Asset Pricing Factor Comparison in American and Chinese Stock Markets with Neural Networks

MA YUE

Abstract

We conduct a comparative factor analysis through the neural network approach between the U.S. market and China market. The neural network bases upon a selected factor zoo, which has 30 verified and grouped factors. With neural networks, we successfully identify that the factors influence varies between the markets from the prediction of individual stock returns. A further discussion for the sub-groups is conducted from the portfolio-level prediction test. The factor importance rank also verifies our study results.

1 Introduction

As many researches have discussed, early noted by Cochrane (2011) and recently by Hou, Xue, and Zhang (2017), hundreds of possible candidate factors that may explain the cross section of equity risk premiums have been proposed. One general problem faced the by the asset pricing field is a continuing factor proliferation which raised a fundamental task to judge and verify the explanatory power of the newly added factor comparing to the exiting factor zoo. Meanwhile, having such an abundant factors zoo, factor selecting methods for asset pricing and model choosing for portfolio constructing is another worth discussing challenge. Moreover, with significant distinctive environments, the U.S. and China equity markets have show plenty differences in terms of market efficiency (Tian, Wan, and Guo 2002), individual trading bias (Ng and Wu 2006), and financial related factors (Narayan and Zheng 2010). One problem. would also be raised: how would the differences in the market natures change the effective factors in one verified model?

In this article, we choose one machine learning method, Neural Network, with specified hidden layers of three, to conduct a comparative analysis between the U.S. and China equity market. In particular, we first construct a zoo containing 30 verified factors through back test. Factors are re-classified into four categories depending on its relation with the stock nature. Then we construct a Back Propagation Neural Network with three hidden layers and train it with four groups of neutralized and normalized factor signals in both U.S and China equity markets. After validation, the neural network is used to construct a portfolio with given weight allocation and turnover rate. Besides measuring the equity risk premiums, we also focus on the factor differences between these two markets and discuss the possible causes of the gap from aspects of market maturity, financial market structure, and economic environments.

The zoo is constructed from a wide-ranging set of factors which are the most frequently used in the last 30 years. We gather data from factor researches focusing separated aspects. For instance, Cochrane (1996) has examined the investment-based asset pricing model through the CAPM and the Chen, Roll, and Ross factor model comparing to a simple consumption-based model. Besides traditional factors, we also included momentum and trading related factors that are widely used by the industry as was discussed by Schmidt et al. (2015). However, a simple expansion of the factor zoo is not our aim and has little contribution to an improved understanding about the nature of the U.S. and China markets. In our discussion, all the selected factors are strictly classified to one of four groups, which represent different dimensions of one equity. An idea factor zoo should have each factor indicating one distinctive aspect of the risk premium. We try to minimize information overlap of signals in the selection procedure through eliminating factors that have similar characteristics. As an example, we only select one factor to represents the company's capacity of generating cash flow. All factors must be statistically significant under benchmark model to be concluded as a effective factor. Here we choose the Fama-French 5-factor model proposed by Fama and French (2015), and test them in both U.S. and China Equity market.

In most asset pricing model, following three models are commonly used: the simple linear regression model via ordinary least squares (OLS), Extension with robust objective function (Fama and French 2008), dimension reduced penalized linear models (PCR and PLS)(Kelly and Pruitt 2013), and generalized linear model(White 1980).However, Gu, Kelly, and Xiu (2020) gave a conclusive statement on empirical literature: Traditionally, the studies on the stock return prediction were strained into two streams. The first one establishes the stock return model from combination of stock-level characteristics, which could be exemplified by

Fama and French (2008) and Lewellen (2014). One typical method used is cross-sectional regression. The second stream focuses on the time series of returns. In the study of Welch and Goyal (2008), Koijen and Van Nieuwerburgh (2011), and Rapach and Zhou (2013), they conducted time-series regressions of aggregate portfolio returns with macroeconomic predictor variables. Gu, Kelly, and Xiu (2020) verified that machine learning methods could improve empirical understanding of asset pricing and also overcome potentially severe limitations of those traditional methods, especially with neural networks and regression trees. We continue our study based on the findings of Gu, Kelly, and Xiu (2020), and choose one machine learning method, Neural Networks, as our model.

Machine learning techniques are extensively used in high-dimensional mathematical analysis involving computational algorithms application by helping machine develop a unique solution toward some advanced problems, which are beyond human computing capacity. Through different training approaches, the machine could be modified to processing diverse data sets on the basis of the problem nature, for instance: supervised learning, unsupervised learning, reinforcement learning and so on(Mitchell et al. 1997). Comparing to traditional methods, machine learning approaches have the potentia to break limitations and overcome difficulties. Traditional regressions method may work well in situation with limited number of variables. However the emergence of new factors multiples the complexity that one model should interpret. Traditional models are challenged by the expanding amount of predictive information behind the burgeoning indicator zoo. Our research investigated over 400,000 lines of data covering 1000 individual stocks in the U.S. and China market form 1970 to 2019 and 1997 to 2019 respectively. Our factors zoo contained 30 factors classified into 4 distinctive groups to verify the differences between this two market. Though machine learning technique, we were able to analysis those data through a more comprehensive model. Referring to Gu, Kelly, and Xiu (2020), after investigating several machine learning algorithms supported by a large amount of test data along with traditional methods, the author demonstrated the capacity of machine learning in terms of performance improving on the return prediction. Moreover, neural networks and regression trees are found out to be the best performing methods.

Neural Network is defined as a supervised learning scheme implemented with a database which is constructed with patterned inputs along with corresponding targets. The pairs of inputs and targets should follow certain rules of classification. With paired data training, the trainee machine is expected to extract relevant information from the database and classify the future data pair pattern(Hansen and Salamon 1990). Neural networks pay more attention to train the machine to discover the influence of inputs regarding to targets rather than the model relationships inside for having several hidden layers stuck in the middle of the input layer and out layer. Based on Gu, Kelly, and Xiu (2020)'s empirical study on neural networks of different hidden layers, the model is found that "shallow" learning outperforms "deep" learning. The results differs from previous literatures on other field of machine learning application, for instance, computer vision or bioinformatics. A neural network with 3 hidden layers performed the best among the networks having hidden layers from 1 to 5. Although a higher-dimensional computing model may improve the performance of the explanatory models, the analysis cannot bypass or ignore the basic nature of asset pricing, that the indicators and results should be supported by their financial meaning rather than working as a pure number in an over abstracted mathematical problem. Thus, the choosing of model layer should still follow the principle of appropriation. This article follows the research results of Gu, Kelly, and Xiu (2020) and chooses the neural networks of 3 hidden layers.

With all the advantages, machine learning methods have already been adventitiously raised and used by previous asset pricing literature before the comprehensive research of Gu, Kelly, and Xiu (2020). A global lagged return prediction study constructed based on neural networks is proposed by Rapach, Strauss, and Zhou (2013). With specification of an empirical new-diffusing model, the authors are able to highlight the role of information frictions and also recognize the alternative explanations supporting the prediction ability of U.S. lagged returns. Khandani, Kim, and Lo (2010) chose to solve the problem of consumer credit and default risk valuation through regression trees, successfully develop a machine-learning model which could accurately forecast future credit risk events, providing ahead warning for banks' preparation for solving default issues. More recently, a deep neural network is implemented in the field of mortgage prepayment and foreclosure by Sirignano, Sadhwani, and Giesecke (2016). This study build from implementation of MLE to measuring of finite-difference approximation of sensitivities model, constructs a comprehensive valuation model of mortgage area. And Freyberger, Neuhierl, and Weber (2020) focus on the stock returns prediction through approximating a stochastic model with a linear function by shrinkage and selection methods.

In this article, we further extend machine learning to China equity markets for construct a comparative study. As a developing market, the China market has differences in many aspects, from government policy and market regulation, to investors' preference and market maturity. For example, the common used trading technique, short selling and margin trading are forbidden in Chinese equity markets. Based on a study of Covrig, Lau, and Ng (2006), investors of the institutional level show various stock preferences of domestic and foreign markets. The geographic asset allocation variantions are concluded to be driven by the international markets difference. Their results suggest the differences among markets also influence the large investors. As Ng and Wu (2006) stated in their research, over 99.5% of the investing population in Mainland China is constructed by individual investors, who are relatively less sophisticated investors compared to institutional investors who are close to the rational investors in the model of a developed market. Though the majority of the individual investors are lack of with full-scale financial knowledge, they still form their unique stock pitching preference, making a huge impact of the markets itself. In the behavioral preference study of Chinese individual investors, Chen et al. (2004) conclude the Chinese investors tend to be "overconfident", exhibit a representativeness bias on the stock trading, and pay less attention to reduce cognitive errors. However, Ng and Wu (2006) disclose in their study that besides behavioral biases, Chinese individual investors are influenced by the stock fundamentals to a large extent.

Our key contribution in this study is to demonstrate measuring of the grouped factors importance in both of the U.S. and China equity markets. Through IC importance test, all 30 factors are verified for its efficiency in the Chinese market by a benchmark mode of Fama-French 5-factor. With a neural networks of 3 layers, our model is able to extract the importance comparison through a large complex database. With grouping of profitability, valuation, momentum and market, 30 factors constructed four NN3 prediction models for both U.S. market, showing significant differences. For instance, the momentum group fails in the Chines equity market while it performaned the best in U.S. market, and the Chinese market shows a significant bias on the valuation and trading factors. This article also continued with portfolio performance valuation for all the characteristic-grouped and market-grouped NN3 models, and compared their out-of-sample predictive R^2 value and portfolio Sharpe ratio. We successfully examed the performance of neural networks in the asset pricing area with more detailed grouped factors, and also shows that the Chinese market is still in a respectively developing and inefficient stage. With well demonstrated comparison, this research could be contribute to an in-depth investigation on the special characteristics of Chinese market, and tothe development of trading strategy based on factor asset pricing model in a developing market.

2 Methodology

In this part, we would describe the all the methods and models applied in our study, including factors zoo construction and the Neural Network model. In each subsection, we would introduce the models in terms of its theoretical foundation, statistical model, and target results in detail.

Our research aim is to perform the asset pricing though machine learning techniques and provide an inproved and in-depth model discerption to understand the relationship between equity market and factors from all dimension in both markets. More specifically, the equity is evaluated by its monthly log return:

$$r_{i,t} = log(\frac{P_{i,t}}{Pi,t-1})$$

And we describe the asset return in as the following model:

$$r_{i,t} = E\left[r_{i,t}|t_{t-1}\right] + \epsilon_{i,t}$$

where

$$E\left[r_{i,t}|t_{t-1}\right] \quad = \quad f(z_{i,j,t})$$

Stocks are labeled by index $i = 1 \ldots N$, and time(months) is labeled by index $t = 1...N$. The return is generally understood as a basic function $E\left[r_{i,t}|t_{t-1}\right]$ with unpredictable noise$\epsilon_{i,t}$. Our target is to isolate the $E\left[r_{i,t}|t_{t-1}\right]$ as a function f of factors and analyze its structure to maximize its out-of-sample predictive ability for the real equity return. The factors from various dimensions are indexed as $j = 1...N$. The function is unrelated to both stock i and time t, and only relies on corresponding z (factors), which means our analysis on function f pays no attention to the information from the historical time t, or information beyond the chosen individual stock.

2.1 Factor Zoo

The factors zoos are constructed based on pre-selection from the most used factors from research and the industry application from two main categories, fundamentals-based factors and trading information-based factors. We further divide this two large group into four more specified sections to describe dimensions of the stock.

Table 1: Factor Zoo Class

Factor Group			
Prof&Growth	Valuation	Momentum	Market

agr	bm	chomom	baspread
chinv	cahspr	mom1m	dolvel
lgr	cfp	mom6m	ill
cashdebt	depr	mom12m	std_turn
operprof	dy	mom36m	maxret
roic	ep		retvol
saleinv	lev		
	mvel1		
	sp		

Followed this grouping principal, we use the information coefficient (IC) to evaluate the predictive efficiency of each factor and select the best performed indicators to construct a final factor zoo. Information coefficient is defined as a measure of the correctness of a predicted value, which is wildly used in financial an analysis to understand the relationship between actual values and predictions. The information coefficient ranges from 0 to 1, where a higher value denotes a stronger linear relationship.

In this selection process, we choose to apply rank IC, which is also called Spearman rank correlation, rather than the general IC (Pearson correlation). The rank IC is calculated as the correlation between the ranked Alpha vectors and the ranked forward equity returns for one time period.

$$r_s(X_1, X_2) = \frac{Cov(r(X_1), r(X_2))}{stdev(r(X_1)) \times stdev(r(X_2))}$$

Then

$$rankIC_{t-1} = r_s(\alpha_{t-1}, r_{t,t+1})$$

Here we select the factors based on its lagged IC form lag 1 to lag 12 and the following are the final factor zoo components with their rank IC.

Table 2: Factors Rank IC

	Lag Index											
agr	.014	.014	.016	.014	.015	.015	.016	.015	.014	.014	.014	.014
baspread	-.102	-.030	-.087	-.155	-.105	-.066	-.016	.104	.1641	.212	.147	.114
bm	.113	.086	.068	.067	.070	.069	.071	.083	.080	.079	.076	.080
cashdebt	.018	.018	.011	.016	.015	.012	.013	.011	.012	.008	.009	.009
cashpr	.013	.009	.012	.016	.023	.021	.020	.026	.023	.028	.024	.028
cfp	.021	.020	.017	.013	.012	.020	.023	.021	.015	.013	.011	.012
chinv	.009	.004	-.003	.002	-.005	-.007	-.005	-.001	-.007	-.004	-.006	0
turh	-.032	-.036	-.107	-.042	-.023	.016	.017	.035	.012	.027	.026	.075

Factor	1	2	3	4	5	6	7	8	9	10	11	12
chmom	-.072	-.039	.014	.043	.026	-.001	-.044	-.020	-.035	-.013	-.015	.087
depr	.021	.026	.021	.022	.022	.021	.023	.021	.013	.012	.019	.015
dolvol	-.032	-.036	-.101	-.042	-.022	.016	.017	.035	.013	.027	.036	.075
dy	-.016	-.063	-.069	-.006	-.049	.026	-.038	-.072	-.122	-.104	-.077	-.038
ep	.02	-.01	-.01	0	.01	-.01	0	.03	.03	.02	.06	.06
ill	.021	.011	.006	.002	.001	-.031	-.003	.041	.001	.007	.008	.010
lev	-.015	.018	.018	.019	.017	.017	.017	.014	.012	.008	.003	.004
lgr	.038	.041	.046	.043	.045	.043	.048	.053	.049	.042	.04	.041
maxret	.224	.122	.051	.050	.069	.028	.026	.063	.025	.076	.040	.041
mom1m	-.002	.002	-.003	-.003	-.001	.0003	0	.004	.003	.004	.001	.001
mom6m	.012	.014	.014	.012	.016	.015	.011	.084	.051	.047	-.023	-.014
mom12m	.014	.014	.012	.011	.009	.007	.004	.005	.004	.005	.003	.004
mom36m	.005	.005	.004	.003	.002	0	-.001	-.001	-.002	-.003	-.004	-.003
mvel1	-.004	-.138	-.167	-.058	-.105	-.049	-.105	-.137	-.128	-.118	-.047	.066
operprof	.117	.110	.097	.094	.089	.087	.082	.079	.086	.072	.065	.063
retvol	-.002	.091	.057	.010	.078	.073	.126	.144	.103	.118	.056	-.006
roaq	.101	.086	.071	.065	.056	.054	.054	.047	.052	.046	.043	.034
roegr	-.088	-.059	-.047	-.027	-.027	-.030	-.045	-.042	-.040	-.025	-.026	-.011
roic	-.176	-.143	-.147	-.171	-.139	-.142	-.043	.032	.062	.032	.033	.022
saleinv	.224	.079	.048	.016	-.037	-.044	-.048	.014	.114	.093	.147	.130
sp	.106	.070	.061	.065	.072	.073	.079	.086	.087	.091	.089	.094
std_turn	-.014	-.022	-.078	-.015	.003	.039	.046	.015	.025	.036	.038	.076
turh	-.032	-.036	-.107	-.042	-.023	.016	.017	.035	.012	.027	.026	.075

2.2 Neural Networks (Layers of 3)

Our main research tool is a nonlinear machine learning method called artificial neural network which is a computing model inspired by the biological neural networks. This kind of system is designed to learn from given examples by itself without any task-specified rule (Chen et al. 2019). Now it's almost the most preferred machine learning approach in solving complex problems, such as natural language processing, image recognition, and game-playing. A neural network is constructed on a collection of layers of units connected by preset links. In the implementation, the network will accept real number signals from the input layer, process the signals and pass them to the next layer of units. One neuron could receive information from multiple neurons in the upper layer, and also pass information to several different units in the next layer.

This study focuses on a traditional "feed-forward" network. A feed-forward network con-

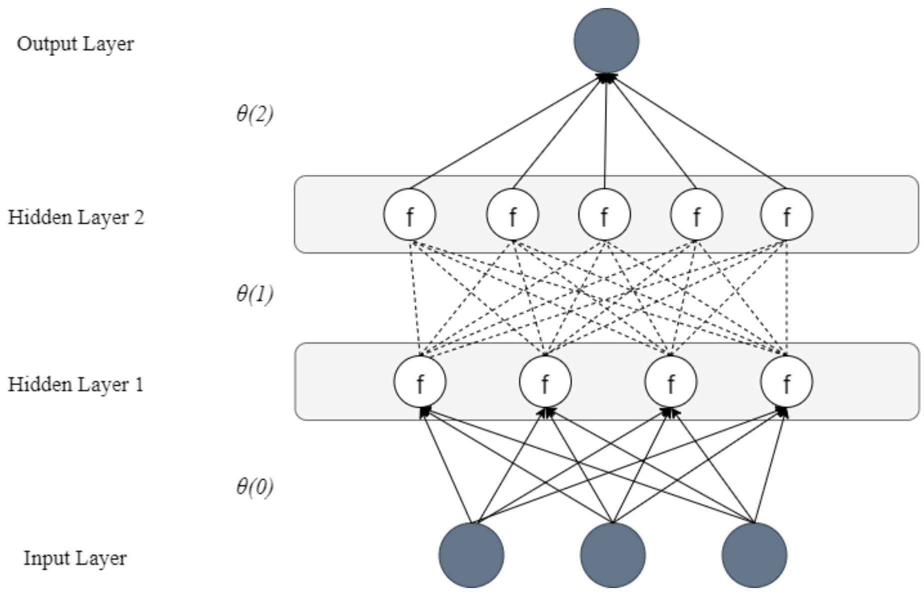


Figure 2.1:

tains an input layer where accepts the real number signals and an output layer to aggregates all the processed information and generates a prediction value just as stated before. Between the input and output layers, the network also has one or more hidden layers consisting neurons where the input signals are analyzed layer by layer through nonlinear transformation. However, the inner relationship between layers and neurons are sealed.

The input layer has the number of units that equals to the number of the dimension of the indicators, which we assume to be three in this demonstration. The following diagram shows how a three-inputs, one-output neural network with layers of two works.

From bottom up, the network will receive three number indicators from the input layer and then pass the information to the first layer. The first hidden layer is set to have four neurons, where each neuron will receive all the information from the input units but with unknown weight. Every neuron has a nonlinear collection function, which is called the step function, to aggregate the received information and prepare the signal that will be passed to the next layer. Then the next layer will repeat all the process conducted in the first layer again and pass these processed signals to the output layer to aggregate a final prediction. For example, for the first neuron in the second layer, it will transform the information collected from the input layer as

$$x_i^{(2)} = g(\theta_{1,0}^{(1)} + \sum_{j=1}^{4} z_j \theta_{1,j}^{(1)})$$

Then, the results of an aggregate forecast result from a linear output layer is:

$$f(z;\theta) = \theta_0^{(2)} + \sum_{j=1}^{5} x_j^{(2)} \theta_j^{(2)}$$

To construct a neural network, all factors of the networks should be decided carefully, including the number of hidden layers, the number of neurons in every layer, and the choice of which input, output and step function. Following to Gu, Kelly, and Xiu (2020), we choose a neural network with hidden layer of three for its best performance in the past study. For the units number of each layer, based on research of Bishop et al. (1995) and considering our factor zoo size, we set the neurons number as following.

First and second hidden layer:

$$N_i = N_{i-1} + 1$$

Third hidden layer:

$$N_i = \frac{1}{2}(N_{i-1} + 1)$$

where N indicates the number of neurons in each layer.

From many choices of transfer function, here we also follow the study of Gu, Kelly, and Xiu (2020), and use the same transfer function for all units in the hidden layers. The chosen function is known as the rectified linear unit (ReLU) and is defined by:

$$ReLU(x) = \begin{cases} 0 & x < 0 \\ x & otherwise \end{cases}$$

This function could speed up the derivative evaluation by encouraging a reduction of the number of active neurons. And we simply choose a linear combination as the output layer function. Then we have a general form of our neural network with implementation of ReLU transfer function. The layer is labeled by index $l = 1 \dots L$. N denotes the number of units in each layer. Output of each layer is defined as a column vector containing N rows $x^{(l)} = (1, x_1^{(1)}, \cdots x_{N^{(l)}}^{(l)})'$, where the input layer is defined similarly as $x^{(0)} = (1, z_1, \cdots z_N)'$. The output function for each layer is defined as (Jarrent,2009):

$$x_N^{(l)} = ReLU(x^{(l-1)'}\theta_N^{(l-1)})$$

while the final forecast result is:

$$f(z;\theta) = x^{(L-1)'}\theta^{(L-1)}$$

The model would have $N^{(l)}(1 + N^{(l-1)})$ weight parameters in evert hidden layer along with $1 + N^{(L-1)}$ weight parameters for the output layer.

The learning process of the neural network could be considered as letting the machine adapting itself to one task from the sample observations. The main task is to adjust the weight parameters in the transfer functions among layers to improve the prediction accuracy by minimizing the prediction errors. The learning rate is set to be the corrective steps in each minimization of errors. A higher learning rate will shorter the training time and speed up the error convergence, but also lead to a lower accuracy for fewer observations being considered. After several test, we set the learning rate to be 1×10^{-6} to reach a suitable training result. For training paradigms, we choose the most popular approach, supervised learning. In this method, the network is trained by paired input and target output. The learning task is to approximate the forecast output to the desired actual value. A cost function of mean-squared error (MSE) would be monitored to reduce the gap:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

For a better model performance, we further apply the backpropagation (BP) algorithm to the feed-forward networks. Backpropagation focuses on the gradient of the cost function regarding to the weight parameters in order to help the network learn the inner representation of its layers better. In general, the network could be understood as a matrix multiplication function and with backpropagation, it could compute the gradients of weights faster with fewer matric multiplications for each level.

3 Study on U.S. and China Equities

3.1 Data

Data Source

We conduct our empirical study on a broad set of the most representative stocks listed in the U.S. and Chinese stock exchanges, which could express the characteristics of the whole equity market better. For the American market, we target the S&P 500 index component stocks while the SCI 500 index component stocks are chosen for the Chinese market. And we select U.S. monthly data from 1979-01-01 to 2019-12-31 and China monthly data from 1997-01-01 to 2019-12-31. All the equity data is acquired from WRDS database[1] and Tushare data interface[2].

Data Processing

We first strikeout all delisted stocks in the current month. All the factor variables are processed with Winsorize method (Hastings, 1947) with respect to a confidence interval of 95%. Each indicator is also normalized into the range [0,1].

The Winsorize Method is commonly used in data processing to handle the possible spurious outliers. This method of computation improves prediction accuracy and eliminate statistical error by limiting extreme values. For a 95% confidence interval, the model treats all data group below the 2.5th percentile as the 2.5th percentile, and data group above the 97.5th percentile as the 97.5th percentile. Normolization means adjusting all signals in different scales to the unitive scale, which is range from 0 to 1 here. These normalized values allow the dataset to merge into a same scale and eliminate the possible effects of gross influence, which is an comon issue in time series problems. We choose the min-max feature scaling to restrict the signals in the dataset into the same data scale.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where X denotes the signal value.

From the price data abstract from the market, we calculate the monthly log return for each stock and combine it with its factors value. Each data line contains one monthly log stock return and 30 factors variables.

3.2 Performance Evaluation and Factor Importance

Performance Evaluation

For assessing the prediction accuracy for individual equity return and portfolio level return forecasts, we use the out-of-sample R to evaluate the model performance.

[1] https://wrds-www.wharton.upenn.edu/login/?next=/pages/support/getting-started/3-ways-use-wrds/
[2] https://tushare.pro/

Comparing to the in-sample test, the OOS test could avoid understating forecasting error. The nuances of history are unlikely to persist into the future. With development of the equity market, the nuances of future value might not be revealed in the past (Tashman 2000). Besides, also as stated by Makridakis et al. (1982), in-sample test also pays little attention to the overfitting and structural changes, and tend to narrow the prediction error and provide inaccurate error evaluation. The calculation is restricted in the testing subsample set, where data will not affect the model estimation or tuning process. Though many studies conducted out-of-sample test compared against historical mean returns, and this approach does perform well in prediction for aggregate index or long-short portfolios, the historical mean stock return could be "noisy" for machine to learn and also lead to underperform comparing to a pure forecast of zero (Gu, Kelly, and Xiu 2020).

$$R^2_{OS} = 1 - \frac{\sum_{t=1}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=1}^{T}(r_t - \bar{r}_t)^2}$$

Table 3: Monthly Stock-level Out-of-sample predictive R^2

	America	China
Fundamentals	0.03	0.84
Trading	0.43	0.68
Prof&Grow	0.34	0.92
Valuation	0.68	0.70
Momentum	0.97	0.54
Market	0.81	0.73

Factor Importance

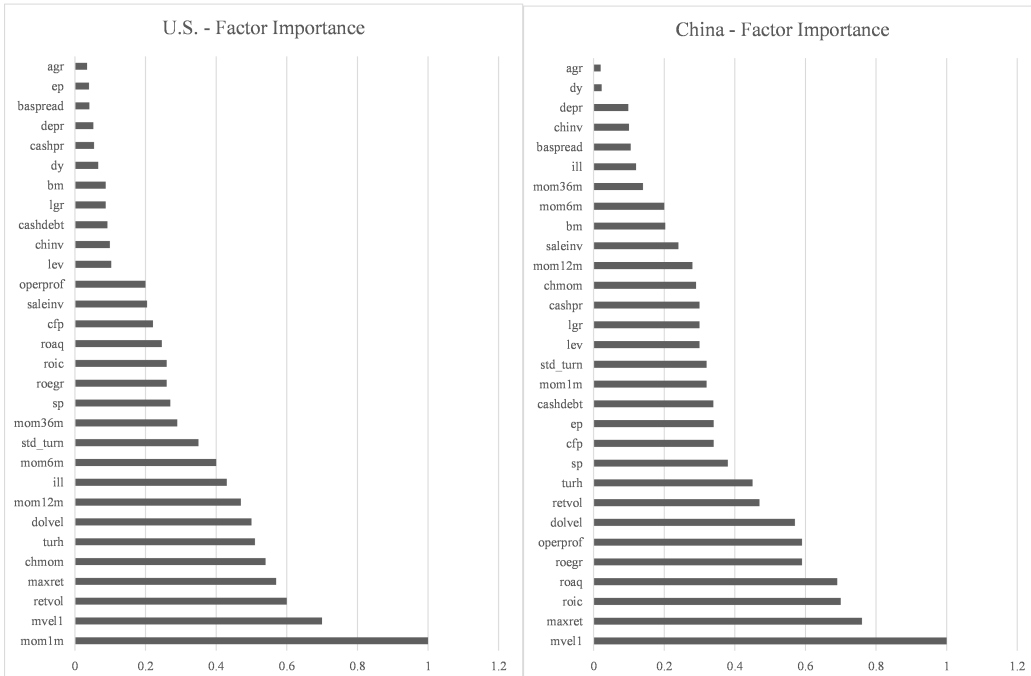
One of our primary purpose is to analyze the influence power of each indicator toward the excess stock return. Through the neural network model, we are able to make a prediction on individual stock return. Still the importance of factors within the forecast model is also crucial for us to understand the driving force of stock price. Thus, we implement the sum of squared partial derivatives of the model in each input indicator j to identify and manifest the factor importance (Dimopoulos, Bourret, and Lek 1995). This approach summarized the sensitivity of the model fits to changes in corresponding variable.

$$SSD = \sum_{i,t \in \mathcal{H}} \left(\frac{\partial g(z;\theta)}{\partial z_j}\Big|_{z=z_{i,t}}\right)^2$$

where the $\mathcal{H}$indicates the training dataset and the input varibales are labeled by $j = 1...N$.

In the table, all the factor importance variables are generalized into the range of [0,1], shift the value of the most essential factor as margin 1, and the zero importance as the margin 0.

In the shown chart, we could notice that in the U.S. market, factors in the momentum group (mom1m, chmom, mom12m, mom36m) are in the head leading position and make a great influence on the individual stock returns, while the market (turh, maxret, retvol,



(a) U.S.-Factor Importance (b) China-Factor Importance

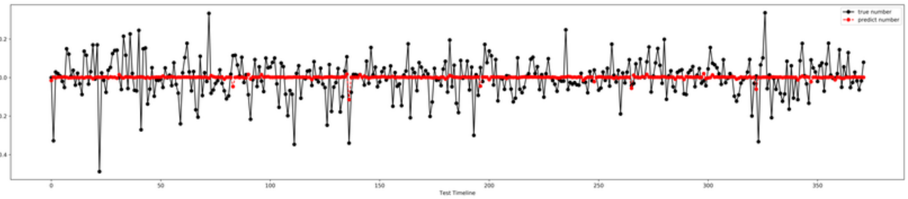
etc) factors also have a respectively strong position of influence. And the fundamentals factors are usually making less decision power in this model. Comparatively speaking, the valuation factors (bm, cashpr, dy, ep) are in the weakest state. For the China market, the failure of the momentum factors that are powerful in the U.S. market indicates the existence of a distinction between these two markets. However, the market factors still play an essential role in the model just as it in the U.S. market. The equity returns in the China market are seemed to be more sensitive to the company's profitability and growth potential (roegr, roic, roaq). In both markets, the mvel1 is always a strong and effective factor, which may indicate the size of one company have a considerable impact on the equity return and investors' decision making. Noticing the domain position of maxret (max return in the past 12 months), we could also recognize that the historical performance does influence the future stock equity a lot. At least it would affect the investors a lot in both U.S. and China market.

The marginal relationship between predicted return and each indicator is also marked in our analysis.

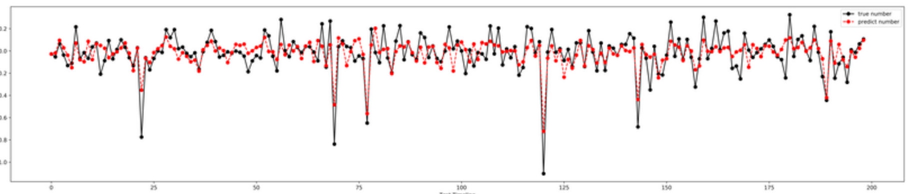
3.3 Model Perfomance Validaton

Neural Networks (Layer 3)

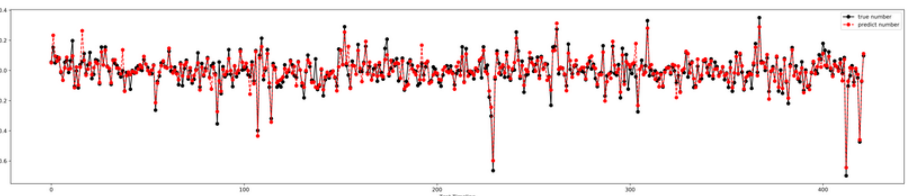
Our feed-forward neural network is built with one input layer, one linear output layer, and three hidden layers. For each hidden layer, it has the number of neurons, as stated in section 2.2. Inside the network, units from near layers are connected by transfer function following the rectified linear unit (ReLU) model. The total dataset is split into two groups, training dataset and test dataset, as the proportion of 0.8:0.2. The neural network is trained by the backpropagation (BP) algorithm using the mean-squared error (MSE) as the cost function. Then the network is compared and tested on the testing sample set and valuated its prediction accuracy. And the program is implemented by Python with the modular



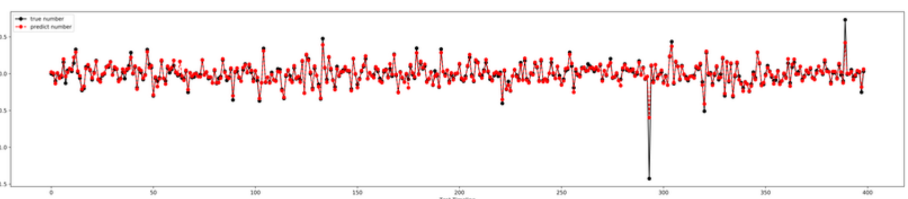
(a) U.S.-Fundamentals



(b) China-Fundamentals



(c) U.S.-Trading



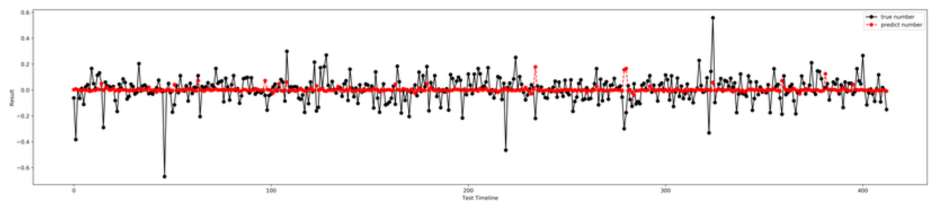
(d) China-Trading

Machine Learning Library Pybrain[3]. For each market, total of six networks are built and tested on the two big groups and four sub-groups.

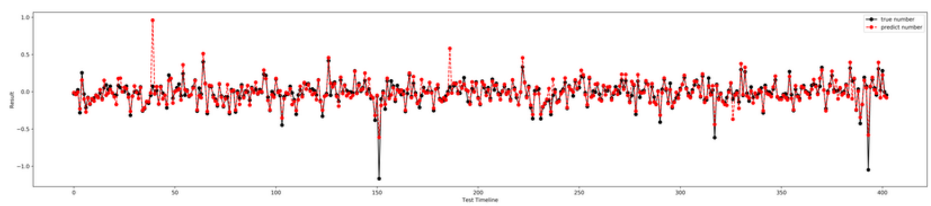
Perfomance Validation

All the models are trained for 101 epochs, and the total error denotes their final epoch error in the training sample set. For all error smaller than or equal to 0.05, we can conclude that our neural networks fit the dataset well in the training datasets. Most of the model performs well in the training process with a total error of less than 0.00%. For Fundamentals, Prof, Grow, and Valuation groups in China market, the little raised error could be due to the respectively less data for the small learning rate of 1×10^{-6}, which are still well fitted.

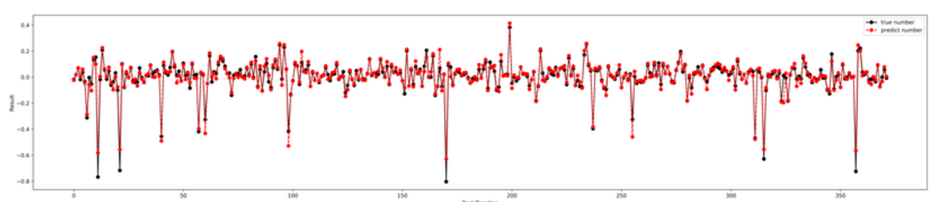
[3] http://pybrain.org/



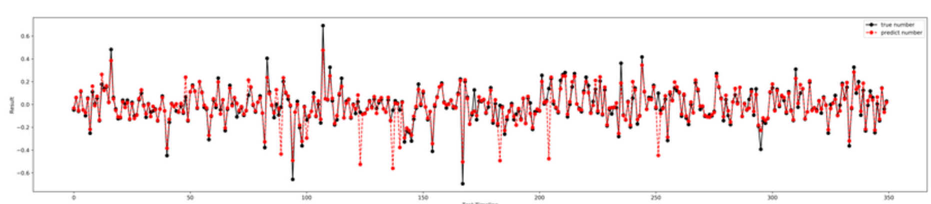
(a) U.S.-Prof&Grow



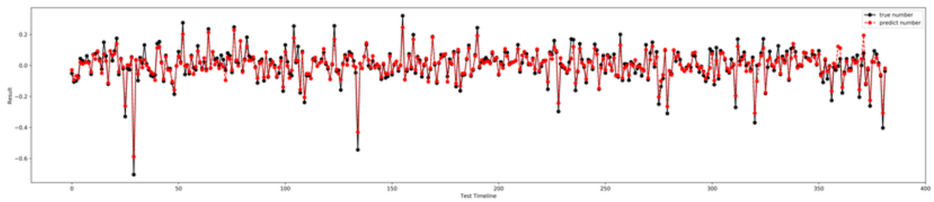
(b) China-Prof&Grow



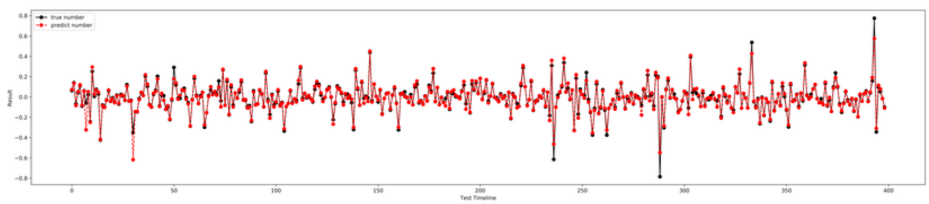
(c) U.S.-Valuation



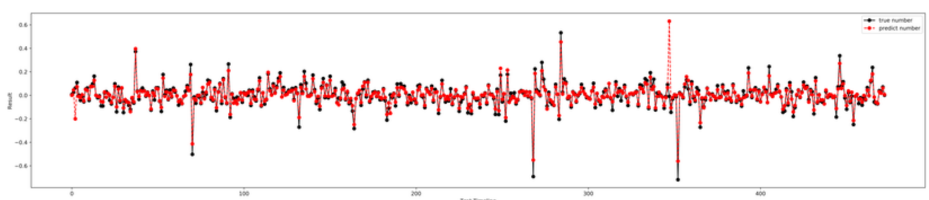
(d) China-Valuation



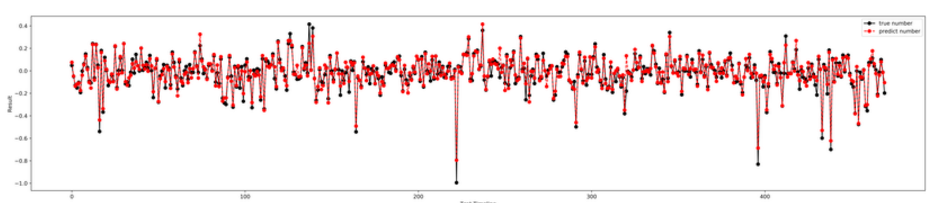
(a) U.S.-Momentum



(b) China-Momentum



(c) U.S.-Market



(d) China-Market

	America		China	
	Epochs	Totol Error	Epochs	Totol Error
Fundamentals	101	0.00%[4]	101	0.05%
Trading	101	0.00%	101	0.00%
Prof&Grow	101	0.00%	101	0.02%
Valuation	101	0.00%	101	0.03%
Momentum	101	0.00%	101	0.00%
Market	101	0.00%	101	0.00%

Considering the size of data, here we only plot around the latest 400 predictions and actual value in the graph for comparison.

We could observe from the graphic (a) that our network fails to predict the latest 400 individual stock returns in the U.S. market based on the fundementals group of factors, which correspond with the corresponding out-of-sample R results. Though with enough data, the machine could perform well in the training data prediction, it could not predict future returns. Comparatively, the fundamentals group of factors could capture the returns trend, including the extreme values well in the China equity market. The same situation also happens in the Pro&Grow group, which is the sub-group of the fundamentals group. Conversely, the trading group fit the U.S. market better. For the valuation and momentum groups of factors, they could carry out a respectively satisfying prediction. On both of these two markets, we could find that the implementation of the market group of predictor perform better in predicting the ordinary returns while fails a lot on the extreme values more often.

3.4 Portfolio Forecast

Till now, we have already discuss and verified the predictability of our neural networks in terms of individual stock returns. In the next step, we shall continue analysis on its forecasting performance of the aggregate portfolio returns. The further study on the portfolio-level returns is essential for us to verify the model efficiency and understand the market better.

First of all, as we have optimized our prediction on the individual stock returns, a study on the aggregate portfolio-level returns could help us test the model accuracy again. Besides, shifting focus from individual stocks to an aggregate portfolio could include broader economic interest and expand our study to the investors trading institutional assets, such as mutual funds, ETFs, and hedge funds. Moreover, the dependence relationship between stocks could be expressed through portfolio returns analysis sensitively, where a well-performed prediction on individual stock returns may not necessarily lead to success in the forecasting of portfolio returns.

Portfolio Construction

The portfolio is formed based on the monthly prediction of our neural network, which could directly reflect the prediction accuracy of our machine learning model. At the end of each month, which means for each line of our test data, the machine is asked to calculate 1-month ahead (next line) out-of-sample equity return prediction for each factor group. And

[4] 0.00% means that the total error is less than 0.00%

all portfolios are formed based on the data source stock baskets, which are S&P 500 index component stocks and SCI 500 index component stocks. Based on the predicted return value, the stocks are sorted from the smallest return to the biggest into 10 equally divided groups. Since we could not short sell in the China market, we set a fixed initial investment for only holding the highest expected return stocks (decile 10) using value weights. We choose value weighting rather than equal weighting for its benefit on minimizing equally weighted forecast error.

Forecast Performance

Table 5: Perforamnce of Neural Network Porfolio

	America			China		
	Pred.[5]	Avg.[6]	SR[7]	Pred.	Avg.	SR
Fundamentals	1.56	0.98	0.76	1.66	1.78	1.14
Trading	2.01	2.26	1.35	1.36	1.56	1.00
Prof&Grow	0.83	0.39	065	1.89	1.75	1.31
Valuation	1.97	1.54	1.20	1.45	1.63	1.14
Momentum	2.22	2.26	0.89	1.01	1.45	0.51
Market	1.98	1.62	0.67	1.83	1.53	0.83

The portfolio-level result slightly deviates from the individual-level stock returns analysis. In general, among the four sub-groups of factors, the momentum group generates the best prediction in the U.S. market which returns on average 2.22% per month (prediction) and 2.26 per month (realized average), and the prof&grow group domains the rest with a prediction of 1.89% monthly return and a realize average monthly return of 1.75%. The difference is highlighted on the group of pro&grow and momentum factors. Though they show no defect in the previous study of individual-level of stock returns, their portfolio-level performance is a little disappointing on the U.S and China market, respectively.

4 Concludsion

Based on the constructed factor zoo, we conducted a comparative study of factor asset pricing for the U.S. and China equity markets through the neural network approach. We successfully verified the good performance of neural networks with three hidden layers. With almost all groups, including the two primary teams and four sub-groups, the network could learn through training samples and make a rational prediction in the out-of-sample test. Though the model generally works, its outcome variate among the factor groups, which indicates the different predictive ability of each group of factors in these two equity markets. Through the distinctive performances of factors, we could summarize that the fundamentals information plays less role in the asset pricing in the U.S. market and is on a more domain

[5] Predicted monthly returns
[6] Average realized monthly returns
[7] Sharpe ratio

role for the China market. While the signals relating to the movements and other aspects of the stock market itself reveal its vital function among American stock returns, they fail to capture the future returns in China. We continue our study further on the portfolio-level prediction, which provided us with more precise guidance on the performance of factors in sub-groups. The valuation of factor importance also validates our findings on the factors influence differences between U.S. and China equity markets. For example, the most effective factors in U.S. is momentum factors or factors reflecting market trading information, such as return volatility (retvol), dollar volume (dolvol), and max return (maxret). However, in China, the equity returns react more towards company-related factors, especially those relating to company growth and profitability ability.

To sum up, the performance gap of the same factor zoo in these two markets significantly indicates the level of market sophistication differs between them. On the one hand, the U.S. market is more mature than the China market for a more extended history, opener market regulation, and freer trading environment. On the other hand, the components of the investors also vary. U.S. market has a lot more institution-level investors who are more ration and better at collecting and processing public information. The investors in the China market are the join of individual investors who tend to have more behavioral preferences, pay more attention to the stock companies themselves, and also harder to predict.

Overall, based on previous literature, our study shows the feasibility of neural work application on asset pricing in the China equity and provides valuable comparative factor research between these two markets. Our findings build a foundation for further research on the comparison of the U.S. market and China market digging from the factor's differences.

Bibliography

Azoff, E Michael (1994). *Neural network time series forecasting of financial markets.* John Wiley & Sons, Inc.

Bishop, Christopher M et al. (1995). *Neural networks for pattern recognition.* Oxford university press.

Campbell, John Y and Samuel B Thompson (2008). "Predicting excess stock returns out of sample: Can anything beat the historical average?" In: *The Review of Financial Studies* 21.4, pp. 1509–1531.

Chen, Gong-Meng et al. (2004). "Behavior and performance of emerging market investors: Evidence from China". In: *unpublished Washington State University Working paper (January).*

Chen, Luyang, Markus Pelger, and Jason Zhu (2019). "Deep learning in asset pricing". In: *Available at SSRN 3350138.*

Chen, Yung-Yao et al. (2019). "Design and implementation of cloud analytics-assisted smart power meters considering advanced artificial intelligence as edge analytics in demand-side management for smart homes". In: *Sensors* 19.9, p. 2047.

Cochrane, John H (1996). "A cross-sectional test of an investment-based asset pricing model". In: *Journal of Political Economy* 104.3, pp. 572–621.

— (2011). "Presidential address: Discount rates". In: *The Journal of finance* 66.4, pp. 1047–1108.

Covrig, Vicentiu, Sie Ting Lau, and Lilian Ng (2006). "Do domestic and foreign fund managers have similar preferences for stock characteristics? A cross-country analysis". In: *Journal of International Business Studies* 37.3, pp. 407–429.

Dimopoulos, Yannis, Paul Bourret, and Sovan Lek (1995). "Use of some sensitivity criteria for choosing networks with good generalization ability". In: *Neural Processing Letters* 2.6, pp. 1–4.

Eldan, Ronen and Ohad Shamir (2016). "The power of depth for feedforward neural networks". In: *Conference on learning theory*, pp. 907–940.

Fama, Eugene F and Kenneth R French (1992). "The cross-section of expected stock returns". In: *the Journal of Finance* 47.2, pp. 427–465.

— (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of.*

— (2008). "Dissecting anomalies". In: *The Journal of Finance* 63.4, pp. 1653–1678.

— (2015). "A five-factor asset pricing model". In: *Journal of financial economics* 116.1, pp. 1–22.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu (2017). "Taming the factor zoo". In: *Chicago Booth research paper* 17-04.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber (2020). "Dissecting characteristics nonparametrically". In: *The Review of Financial Studies* 33.5, pp. 2326–2377.

Grinold, Richard C and Ronald N Kahn (2000). "Active portfolio management". In:

Gu, Shihao, Bryan Kelly, and Dacheng Xiu (2020). “Empirical asset pricing via machine learning”. In: *The Review of Financial Studies* 33.5, pp. 2223–2273.

Hansen, Lars Kai and Peter Salamon (1990). “Neural network ensembles”. In: *IEEE transactions on pattern analysis and machine intelligence* 12.10, pp. 993–1001.

Harvey, Campbell R, Yan Liu, and Heqing Zhu (2016). “¡ and the cross-section of expected returns”. In: *The Review of Financial Studies* 29.1, pp. 5–68.

Hastings, Cecil et al. (1947). “Low moments for small samples: a comparative study of order statistics”. In: *The Annals of Mathematical Statistics* 18.3, pp. 413–426.

Haykin, Simon (1994). *Neural networks: a comprehensive foundation.* Prentice Hall PTR.

Hecht-Nielsen, Robert (1992). “Theory of the backpropagation neural network”. In: *Neural networks for perception.* Elsevier, pp. 65–93.

Hornik, Kurt, Maxwell Stinchcombe, Halbert White, et al. (1989). “Multilayer feedforward networks are universal approximators.” In: *Neural networks* 2.5, pp. 359–366.

Hou, Kewei, Chen Xue, and Lu Zhang (2017). “Replicating anomalies”. In: *The Review of Financial Studies.*

Jarrett, Kevin et al. (2009). “What is the best multi-stage architecture for object recognition?” In: *2009 IEEE 12th international conference on computer vision.* IEEE, pp. 2146–2153.

Kelly, Bryan and Seth Pruitt (2013). “Market expectations in the cross-section of present values”. In: *The Journal of Finance* 68.5, pp. 1721–1756.

Khandani, Amir E, Adlar J Kim, and Andrew W Lo (2010). “Consumer credit-risk models via machine-learning algorithms”. In: *Journal of Banking & Finance* 34.11, pp. 2767–2787.

Koijen, Ralph SJ and Stijn Van Nieuwerburgh (2011). “Predictability of returns and cash flows”. In: *Annu. Rev. Financ. Econ.* 3.1, pp. 467–491.

Lewellen, Jonathan (2014). “The cross section of expected stock returns”. In: *Forthcoming in Critical Finance Review.*

Liew, Jimmy and Maria Vassalou (2000). “Can book-to-market, size and momentum be risk factors that predict economic growth?” In: *Journal of Financial Economics* 57.2, pp. 221–245.

Makridakis, Spyros et al. (1982). “The accuracy of extrapolation (time series) methods: Results of a forecasting competition”. In: *Journal of forecasting* 1.2, pp. 111–153.

Mitchell, Tom M et al. (1997). *Machine learning.*

Narayan, Paresh Kumar and Xinwei Zheng (2010). “Market liquidity risk factor and financial market anomalies: Evidence from the Chinese stock market”. In: *Pacific-Basin finance journal* 18.5, pp. 509–520.

Ng, Lilian and Fei Wu (2006). “Revealed stock preferences of individual investors: Evidence from Chinese equity markets”. In: *Pacific-Basin Finance Journal* 14.2, pp. 175–192.

Rapach, David and Guofu Zhou (2013). “Forecasting stock returns”. In: *Handbook of economic forecasting.* Vol. 2. Elsevier, pp. 328–383.

Rapach, David E, Jack K Strauss, and Guofu Zhou (2013). “International stock return predictability: what is the role of the United States?” In: *The Journal of Finance* 68.4, pp. 1633–1662.

Schaul, Tom et al. (2010). “PyBrain”. In: *Journal of Machine Learning Research* 11.24, pp. 743–746.

Schmidt, Peter S et al. (2015). “On the construction of common size, value and momentum factors in international stock markets: A guide with applications”. In: *CCRS Working Paper Series* 01/11.

Sirignano, Justin, Apaar Sadhwani, and Kay Giesecke (2016). “Deep learning for mortgage risk”. In: *arXiv preprint arXiv:1607.02470*.
Tashman, Leonard J (2000). “Out-of-sample tests of forecasting accuracy: an analysis and review”. In: *International journal of forecasting* 16.4, pp. 437–450.
Tian, Gary Gang, Guang Hua Wan, and Mingyuan Guo (2002). “Market efficiency and the returns to simple technical trading rules: new evidence from US equity market and Chinese equity markets”. In: *Asia-Pacific Financial Markets* 9.3-4, pp. 241–258.
Welch, Ivo and Amit Goyal (2008). “A comprehensive look at the empirical performance of equity premium prediction”. In: *The Review of Financial Studies* 21.4, pp. 1455–1508.
White, Halbert (1980). “Using least squares to approximate unknown regression functions”. In: *International Economic Review*, pp. 149–170.