 7	9.7%	19.9%	0.49
Fund 8	8.2%	19.8%	0.42

Note: The funds listed in the above table are the index futures buy-and-hold strategy and typical Japanese equity mutual funds. AR is the annualized average return, SD is the annualized standard deviation, and SR is the Sharpe ratio. These metrics are calculated without considering management fee and fund distribution.

[1] TOPIX is a Japanese stock market index calculated as a market capitalization-weighted average of about 2000 companies listed on the prime section of the Tokyo Stock Exchange.

Table 2: Performance comparison (01/04/2012-12/29/2023) including management fee and fund distribution

	AR(Net)	SD(Net)	SR(Net)
Fund 3	12.7%	20.4%	0.62
Index futures buy-and-hold	12.0%	19.5%	0.62
Fund 1	11.8%	19.4%	0.61
Fund 7	11.5%	19.2%	0.60
Fund 2	11.0%	19.0%	0.58
Fund 4	11.0%	19.4%	0.57
Fund 6	11.6%	20.6%	0.56
Fund 8	10.6%	19.2%	0.55
Fund 5	10.7%	19.6%	0.55

Note: AR(Net), SD(Net), and SR(Net) are the same as AR, SD, and SR in Table 1, respectively, but are calculated including management fee and fund distribution. Index futures buy-and-hold is assumed to have no management fee and fund distribution.

However, using complex AI models such as deep neural networks in investment strategies raises concerns regarding the interpretability. For instance, if AI-based investment strategies suffer substantial losses with no intuitive explanation, most investors will withdraw their money from the fund.

Thus, we develop a novel investment model that demonstrates high performance by effectively incorporating AI and technical indicators with a multi factor model in finance. In fact, there exist multi factor models capable of explaining asset returns using common factors, which are acceptable in both practice and academia. Among them, one of the most famous models, the Fama-French five (FF 5) factor model (Fama & French [1]) decomposes a stock's return into five factors as follows:

$$R_{i,t} = \alpha_i + \beta_i^m MKT_t + \beta_i^s SMB_t + \beta_i^v HML_t + \beta_i^p RMW_t + \beta_i^j CMA_t + \epsilon_{i,t}, \tag{1}$$

where $R_{i,t}$ is a return of a stock i at t and explanatory variables $(MKT_t, SMB_t, HML_t, RMW_t, CMA_t)$ are the market factor, the size factor, the value factor, the profitability factor, and the investment factor, respectively. The brief descriptions of each factor are as follows:

- MKT stands for the market, which shows a return of the market index.
- SMB stands for small minus big, which shows a spread of the returns between small market capitalization (bottom 50%) and big market capitalization (top 50%) stocks.
- HML stands for high minus low, which shows a spread of the returns between high book-to-market denoted by B/M (top 30%) and low B/M (bottom 30%) stocks. Here, B/M is defined as the ratio of book value of shareholders' equity to market capitalization.
- RMW stands for robust minus weak, which shows a spread of the returns between robust profitability (top 30%) and weak profitability (bottom 30%) stocks.
- CMA stands for conservative minus aggressive, which shows a spread of the returns between conservative investment (bottom 30%) and aggressive investment (top 30%) stocks. Here, investment is measured as the change in total assets from the previous year.

Multi factor models are easy to interpret and often used in academia, but are rarely used to construct a portfolio in asset management practice. Since the multi factor model explains asset returns at t using factors at the same time point t, it is not possible to predict future returns.

To address this issue, the current study proposes a novel equity long (buying) model by extending a multi factor model to predict future returns. In particular, two main components in a multi factor model, namely factor exposures (β_i^k, $k = m, s, v, p, j$) and factor returns are sequentially predicted and updated using a gradient boosting machine (GBM) and a state space model. Also, to hedge against downside risk, an index futures short (selling) model is developed based on IF-THEN rules and three technical indicators.

By combining those equity long and index futures short models, this work achieves a new investment strategy that outperforms typical mutual funds and the index buy-and-hold strategy as shown in Table 3.

Table 3: Performance of our strategies and representative equity funds

	AR	SD	SR	AR(Net)	SD(Net)	SR(Net)
Equity long and futures short model	19.6%	21.0%	0.93	19.4%	21.0%	0.92
Fund 1 (top mutual fund in Table 1)	13.0%	19.4%	0.67	11.8%	19.4%	0.61
Index futures buy-and-hold	12.0%	19.5%	0.62	-	-	-
Fund 3 (top mutual fund in Table 2)	12.7%	20.6%	0.62	12.7%	20.4%	0.62

Note: The index futures buy-and-hold strategy is assumed to have no management fee and fund distribution. The other strategies are assumed to have the management fee and fund distribution estimated as a median of the typical mutual funds.

The organization of this paper is as follows. Section 1.2 reviews the related works. Section 2.1 describes the data used in this research. Section 2.2 introduces the equity long model. Section 2.3 describes the futures short model for downside hedging. Section 3 conducts an empirical analysis to evaluate the performance of the models. Finally, Section 4 concludes the paper.

1.2 Related works

As studies on factor models, Banz (1981) [2] reported that portfolios consisting of small-cap stocks exhibited greater returns than portfolios with large-cap stocks, which is known as the size effect. Rosenberg et al. (1985) [3] found that U.S. stocks' returns were positively related to the ratio of a firm's book value of equity. Integrating these findings, Fama & French [4] proposed the three factor model, which explains the return of an equity using the market return, the size factor, and the value factor. Carhart [5] showed that momentum was also an important factor. Titman et al. [6] demonstrated that companies that substantially increased capital investments subsequently achieved negative benchmark-adjusted returns, which implies that the investment factor is also an important factor. Novy-Marx [7] showed that profitability was also an important factor. Hou et al. [8] proposed the q-factor model, which explains the return of an asset using the market return, the size factor, a profitability factor and an investment factor. Incorporating these findings, Fama & French [1] proposed the five factor model (FF5 factor model), which explains the return using the market return, the size factor, the value factor, the profitability factor, and the investment factor.

Also, there are many studies on the application of AI techniques to asset management. Nakano et al. [9, 10, 11] developed new investment models utilizing anomaly detection and neural networks. Gu et al. [12] performed a comparative analysis of machine learning methods for empirical asset pricing and identified that trees and neural networks are the best-performing methods. Nakano & Takahashi [13] proposed a novel approach for downside hedging based on the factors extracted by AutoEncoder. Takahashi & Takahashi [14] developed a new interval type-2 fuzzy logic system for financial investment with timevarying parameters adaptive to real-time data streams by using an online learning method based on a state-space framework. Khodaee et al. [15] constructed a hybrid Convolutional Neural Network and Long Short-Term Memory model forecasting turning points of stock price. Takahashi & Takahashi [16] and Mita & Takahashi [17] proposed new multi agent models based on state space models, which can improve investment performance by predicting market crashes. Dezhkam & Manzuri [18] introduced a new model called HHT-XGB to predict the changing trends in the next closing price of stocks, which combines the Hilbert-Huang Transform as the feature engineering part and the extreme gradient boost as the closing price trend classifier. Zhang et al. [19] put forward a new LSTM network combined with residual-driven ν support vector regression for index and stock price prediction.

To summarize the above research, while studies on multi factor models focuses on discovering new factors to interpret asset returns and enhancing explanatory power, many AI literatures focus to improve prediction accuracy or investment performance. Apparently, both are important in asset management practice. Thus, this study effectively combines AI techniques, multi factor model and technical indicators to develop a new investment strategy that achieves high performance without losing interpretability.

2 Data and prediction models

This section describes the data and investment models. Section 2.1 explains the dataset used in our analysis. Sections 2.2 and 2.3 introduce equity long and index futures short models, respectively.

2.1 Data

This subsection briefly explains the dataset used in our analysis.

1. Opening prices of Japanese stocks in TOPIX,
2. Opening prices of TOPIX,
3. Opening prices of TOPIX futures[2],
4. Financial data of Japanese companies constituting TOPIX:
    - Book-to-market (B/M) ratio,
    - Operating profit,
    - Shareholders' equity,
    - Total asset,
5. Macro data:
    - 2 year government bond yield in Japan and the United States,
    - 10 year government bond yield in Japan and the United States,
    - 10 year inflation-indexed bond yield in Japan and the United States,
    - Exchange rate between the Japanese yen and the US dollar,
    - Exchange rate between the Japanese yen and the Chinese yuan,
    - Commodity price index.

All data are available from Bloomberg. The Bloomberg tickers of TOPIX, TOPIX futures and macro data are following: TPX Index, TP1 Index, GTJPY2Y Govt, GTJPY10Y Govt, GTJPYII10Y Govt, USGG2YR Index, USGG10YR Index, USGGBE02 Index, USGGBE10 Index, USDJPY Curncy, USDCNY Curncy, and CL1 Comdty. Also, price and financial data are available from JPX data cloud[3]. The data period is from 01/04/2010 to 12/29/2023. This study does not adopt closing prices but opening prices to utilize the previous day's U.S. market information as quickly as possible and the proposed strategy discussed later is also executed at the opening price.

2.2 Equity long model

Section 2.2 proposes a novel equity long model which calculates the expected return ranking for Japanese equities and invests in 40 equities with equal weights, which are almost the top 5% of investable stocks in our analysis. Current work adopts TOPIX 1000 as the investment universe due to its rich liquidity, from which approximately 750 stocks with no missing data are selected for analysis in Section 3. Here, TOPIX 1000 consists of the top 1000 stocks from TOPIX based on market capitalization and liquidity. The return ranking is estimated by effective use of FF5

[2] TOPIX futures are tradable at regular and night sessions. This analysis uses opening price of regular session.
[3] `https://db-ec.jpx.co.jp/?__lang=en`

factors, gradient boosting machine (GBM) and state space model. The architecture of the model is shown in Figure 1.

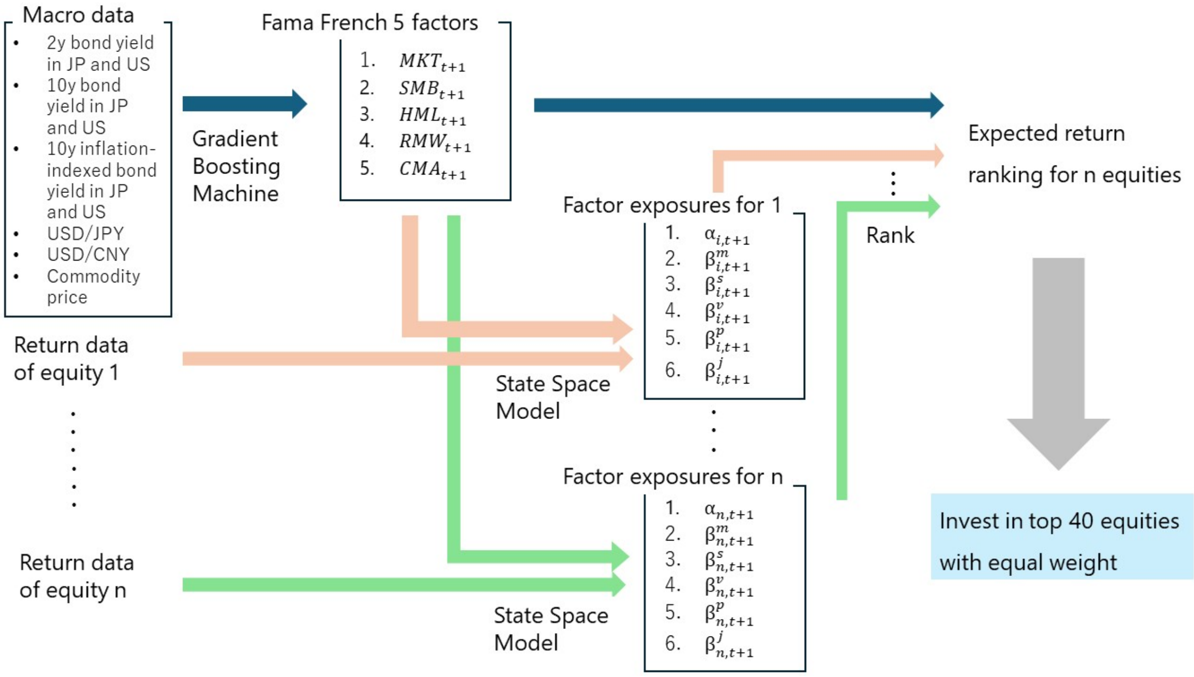

Figure 1: Architecture of equity long model

This work extends the FF5 model represented by equation (1) to predict future returns as follows:

$$
\begin{aligned}
R_{i,t+1} = \alpha_{i,t+1} + \beta^m_{i,t+1} MKT_{t+1} + \beta^s_{i,t+1} SMB_{t+1} + \beta^v_{i,t+1} HML_{t+1} \\
+ \beta^p_{i,t+1} RMW_{t+1} + \beta^j_{i,t+1} CMA_{t+1}, \qquad (2)
\end{aligned}
$$

$$MKT_{t+1} = GBM_{1,t}(macro_t), \qquad (3)$$

$$SMB_{t+1} = GBM_{2,t}(macro_t), \qquad (4)$$

$$HML_{t+1} = GBM_{3,t}(macro_t), \qquad (5)$$

$$RMW_{t+1} = GBM_{4,t}(macro_t), \qquad (6)$$

$$CMA_{t+1} = GBM_{5,t}(macro_t), \qquad (7)$$

where $R_{i,t+1}$, stock i's return at $t+1$, is decomposed into FF5 factors and their coefficients called factor exposures (β^k_i, $k = m, s, v, p, j$) as in equation (2). Future factor returns at $t+1$ are calculated by the gradient boosting machine (GBM) denoted by $GBM_{k,t}$ functions and the macro data available at t. Factor exposures at $t+1$ are estimated by a state space model defined by equations (8) to (10), which are described later.

First, to train the GBM, historical FF5 factors are required. While following Fama & French [1] in factor calculations, we cannot obtain full market capitalization data necessary in precise treatment due to limitation of our accessibility to data sources. Hence, when rigorously speaking full market capitalization data is necessary, we make some simplification in each factor calculation. Nonetheless, we are able to draw meaningful results reported below.

- MKT (market factor): The market factor is calculated as the return of the market index (TOPIX).
- SMB (small minus big, size factor): The size factor is calculated as the spread between the average returns of stocks not included in TOPIX 1000 (small) and those in TOPIX 1000 (big).

- HML (high minus low, value factor): The value factor is computed by the following steps.
  1. Divide stocks into two size groups (small and big as in SMB calculation),
  2. Independently of step 1, divide stocks into three groups based on B/M (book-to-market ratio), where groups are defined as the bottom 30%, the middle 40%, and the top 30% of B/M ratios (low, middle, and high B/M stocks),
  3. Combine the size and B/M classifications to construct six portfolios with equal weights: (1) small-low (SL), (2) small-middle (SM), (3) small-high (SH), (4) big-low (BL), (5) big-middle (BM), and (6) big-high (BH) B/M portfolios,
  4. Calculate the spread between the average returns of the two high and low B/M portfolios: $HML = (R_{SH} + R_{BH})/2 - (R_{SL} + R_{BL})/2$, where R_x means the return of the portfolio x listed in step 3.
- RMW (robust minus weak, profitability factor): The profitability factor is computed by the following steps.
  1. Divide stocks into two size groups (small and big as in SMB calculation),
  2. Independently of step 1, divide stocks into three groups based on profitability, which is measured with dividing operating profit in Profit and Loss Statement by shareholders' equity in Balance Sheet, where groups are defined as the top 30%, the middle 40%, and the bottom 30% of profitability ratios (robust, neutral, and weak profitability stocks),
  3. Combine the size and profitability classifications to construct six portfolios with equal weights: (1) small-robust (SR), (2) small-neutral (SN), (3) small-weak (SW), (4) big-robust (BR), (5) big-neutral (BN), and (6) big-weak (BW) profitability portfolios,
  4. Calculate the spread between the average returns of the two robust and weak profitability portfolios: $RMW = (R_{SR}+R_{BR})/2-(R_{SW}+R_{BW})/2$, where R_x means the return of the portfolio x listed in step 3.
- CMA (conservative minus aggressive, investment factor): The investment factor is computed by the following stepss.
  1. Divide stocks into two size groups (small and big as in SMB calculation),
  2. Independently of step 1, divide stocks into three groups based on investment, which is defined as changes of total assets from those in the previous year, where groups are defined as the bottom 30%, the middle 40%, and the top 30% of investment ratios (conservative, neutral, and aggressive investment stocks),
  3. Combine the size and investment classifications to construct six portfolios with equal weights: (1) small-conservative (SC), (2) small-neutral (SN), (3) small-aggressive (SA), (4) big-conservative (BC), (5) big-neutral (BN), and (6) big-aggressive (BA) investment portfolios,
  4. Calculate the spread between the average returns of the two conservative and aggressive investment portfolios: $CMA = (R_{SC}+R_{BC})-(R_{SA}+R_{BA})$, where R_x means the return of the portfolio x listed in step 3.

Second, equations (3)-(7) use macro data and GBM to calculate future factor returns (MKT_{t+1}, SMB_{t+1}, HML_{t+1}, RMW_{t+1}, CMA_{t+1}). GBM builds an ensemble of decision trees to identify nonlinear relationship between inputs and output, and generally outperforms traditional linear regression models in many machine learning tasks, such as Kaggle[4] competitions. This paper uses the LightGBM[5] library and Python to conduct GBM and the parameters are set as shown in Table 4.

[4]Kaggle is a data science competition platform and online community of data scientists and machine learning practitioners under Google LLC.

[5]LightGBM is a gradient boosting framework developed by Microsoft.

Table 4: Parameter setting for LightGBM

Parameter	Values
objective	'regression'
metrics	'rmse'
learning_rate	0.01
other parameters	default values

In Table 4, most parameters are set as default values to exclude data mining. The parameters we set are objective, metrics, and learning_rate. The objective and metrics are set as 'regression' and 'rmse' to minimize the root mean square error (RMSE) between the predicted and actual factor returns. The learning_rate means the step size of the gradient descent algorithm. Although setting learning_rate to a small value takes more time, we can obtain a more accurate model.

Third, given that the factor returns are calculated by equations (3)-(7), the exposures $\alpha_{i,t+1}, \beta^k_{i,t+1}$ are estimated by the following state space model.

$$\begin{aligned}
&\text{[observation equation]} \quad && R_{i,t+1} = \alpha_{i,t+1} + \beta^m_{i,t+1} MKT_{t+1} + \beta^s_{i,t+1} SMB_{t+1} + \beta^v_{i,t+1} HML_{t+1} \\
& && \quad + \beta^p_{i,t+1} RMW_{t+1} + \beta^j_{i,t+1} CMA_{t+1} + \epsilon_{i,t}, \epsilon_{i,t} \sim N(0,1), \qquad (8) \\
&\text{[state equations]} \quad && \alpha_{i,t+1} = \alpha_{i,t} + \eta_{i,t}, \eta_{i,t} \sim N(0,1), \qquad (9) \\
& && \beta^k_{i,t+1} = \beta^k_{i,t} + \xi_{i,t}, k = m, s, v, p, j, \xi_{i,t} \sim N(0,1), \qquad (10)
\end{aligned}$$

where $N(0,1)$ is the normal distribution with mean 0 and variance 1. Although factor exposures are often calculated by OLS, using the state space model can avoid arbitrariness regarding the data period. By solving the state space model using Kalman filter, we obtain the expected factor exposures at $t+1$ ($\alpha_{i,t+1}$ and $\beta^k_{i,t+1}$).

Then, substituting the expected factor returns and factor exposures into equation (2), the future stock returns can be derived. However, since returns are very noisy, we convert expected returns to rankings as follows:

$$A_{i,t+1} = rank(R_{i,t+1} | R_{1,t+1}, ..., R_{n,t+1}), \qquad (11)$$

where $A_{i,t+1}$ is the estimated return rank of equity i in the investment universe at $t+1$ and $rank$ is the rank function. Finally, this model invests in the top 40 equities with equal weights.

2.3 Futures short model for hedging

Although the equity long model is developed in Section 2.2, if the market crashes and all stocks go down, the investment will not work well. To solve this problem, an index futures short model for downside protection is developed based on the following IF-THEN rules and three technical indicators.

1. IF $X_{1,t}$ is positive, $X^+_{2,t}$ is positive, or $X_{3,t}$ is positive, THEN the market return is positive.

2. IF $X_{1,t}$ is negative, $X^-_{2,t}$ is negative, or $X_{3,t}$ is negative, THEN the market return is negative.

3. Otherwise, the market return is approximately zero.

where $X_{1,t}$ is a trend variable, $X_{2,t}$ is a Bollinger variable, and $X_{3,t}$ is a moving average convergence divergence (MACD) variable defined below.

Firstly, $X_{1,t}$ is defined as follows:

$$\begin{aligned}
X_{1,t} &= L_{5,t} - L_{25,t}, \\
L_{d,t} &= \frac{\sum_{i=1}^{d} (P_{t-i} - \bar{P})(D_i - \bar{D})}{\sum_{i=1}^{d} (D_i - \bar{D})^2}, D_i = d - i, \bar{d} = (d-1)/2,
\end{aligned} \qquad (12)$$

where P_t is the index (TOPIX futures) price at t, $\bar{P}$ is the average price of the past d days, D_i is the number of days ago, and $\bar{d_i}$ is the average of d_i. Then, L_d means the linear regression coefficient of the following equation.

$$\begin{aligned} P_t &= a + bD_t + e_t \\ D_t &= t \end{aligned} \tag{13}$$

If $X_{1,t}$ is positive (negative), the price is in an uptrend (downtrend).

Secondly, $X_{2,t}$ is defined as following:

$$\begin{aligned} X_{2,t}^{+} &= Bol^{-}(P_{t-1}) - P_{t-1}, \\ X_{2,t}^{-} &= Bol^{+}(P_{t-1}) - P_{t-1}, \\ Bol^{-}(P_{t-1}) &= \bar{P}_{t-1} - 2\sqrt{\frac{1}{25}\sum_{i=1}^{25}(P_{t-i} - \bar{P}_{t-1})^2}, \\ Bol^{+}(P_{t-1}) &= \bar{P}_{t-1} + 2\sqrt{\frac{1}{25}\sum_{i=1}^{25}(P_{t-i} - \bar{P}_{t-1})^2}, \\ \bar{P}_{t-1} &= \frac{1}{25}\sum_{i=1}^{25} P_{t-i}, \end{aligned} \tag{14}$$

where $X_{2,t}^{+}$ ($X_{2,t}^{-}$) is the difference between the lower (upper) Bollinger band and P_{t-1}. This study uses the 25-day moving average and standard deviation to calculate the Bollinger band. The 25-day setting is one of the standard settings. If $X_{2,t}^{+}$ ($X_{2,t}^{-}$) is positive (negative), P_{t-1} is under (over) the lower (upper) Bollinger band and expected to go up (down).

Finally, $X_{3,t}$ is based on the well-known technical indicator called MACD and is defined as follows:

$$\begin{aligned} X_{3,t} &= MACD_{t-1} - \frac{1}{\alpha}\sum_{i=1}^{\alpha} MACD_{t-i}, \\ MACD_{t-1} &= \bar{p}_{t-1}^{\beta_1} - \bar{p}_{t-1}^{\beta_2}, \\ \bar{p}_{t-1}^{\beta} &= \beta P_{t-1} + (1-\beta)\bar{p}_{t-2}^{\beta}, \beta = \beta_1, \beta_2, \end{aligned} \tag{15}$$

where $\alpha = 9$, $\beta_1 = 2/(1+12)$ and $\beta_2 = 2/(1+26)$. The choices of these parameters are standard, as shown in Investopedia. IF $X_{3,t}$ is positive (negative), the MACD indicates that the price will go up (down).

3 Empirical Analysis

Section 3 presents an empirical analysis to evaluate the performance of the models introduced in Sections 2.2 and 2.3.

Concretely, this study compares the following three strategies:

- (i) Index futures buy-and-hold strategy.
- (ii) Equity long model: long-only investment strategy using the model in Section 2.2, which is developed with the FF5 factors, GBM and state space model.
- (iii) Equity long and futures short models: long-short strategy using the models in Sections 2.2 and 2.3. Long strategy is the same as strategy (ii) and futures short strategy for hedging is developed with IF-THEN rules and technical indicators.

Strategy (i) is a baseline strategy, which invests in TOPIX futures and holds for the entire simulation period from 01/04/2012 to 12/29/2023. Strategy (ii) is a long-only strategy that invests in the top 40 equities based on the return ranking model in Section 2.2. The portfolio weight is equally distributed among the 40 equities and the portfolio is rebalanced daily. Strategy (iii) is a long-short strategy that enhances strategy (ii) by incorporating the downside protection model proposed in Section 2.3. Specifically, this strategy invests in the top 40 equities when the expected return of the top 40 portfolio is positive and also the market is expected to go up. Otherwise, it shorts TOPIX futures. By comparing the performance of the strategies, we can evaluate the effectiveness of the models in Sections 2.2 and 2.3. Trading costs are set to 1 bp for investing in futures and 7 bp for an individual stock[6]. The investment universe consists of TOPIX 1000, from which approximately 750 stocks with no missing data are selected for this analysis.

Also, the following evaluation metrics are used:

1. Average return (AR): an annualized average return of the strategy.

$$AR = \frac{250}{T}\sum_{t=0}^{T} r_t. \tag{16}$$

2. Standard deviation (SD): a risk measure defined as the annualized standard deviation of the return.

$$SD = \left\{\frac{250}{T}\sum_{t=0}^{T}(r_t - \bar{r})^2\right\}^{1/2}, \bar{r} = \frac{1}{T}\sum_{t=0}^{T} r_t. \tag{17}$$

3. Sharpe ratio (SR): a risk adjusted return measure defined as the average return (AR) divided by the SD.

$$SR = AR/SD. \tag{18}$$

The results are shown in Figure 2 and Table 5.

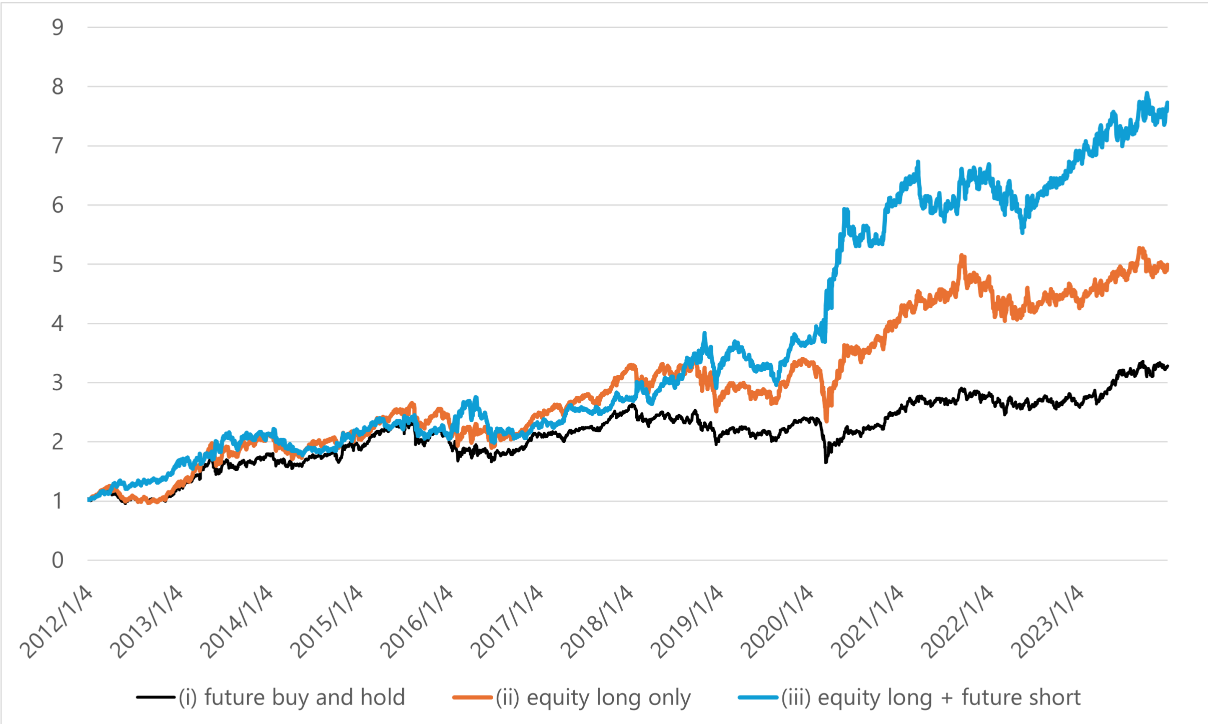

Figure 2: Equity curve of each strategy.

[6]Kudo & Sato [20] revealed that transaction cost in the Japanese equity market had decreased to approximately 7 basis points.

Table 5: Performance of the strategies

	AR	SD	SR
(i) Index futures buy-and-hold	12.0%	19.5%	0.61
(ii) Equity long model	16.0%	21.3%	0.74
(iii) Equity long and futures short model	19.6%	21.0%	0.93

First, as shown in Figure 2 and Table 5, strategies (ii) and (iii) outperform the baseline strategy (i) in terms of AR and SR. Especially, strategy (ii) has consistently outperformed the baseline strategy and strategy (iii) has considerably outperformed during the COVID-19 period.

Second, to investigate the robustness of our models, the same experiments are conducted for other periods. Concretely, simulations during the last 5 years (from 01/04/2019 to 12/29/2023) and 10 years (from 01/04/2014 to 12/29/2023) are shown in Tables 6 and 7, respectively.

Table 6: Performance of the strategies since 2019

	AR	SD	SR
(i) Index futures buy-and-hold	11.3%	18.0%	0.63
(ii) Equity long model	14.7%	20.4%	0.72
(iii) Equity long and futures short model	19.8%	19.8%	1.00

Table 7: Performance of the strategies since 2014

	AR	SD	SR
(i) Index futures buy-and-hold	8.0%	18.9%	0.42
(ii) Equity long model	11.0%	21.0%	0.52
(iii) Equity long and futures short model	15.2%	20.4%	0.75

As shown in Tables 6 and 7, strategy (ii) outperform the baseline strategy (i) in terms of AR and SR. Moreover, strategy (iii) consistently achieves higher AR and SR regardless of the period, which is consistent with the results in Table 5.

Third, to evaluate the advantage of our factor return prediction model based on GBM more precisely, we compare the performance of strategies (ii) and (iii) with those using the previous factor returns as predicted factor returns. The results are shown in Table 8.

Table 8: Performance of the strategies using the return from two days ago to the previous day

	AR	SD	SR
(ii) Equity long model	16.0%	21.3%	0.74
(ii)' Equity long model (without GBM)	12.1%	21.3%	0.57
(iii) Equity long and futures short model	19.6%	21.0%	0.93
(iii)' Equity long and futures short model (without GBM)	15.9%	21.0%	0.76

Table 8 demonstrates that the strategies (ii) and (iii) using GBM outperform the strategies (ii)' and (iii)' in terms of AR and SR, respectively. This result indicates that GBM is effective in enhancing the performance.

Finally, we compare the performance of our strategies with that of typical active mutual funds for Japanese equities. The mutual funds are selected as follows:

1. We choose top two asset management companies each in public and private investment trusts based on assets under management (AUM) as of 04/06/2024. Each asset management company's AUM data is disclosed on the Investment Trust Association's website[7].

[7] `https://www.toushin.or.jp/statistics/` (in Japanese)

2. Next, for each asset management company, we sort Japanese equity mutual funds based on AUM by using SBI securities' power search[8] as of 04/06/2024.

3. Then, from the sorted funds in each company we select top two AUM active funds which do not focus on a specific sector nor high dividend stocks, because our investment strategy is designed to be applicable to a broad class of individual stocks with large capitalization and high liquidity enough to be widely used by institutional investors.

Each fund's performance is calculated based on its net asset value per share data disclosed on its website. For more practical performance evaluation, the current work adds the evaluation metrics AR(Net), SD(Net), and SR(Net), which take management fee and fund distribution into account.

As shown in Tables 9 and 10, our strategies (ii) and (iii) outperform not only (i) futures buy-and-hold but also the typical active mutual funds in terms of AR and SR.

Table 9: Performance comparison

	AR	SD	SR
(iii) Equity long and futures short model	19.6%	21.0%	0.93
(ii) Equity long model	16.0%	21.3%	0.74
Fund 1	13.0%	19.4%	0.67
Fund 2	12.6%	19.0%	0.66
Fund 3	12.7%	20.6%	0.62
(i) Index futures buy-and-hold	12.0%	19.5%	0.62
Fund 4	11.3%	19.6%	0.58
Fund 5	11.3%	19.7%	0.57
Fund 6	11.3%	20.7%	0.55
Fund 7	9.7%	19.9%	0.49
Fund 8	8.2%	19.8%	0.42

Table 10: Performance comparison including management fee and fund distribution

	AR(Net)	SD(Net)	SR(Net)
(iii) Equity long and futures short model	19.4%	21.0%	0.92
(ii) Equity long model	15.8%	21.3%	0.74
Fund 3	12.7%	20.4%	0.62
(i) Index futures buy-and-hold	12.0%	19.5%	0.62
Fund 1	11.8%	19.4%	0.61
Fund 7	11.5%	19.2%	0.60
Fund 2	11.0%	19.0%	0.58
Fund 4	11.0%	19.4%	0.57
Fund 6	11.6%	20.6%	0.56
Fund 3	10.6%	19.2%	0.55
Fund 5	10.7%	19.6%	0.55

Note: The management fee and fund distribution for strategies (ii) and (iii) are estimated as a median of the eight representative equity funds. The baseline strategy (i) is assumed to have no management fee and fund distribution.

4 Conclusion

This paper has proposed a novel equity investment strategy developed by the effective combination of Fama-French five (FF5) factors, AI techniques, and technical indicators. Concretely, it has

[8] `https://site0.sbisec.co.jp/marble/fund/powersearch/fundpsearch.do?` (in Japanese)

extended the well-known FF5 model to predict future returns using a gradient boosting machine (GBM) and a state space model for an individual stock long (buying) model. In addition, the study has developed an index futures short (selling) model for downside hedging based on IF-THEN rules and technical indicators. By integrating these two models, the proposed strategy outperforms the baseline strategy (TOPIX futures buy-and-hold strategy) in terms of average return (AR), risk and Sharpe ratio (SR), and consistently achieves high performance regardless of the period. Moreover, comparing the performance of our strategies with that of typical equity mutual funds for the Japanese stocks shows that our strategy outperforms those funds in terms of AR and SR.

However, the current study has some limitations. First, although the proposed strategy adopts the FF5 factor model, there exist many other multi factor models. While the model is a de facto standard for explaining the asset returns, it is not necessarily superior to the others in terms of investment performance. Thus, applying the proposed methods to other multi factor models may improve the risk and return reported in this paper.

Second, this study ignores the intervention of the public sector such as the central bank and the government, which has significant impacts on the market. For instance, our previous study Mita & et al. [21] investigated the effect of the Bank of Japan's (the central bank of Japan) ETF purchase on the Japanese equity market and found that the ETF purchase has a significant impact on the market. Thus, by incorporating the effect of such a public sector into the proposed model, we may improve the strategy's performance.

Acknowledgement

This work was supported by CARF (Center for Advanced Research in Finance) and CIRJE (Center for International Research on the Japanese Economy), The University of Tokyo.

References

[1] E. F. Fama, K. R. French, A five-factor asset pricing model, Journal of Financial Economics vol.116 (2015) pp.1–22. doi:https://doi.org/10.1016/j.jfineco.2014.10.010.

[2] R. W. Banz, The relationship between return and market value of common stocks, Journal of Financial Economics vol.9 (1) (1981) pp.3–18. doi:https://doi.org/10.1016/0304-405X(81)90018-0.

[3] B. Rosenberg, K. Reid, R. Lanstein, Persuasive evidence of market inefficiency, The Journal of Portfolio Management vol.11 (3) (1985) pp.9–16. doi:https://doi.org/10.3905/jpm.1985.409007.

[4] E. F. Fama, K. R. French, Common risk factors in the returns on stocks and bonds, Journal of financial economics vol.33 (1) (1993) pp.3–56. doi:https://doi.org/10.1016/0304-405X(93)90023-5.

[5] M. M. Carhart, On persistence in mutual fund performance, The Journal of Finance vol.52 (1) (1997) pp.57–82. doi:https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

[6] S. Titman, K. C. J. Wei, F. Xie, Capital investments and stock returns, Journal of Financial and Quantitative Analysis vol.39 (4) (2004) pp.667–700. doi:https://doi.org/10.1017/S0022109000003173.

[7] R. Novy-Marx, The other side of value: The gross profitability premium, Journal of Financial Economics vol.108 (1) (2013) pp.1–28. doi:https://doi.org/10.1016/j.jfineco.2013.01.003.

[8] K. Hou, C. Xue, L. Zhang, Digesting anomalies: An investment approach, The Review of Financial Studies vol.28 (3) (2015) pp.650–705. doi:https://doi.org/10.1093/rfs/hhu068.

[9] M. Nakano, A. Takahashi, S. Takahashi, Generalized exponential moving averagge (ema) model with particle filtering and anomaly detection, Expert Systems with Applications vol.73 (2017) pp.187–200. doi:https://doi.org/10.1016/j.eswa.2016.12.034.

[10] M. Nakano, A. Takahashi, S. Takahashi, Creating investment scheme with state space modeling, Expert Systems with Applications vol.81 (2017) pp.53–66. doi:https://doi.org/10.1016/j.eswa.2017.03.045.

[11] M. Nakano, A. Takahashi, S. Takahashi, Fuzzy logic-based portfolio selection with particle filtering and anomaly detection, Knowledge-Based Systems vol.131 (2017) pp.113–124. doi:https://doi.org/10.1016/j.knosys.2017.06.006.

[12] S. Gu, B. Kelly, D. Xiu, Empirical asset pricing via machine learning, The Review of Financial Studies vol.33 (5) (2020) pp.2223–2273. doi:https://doi.org/10.1093/rfs/hhaa009.

[13] M. Nakano, A. Takahashi, A new investment method with autoencoder: Applications to crypto currencies, Expert Systems with Applications vol.162 (2020) pp.1–11. doi:https://doi.org/10.1016/j.eswa.2020.113730.

[14] A. Takahashi, S. Takahashi, A new interval type-2 fuzzy logic system under dynamic environment: Application to financial investment, Engineering Applications of Artificial Intelligence vol.100 (2021) pp.1–11. doi:https://doi.org/10.1016/j.engappai.2021.104154.

[15] P. Khodaee, A. Esfahanipour, H. M. Taheri, Forecasting turning points in stock price by applying a novel hybrid cnn-lstm-resnet model fed by 2d segmented images, Engineering Applications of Artificial Intelligence vol.116 (2022) pp.1–22. doi:https://doi.org/10.1016/j.engappai.2022.105464.

[16] A. Takahashi, S. Takahashi, A state space modeling for proactive management in equity investment, International Journal of Financial Engineering vol.9 (No.4, 2250025) (2022) pp.1–29. doi:https://doi.org/10.1142/S2424786322500256.

[17] D. Mita, A. Takahashi, Multi-agent model based proactive risk management for equity investment, Engineering Applications of Artificial Intelligence vol.125. doi:https://doi.org/10.1016/j.engappai.2023.106701.

[18] A. Dezhkam, M. T. Manzuri, Forecasting stock market for an efficient portfolio by combining xgboost and hilbert-huang transform, Engineering Applications of Artificial Intelligence vol.118 (2023) pp.1–13. doi:https://doi.org/10.1016/j.engappai.2022.105626.

[19] Y. Zhang, Y. Song, G. Wei, A feature-enhanced long short-term memory network combined with residual-driven v support vector regression for financial market prediction, Engineering Applications of Artificial Intelligence vol.118 (Article 105663) (2023) pp.1–10. doi:https://doi.org/10.1016/j.engappai.2022.105663.

[20] H. Kudo, K. Sato, An empirical analysis of transaction costs in the japanese equity market (in japanese), Securities Analysts Journal.

[21] D. Mita, K. G. Nishimura, T. Saito, A. Takahashi, A dynamic analysis of the bank of japan's etf/reit purchase program, Expert Systems with Applications vol.235. doi:https://doi.org/10.1016/j.eswa.2023.121091.