A Robust Transferable Deep Learning Framework for Cross-sectional Investment Strategy

Kei Nakagawa[1], **Masaya Abe**[1], **and Junpei Komiyama**[2]**.**
[1]Nomura Asset Management Co., Ltd.
[2]New York University.
k-nakagawa, m-abe@nomura-am.co.jp, junpei.komiyama@stern.nyu.edu.

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" (Fama and French 1992) have been proposed to summarize the essence of predictive stock returns. Although machine learning methods are increasingly popular in stock return prediction (Rasekhschaffe and Jones 2019), 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.

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" (Fama and French 1992), are proposed to summarize the essence of predictive stock returns, and a large portion of investors develop their portfolio strategies based on these 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 (Atsalakis and Valavanis 2009; Cavalcante et al. 2016; Rasekhschaffe and Jones 2019). 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. 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.

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, (Chong, Han, and Park 2017) 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. Moreover, we further propose a (3) deep multi-task learning framework that enables inter-regional learning. 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. (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 align with the idea of the asymmetric structure between the two markets (Rejeb and Arfaoui 2016).

Related Work

There are two major strategies in stock trading: Namely, the one based on time-series analysis (Cavalcante et al. 2016) and the one based on cross-sectional analysis (Subrahmanyam 2010).

The methods of the latter strategy, which include the work in this paper, perform a regression analysis using cross-

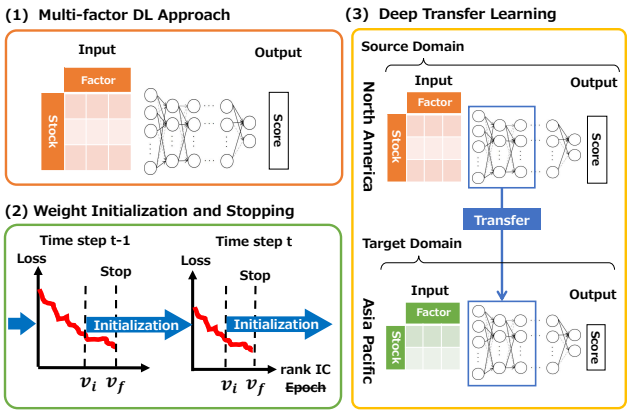

Figure 1: Our approach: RIC-NN.

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. The Fama-French (Fama and French 1992) 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: See (Harvey, Liu, and Zhu 2016) for the history of the proposed factors.

Machine learning approaches, which can capture the nonlinear relationship among multiple factors, are recently applied to a cross-sectional analysis. (Chinco, Clark-Joseph, and Ye 2019) applied the LASSO (Tibshirani 1996) in the U.S. stock market, (Heaton, Polson, and Witte 2016) applied an auto-encoder based nonlinear model into a U.S. biotechnology market, and (Abe and Nakayama 2018; Sugitomo and Minami 2018; Nakagawa, Uchida, and Aoshima 2018) 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 (Abe and Nakayama 2018; Sugitomo and Minami 2018; Nakagawa, Uchida, and Aoshima 2018) adopted epoch-based stopping, which is sensitive to the number of epochs.

Method

Cross-sectional Investment

We consider a medium-term investment cycle, where an investment is done on a monthly basis. Namely, let $t = 1, \ldots, T$ 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.

We consider an equally-weighted (EW) long portfolio, which is simple yet sometimes outperforms more sophisticated alternatives (Demiguel, Garlappi, and Uppal 2009). 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: $R^L_t = \frac{1}{|L_t|}\sum_{i \in L_t} R_{i,t}$. We build this portfolio by ranking the stocks in terms of their expected return. Namely, let $\mathbf{o}_t \in \mathbb{N}^{|U_t|}$ be the ground-truth ranking with its element $o_{i,t} \in \{1, 2, \ldots, |U_t|\}$ denotes the corresponding place for each $i \in U_t$. At each round t, we build the estimated ranking $\hat{\mathbf{o}}_t$ and choose L_t on the basis of $\hat{\mathbf{o}}_t$. 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 $[-1, 1]$. The larger the value of the rank IC is, the better a portfolio strategy is.

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

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 $\mathbf{v}_{i,t} \in \mathbb{R}^{180}$: Given that many of the factors are updated in quarterly basis (i.e., each 3 time steps), we define $\mathbf{v}_{i,t} = (\mathbf{x}_{i,t}, \mathbf{x}_{i,t-3}, ..., \mathbf{x}_{i,t-12}, \mathbf{x}_{i,t}/^R\mathbf{x}_{i,t-3}, ..., \mathbf{x}_{i,t}/^R\mathbf{x}_{i,t-12}) \in \mathbb{R}^{180}$ 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 (Rosenberg and McKibben 1973) with its each element is defined by $2 \times (x - y)/(|x| + |y|)$.

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,

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

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. For the class of functions $f(\cdot, \boldsymbol{\theta})$, we adopt a seven-layer feed forward neural network with Rectified linear function (ReLU) activation function and denote its weight parameter as $\boldsymbol{\theta}$. Details on the hypeparameters are in our preprint (Nakagawa, Abe, and Komiyama 2019).

Initialization and Stopping Criteria

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 $\hat{\mathbf{o}}_{t+1}$ 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.

Performance Measures

We use the following measures that are widely used in the field of finance (W. Brandt 2010). These measures evaluate not only the actual return of the portfolio but also the magnitude of the risk taken.

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-normalized return is measured by information ratio (IR).

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

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

$$\mathbf{IR} = \mathbf{Alpha}/\mathbf{TE}, \tag{4}$$

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

Experiments

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 (Chen et al. 2019). We use 20 popular factors. Details of the data aggregation process and machine-learning models are in our preprint (Nakagawa, Abe, and Komiyama 2019).

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 (Tibshirani 1996), random forest (RF), and standard Neural Network (NN). LASSO and RF are implemented with scikit-learn, and NN is implemented with TensorFlow. These methods are used to learn the relation between $v_{i,t}$ and $r_{i,t+1}$. 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 Pacific[1].

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.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.

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. In other words, the result implies the efficiency of the NA market compared with the AP market.

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 AP to NA is not very successful, whereas 2 shows the transfer from NA to AP 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 (Rejeb and Arfaoui 2016). The experiment here demonstrates the capability of RIC-NN to exploit highly non-trivial causal structure among multiple markets.

Comparison with Actual Investment Funds

Furthermore, we also compared the performance of RIC-NN with major funds where the investments involve decision-making by human experts and verified the return of RIC-NN outperforms the averaged return of these funds. We select the top 5 funds in terms of the total assets (US dollar) excluding index funds. 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. Due to the space limitation, we omit the results of the experiments: See our preprint (Nakagawa, Abe, and Komiyama 2019) for the details. In other words, RIC-NN outperforms major funds that define the expected return of the markets.

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

Table 1: Experimental Results of Long portfolio in MSCI North America. Bold characters indicate the best ones among each category. Alpha measures return, TE measures risk, and IR measures a risk-normalized return.

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

Table 2: Experimental Results of Long portfolio in MSCI Pacific.

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**

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 comparison showed that RIC-NN outperforms off-the-shell machine learning methods.

References

Abe, M., and Nakayama, H. 2018. Deep learning for forecasting stock returns in the cross-section. In *PAKDD*, 273–284. Springer.

Atsalakis, G. S., and Valavanis, K. P. 2009. Surveying stock market forecasting techniques–part ii: Soft computing methods. *Expert Systems with Applications* 36(3):5932–5941.

Cavalcante, R. C.; Brasileiro, R. C.; Souza, V. L.; Nobrega, J. P.; and Oliveira, A. L. 2016. Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications* 55:194–211.

Chen, H.-L.; Shiu, C.-Y.; Wei, H.-S.; et al. 2019. Price effect and investor awareness: Evidence from msci standard index reconstitutions. *J. of Empirical Finance* 50(C):93–112.

Chinco, A.; Clark-Joseph, A. D.; and Ye, M. 2019. Sparse signals in the cross-section of returns. *J. of Finance* 74(1):449–492.

Chong, E.; Han, C.; and Park, F. C. 2017. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications* 83:187–205.

Demiguel, V.; Garlappi, L.; and Uppal, R. 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Rev. of Financial Studies* 22.

Fama, E. F., and French, K. R. 1992. The cross-section of expected stock returns. *J. of Finance* 47(2):427–465.

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

Heaton, J. B.; Polson, N. G.; and Witte, J. H. 2016. Deep portfolio theory. *CoRR* abs/1605.07230.

Nakagawa, K.; Abe, M.; and Komiyama, J. 2019. A robust transferable deep learning framework for cross-sectional investment strategy. *CoRR* abs/1910.01491.

Nakagawa, K.; Uchida, T.; and Aoshima, T. 2018. Deep factor model. In *ECML PKDD 2018 Workshops*, 37–50. Springer.

Rasekhschaffe, K. C., and Jones, R. C. 2019. Machine learning for stock selection. *Financial Analysts Journal* 75(3):70–88.

Rejeb, A. B., and Arfaoui, M. 2016. Financial market interdependencies: A quantile regression analysis of volatility spillover. *Research in International Business and Finance* 36:140–157.

Rosenberg, B., and McKibben, W. 1973. The prediction of systematic and specific risk in common stocks. *J. of Financial and Quantitative Analysis* 8(2):317–333.

Subrahmanyam, A. 2010. 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.

Sugitomo, S., and Minami, S. 2018. Fundamental factor models using machine learning. *J. of Mathematical Finance* 8:111–118.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *JRSS: Series B* 58(1):267–288.

W. Brandt, M. 2010. Portfolio choice problems. *Handbook of Financial Econometrics, Vol 1* 1.