
A ROBUST TRANSFERABLE DEEP LEARNING FRAMEWORK FOR CROSS-SECTIONAL INVESTMENT STRATEGY

A PREPRINT

Kei Nakagawa

Innovation Lab

Nomura Asset Management Co.,

1-11-1 Nihonbashi, Chuo-ku, Tokyo, 103-8260, Japan

kei.nak.0315@gmail.com

Masaya Abe

Innovation Lab

Nomura Asset Management Co.,

1-11-1 Nihonbashi, Chuo-ku, Tokyo, 103-8260, Japan

masaya.abe.428@gmail.com

Junpei Komiyama

Leonard N. Stern School of Business

New York University

44 West 4th Street, New York, NY 10012

junpeikomiyama@gmail.com

October 4, 2019

ABSTRACT

Stock return predictability is an important research theme as it reflects our economic and social organization, and significant efforts are made to explain the dynamism therein. Statistics of strong explanative power, called “factor” [1] have been proposed to summarize the essence of predictive stock returns. Although machine learning methods are increasingly popular in stock return prediction [2], an inference of the stock returns is highly elusive, and still most investors, if partly, rely on their intuition to build a better decision making. The challenge here is to make an investment strategy that is consistent over a reasonably long period, with the minimum human decision on the entire process. To this end, we propose a new stock return prediction framework that we call Ranked Information Coefficient Neural Network (RIC-NN). RIC-NN is a deep learning approach and includes the following three novel ideas: (1) nonlinear multi-factor approach, (2) stopping criteria with ranked information coefficient (rank IC), and (3) deep transfer learning among multiple regions. Experimental comparison with the stocks in the Morgan Stanley Capital International (MSCI) indices shows that RIC-NN outperforms not only off-the-shelf machine learning methods but also the average return of major equity investment funds in the last fourteen years.

1 Introduction

Stock return predictability has been an important research theme as it reflects our economic and social organization. Although the dynamic nature of our economic activity makes it harder to predict the future returns of the stocks, significant efforts are made to explain the dynamism therein. Statistics of strong explanative powers, called “factor” [1], are proposed to summarize the essence of predictive stock returns, and a large portion of investors develop their portfolio strategies based on these factors. For example, Book-value to Price ratio (net asset of a company divided by the market value of the corresponding stock) is one of the nominal factors, and this factor combined with simple sorting portfolio yields a positive return [1]. Due to their predictive power and robustness, investment decisions by professional investors are heavily dependent on the factors.

Machine learning is an increasingly popular tool for predicting unknown target variables; the last decades saw many attempts to apply machine learning algorithms to support smart decision-making in different financial segments [3, 4, 2]. Still, its highly elusive nature makes it harder to make a consistent inference: Most investors, if partly, rely on their intuition to build a better decision-making.

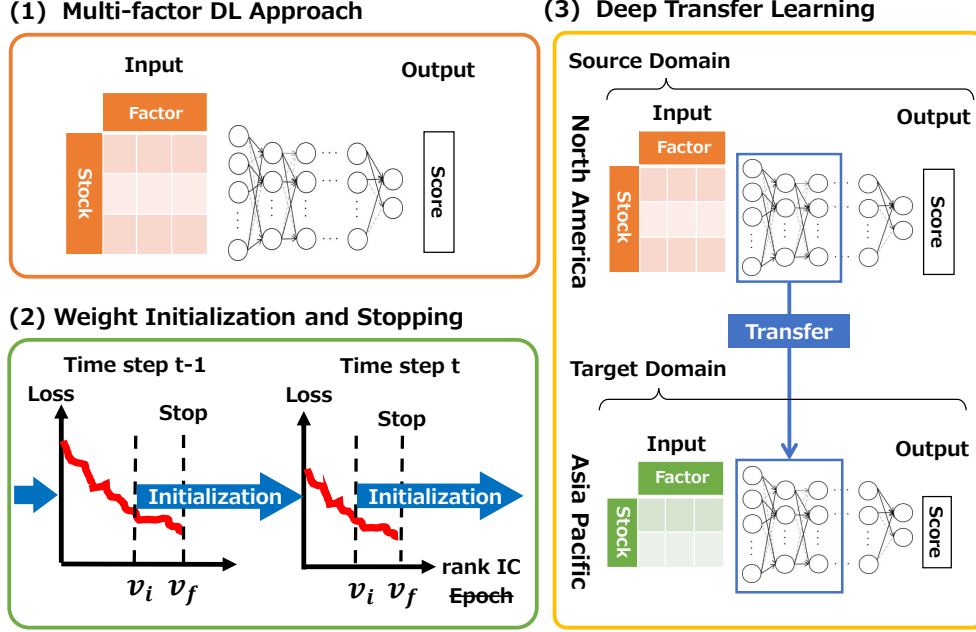


Figure 1: Our approach: RIC-NN.

The challenge in this paper is to make an investment strategy that is consistent over a fairly long period, with the smallest human intervention on the entire process. We propose a novel approach, called Rank Information Coefficient Neural Net (RIC-NN) for developing an investment strategy. Most of the quantitative investment strategies require a ranking over the stock returns, and we made a ranking by using a deep learning (DL) approach. In particular, the largest advantage of the machine learning lies in its capability to learn the nonlinear relationship between the factors and the stock returns [5, 6], and the DL approach is recently reported to outperform other more traditional approaches in many domains, such as natural language processing [7], media recognition [8], and time-series forecasting.

Due to the dynamic nature of our economic activity, naive use of the off-the-shelf machine learning tool easily overfits the existing data, and thus it fails to predict the future stock returns. For example, [9] applied deep learning to stock market prediction: They reported that the advantage of a deep learning model over a linear autoregressive model has mostly disappeared in the test set. We show that the proposed RIC-NN consistently outperforms other methods based on off-the-shelf machine learning algorithms. Our framework involves three novel ideas (Figure 1): Namely, (1) we propose a deep learning multi-factor approach that enables cross-sectional prediction, and (2) the approach involves a novel training method of neural network based on the rank IC. Our framework is very practical: We conducted a comprehensive evaluation of our approach based on the stocks in the Morgan Stanley Capital International (MSCI) indices. Our evaluation demonstrated that a neural network with a standard training method performs poorly, whereas our RIC-NN alleviates overfitting and outperformed linear models and ensemble-based models. Moreover, the average return of RIC-NN over fourteen years surpasses the ones of major equity investment funds. (3) We furthermore considered an information aggregation among several different markets in MSCI indices: Namely, a transfer learning between the North America (NA) region and the Asia Pacific (AP) region. The experimental results imply that one can utilize the NA data to predict the future returns of the AP market, but not vice versa. The results verify the asymmetric causal structure between the two markets [10, 11].

2 Related Work

There are two major strategies in stock trading: Namely, the one based on time-series analysis [4] and the one based on cross-sectional analysis [12].

The methods of the former strategy analyze past stock prices as time-series data [3] and are applied to a practical trading strategy that focuses on a particular stock. Indeed, financial time-series forecasting can be considered one of the significant challenges in time series and machine learning literature [13]. The study of financial time-series was originally started from a linear model, such as the autoregressive (AR) model in which the parameters are

uniquely determined [14]. Introduction of machine learning techniques to the literature enabled us to capture nonlinear relationship among relevant factors without prior knowledge about the input data distribution [3]. Still, an application of nonlinear methods to time-series data is highly non-trivial [9].

The methods of the latter strategy, which include the work in this paper, perform a **regression analysis using cross-sectional data of corporate attributes**. Such a strategy aims to build a portfolio for investing as a subset of a large bucket of stocks and is applied to a practical quantitative investment strategy [15]. One of the most significant interests in a cross-sectional analysis lies in finding “factors” that have strong predictive powers to the expected return of a cross-sectional trading strategy: The Fama-French three-factor model [1, 16] is one of the nominal works in this field. They argued that the cross-sectional structure of the stock price can be explained by three factors: Namely, the beta (market portfolio), the size (market capitalization), and the value (Book-value to price ratio; BPR). This argument inspires many subsequent research papers that propose more sophisticated versions of factors. [17] surveyed the history of the proposed factors and argued that the number of reported factors shows a rapid increase in the last two decades: Each year from 1980 to 1991 we saw a new factor, whereas each year from 1991 to 2003 we saw five new factors. In the period from 2003 to 2012, the number of factors proposed at each year rose sharply to around 18. As a result, over 300 factors were discovered until 2012.

While each factor shows a positive correlation to the investment strategy, the **effectiveness of the factors significantly varies over time and among different markets**. Figure 2 and 3 show the cumulative returns of the long-short portfolio strategy based on each single factor in Appendix A, from which one can see that the return of each single factor varies largely over time. These results motivate a multi-factor approach where more than one factor is taken into consideration to the aim of better returns [5, 6].

In particular, machine learning approaches, which can capture the nonlinear relationship among multiple factors, are recently applied to a cross-sectional analysis. [18] applied the LASSO [19] in the U.S. stock market, [20] applied an auto-encoder based nonlinear model into a U.S. biotechnology market, and [21, 22, 23] applied deep learning in the Japanese stock market. However, these results are not universal: their experiments are performed only in a single market. Note also that, the neural nets by [21, 22, 23] **adopted epoch-based stopping**, which we show in the Experiments Section to be sensitive to the number of epochs.

3 Method

This section describes RIC-NN, a deep learning based investment strategy.

3.1 Cross-sectional Investment

We consider a medium-term investment cycle, where an investment is done on a monthly basis. Namely, let $t = 1, \dots, T$ be the time step, and each step corresponds to the end of a month between December 1994 and December 2018. We use the term “stock universe” (or simply universe) U_t to represent all the stocks of interest at time step t : In the case of the North America stock market, the number of stocks in the each U_t is about 700. Note that U_t gradually changes over the time step to reflect economic activities among different sectors. At each time step, let $i \in U_t$ be an index denoting each stock in the universe. Let $R_{i,t} \in \mathbb{R}$ be the (unit) return of the stock i between the time step $t - 1$ and t . Let $\mathbf{x}_{i,t} \in \mathbb{R}^{20}$ be the 20 factors associated with the stock i at t .

In this paper, we consider investment strategies that are widely used in the literature of finance [1, 24]. Namely, (i) the long portfolio strategy, and (ii) the long-short portfolio strategy. We consider an equally-weighted (EW) portfolio, which is simple yet sometimes outperforms more sophisticated alternatives [25]. (i) The long portfolio strategy considered here buys the **top quintile (i.e., one-fifth) of the stocks with equal weight** aiming to outperform the average return of all the stocks. Namely, let $L_t \subset U_t : |L_t| = 1/5|U_t|$ be the long portfolio. The return from the portfolio is defined as the average return of L_t .

$$R_t^L = \frac{1}{|L_t|} \sum_{i \in L_t} R_{i,t}$$

(ii) The long-short portfolio **strategy not only buys the top quintile of the stocks but also sells the bottom quintile of the stocks**. Namely, let $S_t \subset U_t : |S_t| = 1/5|U_t|$ be the short portfolio. Let $R_t^S = \frac{1}{|S_t|} \sum_{i \in S_t} R_{i,t}$. The average return in this strategy is defined as $R_t^{LS} = R_t^L - R_t^S$. While the long-short portfolio cannot take advantage of the stock market growth, **it is robust against a large market crisis** (i.e., the financial crisis during 2007-2008) because of its neutral position.

Essentially, both of the strategies above **requires a ranking over the expected return** of the stocks in the universe since we invest on the most promising stocks. Namely, let $\mathbf{o}_t \in \mathbb{N}^{|U_t|}$ be the ground-truth ranking with its element

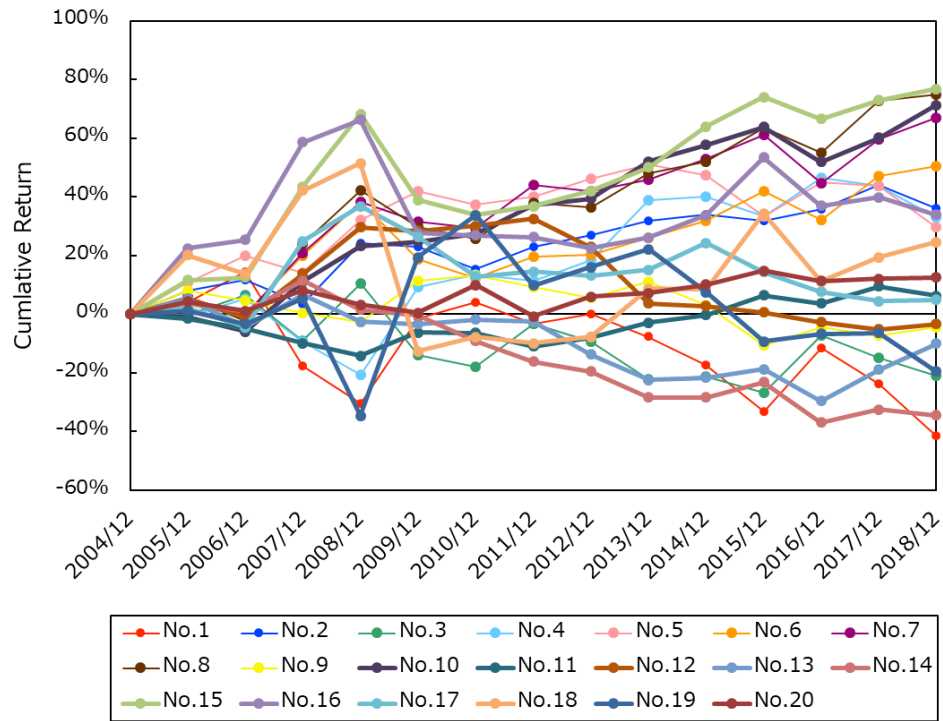


Figure 2: Cumulative portfolio returns in MSCI North America based on each single factor. Factors are listed in Appendix A.

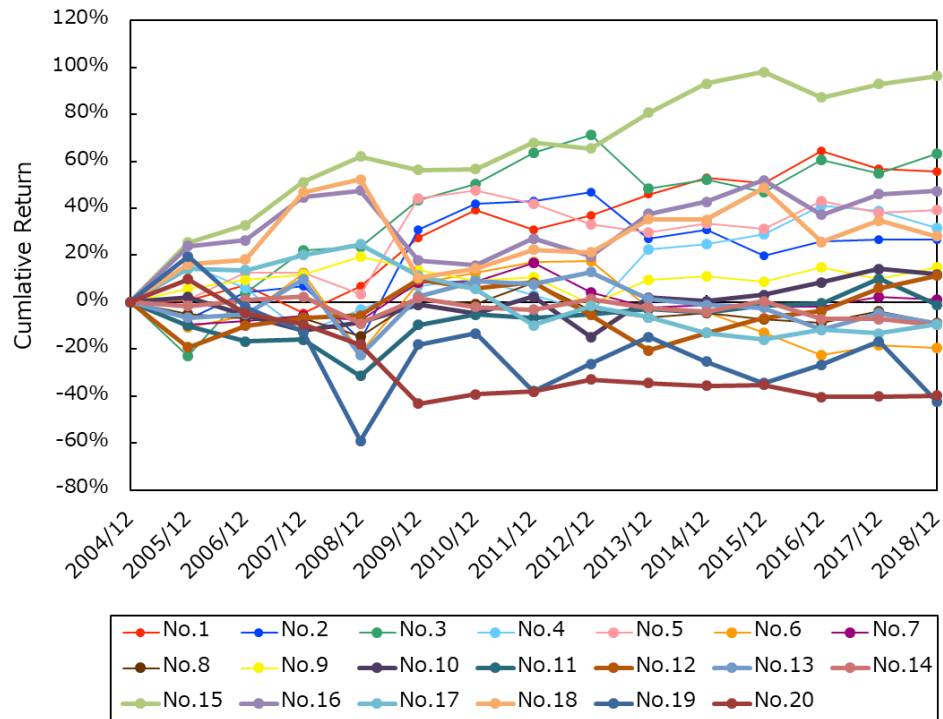


Figure 3: Cumulative portfolio returns in MSCI Pacific based on each single factor.

$o_{i,t} \in \{1, 2, \dots, |U_t|\}$ denotes the corresponding place for each $i \in U_t$. At each round t , we build the estimated ranking \hat{o}_t . We choose L_t and U_t be the top and bottom quintile on the basis of \hat{o}_t , respectively. Ranked information coefficient (rank IC), which is also referred as the Spearman's correlation coefficient, between two rankings $\mathbf{o}_t, \hat{\mathbf{o}}_t$ is defined as

$$\text{rank IC}(\mathbf{o}_t, \hat{\mathbf{o}}_t) = 1 - \frac{6 \sum_{i \in U_t} (o_{i,t} - \hat{o}_{i,t})^2}{|U_t|(|U_t|^2 - 1)},$$

which takes the value in $[-1, 1]$ and is widely used in the field of finance [15]. The larger the value of the rank IC is, the better a portfolio strategy based on the ranking is.

We consider a rolling-horizon setting: Namely, at each time step t , we estimate the ranking of the next time step \hat{o}_{t+1} . The following sections introduce RIC-NN, our DL-based method to build \hat{o}_{t+1} .

3.2 Feature Augmentation

The normalized rank of the stock i at time step t is denoted as $r_{i,t} \in \mathbb{R}$: Namely, we rank the stocks in accordance with their return $\{R_{i,t}\}$ and normalize them so that $r_{i,t} \in [0, 1]$ (i.e., $r_{i,t}$ for the stock of the largest return at each t is 1, whereas $r_{i,t}$ for the stock of the median return is 0.5).

At each time step t , we build an estimator $\hat{r}_{i,t}$ of $r_{i,t}$ by using the following augmented feature vector $\mathbf{v}_{i,t} \in \mathbb{R}^{180}$: Namely, given that many of the factors are updated in quarterly basis (i.e., each 3 time steps), we define $\mathbf{v}_{i,t} = (\mathbf{x}_{i,t}, \mathbf{x}_{i,t-3}, \dots, \mathbf{x}_{i,t-12}, \mathbf{x}_{i,t}/^R \mathbf{x}_{i,t-3}, \dots, \mathbf{x}_{i,t}/^R \mathbf{x}_{i,t-12}) \in \mathbb{R}^{180}$ using the past five time steps, where $\mathbf{x}/^R \mathbf{y}$ over two vectors \mathbf{x} and \mathbf{y} denotes an element-wise differentiation operator with its each element is defined by $2 \times (x - y) / (|x| + |y|)$, which is popularly used in finance [26].

3.3 Prediction Model: Neural Net Architecture

We adopt a seven-layer feed forward neural network with Rectified linear function (ReLU) activation function [27] to learn the relationship between $v_{i,t}$ and $r_{i,t+1}$. The hidden layer size is set to be (150 – 150 – 100 – 100 – 50 – 50), and the dropout rate for each layer is set to be (50% – 50% – 30% – 30% – 10% – 10%).

3.4 Loss Function and Optimization

We adopt the standard mean squared error (MSE) as the loss function and train our deep learning model by using the data of the latest 120 time steps from the past 10 years. Namely,

$$\text{MSE}_t = \frac{1}{K} \left\{ \sum_{t'=t-N}^{t-1} \sum_{i \in U_{t'}} (r_{i,t'+1} - f(\mathbf{v}_{i,t'}; \boldsymbol{\theta}_{t'}))^2 \right\}, \quad (1)$$

where $N = 120$ (i.e., ten years) is the size of sliding window to consider and $K = \sum_{t'=t-N}^{t-1} |U_{t'}|$ is the number of all training examples, and $f(\cdot, \boldsymbol{\theta})$ is our neural net with weight parameter $\boldsymbol{\theta}$. We adopt the Adam [28] optimizer and batch normalization [29]. The mini-batch size is set to be 300.

3.5 Initialization and Stopping Criteria

A fundamental challenge in the cross-sectional analysis lies in its dynamism: Standard machine learning methods focus on the generalization performance on the i.i.d. assumption where the training dataset and the test dataset are drawn from the same (unknown) underlying distribution. However, a straightforward application of deep neural networks leads to overfitting to the current time window, which compromises the performance as a predictor of the next time step.

To avoid overfitting, we initialize and terminate the training in the following criterion ((2) in Figure 1): Namely, we define the initialization rank IC $v_i \in [0, 1]$ and stopping rank IC $v_s \in [0, 1]$, and conducts the training as follows. Let $\boldsymbol{\theta}_{t,v}$ is the weights of the RIC-NN at time step t during the training when the average from rank IC in the training window reaches v . We used (i) $\boldsymbol{\theta}_{t-1,v_i}$ as the initial parameters to train model at time step t and (ii) adopts $\boldsymbol{\theta}_{t,v_s}$ as the final model parameter $\boldsymbol{\theta}_t$. We estimate \hat{o}_{t+1} by ranking the stocks in accordance with $f(\mathbf{v}_{i,t}; \boldsymbol{\theta}_t)$, which, combined with the long or the long-short portfolio, defines our RIC-NN.

The value of v_i, v_f are set to be 0.16, 0.20. These values are optimized by the performance in the three years from 2005 to 2008. Essentially, these values “moderately overfit” to the model: The value of the rank IC of a fairly good portfolio is around 0.1, and thus these values are large enough to exploit current data while it stops before overfitting to the dataset in the current window. The experiment section shows that these values consistently performs well in multiple markets of very different natures.

3.6 Performance Measures

In evaluating an investment strategy, we use the following measures that are widely used in the field of finance [30]. These measures evaluate not only the actual return of the portfolio but also the magnitude of the risk taken: A sample from highly fluctuated series can cause a large variance, and thus **a return normalized by a risk yields a more reliable evaluation.**

Regarding the long portfolio strategy, the annualized return is the excess return (Alpha) against the average return of all stocks in the universe, the risk (tracking error; TE) is calculated as the standard deviation of Alpha and risk/return is Alpha/TE (information ratio; IR).

$$\text{Alpha} = \prod_{t=1}^T (1 + \alpha_t)^{12/T} - 1 \quad (2)$$

$$\text{TE} = \sqrt{\frac{12}{T-1} \times (\alpha_t - \mu_\alpha)^2} \quad (3)$$

$$\text{IR} = \text{Alpha}/\text{TE} \quad (4)$$

Here, $\alpha_t = R_t^L - \frac{1}{|U_t|} \sum_{i \in U_t} R_{i,t}$, $\mu_\alpha = (1/T) \sum_{t=1}^T \alpha_t$.

Likewise, we evaluate the long-short portfolio strategy by its annualized return (AR), risk as the standard deviation of return (RISK), risk/return (R/R) as return divided by risk as for the long portfolio strategy.

$$\text{AR} = \prod_{t=1}^T (1 + R_t^{LS})^{12/T} - 1 \quad (5)$$

$$\text{RISK} = \sqrt{\frac{12}{T-1} \times (R_t^{LS} - \mu^{LS})^2} \quad (6)$$

$$\text{R/R} = \text{AR}/\text{RISK} \quad (7)$$

Here, $\mu^{LS} = (1/T) \sum_{t=1}^T R_t^{LS}$ be the average return of the long-short portfolio.

In summary, the return of the long (resp. long-short) portfolio is evaluated by Alpha (resp. AR), whereas the risk of the long (resp. long-short) portfolio is evaluated by TE (resp. RISK). We use the risk-normalized return (i.e. IR for the long and R/R for the long-short) that gives more reliable measure than the return itself.

We also evaluate maximum drawdown (MaxDD), which is yet another widely used risk measures [31, 32], for both of the long portfolio strategy and the long-short portfolio strategy: Namely, MaxDD is defined as the largest drop from an extremum:

$$\text{MaxDD} = \min_{k \in [1, T]} \left(0, \frac{W_k^{\text{Port}}}{\max_{j \in [1, k]} W_j^{\text{Port}}} - 1 \right) \quad (8)$$

$$W_k^{\text{Port}} = \prod_{i=1}^k (1 + R_i^{\text{Port}}). \quad (9)$$

where $R_i^{\text{Port}} = R_i^L$ (resp. $R_i^{\text{Port}} = R_i^{LS}$) for the long (resp. long-short) strategy. These performance measures are calculated monthly during the prediction period from January 2005 to December 2018 ($T = 168$).

4 Experiments

4.1 Dataset

We prepare a stock dataset corresponding to Morgan Stanley Capital International (MSCI) North America and MSCI Pacific Indices. These MSCI indices comprise the large and mid-cap segments of the North America (NA) and Asia Pacific (AP) markets respectively, and are widely used as a benchmark for the institutional investors investing in each stock market [33]. We use 20 popular factors listed in Appendix A. In calculating these factors, we use the following data sources: Namely, Compustat, WorldScope, Thomson Reuters, the Institutional Brokers' Estimate System (I/B/E/S), and EXSHARE. Combining these sources, we calculate the factors on a monthly basis. As for stock returns, local returns with dividends are acquired.

Table 1: Experimental Results of Long portfolio and Long-Short portfolio in MSCI North America. Bold characters indicate the best ones among each category. The evaluation measures are the ones discussed in the “Performance Measures” Section: Alpha (resp. AR) measures return, TE (resp. RISK) and MaxDD measure risk, and IR (resp. R/R) is a risk-normalized return measure in the long (resp. long-short) portfolio.

Long	Linear	Nonlinear			
	LASSO	RF	NN	RIC-NN	RIC-NN (TF from AP)
Alpha	0.62%	0.79%	0.82%	1.23%	1.20%
TE	5.40%	5.14%	4.48%	4.14%	4.43%
IR	0.11	0.13	0.18	0.30	0.27
MaxDD	-21.84%	-24.57%	-17.41%	-14.37%	-20.57%
Long-Short	Linear	Nonlinear			
	LASSO	RF	NN	RIC-NN	RIC-NN (TF from AP)
AR	2.24%	1.71%	2.10%	3.86%	2.16%
RISK	10.90%	11.51%	9.47%	7.85%	9.52%
R/R	0.21	0.15	0.22	0.49	0.23
MaxDD	-34.73%	-42.21%	-34.49%	-21.26%	-39.35%

By using these sources, we build a dataset comprised 1,194 stocks on average (NA: 702, AP: 492) and 288 time steps from December 1994 to November 2018. The following sections show the performance of the proposed RIC-NN strategy compared with several baselines. This dataset involves a reasonably long period so that we can evaluate a consistently-good investment strategy.

4.2 Comparison with Off-the-Shelf Models

We compare the performance of RIC-NN with major off-the-shelf machine learning algorithms. Namely: LASSO regression (LASSO) model [19], random forest (RF), and standard Neural Network (NN). LASSO and RF are implemented with scikit-learn [34], and NN is implemented with TensorFlow [35]. These methods are used to learn the relation between $v_{i,t}$ and $r_{i,t+1}$. Regarding the hyperparameters, regularization strength ("alpha") of LASSO is set to 0.001, which is the largest value to yield a meaningful ranking. We use the default hyperparameters of RF. Several different hyperparameters are tested, and their results are shown in Appendix B. NN adopted the same framework as our RIC-NN, except for the fact that NN stops the training at Epoch 56 in MSCI North America and 46 in MSCI Pacific¹. We used random numbers as initial weights for the first time step.

Table 1 compares the algorithms in the MSCI NA dataset. RIC-NN outperforms all of the LASSO, RF, and NN in both of the risk and the return measures, regardless of whether the portfolio strategy is the long or the long-short. A notable finding is that RF and NN have smaller returns compared with LASSO. Our hypothesis is that the highly non-stationary nature of the stocks has lead to the overfitting of these nonlinear models. Table 2 shows the results in the MSCI Pacific dataset. Although LASSO yields a larger return than RIC-NN by taking a larger risk, in terms of a risk-normalized return, which is the prominent measure of investment strategy, RIC-NN outperforms the other methods. Regarding the results of the transfer learning ("TF from AP/NA"), we discuss in a later section.

We have also conducted the same experiment with Ridge Regression (RR): The performance of RR is not very different from that of LASSO.

4.3 Stopping Criteria: Rank IC versus Epoch

Table 3 and 4 show the result of NN with different number of training epochs. While NN that stops at epoch 50 performs better in the NA market, NN that stops at epoch 60 performs better in the AP market. One can also find that the performance of NN is very sensitive to the choice of the stopping epoch. On the other hand, RIC-NN that consistently stops at $v_f = 0.20$ outperforms most of (epoch-based) NN. This implies that the rank IC is a consistent measure of the fitness of stock prediction models.

¹These epochs are chosen so that the rank IC reaches 0.20 during the training of the first time step.

Table 2: Experimental Results of Long portfolio and Long-Short portfolio in MSCI Pacific. Bold characters indicate the best ones among each category.

Long	Linear	Nonlinear			
	LASSO	RF	NN	RIC-NN	RIC-NN (TF from NA)
Alpha	5.35%	3.79%	4.34%	5.25%	5.78%
TE	5.17%	5.75%	4.18%	4.20%	3.95%
IR	1.04	0.66	1.04	1.25	1.46
MaxDD	-11.53%	-11.43%	-9.37%	-7.51%	-3.37%
Long-Short	Linear	Nonlinear			
	LASSO	RF	NN	RIC-NN	RIC-NN (TF from NA)
AR	10.27%	7.78%	8.52%	9.81%	10.95%
RISK	9.23%	9.65%	7.78%	7.83%	7.14%
R/R	1.11	0.81	1.10	1.25	1.53
MaxDD	-18.07%	-18.66%	-19.74%	-11.06%	-8.89%

Table 3: Comparison between RIC-NN and NN with different number of training epochs in MSCI North America.

Long	RIC-NN	NN (Epoch)				
		40	50	56	60	80
Alpha	1.23%	0.18%	1.48%	0.82%	1.25%	0.70%
TE	4.14%	4.52%	4.35%	4.48%	4.49%	4.14%
IR	0.30	0.04	0.34	0.18	0.28	0.17
MaxDD	-14.37%	-22.67%	-13.48%	-17.41%	-20.98%	-15.94%
Long-Short	RIC-NN	NN (Epoch)				
		40	50	56	60	80
AR	3.86%	0.67%	3.24%	2.10%	2.02%	3.10%
RISK	7.85%	9.08%	10.06%	9.47%	9.05%	7.73%
R/R	0.49	0.07	0.32	0.22	0.22	0.40
MaxDD	-21.26%	-40.09%	-26.20%	-34.49%	-31.62%	-23.47%

Table 4: Comparison between RIC-NN and NN with different number of training epochs in MSCI Pacific.

Long	RIC-NN	NN (Epoch)				
		40	46	50	60	80
Alpha	5.25%	4.13%	4.34%	4.28%	4.52%	2.99%
TE	4.20%	4.36%	4.18%	4.73%	4.34%	4.06%
IR	1.25	0.95	1.04	0.90	1.04	0.74
MaxDD	-7.51%	-8.08%	-9.37%	-7.16%	-7.45%	-7.52%
Long-Short	RIC-NN	NN (Epoch)				
		40	46	50	60	80
AR	9.81%	8.89%	8.52%	8.97%	9.78%	6.15%
RISK	7.83%	7.63%	7.78%	8.05%	7.73%	7.18%
R/R	1.25	1.16	1.10	1.11	1.26	0.86
MaxDD	-11.06%	-13.07%	-19.74%	-12.17%	-14.70%	-13.44%

4.4 Comparison between NA and AP markets

The value of Alpha (Eq. (2)) indicates the advantage of the long strategy over the average return in the universe, which enables us to infer the possible advantage we can obtain by using machine learning algorithms.

Comparing the Alpha in Table 1 and 2, machine learning algorithms has a smaller advantage in the NA market than they do in the AP market:

A portfolio strategy of a higher return essentially exploits the gap between the market value of the stocks and the true valuation of the companies: The more efficient a market is, the more difficult obtaining a higher return is. In other words, the result implies the efficiency of the NA market compared with the AP market.

4.5 Transfer Learning

To exploit the interdependency between the markets, we further apply transfer learning to our RIC-NN. Namely, we use the weights of the first four layers that are trained in the source region as the initial weight of the target region.

Table 1 shows that the transfer from NA to AP is not very successful, whereas 2 shows the transfer from AP to NA is quite successful. In other words, NA as a source domain is quite informative to enhance the performance of AP, not vice versa. Those results are consistent with the market movements propagate from the NA stock market to the AP stock market [10, 11]. The experiment here shows the capability of RIC-NN to exploit highly non-trivial causal structure among multiple markets by using deep neural networks.

4.6 Comparison with Actual Investment Funds

This section compares the performance of RIC-NN with major funds where the investments involve decision-making by human experts. We select the top 5 funds in terms of the total assets (US dollar) excluding index funds as following criteria and calculate average total return series of these funds, including the trust fees: Namely, we select these funds by querying Bloomberg fund screening search with the following condition:

- Fund Asset Class Focus: Equity
- Fund Geographical Focus: North America Region (resp. Asian Pacific Region)
- Fund Type: Open-End-Funds
- Currency: US dollar
- Market Cap Focus (Holdings Based): Large-cap, Mid-cap
- Inception Date: before 12/31/2004

In both of the NA and AP regions, the correlation coefficient between the performance of the averaged funds above and the MSCI index is larger than 0.9, which implies these funds are based on the long strategy. For comparison, We add the benchmark returns calculated by average return of MSCI North America (resp. MSCI Pacific) constituent to long portfolio strategy performance and convert to US dollars.

Table 5 shows the performance of RIC-NN and the aforementioned stock investing funds from January 2005 to June 2018. The corresponding time-series data is shown in the supplementary material (see Appendix C). Unlike the performance of the machine learning models, the performance of the funds involves the transaction cost: As a conservative baseline, Table 5 shows the performance of RIC-NN where the transaction cost for updating the entire portfolio is deducted (i.e., an overestimated transaction cost²): RIC-NN still outperforms the average performance of the funds.

5 Conclusion

In this paper, we have proposed a new stock price prediction framework called RIC-NN by introducing three novel ideas: (1) a nonlinear multi-factor approach, (2) a stopping criteria based on rank IC and (3) deep transfer learning.

RIC-NN is conceptually simple yet universal: The identical NN architecture and RankIC stopping value yielded a consistently good return for a long timescale and the two different markets of very different structures. Experimental

²We have deducted the cost of rebalancing all the stocks in the portfolio every month. Estimated transaction cost is 0.05% one way in North America and 0.1% one way in Asia Pacific.

Table 5: The upper panel: Performance of RIC-NN in MSCI North America and averaged performance of five investment funds in the NA stock market. The lower panel: Performance of RIC-NN (TF from NA) in MSCI Pacific and averaged performance of five investment funds in the AP stock market.

North America	RIC-NN	RIC-NN (After cost deduction)	Funds
AR	9.09%	7.79%	5.90%
RISK	17.78%	17.78%	14.91%
R/R	0.51	0.44	0.40
Asia Pacific	RIC-NN (TF)	RIC-NN (TF) (After cost deduction)	Funds
AR	12.08%	9.44%	7.88%
RISK	17.23%	17.23%	17.58%
R/R	0.70	0.55	0.45

comparison showed that RIC-NN outperforms off-the-shell machine learning methods and average performance of investment funds in the last decades.

Directions of promising future work includes the followings.

More sophisticated portfolio strategies: In this study, we use a simple equally-weighted (EW) portfolio that maximizing the predictive power of stock returns. On the other hand, the portfolio theory [36] states that explicit consideration of the risk in portfolio selection is important. Regarding this direction, combining our method with more sophisticated portfolio strategies, such as Subset Resampling Portfolio [32] or Ensemble Growth Optimal Portfolio [37] will be an interesting direction for the future work.

Stateful models: This paper considered a rolling-horizon learning of a neural network, whereas there are several other approaches for portfolio selection. In particular, the recurrent neural networks and its variants are stateful neural networks that can capture the time evolution of the stock universe. Note that our RIC-NN model uses quite a large time window (i.e., ten years) for the training, which implies that the long-range interaction is important in the multi-factor machine learning models. While we presume that a straightforward application of recurrent neural network overfits to the data up to the current time horizon, several attempts to capture long-range interactions, such as memory networks and attention mechanisms, can be applied to predict cross-sectional investments.

References

- [1] Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *J. of Finance*, 47(2):427–465, 1992.
- [2] Keywan Christian Rasekhschaffe and Robert C. Jones. Machine learning for stock selection. *Financial Analysts Journal*, 75(3):70–88, 2019.
- [3] George S Atsalakis and Kimon P Valavanis. Surveying stock market forecasting techniques–part ii: Soft computing methods. *Expert Systems with Applications*, 36(3):5932–5941, 2009.
- [4] Rodolfo C Cavalcante, Rodrigo C Brasileiro, Victor LF Souza, Jarley P Nobrega, and Adriano LI Oliveira. Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, 55:194–211, 2016.
- [5] Asriel E Levin. Stock selection via nonlinear multi-factor models. In *NIPS*, pages 966–972, 1996.
- [6] Alan Fan and Marimuthu Palaniswami. Stock selection using support vector machines. In *IJCNN*, volume 3, pages 1793–1798. IEEE, 2001.
- [7] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In *NIPS*, pages 3104–3112, 2014.
- [8] Dan C. Ciresan, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber. Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32:333–338, 2012.
- [9] Eunsuk Chong, Chulwoo Han, and Frank C Park. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83:187–205, 2017.
- [10] Yin-Wong Cheung and Lilian K Ng. A causality-in-variance test and its application to financial market prices. *Journal of econometrics*, 72(1-2):33–48, 1996.

- [11] Aymen Ben Rejeb and Mongi Arfaoui. Financial market interdependencies: A quantile regression analysis of volatility spillover. *Research in International Business and Finance*, 36:140–157, 2016.
- [12] Avanidhar Subrahmanyam. The cross-section of expected stock returns: What have we learnt from the past twenty-five years of research? *European Financial Management*, 16(1):27–42, 2010.
- [13] Francis EH Tay and Lijuan Cao. Application of support vector machines in financial time series forecasting. *omega*, 29(4):309–317, 2001.
- [14] James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, NJ, 1994.
- [15] Richard C Grinold and Ronald N Kahn. Active portfolio management. 2000.
- [16] Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *J. of Financial Economics*, 33(1):3–56, 1993.
- [17] Campbell R Harvey, Yan Liu, and Heqing Zhu. . . . and the cross-section of expected returns. *Rev. of Financial Studies*, 29(1):5–68, 2016.
- [18] Alex Chinco, Adam D Clark-Joseph, and Mao Ye. Sparse signals in the cross-section of returns. *J. of Finance*, 74(1):449–492, 2019.
- [19] Robert Tibshirani. Regression shrinkage and selection via the lasso. *JRSS: Series B*, 58(1):267–288, 1996.
- [20] J. B. Heaton, Nicholas G. Polson, and J. H. Witte. Deep portfolio theory. *CoRR*, abs/1605.07230, 2016.
- [21] Masaya Abe and Hideki Nakayama. Deep learning for forecasting stock returns in the cross-section. In *PAKDD*, pages 273–284. Springer, 2018.
- [22] Kei Nakagawa, Takumi Uchida, and Tomohisa Aoshima. Deep factor model. In *ECML PKDD 2018 Workshops*, pages 37–50. Springer, 2018.
- [23] Seisuke Sugitomo and Shotaro Minami. Fundamental factor models using machine learning. *J. of Mathematical Finance*, 8:111–118, 2018.
- [24] R David McLean and Jeffrey Pontiff. Does academic research destroy stock return predictability? *J. of Finance*, 71(1):5–32, 2016.
- [25] Victor Demiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Rev. of Financial Studies*, 22, 05 2009.
- [26] Barr Rosenberg and Walt McKibben. The prediction of systematic and specific risk in common stocks. *J. of Financial and Quantitative Analysis*, 8(2):317–333, 1973.
- [27] Richard H. R. Hahnloser, Rahul Sarpeshkar, Misha A. Mahowald, Rodney J. Douglas, and H. Sebastian Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405:947–951, 2000.
- [28] Diederik P Kingma and Jimmy Ba. Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [29] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [30] Michael W. Brandt. Portfolio choice problems. *Handbook of Financial Econometrics, Vol 1*, 1, 12 2010.
- [31] Malik Magdon-Ismail and Amir F Atiya. Maximum drawdown. *Risk Magazine*, 17(10):99–102, 2014.
- [32] Weiwei Shen and Jun Wang. Portfolio selection via subset resampling. In *AAAI*, pages 1517–1523, 2017.
- [33] Hung-Ling Chen, Cheng-Yi Shiu, Hui-Shan Wei, et al. Price effect and investor awareness: Evidence from msci standard index reconstitutions. *J. of Empirical Finance*, 50(C):93–112, 2019.
- [34] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *JMLR*, 12(Oct):2825–2830, 2011.
- [35] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: a system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.
- [36] Harry Markowitz. Portfolio selection. *J. of finance*, 7(1):77–91, 1952.
- [37] Weiwei Shen, Bin Wang, Jian Pu, and Jun Wang. The kelly growth optimal portfolio with ensemble learning. In *AAAI*, pages 1134–1141, 2019.

A Appendix A: List of Factors Used in This Paper

Table 6 shows the list of 20 factors used in this paper. The financial data is acquired from Compustat, WorldScope and Reuters Fundamentals (ordered by the priority). Note that the Compustat data source, which is mainly used to build factors for North America, involves a delay of maximum three months, and the other sources, which are mainly used for Asia Pacific, involve a delay of four months. These data sources are used to calculate the factors from No. 1 to No. 14. The Earnings per share (EPS) revisions, which indicate the future value of a company, are obtained from Thomson Reuters Estimates and I/B/E/S Estimates (ordered by the priority).

We can classify these factors into three types; technical, fundamental and both. Technical factors are calculated from historical stock prices, whereas fundamental factors are calculated from the qualitative and quantitative information of a company [4].

Table 6: List of the factors. F (resp. T) in the “Type” column indicates that the corresponding factor is derived from its fundamental (resp. technical) property of the stock, and B in the column indicates the factor derived from both of the fundamental and technical properties.

No	Factor	Description	Type
1	Book-value to Price Ratio	Net Asset/Market Value	B
2	Earnings to Price Ratio	Net Profit/Market Value	B
3	Dividend Yield	Dividend/Market Value	B
4	Sales to Price Ratio	Sales/Market Value	B
5	Cash Flow to Price Ratio	Operating cash flow/Market Value	B
6	Return on Equity	Net Profit/Net Asset	F
7	Return on Asset	Net Operating Profit/Total Asset	F
8	Return on Invested Capital	Net Operating Profit After Taxes/(Liabilities with interest + Net Asset)	F
9	Accruals	-(Changes in Current Assets and Liability-Depreciation)/Total Asset	F
10	Total Asset Growth Rate	Change Rate of Total Assets from the previous period	F
11	Current Ratio	Current Asset/Current Liability	F
12	Equity Ratio	Net Asset/Total Asset	F
13	Total Asset Turnover Rate	Sales/Total Asset	F
14	Capital Expenditure Growth Rate	Change Rate of Capital Expenditure from the previous period	F
15	EPS Revision (1 month)	1 month Earnings Per Share (EPS) Revision	B
16	EPS Revision (3 month)	3 month Earnings Per Share (EPS) Revision	B
17	Momentum (1 month)	Stock Returns in the last month	T
18	Momentum (12-1 month)	Stock Returns in the past 12 months except for last month	T
19	Volatility	Standard Deviation of Stock Returns in the past 60 months	T
20	Skewness	Skewness of Stock Returns in the past 60 months	T

B Appendix B: Additional Experiments

B.1 Different Hyperparameters in Off-the-shelf Models

Tables 9 and 10 show the results of RF with different depths, whereas Tables 7 and 8 show the results of LASSO with different magnitudes of the regularizer. Overall, comparison with Tables 1 and 2 show that (i) RF falls below the competitors, and (ii) LASSO, which outperforms RF, still fall below RIC-NN. We have also confirmed that a LASSO with a regularizer stronger than 0.001 suppresses most of the features, which yields meaningless results.

Table 7: Results of LASSO with different magnitudes of regularizer in MSCI North America.

Long	Regularizer		
	0.001	0.0001	0.00001
Alpha	0.62%	0.53%	1.14%
TE	5.40%	4.60%	4.21%
IR	0.11	0.11	0.27
MaxDD	-21.84%	-17.85%	-16.02%
Long-Short	Regularizer		
	0.001	0.0001	0.00001
AR	2.24%	0.76%	2.00%
RISK	10.90%	9.56%	8.81%
R/R	0.21	0.08	0.23
MaxDD	-34.73%	-35.10%	-34.52%

Table 8: Results of LASSO with different magnitudes of regularizer in MSCI Pacific.

Long	Regularizer		
	0.001	0.0001	0.00001
Alpha	5.35%	5.23%	4.90%
TE	5.17%	4.58%	4.20%
IR	1.04	1.14	1.17
MaxDD	-11.53%	-8.43%	-6.81%
Long-Short	Regularizer		
	0.001	0.0001	0.00001
AR	10.27%	10.42%	8.99%
RISK	9.23%	8.35%	8.08%
R/R	1.11	1.25	1.11
MaxDD	-18.07%	-13.30%	-13.34%

Table 9: Results of RF with different depths in MSCI North America.

Long	Depth		
	3	5	7
Alpha	0.77%	0.85%	0.92%
TE	4.80%	4.61%	4.33%
IR	0.16	0.18	0.21
MaxDD	-21.76%	-23.21%	-21.16%
Long-Short	Depth		
	3	5	7
AR	1.62%	2.44%	2.38%
RISK	11.10%	10.47%	9.86%
R/R	0.15	0.23	0.24
MaxDD	-40.78%	-34.90%	-34.10%

Table 10: Results of RF with different depths in MSCI Pacific.

Long	Depth		
	3	5	7
Alpha	2.48%	3.53%	4.24%
TE	5.28%	4.96%	4.94%
IR	0.47	0.71	0.86
MaxDD	-13.04%	-9.21%	-7.36%
Long-Short	Depth		
	3	5	7
AR	6.48%	7.83%	8.74%
RISK	8.63%	8.61%	8.87%
R/R	0.75	0.91	0.99
MaxDD	-20.09%	-15.13%	-13.56%

C Appendix C: Cumulative Return of RIC-NN and Investment Funds

Figure 4 and 5 show the corresponding return time series in the NA and AP regions, respectively. Regarding the evaluated risk and returns, see Section “Comparison with Actual Investment Funds” in the main paper.

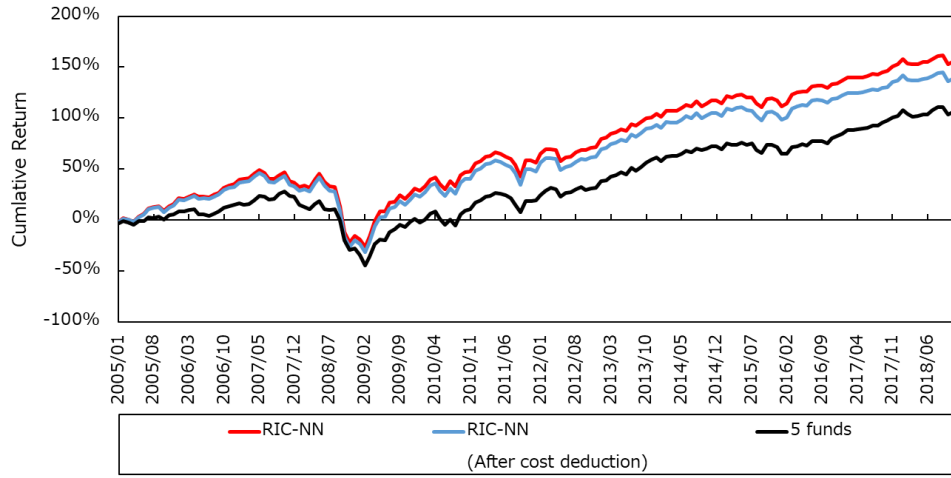


Figure 4: Averaged cumulative returns among the five funds and the cumulative return of RIC-NN in MSCI North America. RIC-NN consistently outperforms the average of the funds. Both of the averaged funds and RIC-NN suffer a large drawdown during the financial crisis in 2007-2008.

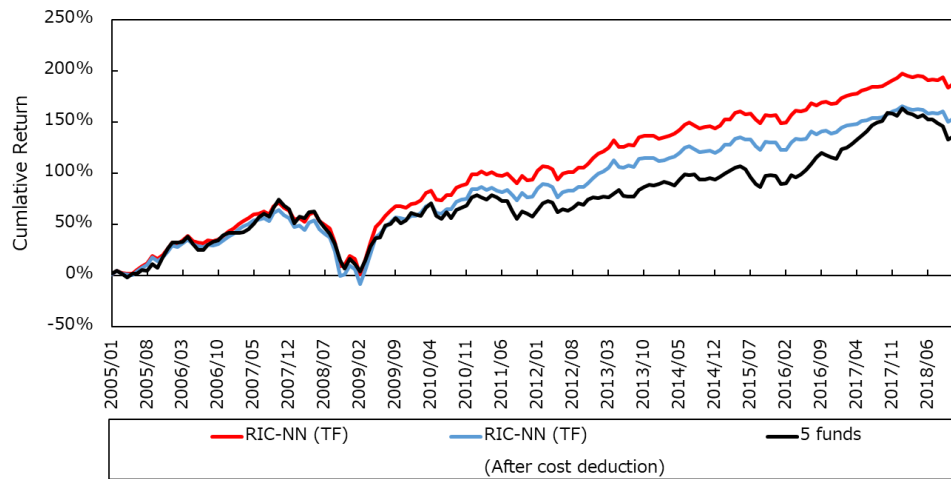


Figure 5: Averaged cumulative returns among the five funds and the cumulative return of RIC-NN in MSCI Pacific.