Factor Integration Based on Neural Networks for Factor Investing

Zhichen Lu[1,2], Wen Long[1,2(✉)], Jiashuai Zhang[2,3], and Yingjie Tian[1,2,3]

[1] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, People's Republic of China
longwen@ucas.ac.cn
[2] Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
[3] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100190, People's Republic of China

Abstract. Factor investing is one kind of quantitative investing methodologies for portfolio construction based on factors. Factors with different style are extracted from multiple sources such as market data, fundamental information from financial statements, sentimental information from the Internet, etc. Numerous style factors are defined by Barra model proposed by Morgan Stanley Capital International(MSCI) to explain the return of a portfolio. Multiple factors are usually integrated linearly when being put to use, which ensures the stability of the process of integration and enhances the effectiveness of integrated factors. In this work, we integrate factors by machine learning and deep learning methodologies to explore deeper information among multiple style factors defined by MSCI Barra model. Multi-factors indexes are compiled using Smart Beta Index methodology proposed by MSCI. The results show non-linear integration by deep neural network can enhance the profitability and stability of the index compiled according to the integrated factor.

Keywords: Neural networks · Deep learning · Factor investing

1 Introduction

The definition of factors of factor investing originates from "Arbitrage pricing theory" proposed by Ross [10], which holds that the expected return of a financial asset can be modeled as a function of various macroeconomic factors or theoretical market indexes. And then researchers have tried to use specific factors to model the return of stocks. Three-factors model [4] was the primary one which modeled excess return of stock by book value, earning. Further researches verified a series of factors can be used to explain the return of investing in stocks, factors can be summarised into three main categories: macroeconomic,

J. M. F. Rodrigues et al. (Eds.): ICCS 2019, LNCS 11538, pp. 286–292, 2019.
https://doi.org/10.1007/978-3-030-22744-9_22

statistical, and fundamental. In risk model developed by Barra team from MSCI company, factor returns are estimated through cross-sectional regression [8]. Factor portfolios were built according to target factors to construct factor returns in Fama-French approach [1,4]. Similarly, Smart Beta Index from MSCI company [2,3] is compiled according to target factors to reflect the style and performance of specific factors under the different market situation. When being put to use, multiple factors usually need to be integrated, a common way to integrate factors is a linearly weighted sum, and weights of each factor are calculated by solving an optimization with subjectively defined target [3]. In recent years, non-linear methods such Support Vector Machine, Logistic Regression, Random Forest, Neural Networks and deep learning methodologies are well used in financial time series modeling, yet most existing works focus on stock price prediction. They learn parameters of models by fitting training samples and presume that the distribution of the training set and test set in the feature space are identical [9,13–15]. In the aspect of cross-section modeling and feature integration, only several works exist [5,6].

In our works, we introduce neural networks into the task of cross-section factor integration, and we extract factors according to the definition from Barra [8]. We use Smart Beta Index methodology to compile factor indexes to reflect performance and style of them on the Chinese market. Experimental results show the index that compiled based on factors integrated by neural networks results in better profitability and stability.

2 Factors and Factor Indexes

The changes of the stock price are not just a result of historical market behavior, but also affected by information from multiple sources such as macroeconomy and financial situation of the corresponding listed company. Indicators can be selected and defined to capture this information for usage on investment practice, and they are called factors. Factors are extracted from three main sources: technical indicators from market samples, fundamental indicators from financial statements and macroeconomic indicators.

When used in market practices, stocks are ranked and selected according to scores calculated by one or multiple factors. Factors that proven to be robust through a long time period are summarized by Barra risk model. Table 1 present the definition of factors. Original indicators are extracted from market data of stocks and financial statement of their corresponding listed companies. Factors are usually sampled in monthly frequency when being used.

To reflect performances of factors on market practices, factor indexes are compiled according to methodologies proposed by MSCI company. At beginning of each season component stocks of benchmark CSI 800 are sorted by factor score, and top 100 are selected as component of factor index and weighted according to their market value. For single factor indexes, component stocks are sorted by single target factor, for multi-factors indexes, weights of component stocks are calculated by solving optimization whose objective are maximizing multiple target factors:

Table 1. BARRA style factors

Factors	Meaning	Indicators
Size	Size of listed company	Market Value of listed company
Momentum	Degrees of trend	Risk adjusted returns of recent 20 days: $\frac{mean(r_{20})}{std(r_{20})}$
Non-linear size	Middle level of size	Residual of the regression between size and third power of size
Volatility	Uncertainty of bias from market	Standard deviation of active return
		The cumulative sum of the active return
		Standard deviation of daily return
Value (BTOP)	Book value to market value	Price earnings ratio (PE)
		Market-to-book ratio (PBR)
		Price-to-sale ratio (PS)
Liquidity	Volume and frequency of trading	Monthly logarithm turnover rate
		Mean value of monthly logarithm turnover rate in recent 3 month
		Mean value of monthly logarithm turnover rate in recent 12 month
Growth	Growth of listed company	Net profit (YoY)
		Total asset (YoY)
		Operating revenue (YoY)
Dividend (Earning Yield)	Profitability of listed company	Dividend yield
		Dividend per share
		Dividend to market value
Quality	Quality of listed company	Debt to equity
		ROE
Leverage	Leverage situation of listed company	Market leverage
		Debt to asset
		Book leverage

$$
\begin{aligned}
\max \quad & \sum_{k=1}^{K}\sum_{i=1}^{n}\omega_i X_{ik}^{target} \\
s.t. \quad & \sum_{i=1}^{n}\omega_i X_{ik}^{non-target} \geq \sum_{i=1}^{n}\omega_i^{benchmark} X_{ik}^{non-target} - 0.25 * std(X_k^{non-target}), \\
& \qquad\qquad k = 1,2,3\ldots,\tilde{K} \\
& \sum_{i=1}^{n}\omega_i X_{ik}^{non-target} \leq \sum_{i=1}^{n}\omega_i^{benchmark} X_{ik}^{non-target} + 0.25 * std(X_k^{non-target}),. \\
& \qquad\qquad k = 1,2,3\ldots,\tilde{K} \\
& max(0, \omega_i^{benchmark} - 2\%) \leq \omega_i \leq \max(10\omega_i^{benchmark}, \omega_i^{benchmark} + 2\%), \\
& \qquad\qquad i = 1,2,3\ldots,n
\end{aligned}
$$

According to this methodology we compile single factor indexes and multi-factor indexes with target on Momentum, Size, Value, Dividend, which follows document from MSCI. Figure 1 is back-test results of factor indexes during 2010 to 2017. Factors present different style among different market situation. Profitability and risk of each factors are evaluated by indicators listed in Table 2, from which we can see that factor indexes reach higher returns and Sharpe ratio than benchmark, which verified the effectiveness of these factors on Chinese market. Moreover, subjectively setting the objective of optimization for factor integration may lead to unsatisfied result on profitability and risk, since factors show different performance in different market.

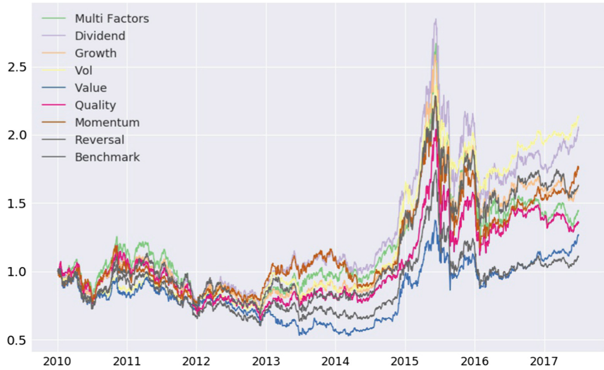

Fig. 1. Smart Beta factor indexes based on CSI 800.

Table 2. Smart Beta Index simulation results based on CSI 800

	Return	Annual Return	Volatility	Downside Beta	VaR	Alpha	Beta	Sharpe	Sortino	Loss rate	MDD	Active Return
Dividend	105.755%	10.425%	26.987%	1.0207	-2.739%	0.0860	1.0466	0.3612	0.2107	43.820%	-48.622%	94.791%
Growth	63.471%	6.988%	27.255%	1.0594	-2.779%	0.0548	1.0627	0.2437	0.0977	43.820%	-47.873%	52.507%
Vol	113.650%	10.998%	20.561%	0.7441	-2.082%	0.0838	0.7400	0.4244	0.3072	43.820%	-33.409%	102.686%
Value	26.687%	3.304%	23.491%	0.8553	-2.410%	0.0146	0.8512	0.0921	-0.0242	48.315%	-37.260%	15.722%
Quality	35.915%	4.308%	26.189%	0.9965	-2.678%	0.0275	0.9840	0.1454	0.0127	38.202%	-45.286%	24.950%
Momentum	76.557%	8.126%	26.901%	1.0443	-2.737%	0.0644	1.0230	0.2831	0.1357	46.067%	-47.984%	65.592%
Reversal	62.528%	6.903%	26.239%	1.0080	-2.676%	0.0524	0.9953	0.2397	0.1005	46.067%	-43.197%	51.564%
Multi Factors	44.600%	5.199%	28.772%	1.1226	-2.934%	0.0387	1.0962	0.1869	0.0375	41.573%	-54.206%	33.636%
Benchmark	10.964%	1.440%	24.670%	1.0011	-2.533%	0.0000	1.0000	0.0249	-0.0901	47.191%	-48.984%	0.000%

3 Neural Networks for Factor Integration

Deep learning methodology is explored on stock price prediction [7,11,12], and deep neural networks are designed to extract features from time series samples for prediction. Portfolio construction is another kind of market practice which provides cross-section level samples. In this work, we introduce Multi-layer Perceptron (MLP) to deal with cross-section factors. Traditional machine learning and linear regression are also applied in the experiment for comparison.

We use factors of each component stock of CSI 800 index from 2008 to 2017 for the experiment. Models are trained at the start of every year using monthly samples $\{\chi_t^i, y_t^i\}$ from previous 3 years, where χ_t^i denotes factors listed in Table 1 of stock i, and y_t^i denotes return of from t to $t+1$. At the start of each month, factors of each stock are integrated by models trained at the start of that year, and stocks are sorted according to integrated factors, and top 100 stocks are used for index compilation and weighted according to their market size.

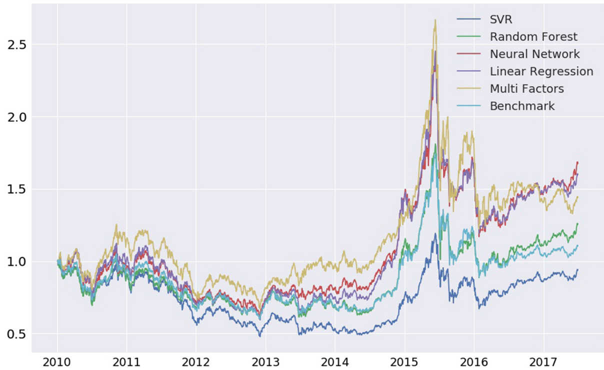

Fig. 2. Model integrated factor indexes based on CSI 800.

Results of indexes compiled based on integrated factors are performed in Fig. 2, from which we can see that the net value of most models based integrated factor indexes outperform benchmarks during most part of the back-test period. We further evaluate each index by the same performance indicators listed in Table 3. From the results of performance indicators, we can conclude that: (1) Factors integrated by neural networks and linear regression show better performance on profitability and stability than the multi-factors index. It implies that the model based integration can potentially mine the relationship between factors of stocks and their future performances. On the one hand, neural networks and linear regression based indexes show higher return than multi-factor indexes, on the other hand, volatility of multi-factor index is higher which means higher risk. Moreover, the higher Sharpe ratio still implies higher stability. (2) Neural networks show better performance than linear regression, which means the non-linear relationship between factors can be used to enhance the performance of integrated factors.

Table 3. Integrated factor indexes simulation results based on CSI 800

	Return	Annual Return	Volatility	Downside Beta	VaR	Alpha	Beta	Sharpe	Sortino	Loss rate	MDD	Active Return
Multi Factors	44.600%	5.199%	28.772%	1.1226	-2.934%	0.0387	1.0962	0.1869	0.0375	41.573%	-54.206%	33.636%
Benchmark	10.964%	1.440%	24.670%	1.0011	-2.533%	0.0000	1.0000	0.0249	-0.0901	47.191%	-48.984%	0.000%
SVR	-5.746%	-0.810%	25.275%	0.9734	-2.606%	-0.0228	0.9867	-0.0589	-0.1696	48.315%	-52.491%	-16.710%
Random Forest	25.914%	3.218%	25.691%	1.0176	-2.630%	0.0173	0.9980	0.1020	-0.0240	44.944%	-51.071%	14.950%
Neural Network	68.440%	7.429%	25.615%	0.9938	-2.602%	0.0563	0.9537	0.2590	0.1145	44.944%	-52.198%	57.476%
Linear Regression	60.386%	6.708%	24.650%	0.9513	-2.508%	0.0487	0.9185	0.2316	0.0953	43.820%	-51.341%	49.421%

4 Conclusion

Factor indexes reflect performances of factors for factor investing so that robust factors can be filtered. Filtered factors need to be further integrated, our work introduces deep neural networks and other supervised models to integrate factors supervised by future return. And indexes are compiled according to integrated factors to evaluate their performance. Experimental results show that supervised integration by the model can enhance the effectiveness of integrated factors compared to integration by optimization with a subjectively defined objective. And Neural network is verified to be more effective since it is able to mine deep non-linear relationship between factors and future performance of stock price.

Acknowledgement. This research was partly supported by the grants from National Natural Science Foundation of China (No. 71771204, 71331005, 91546201).

References

1. Ang, A.: A five-factor asset pricing model. Fama-Miller Working Paper (2014)
2. Bender, J., Briand, R., Melas, D., Subramanian, R.: Foundations of factor investing (2013)
3. Bender, J., Briand, R., Melas, D., Subramanian, R.A., Subramanian, M.: Deploying multi-factor index allocations in institutional portfolios. In: Risk-Based and Factor Investing, pp. 339–363. Elsevier (2015)
4. Fama, E.F., French, K.R.: The cross-section of expected stock returns. J. Finance **47**(2), 427–465 (1992)
5. Gu, S., Kelly, B.T., Xiu, D.: Empirical asset pricing via machine learning. SSRN (2018). https://doi.org/10.2139/ssrn.3159577
6. Krauss, C., Do, X.A., Huck, N.: Deep neural networks, gradient-boosted trees, random forests: statistical arbitrage on the S&P 500. Eur. J. Oper. Res. **259**(2), 689–702 (2017)
7. Long, W., Lu, Z., Cui, L.: Deep learning-based feature engineering for stock price movement prediction. Knowl.-Based Syst. **164**, 163–173 (2019). http://www.sciencedirect.com/science/article/pii/S0950705118305264
8. Menchero, J., Orr, D., Wang, J.: The Barra US equity model (USE4) methodology notes. MSCI Model Insight (2011)
9. Rivest, R.L.: Learning decision lists. Mach. Learn. **2**(3), 229–246 (1987)
10. Ross, S.A.: The arbitrage theory of capital asset pricing. In: Handbook of the Fundamentals of Financial Decision Making: Part I, pp. 11–30. World Scientific (2013)

11. Shen, F., Chao, J., Zhao, J.: Forecasting exchange rate using deep belief networks and conjugate gradient method. Neurocomputing **167**, 243–253 (2015)
12. Singh, R., Srivastava, S.: Stock prediction using deep learning. Multimed. Tools Appl. **76**(18), 18569–18584 (2017)
13. Valiant, L.G.: A theory of the learnable. Commun. ACM **27**(11), 1134–1142 (1984)
14. Xiong, T., Li, C., Bao, Y., Hu, Z., Zhang, L.: A combination method for interval forecasting of agricultural commodity futures prices. Knowl.-Based Syst. **77**(C), 92–102 (2015)
15. Zhou, T., Gao, S., Wang, J., Chu, C., Todo, Y., Tang, Z.: Financial time series prediction using a dendritic neuron model. Knowl.-Based Syst. **105**(C), 214–224 (2016)