Spring 5-2019

Machine Learning and Empirical Asset Pricing

Yingnan Yi

Machine Learning and Empirical Asset Pricing

Yingnan, Yi*

Current Draft: May 6, 2019

Abstract

In this paper, I conduct a comprehensive study of using machine learning tools to forecast the U.S. stock returns. I use three sets of predictors: the past history summarized by 120 lagged returns, the technical indicators measured by 120 moving average trading signals, and the 79 firm fundamentals, which helps to understand the weak-form market efficiency, algorithm trading and fundamental analysis. I find each set independently has strong predictive power, and buying the top 20% stocks with the greatest predicted returns and shorting bottom 20% with the lowest earns economically significant profits, and the profitability is robust to a number of controls. Econometrically, neural network generally improves forecasting over linear models, but makes little difference with firm fundamental predictors. Ensemble method tends to perform the best. However, when combining information from all the predictors, traditional machine learning improves little the performance due to perhaps not enough time series for too large dimensionality. In contrast, simple forecasting combination and portfolio diversification approach provide large gains.

Keywords: Momentum, Machine Learning, Ensemble Learning, Neural Network, Deep Learning.

*Yingnan Yi is a DBA candidate at Washington University in St. Louis, Missouri, United States

§I benefit a great deal from discussions with my advisor Professor Guofu Zhou at Washington University in St. Louis, Professor Fuwei Jiang at School of Finance of Central University of Finance and Economics, Professor Guohao Tang at Hunan University in China, Dr. Xiaming Zeng at Barclays Capital London Office. All errors are mine

‖I gratefully acknowledge the computing support from the Center for High Performance Computing at Washington University in St. Louis. All computations were performed using the facilities of the Washington University Center for High Performance Computing, which were partially provided through grant NCRR 1S10RR022984-01A1.

1. Introduction

Machine learning is receiving increasing attention in various areas, i.e., Heaton, Polson, Witte (2016). Recent studies on U.S. market focus on cross-section predictability of stock returns. For example, McLean and Pontiff (2015) examine various anomalies, and Green, Hand, Zhang (2017) examine the predictive power of firm characteristics. The recent study of Han, He, Rapach, Zhou (2018) use machine learning tools to uncover more stable and greater predictability than previously found. Instead of cross-section predictability, Gu, Kelly, Xiu (2018) apply comprehensive set of machine learning tools to study time series predictability.

In this paper, I use a comprehensive set of machine learning techniques to analyze cross-sectional predictability of U.S. data over 40-year time period from January 1978 to December 2017. I categorize an extended panel of predictors into three subcategories of different economic nature and attempt to speak to different questions under machine learning framework. Specifically, I investigate the portfolio performance using: (1) 120 historical monthly stock returns, (2) 120 technical moving average trading signals (Neely, Rapach, Tu, Zhou (2014)), and (3) 79 fundamental firm characteristics. I also compare model performance during sub-periods, by long- and short-leg portfolios, by size quintiles, and during different investor sentiment periods and business cycles. Following Gu, Kelly, Xiu (2018), I apply major machine learning tools available, including ridge, lasso, elastic net, and 5 neural network models of various architectures. In addition, I apply two commonly used dimensionality-reduction techniques, namely principal component regression (PCR) and partial least square (PLS).

To predict stock return during month t using a given model, I estimate the model on data during month t-1 (training sample). I use 5-fold cross validation or a single holdout validation sample to determine the optimal hyper-parameters or optimal stopping point.[1]. The resulting model is used to predict cross-sectional stock return on the data over month t (testing sample). I independently sort stocks into quintiles based on the predicted return from each model and equal weight stocks within each quintile. The spread portfolio is constructed by buying the predicted winner quintile and selling the predicted loser quintile.

In addition to forming portfolios based on individual models, I construct three ensemble portfolios based on unanimous voting rule. First I split models into two categories: linear and neural network, where linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic Net and neural network family includes 5 neural network models with assorted architectures. At the beginning of each month, each estimated model in a given family casts its votes as to in which quintile the stocks will fall in the coming month. Only stocks that receive unanimous votes of top or bottom quintile from all models within a family will be bought or sold. I also

[1]The validation sample is a 10% holdout sub-sample of the training set.

construct a "total ensemble" portfolio consisting of all 11 models under unanimous voting rule.

Among single-model-based portfolios, our finding shows that in general neural network portfolios perform better than linear ones mainly for price-related predictors, i.e., lagged return and moving average trading signals. For the 120 lagged return setting, the monthly excess return of spread portfolios based on linear models generally falls around 0.8% while the return based on neural networks falls between 0.85% and 1.03%. In addition, combination of linear models (networks) yields monthly return of 1.24% and 1.59%, which are significantly higher than that of any component model in their respective class and represents a 28% (1.59/1.24-1) performance gap between the two classes of models. For the moving average signal setting, the monthly excess return of spread portfolios based on linear models falls around 0.8% while the return based on neural networks falls around 1.0%, about a quarter higher. Like in lagged return setting, combination of linear models (networks) yields monthly return of 1.09% (1.48%), which are approximately 36% (48%) higher than any component model in their respective class and represents a 0.36% (1.48/1.09-1) gap in relative sense. In addition, I observe a significant jump in excess return when linear and networks models are combined. For example, return jumps to 1.74% (from 1.59%) in lagged return setting and to 1.89% (from 1.48%) in moving average signal setting. Risk-adjustment does not weaken our finding.

For fundamental firm characteristics, our results show marginal improvement of neural network models over linear models. Monthly excess return of spread portfolios based on linear models falls between 1.29% to 1.38% while return based on neural networks falls between 1.31% and 1.36%, essentially identical to each other. In addition, combination of linear models yields a monthly return of 1.62% while combination of networks yields a return of 1.75%, again close to each other. Nonetheless, such combination still brings significant boost to excess return relative to the return based on individual models. Furthermore, when linear and network models are combined, return jumps to 1.96%, the highest among all models. However, I only observe improvement in Sharpe ratio in the moving average trading signal setting and not in the other two.

Our finding has two implications. First, the success of neural network in our two price-related settings suggests that the true data generating process underlying the follow-the-trend strategy has significant non-linear component. For example, momentum strategy ranks stocks based return history and buy previous winners and sell previous losers. Such strategy assumes linear influence from past returns and is unnecessarily stringent. Our result suggests that relaxing such assumptions may improve predictive performance of momentum strategy. Our finding also suggests that using models with higher capacity and ensemble methods on market data may provide additional insights into the weak form of efficient market hypothesis. Furthermore, unlike traditional momentum strategy which mainly rely on time-series dependence to pick stocks, our design rely on a cross-sectional design and show that the cross-section of prior returns is also informative.

Second, ensemble method, i.e., unanimous votes, boosts predictive power for both linear and neural network models. I propose two explanations. First, considerable amount of noise exists in predictions from a single model[2] and therefore consensus across models help correct mistakes made by any individual model. It is also possible that I can only approximate part of the true data generating process with any individual model, making it necessary to combine multiple models to finish the puzzle.[3].

Third, I provide evidence that blindly applying neural network models may not be productive. The effectiveness of neural network largely depends on the data in question. In particular, the weak improvement of neural network over linear models on fundamentals suggests that the information in fundamental data may not be sufficiently complicate to warrant models of high capacity. In such case, linear models suffice and complicate models harm predictibility.

I contribute to several streams of literature. First, our paper relates to a growing body of research on the application of machine learning techniques to financial data. A broad range of previous studies on machine learning techniques focuses on extracting factors from a myriad of predictors and correcting the bias in existing methods. Rapach, Strauss, Zhou (2013) uses Lasso to forecast international equity returns using cross-country lagged returns. Rapach, Strauss, Tu, Zhou (2018) use a second-stage OLS to correct for the downward bias in the first-stage Lasso estimation. Rapach, Zhou (2019) use Sparse-PCA to shrink more weights of principal components to zero to enhance interpretability of principal components in forecasting stock return. Freyberger, Neuhierl, Weber (2019) and Feng, Giglio, Xiu (2019) use modified Lasso to select important return predictors from a predictor zoo and to forecast expected stock returns. Light, Maslov, Rytchkov (2017) propose the PLS approach for estimating expected returns on individual stocks from cross-sectional firm characteristics. They start off by modifying the time series PLS adopted by Kelly, Pruitt (2015), and Huang, Chen, Liu (2015) and significantly improve predictive power of firm characteristics. More recently, researchers start to introduce more flexible models to tackle the forecasting task in finance. Butaru, Chen, Clark, Das, Lo (2016) and Sirignano, Sadhwani, Giesecke (2018) apply regression tree and deep neural network to forecast default probability of consumer

[2]Even after I average predictions from the same neural network architecture across 100 random weight initializations.

[3]An analogy is the approximation of a function by its Taylor expansion and I use different models to estimate each terms.

credit card and mortgage loans respectively and show great potential of machine learning techniques in risk management. Heaton, Polson, Witte (2016) introduce a general frame work for deep learning in finance and use auto-encoder to reduce the number of stocks used to replicate NASDAQ Biotechnology Index. Gu, Kelly, Xiu (2018) apply machine learning methods to predict cross-sectional stock returns using a broad set of firm characteristics and show great potential of neural network in financial world. I take one step further and try to measure the informational content of predictors of differential economic nature using machine learning techniques.

Second, our study contributes to the momentum and algorithmic trading literature. Prior returns and price/volume-related signals have the potential to generalize to algorithmic or high frequency trading context due to abundance of data. Plenty studies exist in such area. Huang, Zhang, Zhou, Zhu (2019) combine price and fundamental momentum and construct a twin-momentum portfolio which is more profitable than simple summation of the two momentum portfolios. Another study by Han, Zhou, Zhu (2016) simultaneously consider information in moving average prices over short-, intermediate-, and long-horizon and also identify significantly stronger predictive power and lower risk. I extend these research by applying machine learning tools to an expanded set of momentum and price-related predictors. I create a panel of 120 prior returns and 120 price/volume trading signals and use machine learning tools to forecast returns on the two datasets independently. Our results challenge weak form market efficiency and shows great potential for algorithmic and high frequency trading with machine learning tools. Third, by separating fundamental variables from others I independently examine machine learning's applicability to fundamental analysis. I start from the 94 firm characteristics in Green, Hand, Zhang (2017) and exclude 15 price/volume-related predictors to isolate the predictability of fundamental variables. Our results show that simple neural network models do not outperform linear models on fundamental variables.

The remainder of this paper is organized as follows. Section 2 discusses the design of our tests. Section 3 explains the specification of the models and discusses our selection of hyper-parameters. Section 4 discusses data. Section 5 reports results. Section 6 concludes the paper.

2. Design

In this section I describe the predictors, loss function, data splitting, hyper-parameter tuning, and portfolio construction which apply to all models[4]. Details of models will be

[4]Except for OLS, which needs no parameter tuning or validation.

presented in section 3.

2.1. Predictors

Our target variable is one-month ahead cross-sectional monthly excess returns. I use three sets of predictors and compare their predictive performance using linear and neural network models separately. Our predictors include: (1) 120-month lagged monthly excess return on security level, (2) 60 stock price and 60 trading volume moving average signals (Neely, Rapach, Tu, Zhou (2014)), and (3) 79 fundamental firm characteristics. Since lagged returns are self-explanatory, I explain price/trading volume moving average signals and 79 fundamental firm characteristics in details below.

Following a similar vein in Neely, Rapach, Tu, Zhou (2014), I construct the price and trading volume moving average signals by comparing short-term moving average of monthly closing price and trading volume with their respective long-term moving average. Higher short-term than long-term moving average is interpreted as a buy signal, vice versa for sell signal. Specifically, I calculate the ratio of the short-term and long-term moving average of the two variables and interpret a ratio higher than 1 as a buy signal and a ratio less than 1 as a sell signal. To calculate the price signals, I compute the following:

$$S_{i,t} = \frac{MA_{s,t}}{MA_{l,t}}, \tag{1}$$

where $MA_{j,t} = \frac{1}{j}\sum_{i=0}^{j-1} P_{t-i}$ for $j = s, l$, and P_t is month-end closing price. $s = 1, 2, 3$ and $l = 6, 12, 18, ..., 120$ and therefore I have 60 combinations of short- and long-term comparisons and hence 60 moving average price signals.

For trading volume, I follow Granville (1963) and Neely, Rapach, Tu, Zhou (2014) and first calculate the "on-balance" trading volume as follows:

$$OBV_t = \sum_{k=1}^{t} VOL_k D_k, \tag{2}$$

where VOL_k is trading volume during month k and D_k is a binary variable that takes the value of 1 if $P_k \geqslant P_{k-1}$ and -1 otherwise. Then the moving average signals are constructed using OBV and in the same manner as moving average price signals. Specifically, I compute the following:

$$S_{i,t} = \frac{MA^{OBV}_{s,t}}{MA^{OBV}_{l,t}}, \tag{3}$$

where $MA^{OBV}_{j,t} = \frac{1}{j}\sum_{i=0}^{j-1} OBV_{t-i}$ for $j = s, l$. As in moving average price signals, $s =$

$1, 2, 3$ and $l = 6, 12, 18, ..., 120$ and again I have 60 combinations of short- and long-term comparisons and hence 60 OBV moving average signals. I bundle the 60 price and 60 trading volume moving average signals and form the second set of predictors.

To obtain the 79 firm fundamental characteristics, I start from the 102 fundamental characteristics in table 1 in Green, Hand, Zhang (2017) and exclude 17 price-, trading volume-, or return-related variables, which are chmom (change in 6-month momentum), indmom (industry momentum), maxret (maximum daily return), mom12m (12-month momentum), mom1m (1-month momentum), mom36m (36-month momentum), mom6m (6-month momentum), std_dolvol (volatility of liquidity (dollar trading volume)), retvol (return volatility), std_turn (volatility of liquidity (share turnover)), baspread (bid-ask spread), chcsho (change in shares outstanding), pricedelay (price delay), idiovol (idiosyncratic return volatility), ill (illiquidity), turn (share turnover), and zerotrade (zero trading days). I remove these variables to minimize the overlap of information content with lagged return or moving average signals. In addition, I exclude 6 characteristics whose variance inflation factor (VIF) are greater than 7 (due to multicollinearity concern), including betasq (beta squared), dolvol (dollar trading volume), lgr (growth in long-term debt), pchquick (% change in quick ratio), quick (quick ratio), stdacc (accrual volatility) [5]. Therefore, I end up with 79 (102 - 17 - 6) fundamental characteristics.

2.2. *Loss Function*

I employ mean squared error loss function to measure the fitness of our model on the training set[6]. On the training set during month t and of size N, the loss value is computed in equation 4.

$$L(\theta)_t = \frac{1}{N} \sum_{i=1}^{N} (r_{i,t} - \hat{r_{i,t}})^2, \tag{4}$$

where $r_{i,t}$ is the monthly stock returns for the $\imath^{th}$ firm in month t and $\hat{r_{i,t}}$ is the predicted returns from model.

2.3. *Data Splitting*

To make sure our estimated model generates out-of-sample prediction of stock return $r_{i,t}$ during month t, I train our model on data during month t-1, $S_{train,t-1}$. $S_{train,t-1}$ contains

[5] Per Green, Hand, Zhang (2017), two price-related variables also have VIF greater than 7. They are maxret and mom6m and have already been excluded in the first round of screening

[6] Training set is discussed in 2.3 below

the actual returns during month t-1 as target variable and its aligned (lagged) predictors. Then I use the predictors for month t, which are lagged returns relative to month t, to make predictions. I run a fixed-width rolling window to train our models.

2.4. Hyper-Parameter Tuning

Many machine learning models involve hyper-parameters that need to be provided before fitting the model. Hyper-parameters are parameters that cannot be estimated directly from the data, including the penalty parameters in Ridge regression and the number of hidden layers and nodes in a neural network. Although there is little theoretical guidance as to how to choose hyper-parameters, the majority of machine learning community endorses the use of validation sets[7].

In this study, I implement either five-fold cross validation or holdout validation to determine the optimal hyper-parameter on the validation set. In K-fold cross validation, the training set is first randomly divided into K segments. Then a range of hyper-parametric models are trained on K-1 segments and evaluated on the remaining segment. Each of the K segments takes turns to be evaluated upon and the average loss value on all K segments is assigned as the score of this model. The hyper-parameter that generates the lowest score (loss) are chosen. I apply 5-fold cross validation to PCR, PLS, Lasso, Ridge, and Elastic Net. For neural network, I use a 10% holdout validation sample to pick optimal hyper-parameter.

2.5. Hedge Portfolios

To construct the hedge portfolio for month t using a specific model, I first train the model on data over month t-1 and select hyper-parameters via cross validation or holdout validation. I use the fitted model to predict stock return during month t. Then I sort the monthly cross-section of stocks into quintiles based on their predicted return and form equally weighted portfolios for each quintile.[8] Finally, I construct the hedge portfolio by buying the top quintile and selling the bottom quintile. The hedged return is calculated by subtracting the return of the bottom quintile from that of the top quintile.

2.6. Ensemble Portfolios

I use unanimous voting rule to combine subgroups of models and construct three ensemble portfolios. After training all models, I separate them into linear models and neural networks

[7]Validation should be combined with domain knowledge whenever available.

[8]Stocks whose market capitalization as of the end of month t-1 falls below the monthly NYSE 10% breakpoint are excluded from model estimation and prediction.

and construct three ensemble portfolios by combining models within each subgroup and across the two subgroups. Combination is implemented using unanimous voting rule. For each month, I allow models in each subgroup to cast votes on the quintile assignment of stocks for the next month. I then construct ensemble portfolios by buying the unanimously predicted top quintile and selling the unanimously predicted bottom quintile. The spread of return between the top and bottom quintile represents the return to the associated ensemble portfolio. In the rare cases where no unanimous agreement exists during a given month, I simply stop trading and assume that the return for the corresponding leg(s) is zero. The three ensemble portfolios are constructed by combining linear models, 5 neural networks, and all models in our study separately and I name them linear ensemble, network ensemble, and total ensemble in the analysis below.

3. Models

This sections describes the family of machine learning models used to generate predictions. I provide a brief introduction to each model and its motivation, algorithm, and parameter tuning, if applicable.

3.1. Ordinary Linear Regression

The most commonly used model in empirical finance is Ordinary Linear Regression. It serves as a good benchmark for more sophisticated models.

Linear Regression tries to find the β in equation $Y = X\beta + \epsilon$ that minimize mean squared prediction errors (MSE). It assumes that the linear regression function $E(Y|X)$ is a reasonable approximation of the underlying relation between the target and predictors. For linear regression to be valid, ϵ and Y have to be uncorrelated. As long as this condition holds, the Gauss-Markov Theorem asserts that the $\hat{\beta}$ estimated from Ordinary Least Square has the smallest variance among all linear unbiased estimates, a.k.a. Best Linear Unbiased Estimator or BLUE. Although linear regression allows easy interpretation, it imposes linearity on the underlying relation which limits its predictive power if the true underlying function is other than linear. In addition, OLS typically incurs high variance in predictions when the number of predictors increases because it places no control over the norm of parameters. The generic solution to OLS can be derived through first-order condition on the MSE and its derivation can be found in a classical textbook.

3.2. Principal Component Regression and Partial Least Square

As the number of predictors increases, the variance of prediction from linear regression increases dramatically. Since each estimated parameter contributes randomness into the prediction, noise accumulates and eventually dominates the true relation between predictors and the target. One method to counter the problem is to reduce the dimension of predictors by extracting latent variables embedded in the predictors. The hope is to represent the information in the original predictors with only a few important hidden components and reduce the dimension of predictors. Two methods are commonly used for this purpose: principal component regression and partial least square regression.

3.2.1. Principal Component Regression

Principal component regression (PCR) is merely a regression of the target variable on the hidden components from principal component analysis (PCA). The first principal components are found to be the linear combination of the original predictors that has the highest variance among all such linear combinations. Subsequently, the i^{th} principal components are found in the same manner with the additional requirement that they must be orthogonal to all previously found principal components. Mathematically, the principal components of a data matrix X are given by the eigenvectors of $X^T X$. This is intuitively appealing because $X^T X$ is the variance-covariance matrix of X and its eigenvectors point to the direction in which the variables in X vary the most[9]. In optimization term, the m^{th} component direction v_m solves equation 5.

$$\begin{aligned} \max_{\alpha} \quad & Var(X\alpha) \\ \text{subject to} \quad & ||\alpha|| = 1 \\ & \alpha^T \mathbf{S} v_l = 0 \\ & l = 1, ..., m-1 \end{aligned} \tag{5}$$

where $\mathbf{S}$ is the covariance matrix of the data.

In each month, I determine the optimal number of principal components retained using five-fold cross validation. For implementation, I follow the algorithm called "Non-linear Iterative Partial Least Square" , aka "NIPALS", which repeatedly calculates every component, checks their convergence, and stops until certain low tolerance level is achieved or maximum

[9] Here X must be demeaned before performing PCR. The eigenvector associated with the largest eigenvalue points to the direction in which the linear combination has the largest variance.

iteration is reached[10]. See Algorithm 1 in Appendix for details[11].

3.2.2. Partial Least Squares Regression

Like principal component regression (PCR), partial least squares (PLS) also construct linear combinations of the original predictors to represent the data. However, PLS improves upon PCR by incorporating the correlation between predictors and the target into the calculation of weights in linear combination. That is, PLS finds directions that have both high variance among predictors and high correlation with the target, while PCR only focus on explaining the variance among predictors. In optimization form, the m^{th} PLS component solves equation 6 whose objective function clearly demonstrates the consideration of correlations between predictors and the target.

$$
\begin{aligned}
\max_{\alpha} \quad & Corr^2(y, X\alpha)Var(X\alpha) \\
\text{subject to} \quad & ||\alpha|| = 1 \\
& \alpha^T \mathbf{S} v_l = 0 \\
& l = 1, ..., m-1
\end{aligned}
\tag{6}
$$

where $\mathbf{S}$ is the covariance matrix of the data.

As in PCR, I determine the optimal number of principal components retained using five-fold cross validation. I also follow the NIPALS algorithm as in PCR.

3.3. Penalized Linear Regression

3.3.1. Ridge Regression

Ridge regression shrinks the norm of the coefficient vector by imposing a penalty term on the loss function. Specifically, in our study the optimization algorithm finds $\hat{\beta}$ in equation 7.

$$
\hat{\beta}_{Ridge} = \arg\min_{\beta} \Big\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \Big\}, \tag{7}
$$

where β is the vector of regression coefficients, x_{ij} is the j^{th} predictor of the i^{th} training sample, N is training sample size, and $\lambda \geq 0$ is a penalty parameter that controls the amount of shrinkage. Larger λ imposes more difficulty to the minimization of loss function and thus shrink the β more strongly toward zero. Using matrix notation, the Ridge regression solutions are given in equation 8.

[10] For our study, the tolerance level is 1e-06 and maximum iteration is 500, whichever comes first.

[11] NIPALS applies to partial least squares as well with only minor modification.

$$\hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T y, \tag{8}$$

where X is data matrix and y is the target variable. The method of penalizing loss function by the squared norm of β is called L^2 regularization and is also used in neural network models, which will be discussed in later sections.

An important insight can be gathered by performing singular value decomposition of the centered data matrix X, $X = UDV^T$, where U spans the columns of X, D is the diagonal matrix of singular values, and V spans the row space. Plugging $X = UDV^T$ into equation 8, left-multiply by X, and simplify to get equation 9.

$$X\hat{\beta}^{Ridge} = \sum_{j=1}^{p} u_j \frac{d_j^2}{d_j^2 + \lambda} u_j^T y, \tag{9}$$

where u_j is the j^{th} column of U, λ is the penalty parameter in equation 7, and j is the singular value of X.

Equation 7 shows that Ridge regression proceeds in two steps. First, it computes the coordinates of y relative to the orthogonal basis U. Then it shrinks these coordinates by multiplying them with $\frac{d_j^2}{d_j^2+\lambda}$. Note that the directions along which the coordinates exhibit smaller variance receive larger shrinking[12]. Intuitively, since data does not vary much along such directions, their effects are more difficult to estimate accurately. As a result, Ridge regression assigns a lower weights[13]. The only tuning parameter λ is found by ten-fold cross-validation.

3.3.2. Lasso Regression

Lasso regression follows a similar spirit as Ridge regression in that Lasso also imposes a penalty term to the loss function that is a function of the β vector. However, Lasso calculates the L^1 norm while Ridge calculates L^2 norm (See equation 10). This modification has two major implications. First, the absolute value function precludes a close-formed solution as in Ridge regression. Thus the Lasso estimates have to be solved numerically. Second, Lasso regression tends to shrink some β_i to zero and thus Lasso can be viewed as a variable selection method.

[12]Recall that d_j^2 is the eigenvalue of $X^T X$ and eigenvector associated with larger d_j^2 points to the direction in which data varies the most.

[13]See "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" Chapter 3 for more detailed discussions.

$$\hat{\beta}_{Lasso} = \arg\min_{\beta} \Big\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \Big\}, \tag{10}$$

where all variables are defined similarly as in equation 7 and λ is found by ten-fold cross-validation.

3.3.3. Elastic Net

Ridge and Lasso regressions can be viewed as two extremes. Ridge regression tends to retain all variables while Lasso performs variable selection. Zou and Hastie (2005) proposed a compromise between the two methods, namely Elastic net, which performs some variable selection and in the meanwhile shrinks the coefficients of correlated variables. This is illustrated in equation 11.

$$\hat{\beta}_{Elastic} = \arg\min_{\beta} \Big\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} (\alpha\beta_j^2 + (1-\alpha)|\beta_j|) \Big\}, \tag{11}$$

where all variables are defined similarly as in equation 7 and 10 and λ and α are found by ten-fold cross-validation as well.

3.4. Neural Network

Neural network is by far the most powerful tool in machine learning toolkit. Its development was partly motivated by the failure of traditional algorithms to generalize well to artificial intelligence tasks, i.e., speech and computer vision recognition. As its implementation becomes more user-friendly, neural network has shown great potential in tackling problems in finance, i.e., Gu, Kelly, Xiu (2018). In the following subsections, I go through the building blocks of the neural network and explain our design in details.

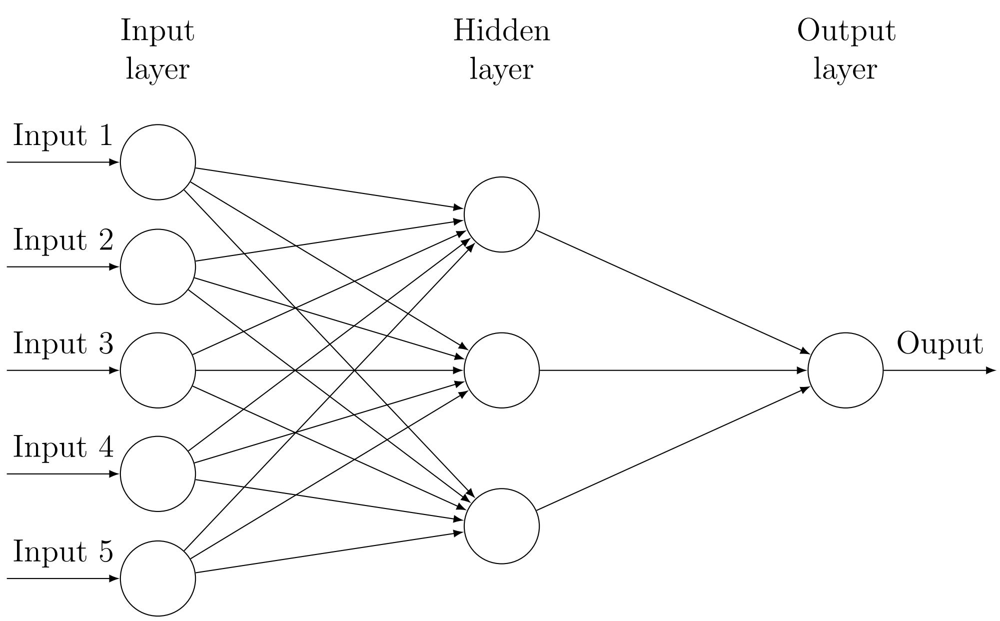
3.4.1. Universal Approximation Property

A key property that makes neural network the pivot of machine learning is the universal approximation property. The universal approximation theorem (Hornik (1989), Cybenko (1989)) states that a feed-forward network with a linear output layer and at least one hidden layer with any squashing activation function (that maps a larger domain into a smaller range, i.e.,sigmoid function) can approximate any Borel-measurable function from one finite-dimensional space to another with any desired nonzero amount of error, provided that the network is given sufficient hidden units. For our mission of predicting stock returns with lagged return, moving average price/trading volume signals, or firm fundamentals, I can safely assume that all of our predictors are bounded during a specific time period and universal approximation theorem applies. Take lagged return for example, universal approximation theorem implies that any continuous function $f : [-1,1]^n \to [-1,1]$ may be approximated by a neural network with large enough capacity, where n is the number of predictors.[14]. The other two settings can be similarly argued.

3.4.2. General Architecture of the Network and Information Flow

Although universal approximation theorem guarantees the existence of a large enough network [15] that approximates any Borel-measurable function with any degree of accuracy, the number of hidden nodes in the single-layer network may be too large to be estimated Barron (1993). In practice, the architecture of a network is more often determined by experimentation. Figure 1 shows a general structure of neural network with five input nodes (predictors), one hidden layer with three hidden nodes (latent features), and one output node (predicted stock return) [16].

Fig. 1. General Architecture of a Neural Network



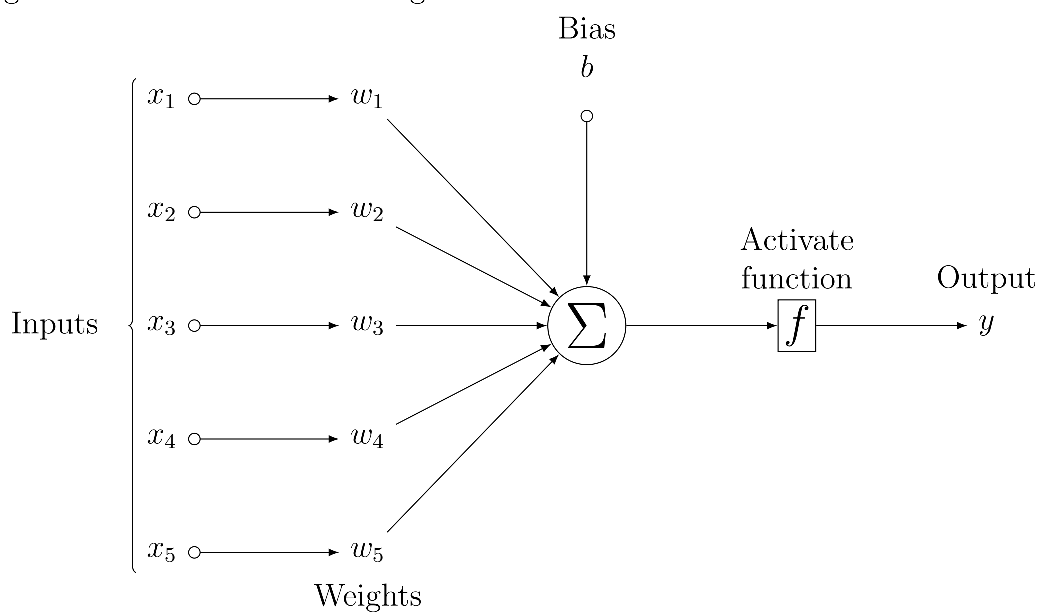
[14]Recall that continuous function on a closed bounded subset of $\mathbb{R}^n$, i.e., compact, is Borel-measurable and $[-1,1]^n$ and $[-1,1]$ are obviously finite dimensional $\forall n < \infty$. Thus the conditions of universal approximation theorem are satisfied.

[15]In the sense of sufficient number of hidden nodes in a single hidden layer

[16]Figure 1 and figure 2 are adapted from the open source codes on https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network.

Zooming in to a specific node in the structure in Figure 1, I obtain a more microscopic view of the information flow within the structure for a single hidden node. Figure 2 shows the mechanics and I briefly introduce the process. First, I take a single observation with five inputs (predictors) and then apply a weighted average on the inputs using weights $w1, w2, w3, w4, w5$ (to be estimated from training) to obtain the summation $\sum$. Second, add a bias parameter b to the $\sum$. So far the procedure resembles regression analysis. Finally, I apply an activation function to the resulting value from previous step and pass the output to following layer, i.e, another hidden layer or output layer. The key point is that all the weights and biases are parameters I try to estimate and all other parts, i.e., hidden layers and nodes, activation functions, penalty terms, regularizations (not shown in the graph) are pre-chosen. I will explain these issues and our choices momentarily. Before that, a more fundamental task is to initialize weights and biases and then to update them, which leads us to weight initialization, forward and backward propagation, and stochastic gradient descent.

Fig. 2. Information Flow Through One Hidden Node of a Feed-forward Network



3.4.3. Weight Initialization

Most optimization algorithms used in training neural networks are iterative in nature and require initial point to start. Initial point has three vital implications. First, most optimization algorithm for neural networks are strongly affected by initial points, i.e., it affects whether iteration converges at all. Second, even if learning does converge, initial

[16]From now on, I omit bias for brevity.

point affects the convergence speed and training time. Third, although different initial points can sometimes converge to points of comparable cost value, these points may have drastically different generalization power. Although these issues are conceptually important, practical strategies are mostly heuristic because neural network optimization itself is not well understood yet.

Glorot, Bengio (2010) shows that for information to flow stably through the network the variance of the outputs of each layer should equal to the variance of its inputs, which is impossible unless $n_{inputs} = n_{outputs}$. As a compromise, I use a popular distribution named "He normal" (He, Zhang, Ren, Sun (2015)) to initialize the weights. This modified normal distribution has been shown to work well with Rectifier Linear Unit (ReLU) activation function and its known variants which I use. Specifically, the initial weights are randomly sampled from the following normal distribution.

$$X \sim \mathcal{N}(0, \sqrt{2}\sqrt{\frac{2}{n_{inputs} + n_{outputs}}}) \tag{12}$$

, where n_{inputs} is the number of inputs nodes and $n_{outputs}$ for outputs. Note that this distribution is also intuitively appealing because as the number of inputs and outputs increases, the initial weights are shrunk to zero to dampen the impact of any individual predictor.

3.4.4. Forward- and Backward-Propagation

After the weights are initialized, I iteratively update them using information from the calculated cost value. The two processes of propagating information forward through the neural network and then using calculated errors to make backward adjustments to parameters are called forward- and backward-propagation respectively. Specifically, given the weights matrices and bias parameters of the model, forward-propagation processes the data and calculate the loss value [17]. Then backward-propagation will calculate the gradients on the activation functions in each layer, starting from the output layer and going backwards to the very first hidden layer. These gradients indicate in which direction the output of each hidden layer should move to reduce the forecast error. Finally, the gradients on weights and biases can be found through chain rule of calculus. (see Algorithm 3 in Appendix for more details)

In practice, the gradients are typically calculated in mini-batches of 32 to 512 randomly selected training samples at a time [18], i.e., LeCun, Bottou, Muller (2012). In this case,

[17] The calculated cost value must be added to a regularizer $\Omega(\omega)$ to obtain the total loss

[18] A batch size of 1 is totally legitimate. In that case it is a pure stochastic gradient descent algorithm and it probably takes a long time to train the model

the gradient from each mini-batch is simply the average gradient of all samples from the same mini-batch. The batch size is another hyper-parameter I can play with and it also has significant impact on model performance. While using larger batch size may yield more accurate estimate of the gradient due to the law of large number, it also requires more memory on your machine when data is large. On the other hand, for some structure of the cost function more accurate locally estimated gradient may be undesirable when your model is following a trajectory to get out of a local minima or saddle point while its gradient is telling it to move downhill [19]. In our study, I balance model performance against training time and choose the batch size of 32.

3.4.5. Stochastic Gradient Descent

The most commonly used optimization algorithm for neural network is stochastic gradient decent (SGD). The central equation for iterative updating of weights is shown in equation 13 (see Algorithm 2 in Appendix for more details).

$$w_t \leftarrow w_{t-1} - \epsilon_{t-1}\hat{\mathbf{g}}_{t-1}, \tag{13}$$

where w_t is the weight matrix after the t^{th} iteration, ϵ_{t-1} is the learning rate after the $(t-1)^{th}$ iteration, and $\hat{\mathbf{g}}_{t-1}$ is the estimated gradient from a mini-batch.

Equation 13 shows that SGD relies on the estimated gradient at a sample point[20] as guidance regarding to which direction the weights should move toward. One drawback of SGD is that convergence may be slow and therefore many modified versions exist to accelerate training. In our study I employ an adaptive learning rate algorithm called ADAM Kingma, Ba (2017). See Algorithm 4 in Appendix for details. ADAM can be roughly viewed as a combination of two adaptive learning rate optimization schemes: (1) momentum, and (2) re-scaling.

Intuitively, momentum represents accumulation of previous estimates of gradients and it will be added to the updating term on weights. With momentum the size of updates will depend on the sum of norms of previous gradient estimates and how aligned their signs are. For example, if the training process experiences a long sequence of either positive or negative gradient estimates then the momentum term will accumulate and the size of updates will be larger over time, shortening the training time. If the signs oscillate and previous gradients cancel out then momentum will be close to zero and the size of updates would be small, elongating the training time.

[19] In this case, an "inaccurate" estimate of gradient may lead us to the right path. Thanks to an anonymous discussant on Stackoveflow to point this out.

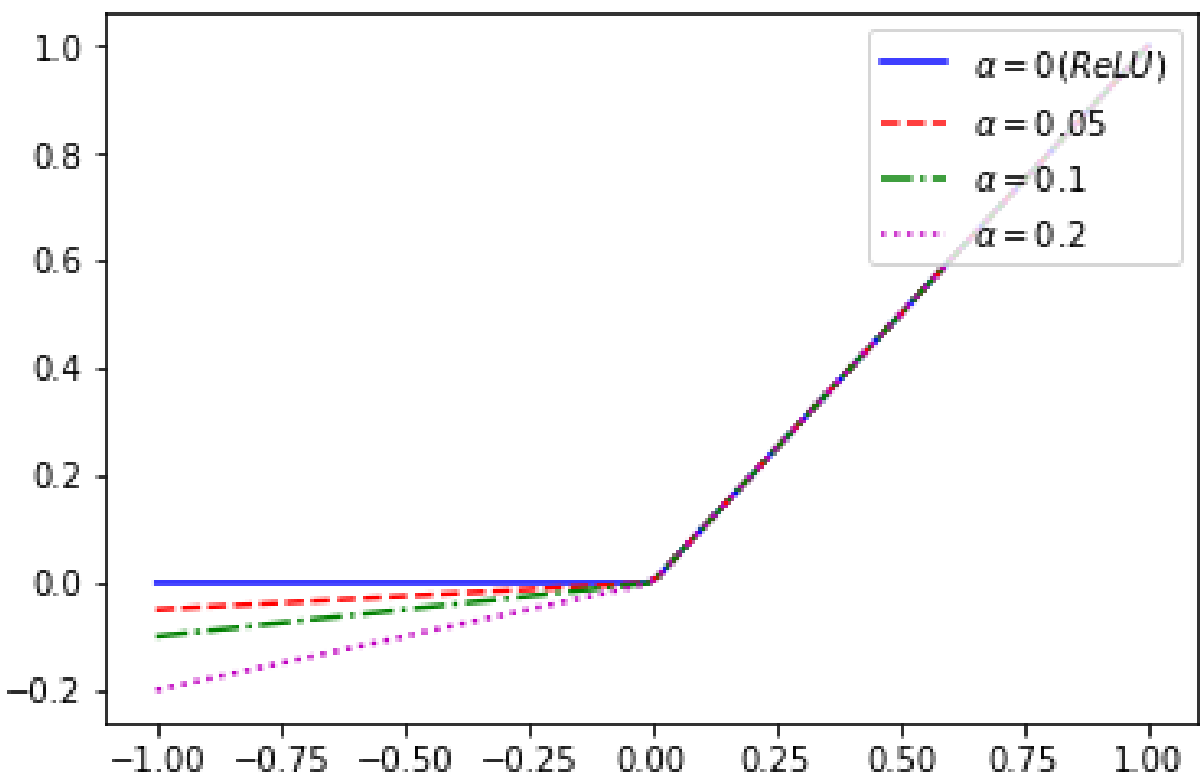
[20] Or a mini-batch of training samples.

3.4.6. Hidden Layers and Activation Function

Now that I are clear on how to initialize weights and biases and how to iteratively update them, I come back to our design choices, i.e., number of hidden layers and nodes and choice of activation functions. As demonstrate in Figure 1, hidden layers and nodes act like latent variables between predictors and target variable and much of the flexibility of neural network model comes from its layered design and non-linear activation function. Due to the lack of theoretical guidance, architectures are usually designed by experiments. Following previous studies, Gu, Kelly, Xiu (2018) and Masters (1993), I explore a series of designs starting from the shallowest network with a single hidden layer and 32 hidden nodes to a 5-layer network with [32, 16, 8, 4, 2] hidden nodes respectively. Although these pyramidal designs seem arbitrary, casual experiments suggest that deviation from them does generate inferior predictive performance most of the time.

As for activation function, I choose leaky rectified linear unit (Leaky ReLU) function for hidden layers and linear function for output layer. Leaky ReLU is a variant of the famous rectifier linear unit (ReLU) activation function used in many recent studies. Comparison of the two functions are in Figure 3.

Fig. 3. ReLU and Leaky ReLU with varying parameters



The functional forms of ReLU and Leaky ReLU function are given below.

$$ReLU(x) = \begin{cases} 0, & \text{if } x \leq 0. \\ x, & \text{otherwise.} \end{cases}$$

$$LeakyReLU(x) = \begin{cases} \alpha x, & \text{if } x \leq 0. \\ x, & \text{otherwise.} \end{cases}$$

where α is a hyper-parameter that can be tuned [21].

The major difference between ReLU and Leaky ReLU lies on the negative part. One drawback of the original ReLU function is that if during training a hidden node's weights get updated such that the weighted average of its inputs is negative, it will output zeros. That is, that node is "dead". To make things worse, future updates to that node is unlikely to bring it back to life because the gradient of ReLU function is zero on the negative part of its domain [22]. To alleviate this issue, Leaky ReLU assigns a small positive slope to the original ReLU on the negative domain and hopes that even if a node dies in one round of training future updates will not always be zero and thus may bring it back to life.

As for output layer, linear activation function is a common choice for regression. Since our target variable is stock return which lies in [-1,1], many other functions with a range between -1 and 1 seem to be legitimate options as well [23]. However, such functions typically saturates for extreme values on their range and thus may cause trouble for optimization algorithms [24].

3.4.7. Regularizations

With tens of thousands of parameters estimated, over-fitting is a big concern. Conceptually, over-fitting refers to a situation where the model family not only includes the true data generating process but also many other generating processes. That is, variance rather than bias dominates the estimation error (Goodfellow et al (2014)). This issue is especially severe

[21] Typical choices of α is 0.01. In our study, I use $\alpha = 0.1$. The choice of α can potentially affect model performance, a fact I observe from experiments. The consideration is to alleviate the dying node problem while preserving the non-linearity of the original ReLU function. No hard rules exist here.

[22] Recall that during backward propagation I use calculus chain rule to calculate the gradients on weights and biases. If the gradient of activation function on its input weights is zero, the whole gradient becomes zero and the weights don't get updated. One direct consequence of dying node is that the neural network model starts to generate identical predictions, which is just the bias parameter.

[23] For example, tangent hyperbolic function.

[24] That is, such activation functions level off for extreme input values and thus their gradients are close to zero, providing little information for updating weights.

with neural network because neural network typically has very high capacity[25]. To address over-fitting, I employ seven regularization schemes: (1) L^1 and L^2 parameter norm penalties, (2) adaptive learning rate shrinkage through stochastic gradient descent, (3) learning rate shrinkage through learning rate scheduler, (4) batch normalization, (5) random dropout, (6) early stopping, and (7) ensemble method. The application of regularization is more of an art than science. Although the seven techniques are very powerful, they impose danger as well. Too much regularization may impose unnecessary constraints on the optimization algorithm and prevent it from heading for a lower cost point. L^1 and L^2 parameter norm penalties are essentially the same as the ones used in Lasso and Ridge regressions respectively and thus are spared from further discussion.

I employ two learning rate shrinking methods, one is embedded in the ADAM algorithm as discussed in 3.4.5 and the other is an explicit scheduler. Although ADAM adaptively shrinks learning rate there is little guarantee that the learning rate will be sufficiently shrunk before the early stopping criterion is met. Large learning rate can cause the weights to oscillate back and forth around a local minima and make it difficult to converge. This issue is especially severe during later phase of training when fine tuning is needed. By imposing an explicit learning rate scheduler, I put a shrinking series of upper limits on the learning rates computed by ADAM at a given step of training and thus make sure that the updating term $\epsilon_{t-1}\hat{\mathbf{g}}_{t-1}$ in equation 13 will converge to zero and thus the training will converge[26].

Batch normalization (Ioffe, Szegedy (2015)) controls the variability of predictors across different layers of the network and across different datasets. The outputs of a mini-batch of data from each layer constitute a "batch". The outputs from preceding layers are cross-sectionally standardized to have zero mean and unity variance before being fed to the next layer. This is done on training and validation samples similarly and on testing sample but in a different manner. To make prediction, I treat the entire testing sample as one batch and feed it to the model in its entirety.[27]. That is, the mean and variance used in normalization is calculated using all the data points in a testing sample.

Dropout (Srivastava et al (2014)) randomly ignores a certain percentage of hidden units during training. All the inputs and outputs connected with the ignored units are omitted from updates as well. Since I use a minibatch-based learning algorithm, units to be ignored are chosen at the beginning of each batch operation and they are chosen independently for

[25]Meaning they include many data generating processes besides the true one, if at all.

[26]As discussed in Deep Learning Chapter 8 by Ian Goodfellow and Yoshua Bengio, gradient descent often does not arrive at any critical point, meaning that the norm of the gradient does not converge to zero at all. As a result, I must shrink learning rate instead to achieve convergence.

[27]As discussed earlier, during training and validating I feed a mini-batch of 32 data points to the model at one time. Batch normalization is performed on this mini-batch of 32 examples during training and validating.

each hidden layer. Conceptually, random dropout is similar to estimating an ensemble of sub-networks of the master network with shared parameters. Predictions are made from accumulating votes from all the sub-networks. Like bagging, such ensemble approach has the advantage of being robust to errors. In addition, since dropout is applied on hidden units rather than the raw inputs, it can be viewed as an intentional omission of some information content rather than the original variable from the learning process. This omission forces the algorithm to complete the task via other useful information in the predictors and to extract as much information from the predictors as possible. In our study, the probability of a hidden node being ignores is set at 10%. No dropout is performed on input layer.

Early stopping is a direct control of model performance on the validation set. Typically, I observe that validation loss first decreases with more updates to the weights and then increases as over-fitting creeps in. One way to choose the optimal hyper-parameters is to train the model for a large number of epochs and then go back along the training history to choose the parameters yielding the lowest validation error. However, such method is time consuming and it is difficult to determine how long the entire training process needs to be to include the true optimal model. Instead, early stopping halts the training as soon as the validation error fails to decrease for a certain number of training rounds. At the end of the training process, the most recently recorded model, rather than the global optimal model trained so far, is returned. In our study, I stop the training if the validation error fails to decrease within 10 epochs. See Algorithm 6 in Appendix for details.

Ensemble learning estimates the same neural network architecture multiple times and averages their predictions. Since each estimation assumes a random matrix of initial weights, it incurs independent estimation error, which supposedly will average away with large number of estimations. For each month, I estimate each neural network 100 times using randomly initialized weight matrices and take average prediction as our final prediction.

4. Data

Market return is collected from monthly CRSP database. Data starts in January 1978 (1978:01) and ends in December 2017 (2017:12), totalling 40 years (480 months). I include all domestic common stocks listed on the NYSE, AMEX, and Nasdaq Stock Exchanges and exclude securities that do not have CRSP share code of 10 or 11. Returns are adjusted by de-listing returns and and stock prices are adjusted by stock dividends and splits. For the 120 lagged returns setting, I require a firm-month to have non-missing past returns for at least 72 months out of 120 months to be included in our sample. For the 60 price (trading volume) moving average signal setting, I require 60% of the previous price (trading volume) data to be non-missing for each calculation of moving average signals. For the 79 fundamental characteristics setting, I follow the same sample selection rule as in Green, Hand, Zhang (2017) and use the SAS code provided on the authors' website to generate the data. For each set of predictors, I winsorize monthly data at 1% and 99% and normalize each variable to have mean 0 and variance 1. After normalization, I fill missing values with zeros. I obtain the Treasury-bill rate as the proxy for risk-free rate from Fama-French Factor database on WRDS.

5. Results

5.1. Average Portfolio Return During 1978:01 - 2017:12

Table 1 panel A reports summary statistics of monthly spread portfolio returns using 120 lagged returns as predictors. Portfolios are constructed on predicted returns from 11 models and their ensembles, including average excess return, t-statistic under zero mean, volatility, Sharpe ratio, skewness, kurtosis, proportion of positive return, minimum return, and maximum drawdown. Over the sample period from 1978:01 to 2017:12, the benchmark OLS model generates an average return of 0.78% (t=5.13) per month while the best performer, Partial Least Square generates a 1.06% (t=7.16). t-statistics are calculated using White Heteroskedasticity robust standard error. Specifically, I draw the following observations from panel A.

First, the 11 portfolios yield highly significant returns. In particular, all mean excess returns are statistically significant at 1% level. The magnitudes of mean excess return fall between 0.78% (Linear) and 1.74% (Ensemble Total) with network models generally outperforming linear ones. Volatility is in the ballpark of 3%-4% and Sharpe ratio varies widely following similar pattern as mean excess returns and volatility. All Skewness and Kurtosis are positive, suggesting that portfolios tend to experience extremely high return in some month. Minimum monthly return fluctuates between -0.18 (PCR) and -0.36(Ensemble Total) and maximum drawdown also varies widely between -0.31 (PCR) and -0.54 (Ensemble Total). Overall, I conclude that model specifications significantly affect their predictive performance and the return-risk profile of the portfolios constructed on them.

Second, different families of models exhibit starkly diverging performance while models within the same family perform similarly. Specifically, linear models (OLS, PCR, PLS, Ridge, Lasso, Elastic net) produce comparable portfolio performance in all statistics reported. In particular, the mean excess returns of all four models fall between 0.65% and 0.68% and are statistically significant at 1% level[28]. Volatility, Sharpe ratio, and other statistics are also similar. Neural network models perform better with all specifications producing a monthly return of more than 1% and t-statistics greater than 3. Even after accounting for the high volatility, neural networks still generate the highest Sharpe ratio among all models (between 0.18 and 0.21).

The three ensemble models significantly outperform their respective families. Specifically, Linear ensemble yields a return of 1.24% which represents a 17% increase to the best single-model portfolio return of 1.06%. Both Skewness (1.51) and Kurtosis (19.73) are slightly higher than those of its component models, suggesting that unanimous voting rule increases the probability of returning extremely positive outcomes. Other statistics show comparable results. Network ensemble significantly improves upon single network models. It generates a return of 1.59% while the best-performing network model (NN2L) only generates 1.03%, a 54% boost of return. As a compensation, network ensemble also exhibits higher variance (6.34% vs 3.95%) and therefore the ensemble Sharpe ratio decreases from 0.26 to 0.25. Network ensemble has about average Skewness and Kurtosis around its component models. Finally, combining both linear and networks further boosts portfolio performance to 1.74%. However, the Sharpe ratio does not improve and remains at 0.25, suggesting that model combination brings little benefit to the trade-off between risk and return.

Table 1 panel B and Panel C reports summary statistics of monthly spread portfolio returns using 60 price and 60 trading volume moving average signals and 79 fundamental firm characteristics, respectively. For moving average signals, the return pattern is similar to the lagged return setting only with minor differences. Like with lagged returns, networks generally outperform linear models and ensemble models outperform their component models. However, one difference is that for moving average signal setting ensemble models slightly improves Sharpe ratio while this is not the case for lagged returns. On the other hand, when 79 fundamental variables are used for predictors, the excess returns of linear models versus network models are indistinguishable and ensemble models again fail to improve the Sharpe ratio beyond their component models.

Observed patterns of the three panels suggest the following conclusions. First, linear models have limited capacity in approximating the true data generating process of stock returns when market data are used for predictors. The fact that adding a penalty on mean squared error loss, as Ridge, Lasso and Elastic net do, does not significantly improve model performance partially spares high variance from being the culprit for poor performance of linear models and invites us to explore a broader family of more flexible models. On the other hand, models with high capacity and strong regularization, i.e., neural networks, take

[28]Except for Ridge regression, which has t=5.13 and significant at 5% level

the upper hand over linear models in such setting. However, networks seem to have limited utility in fundamental-based strategies, as demonstrated in panel C. Second, the superior performance of ensemble models provides evidence that unanimous voting rule is a booster of model performance and such performance boosting applies to both market-based data and firm fundamental characteristics. However, the effectiveness is more salient for lagged return setting (1.74 / 1.06 = 1.64) and moving average signal setting (1.89 / 1.06 = 1.78) than for firm fundamentals (1.91 / 1.37 = 1.39), consistent with the narrow spread of return between linear and network models in firm fundamental setting. Third, neural networks and ensemble method do not seem to bring significant benefit to Sharpe ratio compared to linear models. While slightly boosting Sharpe ratio in moving average setting, most networks and ensemble models in the other two settings either generate similar Sharpe ratio as linear models or sometimes inferior Sharpe ratios. This raises a potential caveat for using network models and ensemble methods. However, it is true that such increased volatility may be caused by upward fluctuation and may not represent a downward risk. Such possibilities are left for future investigations.

5.2. Average Portfolio Return During Sub-Periods

To explore whether our results are driven by any specific period, I present four ten-year-spaced sub-periods in Table 2 to Table 4. Table 2 to Table 4 presents results for (1) 120 lagged return, (2) moving average trading signals, and (3) fundamental variable, respectively. I focus on key observations rather than giving a rundown on each table.

First, consistent with Chordia, Subrahmanyam, Tong (2014) and McLean and Pontiff (2015), excess returns decay over time. For the first 30 years in our sample (1978:01 - 2007:12), most models generate highly significant excess returns on each of the three decade-long subperiods regardless of which set of predictors are used. A swerve occurred in 2008 and most excess returns have became insignificant at 10% level since then.

Second, individual network models generally outperform linear models and examination of the average spread return between linear models and network models does not reveal clear time trend. In the lagged return setting the average spread returns between linear and network models respectively are 0.1383 (1978:01 - 1987:12), -0.07 (1988:01 - 1997:12), 0.3417 (1998:01 - 2007:12), and -0.0267 (2008:01 - 2017:12). The corresponding spreads for moving average signals are 0.0443 (1978:01 - 1987:12), 0.2873 (1988:01 - 1997:12), 0.5157 (1998:01 - 2007:12), and -0.0713 (2008:01 - 2017:12) and for fundamental variables are -0.0683 (1978:01 - 1987:12), -0.0073 (1988:01 - 1997:12), 0.1297 (1998:01 - 2007:12), and 0.0440 (2008:01 - 2017:12). Therefore, network models outperform linear models by large margin in some decades but underperform only slightly in others. Overall, network models outperform and the relative performance of the two families of models does not appear to change monotonically over time.

Third, ensemble models generally outperform their component models and network ensembles outperform linear ensembles with one exception (2008:01 - 2017:12 lagged return setting). I examine the time trend of the ratio of network ensemble returns to linear ensemble returns over the four decades for three sets of predictors. For lagged return, the ratios are 1.45 (1978:01 - 1987:12), 1.08 (1988:01 - 1997:12) , 1.51 (1998:01 - 2007:12),, and 0.80 (2008:01 - 2017:12). So network ensemble outperforms linear ensemble except for the last decade and no clear time trend emerges since the third decade in our sample exhibits a spike of outperformance for network ensemble. For moving average signals, the ratios are 1.15 (1978:01 - 1987:12), 1.53 (1988:01 - 1997:12) , 1.45 (1998:01 - 2007:12),, and 1.36 (2008:01 - 2017:12). Network ensembles consistently beat linear ensembles with high margin. Although the margin has dropped slightly from 0.53 to 0.36, it remains economically significant. For fundamental firm characteristics, the ratios are 1.08 (1978:01 - 1987:12), 1.09 (1988:01 - 1997:12) , 1.10 (1998:01 - 2007:12),, and 1.00 (2008:01 - 2017:12). Although network ensembles still outperform linear ensembles, the margin is much smaller than those for lagged return or moving average signals, suggesting that model capacity may not be a critical issue for fundamental variables.

Another observation is that combining linear models with network models significantly boosts excess returns. I label such model as total ensemble and examine the ratio of its excess return over that of linear ensemble or network ensembles, whichever is higher. For lagged return, the ratios are 1.09 (1978:01 - 1987:12), 1.16 (1988:01 - 1997:12), 0.97 (1998:01 - 2007:12), and 1.16 (2008:01 - 2017:12). For moving average signals, the ratios are 1.14 (1978:01 - 1987:12), 1.10 (1988:01 - 1997:12), 1.37 (1998:01 - 2007:12), and 1.46 (2008:01 - 2017:12). For firm characteristics, the ratios are 1.05 (1978:01 - 1987:12), 1.12 (1988:01 - 1997:12), 1.06 (1998:01 - 2007:12), and 1.27 (2008:01 - 2017:12). Therefore, in terms of model combination, moving average signals seems to reap the largest benefit of model combination and such benefit has increased over past four decades. Our silver medal goes to fundamental characteristics which generate consistent but modest benefit. Lastly, lagged return setting generates positive incremental returns with rare exception, i.e., 1998:01-2007:12.

To visualize the return profiles over time, I report the 12 month-lag moving average monthly return of the three ensemble models and the cumulative portfolio value of one dollar initial investment in figure 6. Figure 4 through Figure 6 show graphs for the three sets of predictors respectively. In each figure, Panel (A) presents moving average return and Panel (B) presents cumulative values for the three sets of predictors with an initial investment of \$1 in January 1, 1978. Consistent with numerical analysis, network ensembles portfolio values outgrow linear ensembles regardless of which predictors are used. In addition, the combination of network and linear models significantly boost model performance especially when price and trading moving average signals are used for predictors. In terms of recent performance, I observe that portfolio performance starts to level off for lagged return and fundamental variables, while remaining strong for moving average trading signals. For fundamental variables, the portfolio value shows little sign of growth after 2009, while lagged return perform relatively better but with large dips in 2014 and 2015. In terms of absolute portfolio value, moving average trading signals also wins out with an ending portfolio value of approximately \$1750.

5.3. Risk-Adjusted Returns

I adjust the spread portfolio returns by existing risk factor model. By regressing portfolio return on factor returns in each model, I filter out the excess return explained by these risk factors and retain the regression intercepts as our estimate of risk-adjusted returns (alpha). Risk-adjustment is done for spread portfolio, long-leg portfolio, and short-leg portfolio. Table 5 to Table 7 reports tne alphas (%) over the full sample period: 1978:01-2017:12 for the three sets of predictors respectively.

Table 5 reports the results for lagged return. Raw excess returns are adjusted by four factor models: CAPM, Fama and French(Fama, French (1993)) three-factor model, Carhart (Carhart (1997)) four-factor model, and Fama and French (Fama, French (2015)) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. First, risk-adjustment does not weaken either statistical or economical significance of our portfolio alpha. The most surprising result is that our results remain unaffected by Carhart 4 factors, which includes a momentum factor. This could be because our model includes month t-1 to month t-120 historical return as predictors while the momentum factor in Carhart model is based on prior return from month t-2 to month t-12. Second, I further decompose spread portfolio into long- and short-leg and reports the results in Panel B and Panel C. Decomposition shows stark contrast between long-leg and short-leg portfolio returns. Long-leg return is almost cut by half while short-leg return is actually enhanced (more negative and statistically significant) after risk-adjustment. This suggests that our long-leg portfolio does a mediocre (but still statistically significant) job at identifying winner stocks than common risk factors [29], while our short-leg portfolio is doing exceptionally better. Third, network models outperform linear models for both long- and short-leg. Take FF5-adjusted return

[29] For ensemble portfolios, at least 50% of raw excess return comes from their exposure to FF-5 factors. For individual models, common risk factors explain 60% to 70% of raw excess returns.

Fig. 4. Portfolio Returns (%) and Cumulative Portfolio Value 1978:01-2017:12 (Predictors: Lagged Returns)

120 lagged returns are used for predictors. Panel (A) reports the monthly spread portfolio return over the time period from 1978:01 to 2017:12. To enhance visualization, I smooth the monthly return with 12-month moving average values. Panel (B) reports the cumulative portfolio value over the time period from 1978:01 to 2017:12 after initial investment of one dollar. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models.

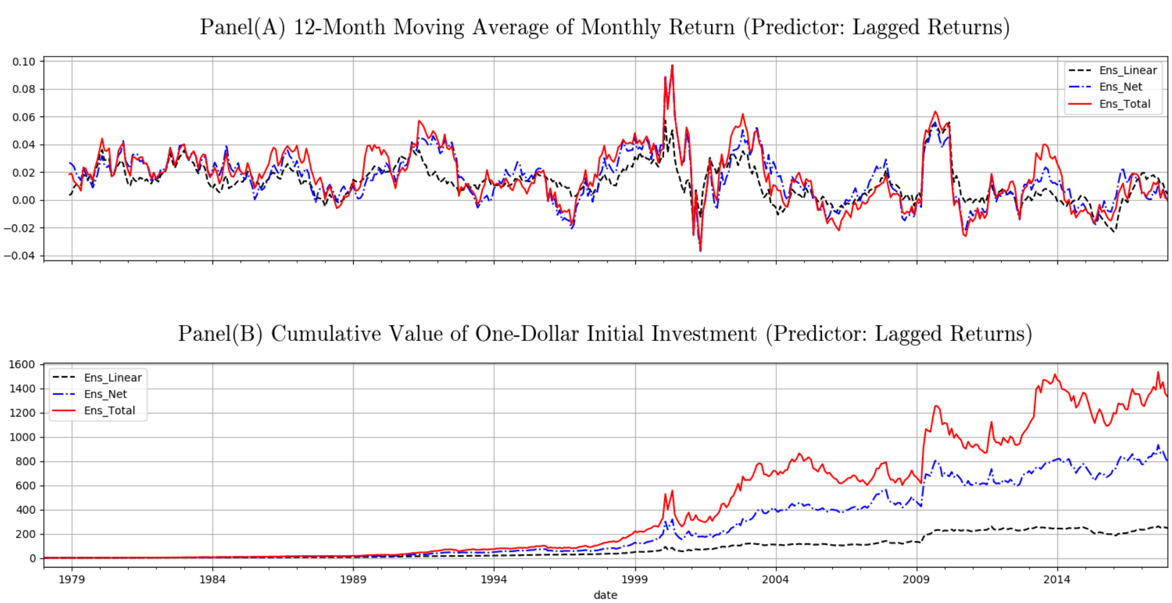


Fig. 5. Portfolio Returns (%) and Cumulative Portfolio Value 1978:01-2017:12 (Predictors: Moving Average Signals)

60 price and 60 trading volume moving average signals are used for predictors. Panel (A) reports the monthly spread portfolio return over the time period from 1978:01 to 2017:12. To enhance visualization, I smooth the monthly return with 12-month moving average values. Panel (B) reports the cumulative portfolio value over the time period from 1978:01 to 2017:12 after initial investment of one dollar. At the beginning of each month, I independently sort stocks into quintiles portfolios on predicted returns from each model and construct spread portfolios by buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models.

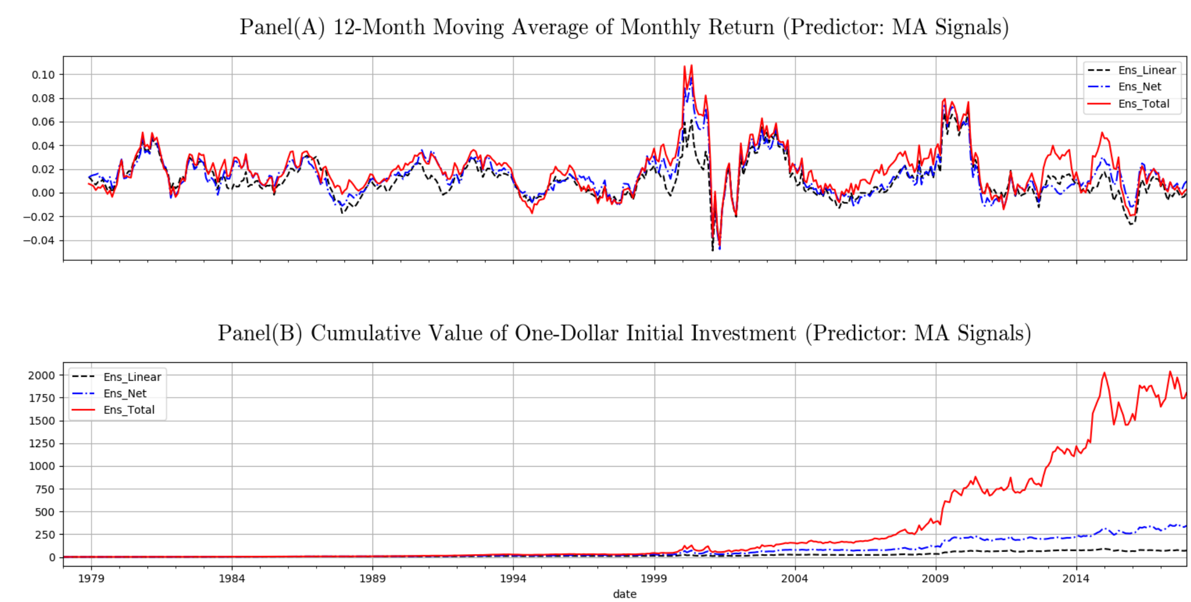
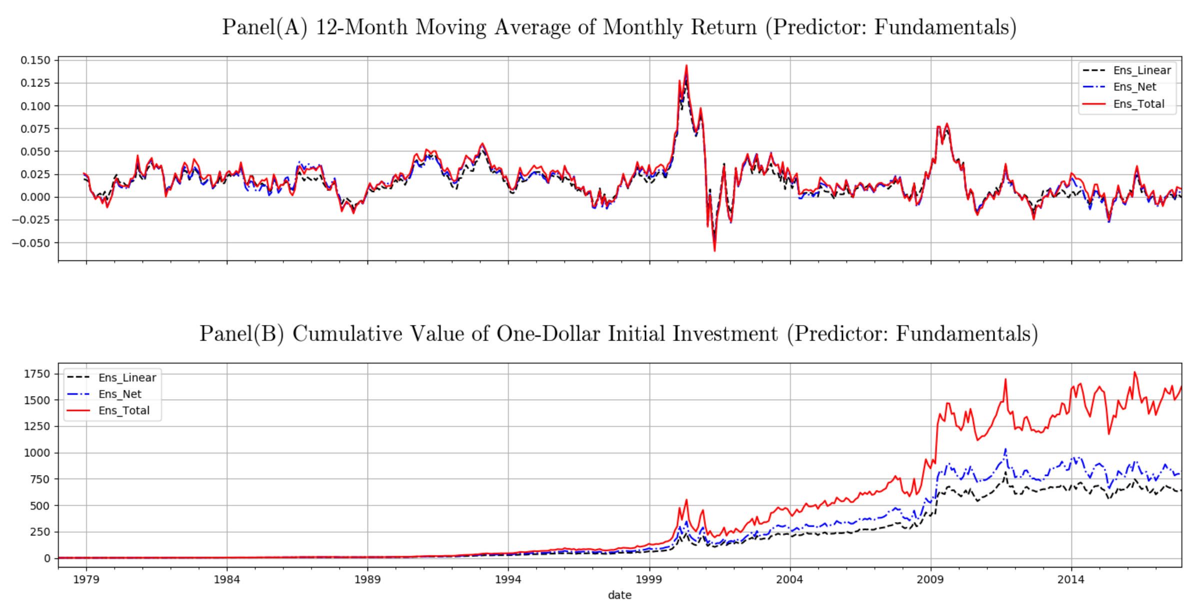


Fig. 6. Portfolio Returns (%) and Cumulative Portfolio Value 1978:01-2017:12 (Predictors: Firm Characteristics)

79 firm fundamentals are used for predictors. Panel (A) reports the monthly spread portfolio return over the time period from 1978:01 to 2017:12. To enhance visualization, I smooth the monthly return with 12-month moving average values. Panel (B) reports the cumulative portfolio value over the time period from 1978:01 to 2017:12 after initial investment of one dollar. At the beginning of each month, I independently sort stocks into quintiles portfolios on predicted returns from each model and construct spread portfolios by buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models.



for example, for long-leg, network ensemble (linear ensemble) model generates a return of 0.73% (0.54%); for short-leg, network ensemble (linear ensemble) model generates a return of -0.90% (-0.63%).

Table 6 reports the result for price and trading volume moving average signals. Results in table 6 resemble those in table 5 except for two points. First, individual models appear to be more sensitive to risk-adjustment than in lagged return setting. Statistical significance are dampened for all individual models except for NN1L and NN2L and economic significance is also reduced by a greater margin than in lagged return setting. On the other hand, the ensemble models are more robust to risk-adjustment than individual models and their risk-adjusted returns are only slightly reduced. Second, the source of increased sensitivity comes from the short-leg. Comparing table 6 panel (C) with the same panel in table 5 shows that the short-leg of moving average signals only marginally outperforms (if at all) common risk factors and statistical significance is poor. As a result, the risk-adjusted spread returns in the moving average signal setting and the lagged return setting are pulled closer than their respective raw returns.

Table 7 reports the result for fundamental firm characteristics. The results appear one-sided. The majority of risk-adjusted return for the fundamental portfolios come from their long-leg and the short-leg contributes only a small portion. For example, the risk-adjusted returns of neural networks long-leg portfolios are approximately 1%, while the corresponding short-leg returns are only -0.4%. In addition, all of the long-leg portfolio returns remain significant at 1% level, while none of the short-leg portfolio returns are significant at 1% level.

Overall, I find that the long-leg for fundamental characteristics performs the best among all long-leg portfolios and is least affected by risk-adjustment and the short-leg for lagged return performs the best among all short-leg portfolios. I also find that the lagged return setting relies on long- and short-leg portfolios somewhat symmetrically for excess return while the other two settings rely more heavily on their long-leg rather than short-leg for excess return. This phenomenon is especially salient for fundamental variable setting, which relies on its long-leg for both economic and statistical significance, and less so for moving average signal setting, whose short-leg contributes a decent portion of economic significance. Finally, I find that most ensemble portfolios survive risk-adjustment even when their component models do not. This reinforces our point that ensemble method helps decouple the correlation of individual models with common risk factors and improves predictive power beyond those risk factors.

Table 8 reports the exposure to common risk factors of ensemble portfolios. Consistent with our finding in table 5 panel (A), ensemble portfolios for lagged return mostly have insignificant exposure to common risk factors, with occasional marginally significant exposure to the investment factor (CMA). For moving average signals, I find consistent significant exposure to the investment factor and insignificant exposure to all other factors. For fundamental variables, market factor has significant negative coefficients and investment factor has significant positive coefficients. Overall, I find that fundamental ensembles could serve as a hedge portfolio for market risk while the two market-data based ensembles are quite neural to market risk.

5.4. Spanning Tests

Significant risk-adjusted return shows that the spread portfolios cannot be spanned by linear combinations of existing risk factor models. This section tests whether spread portfolios lie in the mean-variance frontier of risk factors. This idea was originally proposed by Huberman, Kandel (1987) and was further improved by Kan, Zhou (2012). I follow Kan, Zhou (2012) and provide six test statistics. They are Wald test under conditional homoskedasticity, Wald test under independent and identically distributed (IID) elliptical distribution, Wald test under conditional heteroskedasticity, Bekerart-Urias spanning test with errors-in-variable (EIV) adjustment, Bekerart-Urias spanning test without EIV adjustment, and DeSantis spanning test. p-values are found under asymptotic χ^2 distribution with 2 degrees of freedom and are in the parentheses.

Table 9 presents the results of ensemble portfolios for brevity. In A1 - A3, B1 - B3, and C1 - C3, I report test statistics for the three sets of predictors. The magnitude of statistics dwindles as more risk factors are added to the test, although all statistics are significant at 1% level. Our results show that ensemble portfolios not only yield the highest risk-adjusted returns as shown in previous tables but also add value to a well-diversified portfolio.

5.5. Regression Analysis of Predictive Power

To examine the predictive power of individual models, I estimate a series of monthly regressions of actual stock return on predicted stock return for each model and report the summary statistic of the time series of coefficients. The three panels in table 10 presents summary for lagged return, moving average trading signals, and firm characteristics respectively. From left to right, I report the monthly median of beta coefficients, the average t-statistics, standard deviation of betas, the minimum and maximum of betas, the proportion of positive betas, the proportion of significant betas at 1%, 5%, and 10% level conditional on their positivity, and finally the adjusted R-square.

First, I observe a clear increasing pattern of median beta coefficient as I migrate from top to bottom within each panel, suggesting that an increasing correlation of actual and predicted returns. In all three panels, OLS performs the worst, consistent with our presumption that OLS fails in modeling complex data processes. Specifically, median beta increases from 0.07 (OLS) to 0.61 (NN5L), almost nine-fold increase for the lagged return setting. Similar pattern emerges for moving average trading signal setting and firm characteristic setting, i.e., from 0.02 (OLS) to 0.36 (NN5L) and from 0.10 (OLS) to 0.73 (NN5L) respectively.

Second, among linear models, PCR, PLS, Lasso, and ENet are close in terms of median beta and other statistics. PCR and PLS tend to outperform in the lagged return setting; PCR, PLS, Lasso, and ENet tend to perform similarly in moving average signal setting; Lasso and ENet tend to outperform in firm fundamental setting. Overall, these four models tend to be close in performance and no dominant choice exists. Among neural network models, the median beta increases monotonically from 1-hidden layer to 5-hidden layers, suggesting that more complex network model improves the correlation between actual and predicted returns. Table 10 provides different insights compared to previous tables, i.e., table 1. For example, in table 1 I observe a inverted U-Shaped portfolio return for neural network models in the three panels, but the regression analysis shows a monotonically growing predictive power from NN1L to NN5L.

5.6. Pair-Wise Model Comparison

I further conduct a pair-wise Diebold-Mariano (Diebold, Mariano (2015)) test to compare each pair of models in their out-of-sample predictive accuracy. DM test compares the prediction error on a stock-level and assumes that the resulting statistic follows an asymptotic normal distribution. Table 11 reports the DM statistics with *, **, *** indicating 1%, 5%, 10% statistical significance.

A positive DM statistic suggests that the column model is superior to the row model, i.e., smaller average squared prediction error. In general, neural network models outperform other models as demonstrated by the significant DM statistics under most network models. In addition, performance increases monotonically with the number of hidden layers in network models, confirming our finding in table 10. Overall, our findings in table 10 and table 11 show that network models generate smaller predictive error and larger correlation between actual and predicted returns.

5.7. Size Effects

Smaller firms tend to have more mispricing and higher return. In this section, I check how much of our portfolio returns come from small firms. Specifically, before sorting stocks into quintiles I remove stocks in the bottom $p\%$ of market value at the end of the previous month. I pick $p\%$ to be 20%, 40%, 60%, and 80%. Table 12, table 13, and table 14 report risk-adjusted returns for our three sets of predictors respectively. In following analysis, I mainly focus on raw and risk-adjusted returns of individual models.

First, among the three sets of predictors moving average trading signals are most sensitive to size effect and fundamental characteristics are least sensitive. Specifically, in table 13 (moving average signals) from panel (A) to panel (D), raw and adjusted return of most portfolios decline in magnitude and become insignificant or only marginally significant at 10%. Only NN1L remains significant at 5% with adjusted return of 0.58. In contrast, although in table 14 (firm characteristics) the raw and adjusted return also decline in magnitude as small companies are weeded out, they remain economically more significant than those in table 13 panel (D) and more importantly t-statistic of all portfolios remains significant at 5%. The lagged return setting lies between the other two settings. In table 12 panel (D), although the magnitude of portfolio return for lagged return is comparable to those in table 13 panel (D), the t-statistics are highly significant at 1% or 5% level.

Second, the pattern I observe in table 1 remains unchanged. In particular, the gap of return between linear and network models is most significant for moving average signal setting and this holds for every panel in table 13. Thus this gap is not caused by any specific size quintile of the sample. On the other hand, the gap of return for fundamental variables is quite small. It is true that the highest return typically comes from a network model, but the overall performance of linear models is quite close to that of network models. In fact, such gap of return is hardly recognizable in many panels in table 14. Unlike before, the lagged return setting shows a somewhat inverted U-Shaped performance gap between linear and network models. I notice that linear and network portfolio returns are closer in panel (A) and (D) than in panel (B) and (C) in table 12, suggesting that the out-performance of network models that I observe in table 1 panel (A) likely comes from middle-sized companies.

Third, to further quantify the out-performance of network models over linear models, I calculate the ratio of adjusted network ensemble return over linear ensemble return. In the same order as presented in table 12 through table 14, the ratios are 1.40 (20% excluded), 1.56 (40% excluded), 1.79 (60% excluded), and 1.31 (80% excluded) for lagged return, 1.48 (20% excluded), 1.61 (40% excluded), 1.67 (60% excluded). and 1.48 (80% excluded) for moving average signals, and 1.14 (20% excluded), 1.17 (40% excluded), 1.06 (60% excluded), and 1.04 (80% excluded) for firm fundamentals. The corresponding ratios for the complete sample are 1.39, 1.45, and 1.11 for the three sets of predictors. Consistent with our previous finding, both lagged return and moving average signal exhibit inverted U-Shaped performance gap between network and linear models and fundamental portfolios are relatively insensitive to model choices. Overall, our finding suggests that model choices, i.e., linear versus neural network, matter more for middle-sized companies than small or large companies.

5.8. Impact of Investor Sentiment

Stambaugh et al (2012) shows that portfolio profitability is related to investor sentiment. In this section, I examine how much of our result is driven by investor sentiment. Following Huang, Zhang, Zhou, Zhu (2019), I proxy overall investor sentiment with Baker, Wurgler (2006) sentiment index and label a month as high sentiment if the Baker-Wurgler index over the previous month is above the median of the entire sample period (1978:01 - 2015:09) and low sentiment otherwise[30]. I calculate risk-adjusted return on high- and low-sentiment periods separately and report the results in table 15 through table 17.

First, for lagged return setting the out-performance of network models over linear models is more salient during low investor sentiment period than during high investor sentiment period. In particular, linear models (network models) generate risk-adjusted return of approximately 0.61% - 1.04% (0.70% - 1.01%) during high investor sentiment period, while during low investor sentiment period linear models (network models) generate approximately 0.73% - 0.98% (0.87% - 1.02%). During high investor sentiment period, linear ensemble generates 1.19% and network ensemble generates 1.53% while during low investor sentiment period, linear ensemble generates 1.17% and network ensemble generates 1.74%.

Second, for moving average trading signal setting the out-performance of network models over linear models is also more salient during low investor sentiment period than during high investor sentiment period. Risk-adjusted return becomes insignificant during high investor sentiment period. Although the return of network ensemble (0.71%) is still larger than that of linear ensemble (0.38%), neither of them are significant. On the other hand, during low investor sentiment period risk-adjusted returns of all models remain significant at 1% and network models significantly outperform linear models. In particular, three out of five network models generate more than 1.1% monthly return while most linear models generate a return less than 1%. Ensemble portfolios show similar pattern. Network ensemble generates a monthly return of 1.76% and linear ensemble generates a return of 1.38%. Overall, for moving average signal setting portfolio return concentrate on low sentiment period and network models outperform linear models only during low investor sentiment period.

Finally, fundamental portfolio return in table 17 does not appear to be affected by investor sentiment. During both periods, raw and risk-adjusted returns are significant at 1% for all models and the magnitudes of return are also close with those during low sentiment period

[30]Baker-Wurgler investor sentiment ends in 2015:09.

slight higher. The relative performance between linear and network models remain insensitive to investor sentiment either. Overall, table 17 is consistent with our previous finding that network models do not work better on fundamental variables than linear models.

5.9. *Impact of Business Cycle*

In this section, I explore the impact of business cycle on the performance of our models. Specifically, I examine hedge portfolio return during expansion and recession periods. Our sample months are labeled expansion or recession based on the business cycle definition on National Bureau of of Economic Research (NBER) website.

Both lagged return setting and moving average trading signal setting exhibit similar pattern. For lagged return setting, I find that common risk factors explain more of portfolio return during expansion than during recession periods. During expansion periods most of the portfolio return increases after being adjusted by Fama-French five factors, while during recession periods portfolio return typically declines after being risk-adjusted. In addition, I observe that neural network outperforms linear models by a larger margin during expansion periods than during recession periods. This is evidenced by the large (small) spread between linear and network ensemble returns during expansion (recession) periods. During expansion periods, linear ensemble yields a risk-adjusted return of 1.16% and network ensemble yields 1.75%, a spread of 0.59%. During recession periods, linear ensemble yields a risk-adjusted return of 1.94% and network ensemble yields 1.90%, a trivial spread. Similarly, for moving average trading signal setting, I again find that common risk factors explain more of portfolio return during expansion than during recession periods. During expansion periods most of the portfolio return increases after being adjusted by Fama-French five factors, while during recession periods portfolio return typically declines after being risk-adjusted. In addition, I observe that neural network again outperforms linear models by a larger margin during expansion periods than during recession periods. This is evidenced by the large (small) spread between linear and network ensemble returns during expansion (recession) periods. During expansion periods, linear ensemble yields a risk-adjusted return of 1.00% and network ensemble yields 1.44%, a spread of 0.44%. During recession periods, linear ensemble yields a risk-adjusted return of 1.09% and network ensemble yields 1.42%, neither of which is statistical significant. Overall, our evidence show that market-data based portfolios: (1) typically do a better job at exploiting mispricing during expansion periods than during recession periods, and (2) neural network outperforms linear models when mispricing is more prevalent, i.e., during expansion periods.

For firm fundamentals, I also find that portfolios raw return is more sensitive to risk-adjustment during recession than expansion. For example, return typically increases after risk-adjustment during expansion, while decreases and becomes insignificant during recession. In addition, I find some evidence of out-performance of neural network over linear models during expansion periods where linear ensemble returns 1.71% and network ensemble returns 1.93, a spread of 0.22%. Compared to the spread of 0.59% (lagged return) and 0.44% (moving average trading signal), the spread for firm fundamentals is almost cut by more than 50%. Furthermore, I observe that neural network under-performs linear models by 0.07% during recession periods which represents a non-trivial offset to the already-small out-performance during expansion periods. Summarizing the findings for the three settings, I conclude that in general: (1). neural network does better job at capturing mispricing in the market and this causes its out-performance during expansion periods when mispricing is believed to be more prevalent, (2). point (1) holds for all three sets of predictors but to different extents. From strongest to weakest, the ranking is lagged return >moving average signals >and firm fundamentals, (3). point (1) reverses slightly for lagged return (-0.04%) and firm fundamental (-0.07%) settings. While negligible for lagged return setting, such reverse wipes out 33% of the out-performance for firm fundamentals during expansion, narrowing the gap between neural network and linear models over the entire sample.

5.10. Combination of Predictors

I explore four combinations of the three sets of predictors. They are: (1) lagged return and moving average signals, (2) lagged return and firm characteristics, (3) moving average trading signals and firm characteristics, and (4) all three sets of predictors. Summary statistics are reported in Table 21 Panel (A) through (D).

First, I combine 120 lagged return with the 120 moving average signals and re-estimate the model. Summary statistics are reported in table 21 panel (A). Overall, combining the two sets of market data does not significantly improve spread return. Comparing with the results in table 1 panel (A) and panel (B), I fail to observe big improvement except for a few nuances. This suggests that 120 lagged returns and moving average trading signals may be capturing similar information in return or price momentum and the incremental information content barely offset the increased variance from their combination. Second, large improvement of portfolio return comes from the combination of fundamental variables with either lagged returns or moving average signals. I report such results in panel (B) and panel (C). Excess return of total ensemble portfolio jumps to 2.03% (2.19%) when firm fundamentals are combined with lagged returns (moving average signals). In addition, performance of linear and network models converges as fundamental variables are added to the predictors. Gu, Kelly, Xiu (2018) shows that neural network models significantly improve upon linear models. However, their data only includes a handful of momentum variables. Our finding suggests that such improvement likely comes from market data and not so much from fundamental variables.

I also combine predictors using an alternative method adopted in Han, He, Rapach, Zhou (2018). Specifically, I report the global minimum-variance portfolio, i.e., Basak, Jagannathan, Ma (2009), constructed from the three spread portfolios each of which is estimated using one of the three sets of predictors. First I calculate the spread portfolio return using each set of predictors independently for each model. Then I calculate the global minimum-variance portfolio weights for month t using spread portfolio return up until month t and use the weights to combine the three portfolios. Results are reported in table 22. I draw XXX conclusions. First, generally mean excess return and volatility are lower than those in table 21, consistent with the notion that expected return and volatility are positively correlated along the efficient frontier. Second, in terms of Sharpe ratio I do not observe significant improvement with only a few exceptions, i.e., PLS using prior return and technical predictors. Our results provide preliminary evidence that combining information of different sources generates limited gain.

6. Conclusion

In this paper, I conduct a comprehensive study of using machine learning tools to forecast the U.S. stock returns. I use three sets of predictors: the past history summarized by 120 lagged returns, the technical indicators measured by 120 moving average trading signals, and the 79 firm fundamentals, which helps to understand the weak-form market efficiency, algorithm trading and fundamental analysis. I find each set independently has strong predictive power, and buying the top 20% stocks with the greatest predicted returns and shorting bottom 20% with the lowest earns economically significant profits, and the profitability is robust to a number of controls. Econometrically, neural network generally improves forecasting over linear models, but makes little difference with firm fundamental predictors. Ensemble method tends to perform the best. However, when combining information from all the predictors, traditional machine learning improves little the performance due to perhaps not enough time series for too large dimensionality. In contrast, simple forecasting combination and portfolio diversification approach provide large gains.

Table 1: Summary Statistics of 40-Year Full Sample Period 1978:01 - 2017:12

This table reports summary statistics of spread portfolio returns for the full sample period 1978:01 - 2017:12, including average return (in % per month), t-statistics, volatility (standard deviation), monthly Sharpe ratio, skewness, kurtosis, proportion of positive returns, minimum monthly return, and maximum drawdown. In Panel A, predictors are previous 120 months lagged stock returns; in Panel B predictors are 60 moving average price signals and 60 moving trading volume signals; in Panel C predictors are 79 firm characteristics from Green, Hand, Zhang (2017), excluding price, returns, and trading volume-related variables. t-stat is calculated under the null hypothesis that excess return equals zero using White Heteroskedasticity robust standard error. At the beginning of each month,

Panel A: 120 months lagged returns 1978:01 - 2017:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.78	5.13***	2.96	0.26	0.79	18.34	0.64	-0.19	-0.36
PCR	0.81	6.14***	3.11	0.26	0.83	14.09	0.63	-0.18	-0.31
PLS	1.06	7.16***	3.37	0.31	1.27	15.40	0.66	-0.20	-0.32
Ridge	0.78	5.13***	2.96	0.26	0.79	18.34	0.64	-0.19	-0.36
Lasso	0.82	5.34***	3.04	0.27	0.63	17.14	0.64	-0.20	-0.37
ENet	0.82	5.34***	3.04	0.27	0.63	17.14	0.64	-0.20	-0.37
NN1L	0.91	5.73***	3.29	0.28	0.43	11.64	0.66	-0.18	-0.40
NN2L	1.03	6.06***	3.95	0.26	0.91	13.60	0.66	-0.23	-0.39
NN3L	1.00	5.79***	4.25	0.24	0.86	14.85	0.67	-0.25	-0.44
NN4L	0.92	4.91***	4.37	0.21	1.24	18.84	0.62	-0.28	-0.46
NN5L	0.85	5.21***	4.12	0.21	1.36	19.76	0.62	-0.25	-0.44
Ens_Linear	1.24	6.02***	4.35	0.29	1.51	19.73	0.65	-0.24	-0.41
Ens_Net	1.59	5.68***	6.34	0.25	1.10	13.43	0.64	-0.34	-0.52
Ens_Total	1.74	5.58***	7.00	0.25	1.38	14.60	0.61	-0.36	-0.54

Panel B: 60 moving average price and 60 trading volume signals 1978:01 - 2017:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.68	4.19***	4.26	0.16	-0.07	14.40	0.61	-0.27	-0.45
PCR	0.77	3.68***	5.40	0.14	0.34	12.22	0.59	-0.32	-0.50
PLS	0.75	3.47***	5.47	0.14	0.46	11.13	0.58	-0.30	-0.51
Ridge	0.68	4.19***	4.26	0.16	-0.07	14.37	0.61	-0.27	-0.45
Lasso	0.87	4.09***	5.46	0.16	-0.05	12.51	0.61	-0.32	-0.57
ENet	0.87	4.10***	5.46	0.16	-0.04	12.45	0.61	-0.32	-0.57
NN1L	1.04	4.89***	5.52	0.19	0.09	12.24	0.63	-0.33	-0.53
NN2L	1.06	4.54***	5.97	0.18	0.16	12.85	0.61	-0.38	-0.57
NN3L	0.99	4.15***	6.02	0.16	0.44	13.44	0.59	-0.37	-0.58
NN4L	0.89	3.77***	5.88	0.15	0.58	13.22	0.59	-0.36	-0.55
NN5L	0.84	3.76***	5.86	0.14	0.41	13.72	0.57	-0.37	-0.57
Ens_Linear	1.09	4.23***	6.34	0.17	0.48	10.99	0.61	-0.36	-0.58
Ens_Net	1.48	5.21***	7.21	0.21	0.60	11.28	0.62	-0.41	-0.59
Ens_Total	1.89	5.90***	8.02	0.24	0.76	10.15	0.63	-0.43	-0.58

Table 1: Summary Statistics of 40-Year Full Sample Period 1978:01 - 2017:12 (Cont'd)

I independently sort stocks into quintile portfolios on predicted return from each model and dataset and construct spread portfolios by buying the best-predicted quintile and selling the worst-predicted quintile. For a given family of models, ensemble portfolios are constructed by buying the unanimous winner quintile by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. NYSE 10% market value breakpoint is applied and stocks are equal-weighted within quintile. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: 79 fundamental Firm Characteristics 1978:01 - 2017:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	1.30	4.92***	6.10	0.21	0.85	11.20	0.63	-0.29	-0.51
PCR	1.37	5.18***	6.15	0.22	0.74	10.14	0.64	-0.28	-0.49
PLS	1.33	5.02***	6.31	0.21	0.78	9.74	0.63	-0.29	-0.50
Ridge	1.29	4.91***	6.10	0.21	0.85	11.18	0.63	-0.29	-0.51
Lasso	1.38	4.95***	6.35	0.22	0.77	9.94	0.62	-0.29	-0.52
ENet	1.38	4.95***	6.35	0.22	0.77	9.94	0.62	-0.29	-0.52
NN1L	1.34	4.87***	6.34	0.21	0.81	11.03	0.62	-0.30	-0.50
NN2L	1.36	4.82***	6.56	0.21	0.75	11.09	0.61	-0.31	-0.52
NN3L	1.34	4.79***	6.65	0.20	0.79	11.40	0.62	-0.32	-0.55
NN4L	1.31	4.67***	6.53	0.20	0.82	11.38	0.62	-0.31	-0.55
NN5L	1.31	4.74***	6.38	0.20	0.87	11.88	0.62	-0.29	-0.53
Ens_Linear	1.62	5.41***	6.94	0.23	0.67	7.75	0.63	-0.30	-0.54
Ens_Net	1.75	5.22***	7.98	0.22	0.58	7.37	0.63	-0.32	-0.59
Ens_Total	1.96	5.67***	8.35	0.23	0.55	6.76	0.64	-0.33	-0.61

Table 2: Summary Statistics during 10-Year Sub-Periods (Predictors: 120-month lagged return)

This table reports summary statistics of excess returns of spread portfolios whose construction uses 120-month lagged return as predictors. Statistics are presented for 4 10-year subperiods separately and they include average return (in % per month), t-statistics, volatility (standard deviation), monthly Sharpe ratio, skewness, kurtosis, proportion of positive returns, minimum monthly return, and maximum drawdown. t-stat is calculated under the null hypothesis that excess return equals zero using White Heteroskedasticity robust standard error. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by

Panel A: 120 months lagged return 1978:01 - 1987:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	1.09	5.57***	2.26	0.48	0.05	0.44	0.68	-0.05	-0.05
PCR	0.92	4.36***	2.47	0.37	0.06	0.38	0.66	-0.05	-0.10
PLS	1.21	6.55***	2.85	0.43	-0.09	1.04	0.71	-0.07	-0.07
Ridge	1.09	5.57***	2.26	0.48	0.05	0.44	0.68	-0.05	-0.05
Lasso	1.06	5.36***	2.37	0.45	-0.15	0.48	0.68	-0.06	-0.06
ENet	1.06	5.36***	2.37	0.45	-0.15	0.48	0.68	-0.06	-0.06
NN1L	1.27	6.38***	2.50	0.51	-0.37	1.14	0.72	-0.07	-0.07
NN2L	1.35	6.50***	2.95	0.46	-0.37	1.25	0.73	-0.08	-0.08
NN3L	1.25	5.98***	3.15	0.40	-0.49	2.17	0.72	-0.11	-0.11
NN4L	1.18	5.29***	3.09	0.38	0.03	1.23	0.68	-0.09	-0.09
NN5L	1.00	5.38***	2.70	0.37	-0.15	0.85	0.64	-0.07	-0.07
Ens_Linear	1.53	5.46***	3.40	0.45	-0.13	0.14	0.72	-0.08	-0.09
Ens_Net	2.22	7.74***	4.42	0.50	-0.19	-0.21	0.72	-0.11	-0.11
Ens_Total	2.42	8.54***	4.94	0.49	0.07	0.55	0.68	-0.11	-0.14

Panel B: 120 months lagged return 1988:01 - 1997:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.97	5.84***	1.56	0.62	0.50	0.75	0.72	-0.02	-0.03
PCR	1.00	5.57***	1.92	0.52	0.30	0.74	0.71	-0.05	-0.06
PLS	1.26	5.90***	1.97	0.64	0.22	0.48	0.78	-0.05	-0.05
Ridge	0.97	5.84***	1.56	0.62	0.50	0.75	0.72	-0.02	-0.03
Lasso	1.08	5.49***	1.67	0.65	0.64	0.89	0.76	-0.02	-0.04
ENet	1.08	5.49***	1.67	0.65	0.64	0.89	0.76	-0.02	-0.04
NN1L	1.07	5.53***	1.95	0.55	0.78	1.30	0.72	-0.04	-0.05
NN2L	1.17	5.11***	2.37	0.49	0.84	2.20	0.71	-0.06	-0.06
NN3L	1.01	4.60***	2.50	0.40	0.21	2.43	0.69	-0.08	-0.08
NN4L	0.91	4.90***	2.50	0.37	0.13	2.74	0.66	-0.08	-0.09
NN5L	0.79	3.58***	2.30	0.34	0.19	2.41	0.66	-0.08	-0.11
Ens_Linear	1.52	6.46***	2.34	0.65	0.06	0.08	0.73	-0.05	-0.06
Ens_Net	1.64	3.28***	4.64	0.35	-0.65	1.77	0.67	-0.18	-0.27
Ens_Total	1.91	3.68***	5.49	0.35	-0.79	1.46	0.68	-0.17	-0.24

Table 2: Summary Statistics during 10-Year Sub-Periods (Predictors: 120-month lagged return) (Cont'd)

buying the best-predicted quintile and selling the worst-predicted quintile. For a given family of models, ensemble portfolios are constructed by buying the unanimou winner quintile by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes five neural networks. Total ensemble includes both linear and network families, totalling 11 models. NYSE 10% market value breakpoint is applied and stocks are equal-weighted within quintile. * ,** ,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: 120 months lagged return 1998:01 - 2007:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.57	1.62	3.78	0.15	-1.37	7.58	0.59	-0.19	-0.36
PCR	1.00	2.91***	4.62	0.22	0.56	9.73	0.61	-0.18	-0.31
PLS	1.23	3.37***	4.60	0.27	0.88	12.10	0.62	-0.20	-0.32
Ridge	0.57	1.62	3.78	0.15	-1.37	7.58	0.59	-0.19	-0.36
Lasso	0.68	1.86*	3.89	0.17	-1.34	7.82	0.59	-0.20	-0.37
ENet	0.68	1.86*	3.89	0.17	-1.34	7.82	0.59	-0.20	-0.37
NN1L	0.79	2.12**	4.51	0.17	-0.70	5.13	0.62	-0.18	-0.40
NN2L	1.27	3.08***	5.98	0.21	0.40	7.37	0.65	-0.23	-0.39
NN3L	1.21	2.66**	6.69	0.18	0.59	7.62	0.67	-0.25	-0.44
NN4L	1.20	2.41**	7.01	0.17	0.85	9.41	0.62	-0.28	-0.46
NN5L	1.18	2.67**	6.80	0.17	0.94	8.79	0.63	-0.25	-0.44
Ens_Linear	1.36	2.90***	5.97	0.23	0.41	10.15	0.62	-0.24	-0.41
Ens_Net	2.06	2.95***	9.04	0.23	0.90	9.38	0.64	-0.34	-0.52
Ens_Total	1.99	2.53**	9.58	0.21	1.20	11.44	0.61	-0.36	-0.54

Panel D: 120 months lagged return 2008:01 - 2017:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.49	1.22	3.63	0.13	3.53	23.89	0.56	-0.08	-0.23
PCR	0.33	1.25	2.74	0.12	1.85	8.92	0.53	-0.06	-0.10
PLS	0.51	1.55	3.48	0.15	2.75	17.18	0.54	-0.08	-0.17
Ridge	0.49	1.22	3.63	0.13	3.53	23.89	0.56	-0.08	-0.23
Lasso	0.46	1.18	3.65	0.13	3.37	23.45	0.52	-0.09	-0.22
ENet	0.46	1.18	3.65	0.13	3.37	23.45	0.52	-0.09	-0.22
NN1L	0.50	1.24	3.59	0.14	3.04	19.60	0.57	-0.09	-0.19
NN2L	0.32	0.88	3.47	0.09	3.09	19.90	0.54	-0.09	-0.22
NN3L	0.53	1.39	3.37	0.16	2.16	13.23	0.58	-0.09	-0.12
NN4L	0.38	0.91	3.39	0.11	2.49	14.79	0.52	-0.09	-0.19
NN5L	0.42	1.28	3.01	0.14	1.95	11.71	0.54	-0.08	-0.15
Ens_Linear	0.56	1.04	4.77	0.12	4.08	31.15	0.53	-0.09	-0.29
Ens_Net	0.45	0.84	6.11	0.07	2.34	15.32	0.54	-0.14	-0.26
Ens_Total	0.65	1.00	7.01	0.09	2.96	18.13	0.45	-0.13	-0.31

Table 3: Summary Statistics during Sub-Periods (Predictors: 60 price and 60 trading volume MAs signals)

This table reports summary statistics of excess returns of spread portfolios whose construction uses 60 price and 60 trading volume MAs signals as predictors. Statistics are presented for 4 10-year subperiods separately and they include average return (in % per month), t-statistics, volatility (standard deviation), monthly Sharpe ratio, skewness, kurtosis, proportion of positive returns, minimum monthly return, and maximum drawdown. t-stat is calculated under the null hypothesis that excess return equals zero using White Heteroskedasticity robust standard error. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by

Panel A: 60 moving average price and 60 trading volume signals 1978:01 - 1987:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.99	3.46***	2.84	0.35	-0.28	2.78	0.68	-0.10	-0.12
PCR	0.90	2.53**	3.72	0.24	-0.10	1.57	0.58	-0.13	-0.22
PLS	0.98	2.68**	3.80	0.26	-0.10	1.49	0.63	-0.13	-0.21
Ridge	1.00	3.50***	2.85	0.35	-0.26	2.90	0.68	-0.10	-0.12
Lasso	1.13	3.22***	3.77	0.30	-0.14	2.06	0.64	-0.13	-0.20
ENet	1.13	3.23***	3.77	0.30	-0.14	2.06	0.64	-0.13	-0.20
NN1L	1.19	3.97***	3.61	0.33	-0.08	2.34	0.65	-0.12	-0.12
NN2L	1.17	3.74***	3.86	0.30	-0.00	2.33	0.64	-0.13	-0.13
NN3L	1.07	3.60***	3.89	0.27	0.02	2.58	0.61	-0.14	-0.14
NN4L	0.93	2.90***	3.74	0.25	0.17	2.62	0.61	-0.13	-0.13
NN5L	0.97	3.41***	3.73	0.26	0.17	2.81	0.62	-0.13	-0.13
Ens_Linear	1.28	3.01***	4.32	0.30	-0.36	1.49	0.64	-0.15	-0.23
Ens_Net	1.47	3.87***	4.73	0.31	-0.20	1.71	0.63	-0.17	-0.17
Ens_Total	1.68	4.17***	5.25	0.32	-0.31	1.15	0.66	-0.18	-0.18

Panel B: 60 moving average price and 60 trading volume signals 1988:01 - 1997:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.42	3.00***	1.53	0.28	-0.05	-0.05	0.58	-0.04	-0.07
PCR	0.58	2.24**	2.48	0.23	-0.16	0.03	0.60	-0.07	-0.15
PLS	0.50	2.19**	2.39	0.21	-0.28	0.56	0.58	-0.07	-0.15
Ridge	0.42	2.99***	1.54	0.27	-0.05	-0.07	0.58	-0.04	-0.07
Lasso	0.62	2.64**	2.36	0.26	-0.40	0.73	0.62	-0.08	-0.13
ENet	0.62	2.62**	2.37	0.26	-0.41	0.71	0.62	-0.08	-0.13
NN1L	0.91	4.95***	2.51	0.36	0.14	0.90	0.65	-0.07	-0.10
NN2L	0.96	4.60***	2.73	0.35	0.04	1.22	0.65	-0.09	-0.12
NN3L	0.80	3.67***	2.78	0.29	0.18	0.71	0.61	-0.08	-0.12
NN4L	0.77	3.40***	2.63	0.29	0.06	1.14	0.63	-0.08	-0.12
NN5L	0.63	2.92***	2.56	0.25	-0.10	1.33	0.58	-0.09	-0.14
Ens_Linear	0.83	2.87***	2.86	0.29	-0.21	0.11	0.63	-0.08	-0.13
Ens_Net	1.27	3.76***	3.78	0.34	0.32	1.22	0.63	-0.11	-0.16
Ens_Total	1.40	3.33***	4.02	0.35	-0.13	0.36	0.64	-0.11	-0.21

Table 3: Summary Statistics during Sub-Periods (Predictors: 60 price and 60 trading volume MAs signals) - Cont'd

buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the unanimou winner quintile by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes five neural networks. Total ensemble includes both linear and network families, totalling 11 models. NYSE 10% market value breakpoint is applied and stocks are equal-weighted within quintile. * ,** ,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: 60 moving average price and 60 trading volume signals 1998:01 - 2007:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.77	2.09**	6.63	0.12	-0.83	5.19	0.62	-0.27	-0.45
PCR	0.93	1.87*	8.13	0.11	-0.53	4.28	0.60	-0.32	-0.50
PLS	0.87	1.62	8.37	0.10	-0.26	3.76	0.56	-0.30	-0.51
Ridge	0.77	2.09**	6.63	0.12	-0.83	5.18	0.62	-0.27	-0.45
Lasso	1.02	1.91*	8.39	0.12	-0.80	4.73	0.57	-0.32	-0.57
ENet	1.03	1.92*	8.38	0.12	-0.79	4.70	0.57	-0.32	-0.57
NN1L	1.34	2.30**	8.73	0.15	-0.56	4.41	0.64	-0.33	-0.53
NN2L	1.43	2.31**	9.48	0.15	-0.45	4.97	0.61	-0.38	-0.57
NN3L	1.56	2.39**	9.61	0.16	-0.20	5.35	0.62	-0.37	-0.58
NN4L	1.34	2.07**	9.37	0.14	-0.08	5.03	0.60	-0.36	-0.55
NN5L	1.40	2.18**	9.39	0.15	-0.26	5.26	0.60	-0.37	-0.57
Ens_Linear	1.34	2.08**	9.49	0.14	-0.29	3.96	0.59	-0.36	-0.58
Ens_Net	1.94	2.46**	11.07	0.18	0.05	5.04	0.62	-0.41	-0.59
Ens_Total	2.66	3.16***	11.96	0.22	0.23	5.17	0.63	-0.43	-0.58

Panel D: 60 moving average price and 60 trading volume signals 2008:01 - 2017:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.55	1.33	4.29	0.13	2.74	19.81	0.57	-0.13	-0.20
PCR	0.65	1.25	5.58	0.12	2.80	18.91	0.57	-0.15	-0.22
PLS	0.65	1.25	5.48	0.12	2.69	17.44	0.54	-0.14	-0.22
Ridge	0.55	1.31	4.29	0.13	2.74	19.86	0.57	-0.13	-0.19
Lasso	0.70	1.36	5.45	0.13	2.63	17.92	0.60	-0.13	-0.22
ENet	0.70	1.37	5.45	0.13	2.63	17.92	0.60	-0.13	-0.22
NN1L	0.73	1.42	5.20	0.14	2.75	18.43	0.57	-0.12	-0.18
NN2L	0.66	1.16	5.58	0.12	2.60	16.88	0.55	-0.14	-0.23
NN3L	0.54	0.97	5.50	0.10	2.67	16.87	0.53	-0.12	-0.25
NN4L	0.51	0.94	5.49	0.09	2.69	16.84	0.52	-0.11	-0.26
NN5L	0.37	0.79	5.40	0.07	2.76	17.78	0.49	-0.11	-0.26
Ens_Linear	0.92	1.46	6.68	0.14	2.43	15.20	0.56	-0.16	-0.31
Ens_Net	1.25	1.98*	7.05	0.18	2.09	11.86	0.57	-0.14	-0.23
Ens_Total	1.82	2.52**	8.44	0.22	1.65	7.41	0.58	-0.16	-0.28

Table 4: Summary Statistics during 10-Year Sub-Periods (Predictors: 79 fundamental Firm Characteristics)

This table reports summary statistics of excess returns of spread portfolios whose construction uses 79 fundamental firm characteristics as predictors. Statistics are presented for 4 10-year subperiods separately and they include average return (in % per month), t-statistics, volatility (standard deviation), monthly Sharpe ratio, skewness, kurtosis, proportion of positive returns, minimum monthly return, and maximum drawdown. t-stat is calculated under the null hypothesis that excess return equals zero using White Heteroskedasticity robust standard error. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by

Panel A: 79 fundamental Firm Characteristics 1978:01 - 1987:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	1.20	5.26***	3.96	0.30	-0.31	0.94	0.65	-0.11	-0.15
PCR	1.26	5.41***	4.26	0.30	-0.14	0.93	0.62	-0.11	-0.16
PLS	1.26	5.11***	4.33	0.29	-0.27	0.96	0.62	-0.13	-0.15
Ridge	1.19	5.15***	3.98	0.30	-0.31	0.96	0.64	-0.11	-0.15
Lasso	1.38	5.79***	4.35	0.32	-0.27	0.92	0.65	-0.12	-0.16
ENet	1.38	5.79***	4.35	0.32	-0.27	0.92	0.65	-0.12	-0.16
NN1L	1.27	5.19***	4.37	0.29	-0.34	1.16	0.62	-0.13	-0.17
NN2L	1.24	4.76***	4.59	0.27	-0.26	0.98	0.62	-0.13	-0.18
NN3L	1.25	4.71***	4.57	0.27	-0.20	1.01	0.62	-0.13	-0.16
NN4L	1.19	4.63***	4.56	0.26	-0.21	0.81	0.62	-0.13	-0.16
NN5L	1.17	4.64***	4.35	0.27	-0.22	0.93	0.59	-0.12	-0.15
Ens_Linear	1.53	5.77***	5.03	0.30	-0.12	0.82	0.62	-0.14	-0.21
Ens_Net	1.62	4.64***	6.13	0.26	-0.14	0.73	0.64	-0.17	-0.20
Ens_Total	1.82	4.74***	6.55	0.28	-0.16	0.60	0.65	-0.18	-0.24

Panel B: 79 fundamental Firm Characteristics 1988:01 - 1997:12

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	1.56	5.10***	3.33	0.47	0.07	0.63	0.68	-0.08	-0.10
PCR	1.62	4.78***	3.66	0.44	-0.13	1.01	0.70	-0.10	-0.12
PLS	1.42	3.63***	4.01	0.35	-0.17	0.96	0.65	-0.11	-0.14
Ridge	1.56	5.10***	3.33	0.47	0.07	0.63	0.68	-0.08	-0.10
Lasso	1.60	5.01***	3.63	0.44	-0.02	0.66	0.68	-0.09	-0.11
ENet	1.60	5.01***	3.63	0.44	-0.02	0.66	0.68	-0.09	-0.11
NN1L	1.61	4.87***	3.67	0.44	-0.09	0.76	0.70	-0.10	-0.12
NN2L	1.60	4.64***	3.87	0.41	-0.09	0.89	0.68	-0.10	-0.14
NN3L	1.51	4.39***	3.97	0.38	-0.09	0.87	0.69	-0.10	-0.15
NN4L	1.43	3.92***	3.89	0.37	-0.12	0.84	0.69	-0.10	-0.14
NN5L	1.54	4.79***	3.74	0.41	-0.09	1.22	0.70	-0.10	-0.14
Ens_Linear	1.85	4.10***	4.48	0.41	-0.13	0.36	0.65	-0.12	-0.15
Ens_Net	2.03	4.54***	5.22	0.39	-0.17	0.22	0.68	-0.14	-0.22
Ens_Total	2.17	4.33***	5.56	0.39	-0.16	0.15	0.68	-0.15	-0.24

Table 4: Summary Statistics during 10-Year Sub-Periods (Predictors: 79 fundamental Firm Characteristics) - Cont'd

buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the unanimou winner quintile by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes five neural networks. Total ensemble includes both linear and network families, totalling 11 models. NYSE 10% market value breakpoint is applied and stocks are equal-weighted within quintile. * ,** ,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: 79 fundamental Firm Characteristics 1998:01 - 2007:12									
Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	1.88	2.37**	9.93	0.19	0.51	4.52	0.64	-0.29	-0.51
PCR	1.95	2.47**	9.86	0.20	0.50	4.48	0.66	-0.28	-0.49
PLS	1.92	2.53**	9.97	0.19	0.51	4.44	0.67	-0.29	-0.50
Ridge	1.89	2.38**	9.93	0.19	0.51	4.52	0.64	-0.29	-0.51
Lasso	2.01	2.41**	10.19	0.20	0.46	4.19	0.62	-0.29	-0.52
ENet	2.01	2.41**	10.19	0.20	0.46	4.19	0.62	-0.29	-0.52
NN1L	1.98	2.43**	10.16	0.19	0.57	5.00	0.62	-0.30	-0.50
NN2L	2.01	2.39**	10.53	0.19	0.55	5.12	0.60	-0.31	-0.52
NN3L	2.05	2.43**	10.74	0.19	0.55	5.12	0.62	-0.32	-0.55
NN4L	2.05	2.48**	10.63	0.19	0.56	4.89	0.62	-0.31	-0.55
NN5L	1.97	2.33**	10.49	0.19	0.60	4.89	0.63	-0.29	-0.53
Ens_Linear	2.20	2.56**	10.59	0.21	0.45	3.86	0.66	-0.30	-0.54
Ens_Net	2.54	2.69**	11.78	0.22	0.53	4.72	0.65	-0.32	-0.59
Ens_Total	2.68	2.71**	12.12	0.22	0.52	4.58	0.67	-0.33	-0.61
Panel D: 79 fundamental Firm Characteristics 2008:01 - 2017:12									
Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.55	1.15	4.85	0.11	1.58	8.11	0.56	-0.13	-0.20
PCR	0.63	1.38	4.74	0.13	1.11	4.93	0.56	-0.12	-0.18
PLS	0.73	1.45	5.03	0.14	1.51	7.45	0.57	-0.12	-0.22
Ridge	0.55	1.13	4.85	0.11	1.59	8.11	0.55	-0.13	-0.20
Lasso	0.52	1.02	5.00	0.10	1.54	7.40	0.52	-0.13	-0.26
ENet	0.52	1.02	5.00	0.10	1.54	7.40	0.52	-0.13	-0.26
NN1L	0.50	1.01	4.96	0.10	1.06	4.82	0.52	-0.13	-0.27
NN2L	0.58	1.17	5.02	0.12	0.72	2.99	0.55	-0.13	-0.26
NN3L	0.57	1.17	4.99	0.11	0.87	3.37	0.54	-0.12	-0.28
NN4L	0.56	1.15	4.63	0.12	0.85	2.90	0.56	-0.11	-0.26
NN5L	0.54	1.23	4.44	0.12	0.83	3.01	0.55	-0.11	-0.24
Ens_Linear	0.88	1.61	5.94	0.15	1.32	6.83	0.58	-0.15	-0.24
Ens_Net	0.82	1.32	7.17	0.11	0.55	2.82	0.57	-0.20	-0.38
Ens_Total	1.18	1.96*	7.66	0.15	0.60	3.03	0.57	-0.21	-0.27

Table 5: Risk-Adjusted Returns (%) (Predictors: 120-month lagged return)

Panel A reports risk-adjusted returns of the 14 spread portfolios by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. Panel B and panel C report the risk-adjusted long- and short-leg portfolio returns by the same factor models. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all six neural networks. Total ensemble includes both linear and network families, totalling 12 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel A: Spread Portfolio (Predictors: 120-month lagged return)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.78	0.78	0.75	0.76	0.69
	(5.13***)	(5.49***)	(5.24***)	(4.15***)	(4.53***)
PCR	0.81	0.84	0.82	0.78	0.80
	(6.14***)	(6.30***)	(6.06***)	(5.12***)	(4.75***)
PLS	1.06	1.08	1.03	1.01	1.01
	(7.16***)	(7.26***)	(6.93***)	(5.95***)	(5.60***)
Ridge	0.78	0.78	0.75	0.76	0.69
	(5.13***)	(5.49***)	(5.24***)	(4.15***)	(4.53***)
Lasso	0.82	0.82	0.78	0.79	0.71
	(5.34***)	(5.68***)	(5.44***)	(4.25***)	(4.67***)
ENet	0.82	0.82	0.78	0.79	0.71
	(5.34***)	(5.68***)	(5.44***)	(4.25***)	(4.67***)
NN1L	0.91	0.94	0.91	0.92	0.87
	(5.73***)	(6.38***)	(6.09***)	(4.87***)	(5.35***)
NN2L	1.03	1.08	1.05	1.04	0.97
	(6.06***)	(6.62***)	(6.25***)	(5.62***)	(4.90***)
NN3L	1.00	1.06	1.04	1.00	0.98
	(5.79***)	(6.47***)	(6.10***)	(5.05***)	(4.49***)
NN4L	0.92	0.99	0.96	0.93	0.90
	(4.91***)	(5.58***)	(5.13***)	(4.29***)	(3.68***)
NN5L	0.85	0.92	0.89	0.86	0.83
	(5.21***)	(5.79***)	(5.43***)	(4.52***)	(3.90***)
Ens_linear	1.24	1.26	1.21	1.21	1.17
	(6.02***)	(6.33***)	(6.02***)	(4.82***)	(4.91***)
Ens_Net	1.59	1.69	1.65	1.64	1.63
	(5.68***)	(6.23***)	(5.67***)	(5.19***)	(4.53***)
Ens_total	1.74	1.83	1.75	1.75	1.73
	(5.58***)	(5.95***)	(5.48***)	(4.91***)	(4.42***)

Table 5: Risk-Adjusted Returns (%) (Predictors: 120-month lagged return) - Cont'd

Panel A reports risk-adjusted returns of the 15 spread portfolios by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. Panel B and panel C report the risk-adjusted long- and short-leg portfolio returns by the same factor models. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all six neural networks. Total ensemble includes both linear and network families, totalling 12 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Long-Leg (Predictors: 120-month lagged return)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.25	0.47	0.32	0.43	0.32
	(5.05***)	(3.25***)	(3.78***)	(3.79***)	(4.01***)
PCR	1.26	0.49	0.35	0.43	0.39
	(5.12***)	(3.34***)	(4.37***)	(4.43***)	(4.73***)
PLS	1.41	0.64	0.47	0.57	0.51
	(5.61***)	(3.98***)	(5.14***)	(5.10***)	(5.45***)
Ridge	1.25	0.47	0.32	0.43	0.32
	(5.05***)	(3.25***)	(3.78***)	(3.79***)	(4.01***)
Lasso	1.28	0.50	0.34	0.45	0.34
	(5.16***)	(3.34***)	(3.96***)	(3.94***)	(4.24***)
ENet	1.28	0.50	0.34	0.45	0.34
	(5.16***)	(3.34***)	(3.96***)	(3.94***)	(4.24***)
NN1L	1.32	0.56	0.41	0.51	0.42
	(5.38***)	(3.93***)	(4.76***)	(4.55***)	(5.04***)
NN2L	1.35	0.61	0.45	0.56	0.46
	(5.49***)	(3.76***)	(4.66***)	(4.75***)	(4.87***)
NN3L	1.36	0.63	0.48	0.56	0.50
	(5.53***)	(3.98***)	(4.91***)	(4.58***)	(4.72***)
NN4L	1.29	0.57	0.42	0.51	0.43
	(5.28***)	(3.66***)	(4.25***)	(3.99***)	(4.00***)
NN5L	1.25	0.53	0.38	0.46	0.40
	(5.13***)	(3.36***)	(4.01***)	(3.96***)	(3.98***)
Ens_linear	1.47	0.66	0.49	0.61	0.54
	(5.31***)	(3.64***)	(4.18***)	(4.01***)	(4.50***)
Ens_Net	1.58	0.83	0.66	0.78	0.73
	(5.40***)	(3.96***)	(4.24***)	(4.15***)	(4.33***)
Ens_total	1.72	0.92	0.74	0.87	0.83
	(5.35***)	(3.91***)	(4.13***)	(3.97***)	(4.17***)

Table 5: Risk-Adjusted Returns (%) (Predictors: 120-month lagged return) - Cont'd

Panel A reports risk-adjusted returns of the 15 spread portfolios by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. Panel B and panel C report the risk-adjusted long- and short-leg portfolio returns by the same factor models. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all six neural networks. Total ensemble includes both linear and network families, totalling 12 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: Short-Leg (Predictors: 120-month lagged return)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.47	-0.31	-0.43	-0.33	-0.37
	(1.89*)	(-2.15**)	(-4.52***)	(-3.22***)	(-3.39***)
PCR	0.45	-0.34	-0.48	-0.36	-0.41
	(1.83*)	(-2.65**)	(-5.45***)	(-4.29***)	(-3.53***)
PLS	0.35	-0.44	-0.56	-0.44	-0.50
	(1.39)	(-3.38***)	(-6.20***)	(-5.03***)	(-4.26***)
Ridge	0.47	-0.31	-0.43	-0.33	-0.37
	(1.89*)	(-2.15**)	(-4.52***)	(-3.22***)	(-3.39***)
Lasso	0.46	-0.32	-0.44	-0.34	-0.37
	(1.85*)	(-2.22**)	(-4.70***)	(-3.27***)	(-3.43***)
ENet	0.46	-0.32	-0.44	-0.34	-0.37
	(1.85*)	(-2.22**)	(-4.70***)	(-3.27***)	(-3.43***)
NN1L	0.42	-0.38	-0.50	-0.41	-0.45
	(1.64)	(-2.74**)	(-5.37***)	(-4.04***)	(-4.04***)
NN2L	0.33	-0.48	-0.59	-0.49	-0.51
	(1.29)	(-3.67***)	(-6.03***)	(-5.15***)	(-3.87***)
NN3L	0.36	-0.43	-0.55	-0.44	-0.48
	(1.53)	(-3.21***)	(-5.42***)	(-4.18***)	(-3.42***)
NN4L	0.37	-0.42	-0.54	-0.43	-0.47
	(1.52)	(-3.08***)	(-4.79***)	(-3.81***)	(-3.03***)
NN5L	0.41	-0.39	-0.51	-0.40	-0.43
	(1.70*)	(-3.04***)	(-5.36***)	(-4.03***)	(-3.23***)
Ens_linear	0.22	-0.60	-0.72	-0.60	-0.63
	(0.82)	(-3.74***)	(-5.94***)	(-4.66***)	(-4.05***)
Ens_Net	-0.01	-0.87	-0.99	-0.86	-0.90
	(-0.05)	(-4.40***)	(-5.53***)	(-4.96***)	(-3.92***)
Ens_total	-0.02	-0.90	-1.01	-0.88	-0.90
	(-0.08)	(-3.94***)	(-4.81***)	(-4.32***)	(-3.58***)

Table 6: Risk-Adjusted Returns (%) (Predictors: 60 price and 60 trading volume MA signals)

Panel A reports risk-adjusted returns of the 14 spread portfolios by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks of assorted structures. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel A: Spread Portfolio (Predictors: 60 price and 60 trading volume MA signals)

Model Name	Mean	CAPM	FF3	Carhart4	F F5
OLS	0.68	0.74	0.69	0.65	0.52
	(4.19***)	(5.08***)	(4.64***)	(2.82***)	(2.51**)
PCR	0.77	0.82	0.79	0.72	0.59
	(3.68***)	(4.37***)	(4.05***)	(2.46**)	(2.26**)
PLS	0.75	0.82	0.79	0.70	0.60
	(3.47***)	(4.14***)	(3.88***)	(2.41**)	(2.23**)
Ridge	0.68	0.74	0.69	0.65	0.52
	(4.19***)	(5.08***)	(4.63***)	(2.82***)	(2.50**)
Lasso	0.87	0.92	0.90	0.83	0.69
	(4.09***)	(4.92***)	(4.47***)	(2.85***)	(2.58**)
ENet	0.87	0.92	0.90	0.84	0.70
	(4.10***)	(4.94***)	(4.49***)	(2.85***)	(2.59**)
NN1L	1.04	1.12	1.08	1.04	0.89
	(4.89***)	(5.92***)	(5.33***)	(3.55***)	(3.25***)
NN2L	1.06	1.15	1.12	1.05	0.92
	(4.54***)	(5.45***)	(5.06***)	(3.29***)	(3.00***)
NN3L	0.99	1.08	1.06	0.98	0.87
	(4.15***)	(4.92***)	(4.63***)	(3.06***)	(2.73**)
NN4L	0.89	0.97	0.94	0.86	0.74
	(3.77***)	(4.38***)	(4.17***)	(2.75**)	(2.43**)
NN5L	0.84	0.93	0.91	0.84	0.72
	(3.76***)	(4.52***)	(4.19***)	(2.78**)	(2.36**)
Ens_linear	1.09	1.16	1.15	1.04	0.93
	(4.23***)	(4.89***)	(4.71***)	(2.99***)	(2.94***)
Ens_Net	1.48	1.58	1.57	1.47	1.35
	(5.21***)	(6.02***)	(5.74***)	(3.90***)	(3.63***)
Ens_total	1.89	1.97	1.97	1.86	1.81
	(5.90***)	(6.58***)	(6.38***)	(4.54***)	(4.36***)

Table 6: Risk-Adjusted Returns (%) (Predictors: 60 price and 60 trading volume MA signals) - Cont'd

Panel B reports risk-adjusted returns of the 14 long-leg portfolios by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct long-leg portfolios by buying the best-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks of assorted structures. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Long-Leg (Predictors: 60 price and 60 trading volume MA signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.06	0.22	0.14	0.24	0.25
	(3.58***)	(1.31)	(1.55)	(1.92*)	(2.53**)
PCR	1.16	0.33	0.25	0.35	0.33
	(3.93***)	(1.87*)	(2.31**)	(2.21**)	(2.80**)
PLS	1.14	0.32	0.23	0.33	0.33
	(3.90***)	(1.76*)	(2.15**)	(2.13**)	(2.66**)
Ridge	1.06	0.22	0.14	0.24	0.25
	(3.59***)	(1.32)	(1.59)	(1.95*)	(2.55**)
Lasso	1.19	0.36	0.28	0.39	0.38
	(3.99***)	(2.03**)	(2.56**)	(2.48**)	(3.10***)
ENet	1.19	0.36	0.28	0.39	0.38
	(4.00***)	(2.04**)	(2.57**)	(2.48**)	(3.11***)
NN1L	1.30	0.49	0.39	0.53	0.47
	(4.46***)	(2.75**)	(3.34***)	(3.26***)	(3.80***)
NN2L	1.32	0.53	0.43	0.56	0.51
	(4.53***)	(2.88***)	(3.49***)	(3.22***)	(3.74***)
NN3L	1.29	0.50	0.40	0.51	0.47
	(4.40***)	(2.61**)	(3.13***)	(3.02***)	(3.31***)
NN4L	1.25	0.45	0.36	0.47	0.42
	(4.30***)	(2.44**)	(2.90***)	(2.83***)	(3.10***)
NN5L	1.22	0.42	0.33	0.43	0.40
	(4.22***)	(2.40**)	(2.75**)	(2.75**)	(2.93***)
Ens_linear	1.35	0.50	0.43	0.53	0.54
	(4.12***)	(2.37**)	(2.98***)	(2.70**)	(3.30***)
Ens_Net	1.51	0.69	0.60	0.72	0.68
	(4.74***)	(3.27***)	(3.70***)	(3.50***)	(3.80***)
Ens_total	1.69	0.85	0.76	0.89	0.91
	(4.84***)	(3.48***)	(3.92***)	(3.75***)	(4.27***)

Table 6: Risk-Adjusted Returns (%) (Predictors: 60 price and 60 trading volume MA signals) - Cont'd

Panel C reports risk-adjusted returns of the 14 short-leg portfolios by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct short-leg portfolios by selling the poorestly-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks of assorted structures. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: Short-Leg (Predictors: 60 price and 60 trading volume MA signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.37	-0.52	-0.56	-0.41	-0.27
	(1.34)	(-4.18***)	(-6.22***)	(-3.21***)	(-2.00*)
PCR	0.39	-0.49	-0.54	-0.37	-0.26
	(1.40)	(-3.89***)	(-4.99***)	(-2.46**)	(-1.54)
PLS	0.39	-0.50	-0.56	-0.37	-0.27
	(1.37)	(-3.86***)	(-4.78***)	(-2.43**)	(-1.60)
Ridge	0.38	-0.52	-0.55	-0.41	-0.27
	(1.35)	(-4.16***)	(-6.16***)	(-3.18***)	(-1.98*)
Lasso	0.32	-0.56	-0.62	-0.45	-0.32
	(1.15)	(-4.20***)	(-5.35***)	(-2.88***)	(-1.84*)
ENet	0.32	-0.57	-0.62	-0.45	-0.32
	(1.15)	(-4.22***)	(-5.37***)	(-2.89***)	(-1.85*)
NN1L	0.26	-0.62	-0.69	-0.51	-0.42
	(0.94)	(-5.09***)	(-6.37***)	(-3.40***)	(-2.31**)
NN2L	0.27	-0.62	-0.69	-0.49	-0.41
	(0.96)	(-4.90***)	(-5.84***)	(-3.01***)	(-2.08**)
NN3L	0.29	-0.59	-0.66	-0.46	-0.40
	(1.06)	(-4.57***)	(-5.49***)	(-2.77**)	(-1.95*)
NN4L	0.36	-0.51	-0.58	-0.39	-0.32
	(1.30)	(-4.05***)	(-4.93***)	(-2.41**)	(-1.64)
NN5L	0.37	-0.51	-0.58	-0.40	-0.32
	(1.34)	(-4.00***)	(-4.98***)	(-2.48**)	(-1.66)
Ens_linear	0.26	-0.66	-0.72	-0.51	-0.39
	(0.87)	(-4.48***)	(-5.49***)	(-2.92***)	(-2.12**)
Ens_Net	0.03	-0.89	-0.97	-0.75	-0.67
	(0.11)	(-5.97***)	(-6.85***)	(-3.80***)	(-2.93***)
Ens_total	-0.20	-1.13	-1.22	-0.97	-0.89
	(-0.65)	(-6.67***)	(-7.81***)	(-4.71***)	(-3.69***)

Table 7: Risk-Adjusted Returns (%) (Predictors: 79 fundamental firm characteristics)

Panel A reports risk-adjusted returns of the 14 spread portfolios by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the best-predicted portfolio and selling the worst-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks of assorted structures. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel A: Spread Portfolio (Predictors: 79 fundamental firm characteristics)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.30	1.46	1.44	1.33	1.32
	(4.92***)	(5.58***)	(5.20***)	(4.39***)	(3.76***)
PCR	1.37	1.53	1.51	1.39	1.41
	(5.18***)	(5.87***)	(5.42***)	(4.66***)	(4.01***)
PLS	1.33	1.49	1.47	1.36	1.37
	(5.02***)	(5.74***)	(5.34***)	(4.37***)	(3.89***)
Ridge	1.29	1.45	1.44	1.33	1.32
	(4.91***)	(5.58***)	(5.19***)	(4.39***)	(3.75***)
Lasso	1.38	1.55	1.53	1.42	1.42
	(4.95***)	(5.70***)	(5.29***)	(4.51***)	(3.87***)
ENet	1.38	1.55	1.53	1.42	1.42
	(4.95***)	(5.70***)	(5.29***)	(4.51***)	(3.87***)
NN1L	1.34	1.52	1.51	1.36	1.39
	(4.87***)	(5.59***)	(5.21***)	(4.32***)	(3.76***)
NN2L	1.36	1.54	1.53	1.38	1.43
	(4.82***)	(5.51***)	(5.14***)	(4.30***)	(3.77***)
NN3L	1.34	1.53	1.53	1.40	1.44
	(4.79***)	(5.54***)	(5.16***)	(4.44***)	(3.79***)
NN4L	1.31	1.49	1.48	1.36	1.39
	(4.67***)	(5.42***)	(5.00***)	(4.29***)	(3.68***)
NN5L	1.31	1.47	1.48	1.35	1.38
	(4.74***)	(5.39***)	(5.02***)	(4.36***)	(3.68***)
Ens_linear	1.62	1.79	1.77	1.65	1.67
	(5.41***)	(6.14***)	(5.70***)	(4.74***)	(4.25***)
Ens_Net	1.75	1.97	1.96	1.78	1.87
	(5.22***)	(5.96***)	(5.50***)	(4.66***)	(4.21***)
Ens_total	1.96	2.17	2.17	2.00	2.07

Table 7: Risk-Adjusted Returns (%) (Predictors: 79 fundamental firm characteristics) - Cont'd

Panel B reports risk-adjusted returns of the 14 long-leg portfolios by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct long-leg portfolios by buying the best-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks of assorted structures. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Long-Leg (Predictors: 79 fundamental firm characteristics)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.71	0.96	0.87	0.93	0.96
	(5.85***)	(4.74***)	(5.22***)	(5.33***)	(5.15***)
PCR	1.74	1.00	0.91	0.96	1.01
	(5.90***)	(4.92***)	(5.27***)	(5.46***)	(5.30***)
PLS	1.72	0.98	0.88	0.95	0.99
	(5.86***)	(4.72***)	(5.17***)	(5.23***)	(5.24***)
Ridge	1.71	0.96	0.87	0.94	0.96
	(5.87***)	(4.76***)	(5.24***)	(5.35***)	(5.18***)
Lasso	1.73	1.00	0.90	0.97	1.00
	(5.88***)	(4.77***)	(5.20***)	(5.30***)	(5.14***)
ENet	1.73	1.00	0.91	0.97	1.00
	(5.88***)	(4.77***)	(5.20***)	(5.31***)	(5.14***)
NN1L	1.72	0.98	0.89	0.93	0.99
	(5.82***)	(4.68***)	(5.09***)	(5.13***)	(5.01***)
NN2L	1.72	0.99	0.91	0.95	1.01
	(5.84***)	(4.71***)	(5.00***)	(5.11***)	(4.97***)
NN3L	1.70	0.98	0.90	0.95	1.01
	(5.79***)	(4.65***)	(4.95***)	(5.13***)	(4.93***)
NN4L	1.70	0.97	0.89	0.95	1.00
	(5.76***)	(4.68***)	(5.03***)	(5.26***)	(5.06***)
NN5L	1.69	0.95	0.87	0.92	0.98
	(5.67***)	(4.55***)	(4.87***)	(5.10***)	(4.94***)
Ens_linear	1.80	1.05	0.96	1.03	1.08
	(5.75***)	(4.64***)	(4.83***)	(4.83***)	(4.87***)
Ens_Net	1.82	1.08	1.00	1.03	1.12
	(5.43***)	(4.28***)	(4.41***)	(4.39***)	(4.46***)
Ens_total	1.89	1.14	1.06	1.10	1.18
	(5.43***)	(4.32***)	(4.40***)	(4.38***)	(4.38***)

Table 7: Risk-Adjusted Returns (%) (Predictors: 79 fundamental firm characteristics) - Cont'd

Panel C reports risk-adjusted returns of the 14 short-leg portfolios by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct short-leg portfolios by selling the poorestly-predicted portfolio. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks of assorted structures. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: Short-Leg (Predictors: 79 fundamental firm characteristics)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.41	-0.49	-0.57	-0.40	-0.36
	(1.43)	(-3.41***)	(-4.18***)	(-2.54**)	(-1.85*)
PCR	0.37	-0.53	-0.60	-0.43	-0.40
	(1.31)	(-3.72***)	(-4.51***)	(-2.85***)	(-2.10**)
PLS	0.39	-0.51	-0.58	-0.41	-0.37
	(1.36)	(-3.67***)	(-4.40***)	(-2.57**)	(-1.95*)
Ridge	0.42	-0.49	-0.57	-0.39	-0.36
	(1.44)	(-3.38***)	(-4.14***)	(-2.51**)	(-1.84*)
Lasso	0.36	-0.55	-0.63	-0.45	-0.42
	(1.23)	(-3.77***)	(-4.41***)	(-2.82***)	(-2.07**)
ENet	0.36	-0.55	-0.63	-0.45	-0.42
	(1.23)	(-3.77***)	(-4.41***)	(-2.82***)	(-2.07**)
NN1L	0.38	-0.54	-0.62	-0.42	-0.41
	(1.30)	(-3.61***)	(-4.33***)	(-2.60**)	(-2.00*)
NN2L	0.37	-0.54	-0.63	-0.43	-0.42
	(1.27)	(-3.53***)	(-4.30***)	(-2.59**)	(-2.02**)
NN3L	0.36	-0.55	-0.63	-0.46	-0.43
	(1.25)	(-3.66***)	(-4.43***)	(-2.81***)	(-2.09**)
NN4L	0.39	-0.52	-0.59	-0.42	-0.39
	(1.36)	(-3.46***)	(-4.04***)	(-2.50**)	(-1.85*)
NN5L	0.38	-0.53	-0.61	-0.43	-0.40
	(1.32)	(-3.48***)	(-4.23***)	(-2.71**)	(-1.95*)
Ens_linear	0.18	-0.74	-0.81	-0.62	-0.59
	(0.62)	(-4.91***)	(-5.72***)	(-3.74***)	(-2.91***)
Ens_Net	0.06	-0.89	-0.96	-0.75	-0.75
	(0.21)	(-5.12***)	(-5.81***)	(-3.97***)	(-3.20***)
Ens_total	-0.07	-1.03	-1.10	-0.89	-0.89
	(-0.23)	(-5.79***)	(-6.64***)	(-4.68***)	(-3.74***)

Table 8: Risk-Adjusted Spread Portfolio Returns of Ensemble Models (Full Fs-timateion)

This table reports the complete risk-adjusted estimate of ensemble models for all three sets of predictors. Panel A through panel C report estimates for lagged return, moving average trading signals, and firm fundamentals respectively. Raw excess returns are adjusted by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel A1: Linear ensemble w/ 120 lagged returns

Factor	CAPM	FF3	Carhart4	FF5
Alpha(%)	1.26	1.21	1.21	1.17
	(6.33***)	(6.02***)	(4.82***)	(4.91***)
MKT	-0.02	-0.03	-0.03	0.00
	(-0.32)	(-0.33)	(-0.38)	(0.04)
SMB		0.14	0.14	0.11
		(0.73)	(0.73)	(0.88)
HML		0.10	0.10	-0.11
		(0.69)	(0.78)	(-0.61)
MOM			0.00	
			(0.02)	
RMW				-0.10
				(-0.45)
CMA				0.44
				(1.94*)
R2(%)	0.05	1.12	1.12	3.66

Panel A2: Network ensemble w/ 120 lagged returns

Factor	CAPM	FF3	Carhart4	FF5
Alpha(%)	1.69	1.65	1.64	1.63
	(6.23***)	(5.67***)	(5.19***)	(4.53***)
MKT	-0.15	-0.16	-0.16	-0.14
	(-1.54)	(-1.43)	(-1.64)	(-1.21)
SMB		0.19	0.19	0.12
		(0.74)	(0.75)	(0.86)
HML		0.10	0.10	-0.13
		(0.53)	(0.63)	(-0.59)
MOM			0.00	
			(0.01)	
RMW				-0.20
				(-0.65)
CMA				0.51
				(1.65*)
R2(%)	1.08	1.86	1.86	3.84

Table 8: Risk-Adjusted Spread Portfolio Returns of Ensemble Models (Full Fstimateion) - Cont'd

This table reports the complete risk-adjusted estimate of ensemble models for all three sets of predictors. Panel A through panel C report estimates for lagged return, moving average trading signals, and firm fundamentals respectively. Raw excess returns are adjusted by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel A3: Total Ensemble w/ 120 lagged returns

Factor	CAPM	FF3	Carhart4	FF5
Alpha(%)	1.83	1.75	1.75	1.73
	(5.95***)	(5.48***)	(4.91***)	(4.42***)
MKT	-0.12	-0.13	-0.13	-0.10
	(-1.13)	(-1.03)	(-1.16)	(-0.81)
SMB		0.23	0.23	0.17
		(0.8)	(0.8)	(0.96)
HML		0.15	0.16	-0.09
		(0.81)	(0.95)	(-0.37)
MOM			0.00	
			(0.01)	
RMW				-0.20
				(-0.6)
CMA				0.53
				(1.53)
R2(%)	0.60	1.68	1.68	3.39

Panel B1: Linear ensemble w/ 60 price and 60 trading volume MA signals

Factor	CAPM	FF3	Carhart4	FF5
Alpha(%)	1.16	1.15	1.04	0.93
	(4.89***)	(4.71***)	(2.99***)	(2.94***)
MKT	-0.10	-0.11	-0.08	-0.02
	(-0.99)	(-1.01)	(-0.92)	(-0.14)
SMB		0.06	0.05	0.06
		(0.3)	(0.26)	(0.41)
HML		0.02	0.07	-0.38
		(0.12)	(0.4)	(-1.64)
MOM			0.12	
			(0.52)	
RMW				0.03
				(0.12)
CMA				0.91
				(2.72***)
R2(%)	0.49	0.58	1.26	4.57

Table 8: Risk-Adjusted Spread Portfolio Returns of Ensemble Models (Full Fs-timateion) - Cont'd

This table reports the complete risk-adjusted estimate of ensemble models for all three sets of predictors. Panel A through panel B report estimates for lagged return, moving average trading signals, and firm fundamentals respectively. Raw excess returns are adjusted by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B2: Network ensemble w/ 60 price and 60 trading volume MA signals

Factor	CAPM	FF3	Carhart4	FF5
Alpha(%)	1.58	1.57	1.47	1.35
	(6.02***)	(5.74***)	(3.9***)	(3.63***)
MKT	-0.15	-0.16	-0.13	-0.05
	(-1.36)	(-1.34)	(-1.29)	(-0.43)
SMB		0.07	0.05	0.01
		(0.23)	(0.2)	(0.08)
HML		0.03	0.07	-0.46
		(0.12)	(0.36)	(-1.87*)
MOM			0.12	
			(0.47)	
RMW				-0.07
				(-0.19)
CMA				1.12
				(2.8***)
R2(%)	0.84	0.91	1.41	5.72

Panel B3: Total ensemble w/ 60 price and 60 trading volume MA signals

Factor	CAPM	FF3	Carhart4	FF5
Alpha(%)	1.97	1.97	1.86	1.81
	(6.58***)	(6.38***)	(4.54***)	(4.36***)
MKT	-0.13	-0.15	-0.12	-0.06
	(-1.1)	(-1.2)	(-1.09)	(-0.49)
SMB		0.12	0.11	0.06
		(0.4)	(0.37)	(0.31)
HML		-0.02	0.04	-0.47
		(-0.07)	(0.17)	(-1.69*)
MOM			0.14	
			(0.51)	
RMW				-0.14
				(-0.36)
CMA				1.03
				(2.25**)
R2(%)	0.48	0.69	1.22	4.23

Table 8: Risk-Adjusted Spread Portfolio Returns of Ensemble Models (Full Fstimateion) - Cont'd

This table reports the complete risk-adjusted estimate of ensemble models for all three sets of predictors. Panel A through panel B report estimates for lagged return, moving average trading signals, and firm fundamentals respectively. Raw excess returns are adjusted by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C1: Linear ensemble w/ 79 firm characteristics

Factor	CAPM	FF3	Carhart4	FF5
Alpha(%)	1.79	1.77	1.65	1.67
	(6.14***)	(5.7***)	(4.74***)	(4.25***)
MKT	-0.26	-0.27	-0.24	-0.20
	(-2.68***)	(-2.47**)	(-2.4**)	(-1.94*)
SMB		0.09	0.08	0.00
		(0.33)	(0.3)	(0.01)
HML		0.03	0.09	-0.34
		(0.15)	(0.54)	(-1.43)
MOM			0.15	
			(0.72)	
RMW				-0.19
				(-0.55)
CMA				0.86
				(2.41**)
R2(%)	2.72	2.86	3.68	6.37

Panel C2: Neural networks w/ 79 firm characteristics

Factor	CAPM	FF3	Carhart4	FF5
Alpha(%)	1.97	1.96	1.78	1.87
	(5.96***)	(5.5***)	(4.66***)	(4.21***)
MKT	-0.32	-0.34	-0.30	-0.26
	(-3.12***)	(-2.74***)	(-2.62***)	(-2.22**)
SMB		0.10	0.08	-0.03
		(0.3)	(0.26)	(-0.17)
HML		0.00	0.08	-0.41
		(0.01)	(0.48)	(-1.49)
MOM			0.21	
			(0.97)	
RMW				-0.27
				(-0.71)
CMA				0.97
				(2.48**)
R2(%)	3.12	3.25	4.50	6.87

Table 8: Risk-Adjusted Spread Portfolio Returns of Ensemble Models (Full Fstimateion) - Cont'd

This table reports the complete risk-adjusted estimate of ensemble models of network family (panel A) and combined linear and network family (panel B). Raw excess returns are adjusted by CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model. Newey-West t-statistics with 12 lags are reported in parentheses. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 12 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C3: Total ensemble w/ 79 firm characteristics

Factor	CAPM	FF3	Carhart4	FF5
Alpha(%)	2.17	2.17	2.00	2.07
	(6.38***)	(5.89***)	(5.05***)	(4.5***)
MKT	-0.32	-0.33	-0.29	-0.26
	(-2.97***)	(-2.62***)	(-2.51**)	(-2.11**)
SMB		0.10	0.08	-0.02
		(0.3)	(0.26)	(-0.11)
HML		0.00	0.08	-0.41
		(0.01)	(0.47)	(-1.43)
MOM			0.21	
			(0.91)	
RMW				-0.26
				(-0.65)
CMA				0.97
				(2.4**)
R2(%)	2.81	2.93	4.03	6.19

Table 9: Mean-Variance Spanning Test

This table explore whether ensemble portfolios can be spanned by the factors in CAPM, Fama and French (FF3, 1993) three-factor model, Carhart (Carhart, 1997) four-factor model, and Fama and French (FF5, 2015) five-factor model, respectively. Panel A1 through C3 present results for the combinations of three types of ensembles portfolios and three types of predictors. W is the Wald test under conditional homoskedasticity (continued next page)

Panel A1: Linear ensemble w/ 120 lagged returns

	W	W_e	W_a	J_1	J_2	J_3
CAPM	525.94***	71.53***	365.65***	21.25***	21.2***	86.98***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3	66.89***	34.09***	61.6***	18.52***	19.54***	21.07***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3M	53.29***	28.41***	34.75***	22.4***	23.87***	25.42***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF5	35.07***	25.5***	27.19***	17.71***	18.88***	18.6***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Panel A2: Network ensemble w/ 120 lagged returns

	W	W_e	W_a	J_1	J_2	J_3
CAPM	321.47***	65.05***	231.29***	27.35***	27.27***	65.95***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3	48.14***	30.42***	49.18***	20.07***	21.13***	21.75***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3M	39.75***	25.45***	30.2***	19.71***	21.1***	22.17***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF5	31.27***	24.27***	22.64***	17.99***	19.34***	18.88***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Panel A3: Total ensemble w/ 120 lagged returns

	W	W_e	W_a	J_1	J_2	J_3
CAPM	254.14***	49.68***	206.88***	26.35***	26.31***	51.78***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3	37.91***	26.88***	42.52***	20.37***	21.6***	21.09***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3M	32.55***	23.58***	26.86***	20.66***	22.21***	22.09***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF5	27.64***	23.32***	21.95***	19.31***	20.77***	19.81***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Table 9: Mean-Variance Spanning Test - Cont'd

W_e is the Wald test under IID elliptical, W_a is the Wald test under conditional heteroskedasticity, J_1 is the Bekerart-Urias test with the errors-in-variable adjustment, J_2 is the Bekerart-Urias test without the errors-in-variable adjustment, J_3 is the DeSantis test. P-values are reported in paranthesis. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models (continued next page)

Panel B1: Linear ensemble w/ 60 price and 60 trading volume MA signals

	W	W_e	W_a	J_1	J_2	J_3
CAPM	285.92***	68.06***	145.54***	25.82***	25.68***	71.48***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3	41.64***	19.32***	29.84***	12.67***	12.77***	14.49***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3M	23.41***	11.87***	9.99***	8.9**	9.25***	10.67***
	(0.0)	(0.0)	(0.01)	(0.01)	(0.01)	(0.0)
FF5	9.92***	8.32**	8.32**	7.09**	7.31**	7.18**
	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)

Panel B2: Network ensemble w/ 60 price and 60 trading volume MA signals

	W	W_e	W_a	J_1	J_2	J_3
CAPM	246.66***	57.74***	147.94***	26.64***	26.56***	59.98***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3	42.54***	23.9***	36.74***	15.62***	15.99***	17.22***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3M	26.88***	16.5***	13.89***	11.22***	11.82***	13.01***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF5	16.03***	14.29***	13.14***	11.12***	11.7***	11.26***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Panel B3: Total ensemble w/ 60 price and 60 trading volume MA signals

	W	W_e	W_a	J_1	J_2	J_3
CAPM	196.41***	52.8***	130.71***	28.85***	28.95***	53.18***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3	42.32***	27.87***	42.81***	19.22***	20.01***	20.51***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3M	28.58***	20.81***	17.56***	14.89***	15.97***	16.51***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF5	22.86***	20.5***	18.93***	15.86***	17.02***	16.13***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Table 9: Mean-Variance Spanning Test - Cont'd

in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C1: Linear ensemble w/ 79 firm characteristics

	W	W_e	W_a	J_1	J_2	J_3
CAPM	326.25***	116.33***	169.49***	33.39***	33.7***	96.47***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3	56.92***	35.43***	47.92***	22.95***	23.89***	26.32***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3M	35.48***	23.24***	21.8***	17.81***	19.09***	21.13***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF5	28.37***	22.4***	20.72***	17.37***	18.68***	18.51***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Panel C2: Network ensemble w/ 79 firm characteristics

	W	W_e	W_a	J_1	J_2	J_3
CAPM	273.35***	99.31***	147.58***	32.08***	32.43***	80.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3	51.2***	32.66***	43.2***	22.32***	23.22***	25.18***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3M	29.79***	20.44***	20.33***	16.3***	17.37***	18.57***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF5	27.31***	21.27***	19.09***	16.57***	17.81***	17.8***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Panel C3: Total ensemble w/ 79 firm characteristics

	W	W_e	W_a	J_1	J_2	J_3
CAPM	251.77***	97.43***	140.23***	33.47***	33.96***	77.37***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3	51.38***	34.77***	44.68***	23.57***	24.76***	26.36***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF3M	31.75***	23.16***	22.35***	18.2***	19.6***	20.54***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FF5	29.38***	24.11***	21.79***	18.5***	20.07***	19.62***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Table 10: Regression Analysis of Actual Return on Predicted Return

This table reports summary statistics from monthly regression of actual return on predicted return under each model. Statistics include average beta coefficients, t-statistic of the time-series of betas, standard deviation incorporating a Newey-West correction with 12 lags, minimum and maximum, proportion of positive betas, proportion of significant betas conditional on positivity at 1%, 5%, and 10% levels, and adjusted R-square. For each model, I estimate monthly cross-sectional regressions (480 regressions in total) and calculate the relevant statistics. Panel A reports results where 120 lagged returns are used as predictors, Panel B reports results for 60 price and 60 trading volume MA signals, and Panel C reports results for 79 firm characteristics.

Panel A Predictors:120 lagged returns

	Mean($\hat{\beta}$)	t-stat	Std.	Min	Max	Pos.	Sig1%.	Sig5%.	Sig10%.	AdjR2
OLS	0.07	6.90	0.22	-0.95	1.23	0.64	0.44	0.56	0.62	0.01
PCR	0.17	7.15	0.51	-1.46	2.54	0.63	0.46	0.56	0.62	0.01
PLS	0.19	8.24	0.51	-1.40	3.59	0.66	0.49	0.59	0.66	0.01
Ridge	0.07	6.90	0.22	-0.95	1.23	0.64	0.44	0.56	0.62	0.01
Lasso	0.10	7.35	0.29	-1.27	1.53	0.65	0.43	0.56	0.61	0.01
ENet	0.10	7.35	0.29	-1.27	1.53	0.65	0.43	0.56	0.61	0.01
NN1L	0.12	7.41	0.35	-1.41	1.44	0.66	0.46	0.58	0.64	0.01
NN2L	0.27	7.07	0.85	-2.86	4.40	0.64	0.45	0.59	0.65	0.01
NN3L	0.46	6.99	1.44	-5.43	6.90	0.66	0.43	0.58	0.65	0.01
NN4L	0.59	5.91	2.18	-7.81	10.82	0.62	0.48	0.56	0.66	0.01
NN5L	0.72	5.80	2.72	-9.05	15.29	0.62	0.42	0.55	0.61	0.01

Panel B Predictors: 60 price and 60 trading volume MA signals

	Mean($\hat{\beta}$)	t-stat	Std.	Min	Max	Pos.	Sig1%.	Sig5%.	Sig10%.	AdjR2
OLS	0.03	3.33	0.20	-1.15	1.15	0.59	0.29	0.41	0.50	0.01
PCR	0.31	1.93	3.53	-5.71	73.10	0.59	0.59	0.70	0.74	0.01
PLS	0.11	2.96	0.82	-4.28	4.55	0.57	0.60	0.70	0.76	0.02
Ridge	0.03	3.37	0.20	-1.15	1.15	0.59	0.30	0.42	0.51	0.01
Lasso	0.15	4.75	0.69	-2.14	2.84	0.61	0.58	0.66	0.72	0.02
ENet	0.15	4.76	0.69	-2.15	2.84	0.61	0.58	0.67	0.72	0.02
NN1L	0.15	5.62	0.57	-2.43	2.73	0.61	0.61	0.68	0.74	0.02
NN2L	0.20	5.36	0.81	-3.01	4.30	0.59	0.65	0.73	0.76	0.02
NN3L	0.26	4.86	1.17	-4.30	7.07	0.60	0.63	0.71	0.74	0.02
NN4L	0.36	4.48	1.78	-5.50	11.61	0.58	0.60	0.70	0.75	0.02
NN5L	0.48	3.95	2.67	-8.06	21.45	0.56	0.59	0.67	0.71	0.02

Panel C Predictors: 79 firm characteristics

	Mean($\hat{\beta}$)	t-stat	Std.	Min	Max	Pos.	Sig1%.	Sig5%.	Sig10%.	AdjR2
OLS	0.11	6.18	0.38	-1.23	1.93	0.63	0.64	0.74	0.77	0.02
PCR	0.18	5.86	0.66	-1.63	9.04	0.62	0.70	0.78	0.81	0.02
PLS	0.14	5.92	0.51	-1.49	2.22	0.62	0.67	0.78	0.81	0.02
Ridge	0.12	6.62	0.39	-1.23	1.93	0.64	0.65	0.75	0.80	0.02
Lasso	0.15	6.73	0.50	-1.36	2.08	0.64	0.68	0.78	0.80	0.02
ENet	0.15	6.73	0.50	-1.36	2.08	0.64	0.68	0.78	0.80	0.02
NN1L	0.16	6.50	0.54	-1.72	2.06	0.64	0.68	0.76	0.79	0.02
NN2L	0.23	6.08	0.82	-2.50	3.31	0.62	0.71	0.77	0.81	0.03
NN3L	0.34	5.67	1.31	-4.43	4.57	0.62	0.70	0.77	0.82	0.03
NN4L	0.58	5.82	2.18	-8.33	9.54	0.63	0.65	0.73	0.76	0.03
NN5L	0.77	5.72	2.97	-10.50	12.37	0.62	0.65	0.75	0.78	0.02

Table 11: Comparison of Out-of-Sample Prediction using Diebold-Mariano Test

This table reports pairwise Diebold-Mariano test statistics which compares the out-of-sample stock-level prediction performance among 11 models. Positive statistic indicates that the column model outperforms the row model. The test statistic is calculated based on the actual and predicted returns on a 40-year pooled sample. Panel A, B, and C present cases where 120 lagged returns, 60 price and 60 trading volume MA signals, and 79 firm characteristics are used as predictors respectively. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel A Predictors: 120 lagged returns

	PCR	PLS	Ridge	Lasso	ENet	NN1L	NN2L	NN3L	NN4L	NN5L
OLS	84.03***	98.44***	163.34***	142.03***	142.03***	110.72***	119.74***	109.10***	105.67***	103.18***
PCR		-15.41***	-84.03***	-49.71***	-49.71***	-28.85***	47.58***	54.75***	60.32***	59.05***
PLS			-98.44***	-53.90***	-53.90***	-24.19***	60.64***	61.31***	64.78***	63.36***
Ridge				142.03***	142.03***	110.72***	119.74***	109.10***	105.67***	103.18***
Lasso					nan	39.27***	91.56***	83.54***	81.55***	79.37***
ENet						39.27***	91.56***	83.54***	81.55***	79.37***
NN1L							77.51***	71.25***	69.76***	67.80***
NN2L								21.55***	31.02***	30.60***
NN3L									24.69***	25.58***
NN4L										11.70***

Panel B Predictors: 60 price and 60 trading volume MA signals

	PCR	PLS	Ridge	Lasso	ENet	NN1L	NN2L	NN3L	NN4L	NN5L
OLS	25.81***	25.86***	9.46***	27.43***	27.40***	28.19***	32.55***	35.02***	37.06***	37.93***
PCR		0.34	-26.24***	22.66***	22.32***	19.28***	51.94***	66.39***	72.82***	72.69***
PLS			-26.29***	22.49***	22.19***	18.88***	52.01***	67.54***	74.41***	74.22***
Ridge				27.92***	27.90***	28.70***	33.23***	35.78***	37.89***	38.78***
Lasso					-7.16***	8.65***	47.72***	64.79***	71.01***	70.23***
ENet						8.88***	47.90***	64.93***	71.12***	70.33***
NN1L							54.90***	62.49***	63.77***	61.02***
NN2L								40.54***	48.98***	46.87***
NN3L									41.43***	40.45***
NN4L										24.56***

Table 11: Comparison of Out-of-Sample Prediction using Diebold-Mariano Test - Cont'd

Panel C Predictors: 79 firm characteristics

	PCR	PLS	Ridge	Lasso	ENet	NN1L	NN2L	NN3L	NN4L	NN5L
OLS	16.00***	16.00***	16.00***	16.00***	16.00***	16.00***	16.00***	16.00***	16.00***	16.00***
PCR		4.52***	4.52***	4.52***	4.52***	4.52***	4.52***	4.52***	4.52***	4.52***
PLS			-43.38***	47.76***	47.75***	54.17***	100.43***	106.71***	100.07***	93.17***
Ridge				116.70***	116.69***	82.81***	114.46***	116.80***	109.09***	101.81***
Lasso					-11.04***	30.14***	89.07***	99.33***	93.46***	86.75***
ENet						30.16***	89.08***	99.34***	93.47***	86.75***
NN1L							69.15***	81.16***	74.77***	69.65***
NN2L								55.42***	57.42***	53.39***
NN3L									33.34***	35.81***
NN4L										25.73***

Table 12: Size Effect (Predictors: 120 lagged returns)

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 120 lagged returns. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks (in terms of market value at the end of previous month) within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel A: Bottom 20% Excluded (Predictors: 120 lagged returns)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.71	0.71	0.68	0.68	0.61
	(4.70***)	(4.94***)	(4.74***)	(3.98***)	(4.11***)
PCR	0.70	0.73	0.72	0.67	0.70
	(5.55***)	(5.68***)	(5.43***)	(4.84***)	(4.41***)
PLS	0.94	0.95	0.91	0.88	0.89
	(6.46***)	(6.49***)	(6.25***)	(5.46***)	(5.23***)
Ridge	0.71	0.71	0.68	0.68	0.61
	(4.70***)	(4.94***)	(4.74***)	(3.98***)	(4.11***)
Lasso	0.73	0.73	0.69	0.70	0.62
	(4.78***)	(5.01***)	(4.83***)	(4.04***)	(4.17***)
ENet	0.73	0.73	0.69	0.70	0.62
	(4.78***)	(5.01***)	(4.83***)	(4.04***)	(4.17***)
NN1L	0.82	0.85	0.82	0.83	0.78
	(5.23***)	(5.71***)	(5.50***)	(4.63***)	(5.02***)
NN2L	0.89	0.95	0.91	0.90	0.84
	(5.59***)	(6.05***)	(5.69***)	(5.30***)	(4.65***)
NN3L	0.86	0.91	0.89	0.85	0.84
	(5.09***)	(5.63***)	(5.30***)	(4.63***)	(4.00***)
NN4L	0.78	0.85	0.83	0.79	0.77
	(4.43***)	(5.18***)	(4.70***)	(4.01***)	(3.34***)
NN5L	0.75	0.82	0.80	0.77	0.75
	(5.11***)	(5.88***)	(5.47***)	(4.69***)	(3.87***)
Ens_linear	1.10	1.11	1.07	1.05	1.01
	(5.42***)	(5.50***)	(5.21***)	(4.49***)	(4.28***)
Ens_Net	1.37	1.47	1.43	1.42	1.41
	(5.07***)	(5.70***)	(5.19***)	(4.75***)	(4.21***)
Ens_total	1.33	1.39	1.34	1.34	1.30
	(4.28***)	(4.56***)	(4.21***)	(3.80***)	(3.43***)

Table 12: Size Effect (Predictors: 120 lagged returns) - Cont'd

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 120 lagged returns. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks (in terms of market value at the end of previous month) within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Bottom 40% Excluded (Predictors: 120 lagged returns)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.64	0.63	0.60	0.59	0.53
	(4.28***)	(4.39***)	(4.23***)	(3.50***)	(3.63***)
PCR	0.64	0.66	0.65	0.59	0.64
	(5.33***)	(5.47***)	(5.36***)	(4.69***)	(4.49***)
PLS	0.87	0.87	0.84	0.79	0.82
	(6.21***)	(6.06***)	(5.97***)	(5.24***)	(5.28***)
Ridge	0.64	0.63	0.60	0.59	0.53
	(4.28***)	(4.39***)	(4.23***)	(3.50***)	(3.63***)
Lasso	0.67	0.67	0.63	0.63	0.56
	(4.41***)	(4.50***)	(4.40***)	(3.64***)	(3.91***)
ENet	0.67	0.67	0.63	0.63	0.56
	(4.41***)	(4.50***)	(4.40***)	(3.64***)	(3.91***)
NN1L	0.74	0.75	0.73	0.72	0.69
	(4.94***)	(5.21***)	(5.11***)	(4.24***)	(4.78***)
NN2L	0.78	0.81	0.79	0.76	0.73
	(5.14***)	(5.45***)	(5.20***)	(4.74***)	(4.34***)
NN3L	0.77	0.81	0.80	0.75	0.75
	(4.67***)	(5.17***)	(4.97***)	(4.21***)	(3.85***)
NN4L	0.68	0.74	0.73	0.67	0.68
	(4.03***)	(4.74***)	(4.40***)	(3.66***)	(3.16***)
NN5L	0.67	0.73	0.72	0.67	0.67
	(4.78***)	(5.39***)	(5.14***)	(4.33***)	(3.74***)
Ens_linear	0.95	0.96	0.93	0.88	0.86
	(5.24***)	(5.21***)	(4.99***)	(4.23***)	(4.22***)
Ens_Net	1.28	1.36	1.33	1.29	1.34
	(4.82***)	(5.34***)	(4.97***)	(4.48***)	(4.31***)
Ens_total	1.12	1.17	1.13	1.09	1.10
	(3.74***)	(4.00***)	(3.65***)	(3.27***)	(3.14***)

Table 12: Size Effect (Predictors: 120 lagged returns) - Cont'd

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 120 lagged returns. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks (in terms of market value at the end of previous month) within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: Bottom 60% Excluded (Predictors: 120 lagged returns)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.53	0.52	0.48	0.47	0.44
	(3.50***)	(3.56***)	(3.31***)	(2.94***)	(2.97***)
PCR	0.51	0.52	0.51	0.42	0.49
	(3.97***)	(4.08***)	(3.97***)	(3.31***)	(3.41***)
PLS	0.73	0.74	0.70	0.64	0.70
	(5.43***)	(5.29***)	(5.07***)	(4.48***)	(4.65***)
Ridge	0.53	0.52	0.48	0.47	0.44
	(3.50***)	(3.56***)	(3.31***)	(2.94***)	(2.97***)
Lasso	0.56	0.55	0.51	0.49	0.45
	(3.72***)	(3.73***)	(3.50***)	(3.08***)	(3.08***)
ENet	0.56	0.55	0.51	0.49	0.45
	(3.72***)	(3.73***)	(3.50***)	(3.08***)	(3.08***)
NN1L	0.60	0.60	0.57	0.56	0.54
	(4.16***)	(4.32***)	(4.08***)	(3.60***)	(3.82***)
NN2L	0.68	0.70	0.69	0.63	0.63
	(4.76***)	(5.02***)	(4.79***)	(4.41***)	(4.11***)
NN3L	0.63	0.66	0.66	0.57	0.62
	(3.99***)	(4.37***)	(4.10***)	(3.67***)	(3.30***)
NN4L	0.59	0.64	0.64	0.56	0.60
	(3.81***)	(4.56***)	(4.16***)	(3.50***)	(2.99***)
NN5L	0.57	0.64	0.63	0.56	0.60
	(4.60***)	(5.32***)	(5.01***)	(4.36***)	(3.75***)
Ens_linear	0.86	0.85	0.82	0.75	0.76
	(4.54***)	(4.46***)	(4.18***)	(3.82***)	(3.75***)
Ens_Net	1.27	1.33	1.33	1.25	1.36
	(4.71***)	(5.11***)	(4.78***)	(4.43***)	(4.27***)
Ens_total	1.26	1.31	1.30	1.19	1.25
	(4.47***)	(4.65***)	(4.30***)	(3.89***)	(3.63***)

Table 12: Size Effect (Predictors: 120 lagged returns) - Cont'd

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 120 lagged returns. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks (in terms of market value at the end of previous month) within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel D: Bottom 80% Excluded (Predictors: 120 lagged returns)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.47	0.45	0.41	0.38	0.37
	(2.52**)	(2.47**)	(2.28**)	(2.03**)	(2.13**)
PCR	0.39	0.38	0.36	0.26	0.31
	(3.06***)	(3.05***)	(2.91***)	(2.05**)	(2.22**)
PLS	0.65	0.63	0.59	0.50	0.58
	(4.48***)	(4.32***)	(4.14***)	(3.36***)	(4.00***)
Ridge	0.47	0.45	0.41	0.38	0.37
	(2.52**)	(2.47**)	(2.28**)	(2.03**)	(2.13**)
Lasso	0.46	0.43	0.40	0.37	0.35
	(2.63**)	(2.52**)	(2.33**)	(2.16**)	(2.19**)
ENet	0.46	0.43	0.40	0.37	0.35
	(2.63**)	(2.52**)	(2.33**)	(2.16**)	(2.19**)
NN1L	0.45	0.43	0.39	0.36	0.37
	(2.66**)	(2.64**)	(2.50**)	(2.22**)	(2.45**)
NN2L	0.58	0.59	0.57	0.49	0.53
	(3.84***)	(3.99***)	(3.86***)	(3.63***)	(3.62***)
NN3L	0.52	0.51	0.51	0.42	0.50
	(3.29***)	(3.44***)	(3.30***)	(2.80**)	(2.75**)
NN4L	0.50	0.53	0.52	0.44	0.49
	(3.18***)	(3.78***)	(3.38***)	(2.80**)	(2.41**)
NN5L	0.44	0.47	0.46	0.39	0.44
	(3.39***)	(3.79***)	(3.55***)	(3.14***)	(2.74**)
Ens_linear	0.80	0.76	0.71	0.67	0.67
	(3.87***)	(3.76***)	(3.52***)	(3.36***)	(3.31***)
Ens_Net	0.93	0.89	0.86	0.79	0.88
	(3.02***)	(3.04***)	(2.79**)	(2.47**)	(2.62**)
Ens_total	0.90	0.84	0.81	0.70	0.70
	(2.55**)	(2.42**)	(2.25**)	(1.81*)	(1.82*)

Table 13: Size Effect (Predictors: 60 price and 60 trading volume MA signals)

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 60 price and 60 trading volume MA signals. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel A: Bottom 20% Excluded (Predictors: MA Signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.60	0.65	0.61	0.56	0.43
	(3.82***)	(4.81***)	(4.36***)	(2.61**)	(2.25**)
PCR	0.68	0.74	0.71	0.64	0.51
	(3.53***)	(4.25***)	(3.89***)	(2.35**)	(2.08**)
PLS	0.68	0.75	0.72	0.63	0.54
	(3.30***)	(3.98***)	(3.73***)	(2.31**)	(2.12**)
Ridge	0.60	0.65	0.61	0.57	0.44
	(3.82***)	(4.82***)	(4.37***)	(2.62**)	(2.26**)
Lasso	0.76	0.82	0.79	0.74	0.60
	(3.76***)	(4.63***)	(4.20***)	(2.68**)	(2.34**)
ENet	0.76	0.82	0.79	0.74	0.60
	(3.75***)	(4.62***)	(4.19***)	(2.67**)	(2.33**)
NN1L	0.95	1.03	0.99	0.95	0.80
	(4.72***)	(5.75***)	(5.12***)	(3.44***)	(3.05***)
NN2L	0.95	1.05	1.01	0.94	0.81
	(4.34***)	(5.26***)	(4.79***)	(3.16***)	(2.78**)
NN3L	0.89	0.98	0.96	0.88	0.76
	(3.94***)	(4.69***)	(4.37***)	(2.93***)	(2.50**)
NN4L	0.80	0.88	0.85	0.77	0.65
	(3.63***)	(4.21***)	(3.97***)	(2.63**)	(2.22**)
NN5L	0.75	0.84	0.82	0.75	0.62
	(3.56***)	(4.30***)	(3.95***)	(2.64**)	(2.12**)
Ens_linear	0.95	1.03	1.01	0.91	0.81
	(3.94***)	(4.64***)	(4.42***)	(2.84***)	(2.72**)
Ens_Net	1.34	1.44	1.42	1.31	1.20
	(5.09***)	(5.86***)	(5.49***)	(3.80***)	(3.34***)
Ens_total	1.67	1.74	1.72	1.60	1.57
	(5.24***)	(5.74***)	(5.46***)	(4.14***)	(3.75***)

Table 13: Size Effect (Predictors: 60 price and 60 trading volume MA signals) - Cont'd

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 60 price and 60 trading volume MA signals. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Bottom 40% Excluded (Predictors: MA Signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.56	0.61	0.58	0.51	0.41
	(3.62***)	(4.50***)	(4.12***)	(2.52**)	(2.16**)
PCR	0.60	0.65	0.63	0.54	0.44
	(3.11***)	(3.73***)	(3.38***)	(2.03**)	(1.77*)
PLS	0.56	0.63	0.61	0.50	0.44
	(2.79**)	(3.36***)	(3.13***)	(1.88*)	(1.73*)
Ridge	0.56	0.62	0.59	0.52	0.42
	(3.65***)	(4.55***)	(4.16***)	(2.55**)	(2.19**)
Lasso	0.64	0.70	0.68	0.61	0.49
	(3.18***)	(3.92***)	(3.56***)	(2.26**)	(1.94*)
ENet	0.64	0.70	0.68	0.60	0.49
	(3.18***)	(3.92***)	(3.56***)	(2.26**)	(1.93*)
NN1L	0.84	0.92	0.89	0.83	0.70
	(4.29***)	(5.22***)	(4.64***)	(3.11***)	(2.69**)
NN2L	0.80	0.90	0.88	0.78	0.67
	(3.80***)	(4.58***)	(4.16***)	(2.72**)	(2.28**)
NN3L	0.75	0.84	0.83	0.73	0.64
	(3.45***)	(4.16***)	(3.87***)	(2.52**)	(2.09**)
NN4L	0.66	0.75	0.73	0.63	0.54
	(3.21***)	(3.80***)	(3.55***)	(2.30**)	(1.88*)
NN5L	0.66	0.75	0.73	0.64	0.53
	(3.31***)	(4.02***)	(3.72***)	(2.43**)	(1.86*)
Ens_linear	0.83	0.91	0.89	0.76	0.70
	(3.56***)	(4.26***)	(4.02***)	(2.61**)	(2.41**)
Ens_Net	1.27	1.38	1.37	1.23	1.13
	(4.97***)	(5.75***)	(5.36***)	(3.61***)	(3.16***)
Ens_total	1.42	1.52	1.49	1.32	1.33
	(4.25***)	(4.74***)	(4.50***)	(3.43***)	(3.08***)

Table 13: Size Effect (Predictors: 60 price and 60 trading volume MA signals) - Cont'd

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 60 price and 60 trading volume MA signals. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: Bottom 60% Excluded (Predictors: MA Signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.49	0.54	0.51	0.44	0.33
	(3.06***)	(3.74***)	(3.47***)	(2.20**)	(1.75*)
PCR	0.49	0.54	0.52	0.40	0.34
	(2.51**)	(3.06***)	(2.76**)	(1.58)	(1.36)
PLS	0.45	0.50	0.49	0.36	0.32
	(2.22**)	(2.67**)	(2.50**)	(1.40)	(1.29)
Ridge	0.49	0.54	0.52	0.44	0.34
	(3.06***)	(3.76***)	(3.49***)	(2.22**)	(1.77*)
Lasso	0.52	0.58	0.57	0.46	0.38
	(2.54**)	(3.15***)	(2.89***)	(1.80*)	(1.48)
ENet	0.52	0.58	0.57	0.46	0.38
	(2.54**)	(3.16***)	(2.90***)	(1.80*)	(1.49)
NN1L	0.70	0.78	0.76	0.66	0.58
	(3.83***)	(4.78***)	(4.21***)	(2.71**)	(2.31**)
NN2L	0.67	0.77	0.75	0.63	0.56
	(3.31***)	(4.09***)	(3.71***)	(2.36**)	(1.97*)
NN3L	0.61	0.70	0.69	0.55	0.51
	(2.93***)	(3.61***)	(3.37***)	(2.08**)	(1.70*)
NN4L	0.53	0.61	0.60	0.45	0.42
	(2.67**)	(3.17***)	(2.97***)	(1.76*)	(1.47)
NN5L	0.54	0.64	0.63	0.51	0.43
	(2.91***)	(3.70***)	(3.37***)	(2.06**)	(1.55)
Ens_linear	0.69	0.77	0.75	0.61	0.54
	(2.86***)	(3.43***)	(3.23***)	(2.10**)	(1.86*)
Ens_Net	1.02	1.14	1.13	0.93	0.90
	(4.24***)	(4.95***)	(4.64***)	(2.87***)	(2.54**)
Ens_total	1.07	1.16	1.16	0.94	0.97
	(3.40***)	(3.77***)	(3.78***)	(2.57**)	(2.38**)

Table 13: Size Effect (Predictors: 60 price and 60 trading volume MA signals) - Cont'd

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 60 price and 60 trading volume MA signals. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel D: Bottom 80% Excluded (Predictors: MA Signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.47	0.51	0.50	0.38	0.35
	(2.75**)	(3.12***)	(2.97***)	(1.83*)	(1.70*)
PCR	0.43	0.47	0.46	0.32	0.29
	(2.24**)	(2.62**)	(2.40**)	(1.33)	(1.19)
PLS	0.37	0.42	0.41	0.25	0.26
	(1.85*)	(2.20**)	(2.04**)	(1.02)	(1.03)
Ridge	0.47	0.51	0.49	0.37	0.34
	(2.73**)	(3.10***)	(2.96***)	(1.81*)	(1.68*)
Lasso	0.48	0.53	0.52	0.38	0.35
	(2.33**)	(2.77**)	(2.56**)	(1.52)	(1.37)
ENet	0.47	0.52	0.52	0.38	0.35
	(2.31**)	(2.75**)	(2.54**)	(1.51)	(1.36)
NN1L	0.69	0.77	0.75	0.61	0.58
	(3.87***)	(4.61***)	(4.10***)	(2.68**)	(2.36**)
NN2L	0.60	0.67	0.66	0.51	0.50
	(2.93***)	(3.47***)	(3.21***)	(2.07**)	(1.77*)
NN3L	0.54	0.61	0.60	0.43	0.45
	(2.65**)	(3.16***)	(2.92***)	(1.75*)	(1.50)
NN4L	0.52	0.58	0.56	0.38	0.41
	(2.59**)	(2.89***)	(2.72**)	(1.56)	(1.40)
NN5L	0.48	0.55	0.53	0.37	0.35
	(2.55**)	(3.02***)	(2.69**)	(1.64)	(1.24)
Ens_linear	0.55	0.61	0.59	0.42	0.42
	(2.14**)	(2.48**)	(2.35**)	(1.47)	(1.37)
Ens_Net	0.75	0.82	0.83	0.57	0.62
	(2.83***)	(3.10***)	(3.04***)	(1.81*)	(1.71*)
Ens_total	0.87	0.91	0.89	0.64	0.75
	(2.45**)	(2.57**)	(2.55**)	(1.65)	(1.74*)

Table 14: Size Effect (Predictors: 79 firm characteristics)

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 79 firm characteristics. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel A: Bottom 20% Excluded (Predictors: 79 firm characteristics)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.14	1.30	1.29	1.18	1.20
	(4.50***)	(5.21***)	(4.82***)	(4.22***)	(3.49***)
PCR	1.17	1.34	1.33	1.21	1.25
	(4.69***)	(5.42***)	(4.96***)	(4.40***)	(3.65***)
PLS	1.14	1.30	1.29	1.18	1.21
	(4.56***)	(5.31***)	(4.91***)	(4.17***)	(3.57***)
Ridge	1.14	1.30	1.29	1.18	1.20
	(4.51***)	(5.22***)	(4.82***)	(4.23***)	(3.49***)
Lasso	1.19	1.37	1.36	1.25	1.26
	(4.54***)	(5.32***)	(4.90***)	(4.30***)	(3.56***)
ENet	1.19	1.37	1.36	1.25	1.26
	(4.54***)	(5.32***)	(4.90***)	(4.30***)	(3.56***)
NN1L	1.17	1.35	1.35	1.21	1.26
	(4.45***)	(5.21***)	(4.82***)	(4.12***)	(3.50***)
NN2L	1.16	1.34	1.34	1.20	1.25
	(4.32***)	(5.04***)	(4.68***)	(4.02***)	(3.42***)
NN3L	1.13	1.32	1.33	1.20	1.25
	(4.24***)	(5.02***)	(4.65***)	(4.08***)	(3.38***)
NN4L	1.11	1.30	1.29	1.19	1.22
	(4.26***)	(5.07***)	(4.67***)	(4.10***)	(3.39***)
NN5L	1.11	1.28	1.29	1.16	1.21
	(4.33***)	(5.04***)	(4.67***)	(4.17***)	(3.38***)
Ens_linear	1.43	1.60	1.60	1.48	1.52
	(5.02***)	(5.74***)	(5.31***)	(4.61***)	(3.97***)
Ens_Net	1.49	1.70	1.70	1.53	1.63
	(4.66***)	(5.41***)	(4.99***)	(4.36***)	(3.81***)
Ens_total	1.71	1.92	1.92	1.75	1.86
	(5.15***)	(5.85***)	(5.39***)	(4.82***)	(4.13***)

Table 14: Size Effect (Predictors: 79 firm characteristics) - Cont'd

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 79 firm characteristics. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Bottom 40% Excluded (Predictors: 79 firm characteristics)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.95	1.12	1.12	1.00	1.03
	(4.09***)	(4.86***)	(4.50***)	(3.95***)	(3.19***)
PCR	0.96	1.13	1.13	1.00	1.04
	(4.22***)	(5.06***)	(4.66***)	(4.08***)	(3.34***)
PLS	0.94	1.09	1.09	0.97	1.00
	(4.05***)	(4.87***)	(4.52***)	(3.77***)	(3.19***)
Ridge	0.95	1.12	1.12	1.00	1.03
	(4.07***)	(4.86***)	(4.50***)	(3.96***)	(3.20***)
Lasso	0.98	1.16	1.15	1.04	1.05
	(4.06***)	(4.92***)	(4.52***)	(3.95***)	(3.17***)
ENet	0.98	1.16	1.15	1.04	1.05
	(4.06***)	(4.92***)	(4.52***)	(3.95***)	(3.17***)
NN1L	0.97	1.14	1.15	1.01	1.06
	(3.95***)	(4.75***)	(4.41***)	(3.73***)	(3.13***)
NN2L	0.95	1.13	1.15	0.99	1.06
	(3.76***)	(4.55***)	(4.26***)	(3.62***)	(3.06***)
NN3L	0.93	1.11	1.13	0.99	1.04
	(3.70***)	(4.53***)	(4.23***)	(3.66***)	(3.00***)
NN4L	0.90	1.08	1.10	0.97	1.02
	(3.79***)	(4.65***)	(4.29***)	(3.78***)	(3.03***)
NN5L	0.90	1.07	1.09	0.95	1.01
	(3.81***)	(4.58***)	(4.26***)	(3.79***)	(3.00***)
Ens_linear	1.19	1.36	1.37	1.23	1.27
	(4.43***)	(5.23***)	(4.87***)	(4.23***)	(3.55***)
Ens_Net	1.24	1.46	1.47	1.28	1.40
	(4.14***)	(4.97***)	(4.59***)	(4.05***)	(3.46***)
Ens_total	1.44	1.66	1.67	1.47	1.58
	(4.66***)	(5.43***)	(5.01***)	(4.53***)	(3.76***)

Table 14: Size Effect (Predictors: 79 firm characteristics) - Cont'd

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 79 firm characteristics. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: Bottom 60% Excluded (Predictors: 79 firm characteristics)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.78	0.93	0.95	0.83	0.86
	(3.89***)	(4.76***)	(4.50***)	(3.82***)	(3.09***)
PCR	0.79	0.94	0.95	0.82	0.85
	(4.07***)	(4.97***)	(4.65***)	(3.92***)	(3.20***)
PLS	0.76	0.91	0.93	0.80	0.83
	(3.64***)	(4.51***)	(4.30***)	(3.49***)	(2.96***)
Ridge	0.78	0.93	0.95	0.83	0.86
	(3.89***)	(4.77***)	(4.50***)	(3.84***)	(3.10***)
Lasso	0.83	0.99	1.00	0.88	0.89
	(4.01***)	(4.96***)	(4.59***)	(3.90***)	(3.13***)
ENet	0.83	0.99	1.00	0.88	0.89
	(4.01***)	(4.96***)	(4.59***)	(3.89***)	(3.13***)
NN1L	0.82	0.98	1.00	0.85	0.91
	(3.76***)	(4.63***)	(4.36***)	(3.64***)	(3.04***)
NN2L	0.78	0.95	0.97	0.81	0.88
	(3.45***)	(4.29***)	(4.08***)	(3.35***)	(2.82***)
NN3L	0.81	0.98	1.00	0.86	0.92
	(3.56***)	(4.43***)	(4.20***)	(3.53***)	(2.93***)
NN4L	0.72	0.88	0.90	0.77	0.83
	(3.38***)	(4.31***)	(4.00***)	(3.42***)	(2.70**)
NN5L	0.79	0.94	0.98	0.83	0.89
	(3.67***)	(4.45***)	(4.24***)	(3.71***)	(2.92***)
Ens_linear	1.02	1.20	1.23	1.08	1.10
	(4.32***)	(5.19***)	(4.99***)	(4.19***)	(3.56***)
Ens_Net	1.04	1.26	1.28	1.08	1.20
	(3.73***)	(4.55***)	(4.36***)	(3.68***)	(3.33***)
Ens_total	1.17	1.40	1.44	1.23	1.33
	(4.10***)	(4.90***)	(4.73***)	(4.09***)	(3.54***)

Table 14: Size Effect (Predictors: 79 firm characteristics) - Cont'd

This table reports the alphas of the spread portfolios constructed on a sample with certain percentile of the smallest stocks in each month removed. Predictors used are 79 firm characteristics. Panel A to panel D are based on sub-samples with 20%, 40%, 60%, and 80% of the smallest stocks within each month removed. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted returns from each model and construct spread portfolios by buying the top quintile and selling the bottom quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile unanimously agreed by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel D: Bottom 80% Excluded (Predictors: 79 firm characteristics)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.69	0.82	0.83	0.70	0.74
	(3.46***)	(4.22***)	(3.94***)	(3.37***)	(2.72**)
PCR	0.65	0.77	0.78	0.63	0.66
	(3.52***)	(4.25***)	(3.99***)	(3.34***)	(2.68**)
PLS	0.62	0.74	0.76	0.62	0.64
	(3.17***)	(3.94***)	(3.77***)	(3.01***)	(2.50**)
Ridge	0.70	0.82	0.84	0.70	0.74
	(3.49***)	(4.26***)	(3.97***)	(3.41***)	(2.75**)
Lasso	0.75	0.90	0.91	0.76	0.79
	(3.67***)	(4.55***)	(4.26***)	(3.64***)	(2.90***)
ENet	0.75	0.90	0.91	0.76	0.79
	(3.67***)	(4.55***)	(4.26***)	(3.64***)	(2.90***)
NN1L	0.72	0.86	0.87	0.71	0.78
	(3.49***)	(4.26***)	(4.01***)	(3.37***)	(2.82***)
NN2L	0.70	0.84	0.87	0.69	0.76
	(3.23***)	(4.05***)	(3.90***)	(3.19***)	(2.65**)
NN3L	0.68	0.83	0.86	0.69	0.78
	(3.11***)	(3.93***)	(3.79***)	(3.23***)	(2.60**)
NN4L	0.59	0.73	0.76	0.61	0.67
	(2.82***)	(3.59***)	(3.45***)	(2.93***)	(2.34**)
NN5L	0.63	0.75	0.78	0.62	0.69
	(3.09***)	(3.73***)	(3.61***)	(3.09***)	(2.42**)
Ens_linear	0.86	1.01	1.03	0.86	0.91
	(3.81***)	(4.58***)	(4.39***)	(3.73***)	(3.16***)
Ens_Net	0.88	1.06	1.10	0.86	1.03
	(3.18***)	(3.95***)	(3.86***)	(3.29***)	(2.95***)
Ens_total	0.90	1.07	1.12	0.86	1.03
	(3.07***)	(3.77***)	(3.67***)	(3.17***)	(2.78**)

Table 15: Risk-Adjusted Spread Portfolio Returns over High and Low Investor Sentiment Periods (Predictors: 120 Lagged Return)

This table reports risk-adjusted spread portfolio returns for high versus low investment sentiment periods where 120 lagged returns are used for predictors. Panel A and B report high and low investor sentiment respectively. A given month is labeled high sentiment if the Baker-Wurgler sentiment index in the previous month is above the median value of the sample period (1978:01 - 2015:09) and is labeled low sentiment otherwise. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I

Panel A: High Sentiment Period (Predictors: 120 Lagged Return)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.70	0.74	0.72	0.59	0.61
	(3.82***)	(4.17***)	(3.49***)	(3.49***)	(2.81***)
PCR	0.78	0.83	0.78	0.66	0.74
	(4.87***)	(4.99***)	(3.33***)	(3.31***)	(2.80**)
PLS	1.08	1.14	1.08	0.92	1.04
	(5.72***)	(6.10***)	(4.53***)	(4.84***)	(3.90***)
Ridge	0.70	0.74	0.72	0.59	0.61
	(3.82***)	(4.17***)	(3.49***)	(3.49***)	(2.81***)
Lasso	0.72	0.76	0.71	0.58	0.60
	(3.91***)	(4.27***)	(3.46***)	(3.46***)	(2.74**)
ENet	0.72	0.76	0.71	0.58	0.60
	(3.91***)	(4.27***)	(3.46***)	(3.46***)	(2.74**)
NN1L	0.85	0.92	0.89	0.78	0.83
	(4.39***)	(4.85***)	(3.96***)	(4.17***)	(3.41***)
NN2L	1.04	1.13	1.09	0.97	1.01
	(4.94***)	(5.24***)	(4.01***)	(4.20***)	(3.38***)
NN3L	0.89	0.99	0.95	0.81	0.91
	(3.96***)	(4.03***)	(3.00***)	(3.00***)	(2.59**)
NN4L	0.81	0.91	0.83	0.69	0.79
	(3.28***)	(3.34***)	(2.26**)	(2.13**)	(1.91*)
NN5L	0.74	0.83	0.77	0.63	0.70
	(3.27***)	(3.44***)	(2.39**)	(2.27**)	(1.90*)
Ens_linear	1.23	1.29	1.24	1.06	1.19
	(5.37***)	(5.54***)	(3.85***)	(3.90***)	(3.30***)
Ens_Net	1.41	1.56	1.54	1.37	1.53
	(3.66***)	(3.83***)	(2.91***)	(2.90***)	(2.60**)
Ens_total	1.59	1.73	1.68	1.48	1.67
	(3.51***)	(3.65***)	(2.70**)	(2.64**)	(2.46**)

Table 15: Risk-Adjusted Spread Portfolio Returns over High and Low Investor Sentiment Periods (Predictors: 120 Lagged Return) (Cont'd)

independently sort stocks into quintiles on predicted returns from each model and construct spread portfolios by buying the winner quintile and selling the loser quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Low Sentiment Period (Predictors: 120 Lagged Return)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.84	0.72	0.73	0.83	0.73
	(3.32***)	(2.94***)	(2.83***)	(2.92***)	(2.86***)
PCR	0.90	0.83	0.83	0.87	0.85
	(4.38***)	(4.27***)	(4.11***)	(3.93***)	(4.05***)
PLS	1.07	0.95	0.96	1.05	0.98
	(4.66***)	(4.27***)	(4.11***)	(3.99***)	(3.94***)
Ridge	0.84	0.72	0.73	0.83	0.73
	(3.32***)	(2.94***)	(2.83***)	(2.92***)	(2.86***)
Lasso	0.92	0.79	0.80	0.89	0.79
	(3.59***)	(3.15***)	(3.04***)	(3.09***)	(3.06***)
ENet	0.92	0.79	0.80	0.89	0.79
	(3.59***)	(3.15***)	(3.04***)	(3.09***)	(3.06***)
NN1L	0.98	0.87	0.86	0.96	0.87
	(3.65***)	(3.34***)	(3.23***)	(3.29***)	(3.37***)
NN2L	1.11	0.97	0.95	1.04	0.96
	(4.15***)	(3.86***)	(3.71***)	(3.71***)	(3.71***)
NN3L	1.16	1.02	0.98	1.04	0.99
	(4.40***)	(4.15***)	(4.09***)	(3.78***)	(3.96***)
NN4L	1.14	1.03	0.99	1.06	1.02
	(4.06***)	(3.91***)	(3.85***)	(3.65***)	(3.78***)
NN5L	1.03	0.93	0.90	0.97	0.92
	(4.30***)	(4.11***)	(4.06***)	(3.98***)	(4.07***)
Ens_linear	1.32	1.18	1.19	1.32	1.17
	(3.77***)	(3.53***)	(3.40***)	(3.34***)	(3.31***)
Ens_Net	1.90	1.70	1.69	1.80	1.74
	(4.45***)	(4.29***)	(4.23***)	(4.17***)	(4.17***)
Ens_total	2.02	1.80	1.80	1.94	1.83
	(4.23***)	(4.01***)	(3.90***)	(3.84***)	(3.83***)

Table 16: Risk-Adjusted Spread Portfolio Returns over High and Low Investor Sentiment Periods (Predictors: MA Signals)

This table reports risk-adjusted spread portfolio returns for high versus low investment sentiment periods where 60 price and 60 trading volume MA signals are used for predictors. Panel A and B report high and low investor sentiment respectively. A given month is labeled high sentiment if the Baker-Wurgler sentiment index in the previous month is above the median value of the sample period (1978:01 - 2015:09) and is labeled low sentiment otherwise. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each

Panel A: High Sentiment Period (Predictors: MA signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.42	0.51	0.36	0.15	0.14
	(1.92*)	(2.29**)	(1.26)	(0.59)	(0.38)
PCR	0.41	0.51	0.37	0.09	0.13
	(1.59)	(1.89*)	(1.00)	(0.28)	(0.29)
PLS	0.41	0.51	0.39	0.10	0.19
	(1.57)	(1.85*)	(1.03)	(0.33)	(0.41)
Ridge	0.42	0.51	0.36	0.15	0.14
	(1.92*)	(2.29**)	(1.26)	(0.59)	(0.39)
Lasso	0.47	0.57	0.44	0.17	0.20
	(1.70*)	(1.98*)	(1.11)	(0.50)	(0.40)
ENet	0.48	0.57	0.44	0.17	0.20
	(1.70*)	(1.98*)	(1.11)	(0.50)	(0.41)
NN1L	0.73	0.85	0.66	0.42	0.45
	(3.00***)	(3.31***)	(1.87*)	(1.33)	(1.03)
NN2L	0.76	0.88	0.71	0.43	0.51
	(2.90***)	(3.13***)	(1.83*)	(1.22)	(1.03)
NN3L	0.73	0.85	0.70	0.40	0.49
	(2.68**)	(2.92***)	(1.75*)	(1.14)	(0.97)
NN4L	0.63	0.73	0.57	0.29	0.36
	(2.35**)	(2.57**)	(1.46)	(0.83)	(0.74)
NN5L	0.60	0.72	0.57	0.28	0.35
	(2.23**)	(2.51**)	(1.45)	(0.80)	(0.72)
Ens_linear	0.61	0.71	0.61	0.28	0.38
	(1.95*)	(2.18**)	(1.44)	(0.77)	(0.71)
Ens_Net	0.98	1.12	0.91	0.56	0.71
	(2.86***)	(3.03***)	(1.87*)	(1.34)	(1.18)
Ens_total	1.27	1.39	1.24	0.85	1.10
	(3.26***)	(3.38***)	(2.33**)	(1.83*)	(1.68*)

Table 16: Risk-Adjusted Spread Portfolio Returns over High and Low Investor Sentiment Periods (Predictors: MA Signals) (Cont'd)

month, I independently sort stocks into quintiles on predicted returns from each model and construct spread portfolios by buying the winner quintile and selling the loser quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Low Sentiment Period (Predictors: MA signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.01	0.90	0.86	0.96	0.81
	(4.14***)	(4.19***)	(4.10***)	(3.67***)	(3.79***)
PCR	1.19	1.04	0.98	1.08	0.92
	(3.69***)	(3.67***)	(3.64***)	(3.11***)	(3.37***)
PLS	1.14	1.02	0.97	1.05	0.89
	(3.35***)	(3.34***)	(3.32***)	(2.83***)	(2.88***)
Ridge	1.02	0.90	0.86	0.96	0.82
	(4.15***)	(4.20***)	(4.11***)	(3.66***)	(3.78***)
Lasso	1.33	1.18	1.13	1.24	1.08
	(4.11***)	(4.15***)	(4.10***)	(3.62***)	(3.77***)
ENet	1.33	1.18	1.13	1.24	1.08
	(4.12***)	(4.17***)	(4.12***)	(3.63***)	(3.78***)
NN1L	1.45	1.31	1.25	1.38	1.21
	(4.39***)	(4.57***)	(4.61***)	(4.18***)	(4.22***)
NN2L	1.44	1.33	1.25	1.36	1.20
	(3.91***)	(3.90***)	(4.03***)	(3.61***)	(3.68***)
NN3L	1.35	1.24	1.16	1.28	1.12
	(3.62***)	(3.59***)	(3.66***)	(3.32***)	(3.34***)
NN4L	1.23	1.11	1.03	1.13	0.97
	(3.32***)	(3.24***)	(3.31***)	(2.99***)	(2.97***)
NN5L	1.16	1.05	0.97	1.08	0.94
	(3.41***)	(3.47***)	(3.55***)	(3.23***)	(3.11***)
Ens_linear	1.67	1.53	1.46	1.56	1.38
	(4.05***)	(4.07***)	(4.15***)	(3.56***)	(3.66***)
Ens_Net	2.04	1.90	1.80	1.93	1.76
	(4.84***)	(4.88***)	(4.99***)	(4.37***)	(4.69***)
Ens_total	2.65	2.53	2.40	2.54	2.40
	(5.62***)	(5.65***)	(5.79***)	(5.19***)	(5.45***)

Table 17: Risk-Adjusted Spread Portfolio Returns over High and Low Investor Sentiment Periods (Predictors: 79 firm characteristics)

This table reports risk-adjusted spread portfolio returns for high versus low investment sentiment periods where 86 firm fundamental variables are used for predictors. Panel A and B report high and low investor sentiment respectively. A given month is labeled high sentiment if the Baker-Wurgler sentiment index in the previous month is above the median value of the sample period (1978:01 - 2015:09) and is labeled low sentiment otherwise. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I

Panel A: High Sentiment Period (Predictors: 79 Firm Fundamentals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.36	1.50	1.35	1.10	1.27
	(4.43***)	(4.38***)	(3.23***)	(2.95***)	(2.62**)
PCR	1.43	1.58	1.41	1.16	1.35
	(4.47***)	(4.41***)	(3.28***)	(3.00***)	(2.78**)
PLS	1.31	1.46	1.28	1.02	1.24
	(4.13***)	(4.22***)	(3.06***)	(2.71**)	(2.52**)
Ridge	1.36	1.50	1.35	1.10	1.27
	(4.42***)	(4.37***)	(3.23***)	(2.95***)	(2.63**)
Lasso	1.45	1.61	1.46	1.20	1.38
	(4.52***)	(4.47***)	(3.34***)	(3.08***)	(2.76**)
ENet	1.45	1.61	1.46	1.20	1.38
	(4.52***)	(4.47***)	(3.34***)	(3.08***)	(2.76**)
NN1L	1.43	1.58	1.41	1.13	1.36
	(4.37***)	(4.32***)	(3.12***)	(2.83***)	(2.63**)
NN2L	1.37	1.52	1.38	1.10	1.33
	(4.03***)	(3.94***)	(2.89***)	(2.57**)	(2.49**)
NN3L	1.35	1.51	1.38	1.13	1.35
	(3.89***)	(3.90***)	(2.90***)	(2.65**)	(2.55**)
NN4L	1.34	1.51	1.38	1.14	1.34
	(3.83***)	(3.86***)	(2.90***)	(2.66**)	(2.58**)
NN5L	1.28	1.43	1.32	1.08	1.28
	(3.75***)	(3.71***)	(2.84***)	(2.57**)	(2.49**)
Ens_linear	1.59	1.75	1.56	1.28	1.52
	(4.54***)	(4.49***)	(3.37***)	(3.09***)	(2.84***)
Ens_Net	1.69	1.87	1.70	1.38	1.74
	(4.11***)	(4.03***)	(3.09***)	(2.84***)	(2.86***)
Ens_total	1.89	2.07	1.88	1.56	1.92
	(4.43***)	(4.32***)	(3.32***)	(3.08***)	(3.05***)

Table 17: Risk-Adjusted Spread Portfolio Returns over High and Low Investor Sentiment Periods (Predictors: 79 firm characteristics) (Cont'd)

independently sort stocks into quintiles on predicted returns from each model and construct spread portfolios by buying the winner quintile and selling the loser quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Low Sentiment Period (Predictors: 79 Firm Fundamentals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.42	1.40	1.31	1.36	1.33
	(3.90***)	(3.80***)	(4.03***)	(3.72***)	(3.88***)
PCR	1.46	1.45	1.36	1.40	1.40
	(4.09***)	(4.13***)	(4.36***)	(3.95***)	(4.15***)
PLS	1.50	1.48	1.38	1.43	1.40
	(4.09***)	(4.02***)	(4.21***)	(3.79***)	(3.98***)
Ridge	1.41	1.40	1.31	1.36	1.33
	(3.86***)	(3.80***)	(4.02***)	(3.71***)	(3.87***)
Lasso	1.49	1.49	1.39	1.45	1.41
	(3.84***)	(3.89***)	(4.11***)	(3.80***)	(3.85***)
ENet	1.49	1.49	1.39	1.45	1.41
	(3.84***)	(3.89***)	(4.11***)	(3.80***)	(3.85***)
NN1L	1.43	1.44	1.33	1.36	1.35
	(3.85***)	(3.79***)	(3.99***)	(3.56***)	(3.79***)
NN2L	1.51	1.53	1.42	1.44	1.45
	(3.91***)	(3.89***)	(4.12***)	(3.66***)	(3.90***)
NN3L	1.49	1.51	1.41	1.43	1.45
	(3.96***)	(3.92***)	(4.17***)	(3.74***)	(3.90***)
NN4L	1.43	1.41	1.31	1.35	1.37
	(3.79***)	(3.67***)	(3.90***)	(3.53***)	(3.69***)
NN5L	1.48	1.46	1.37	1.39	1.41
	(4.02***)	(3.84***)	(4.14***)	(3.73***)	(3.84***)
Ens_linear	1.84	1.83	1.73	1.79	1.75
	(4.38***)	(4.36***)	(4.58***)	(4.09***)	(4.33***)
Ens_Net	1.97	2.01	1.90	1.91	1.88
	(4.11***)	(4.05***)	(4.19***)	(3.72***)	(3.88***)
Ens_total	2.16	2.20	2.08	2.10	2.06
	(4.42***)	(4.40***)	(4.57***)	(4.04***)	(4.23***)

Table 18: Risk-Adjusted Spread Portfolio Returns over Expansion and Recession Periods (Predictors: 120 Lagged Return)

This table reports risk-adjusted spread portfolio returns (alpha in %) for expansion and recession periods where 120 lagged return are used for predictors. A given month is labeled expansion or recession based on the business cycle definition on National Bureau of of Economic Research (NBER) website. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted return from each model and construct spread portfolios by buying

Panel A: Expansion Period (Predictors: 120 Lagged Return)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.65	0.68	0.69	0.62	0.66
	(4.46***)	(4.72***)	(4.68***)	(4.32***)	(4.23***)
PCR	0.75	0.81	0.83	0.75	0.82
	(5.41***)	(5.54***)	(5.37***)	(5.17***)	(4.71***)
PLS	0.97	1.03	1.05	0.96	1.04
	(6.42***)	(6.64***)	(6.40***)	(6.37***)	(5.64***)
Ridge	0.65	0.68	0.69	0.62	0.66
	(4.46***)	(4.72***)	(4.68***)	(4.32***)	(4.23***)
Lasso	0.69	0.72	0.72	0.65	0.69
	(4.68***)	(4.95***)	(4.95***)	(4.47***)	(4.47***)
ENet	0.69	0.72	0.72	0.65	0.69
	(4.68***)	(4.95***)	(4.95***)	(4.47***)	(4.47***)
NN1L	0.76	0.83	0.84	0.79	0.82
	(5.15***)	(5.79***)	(5.52***)	(5.28***)	(5.08***)
NN2L	0.90	1.00	1.02	0.97	0.97
	(5.22***)	(5.83***)	(5.33***)	(5.50***)	(4.67***)
NN3L	0.89	0.99	1.03	0.97	1.00
	(5.23***)	(6.00***)	(5.41***)	(5.35***)	(4.48***)
NN4L	0.80	0.92	0.94	0.86	0.90
	(4.30***)	(4.93***)	(4.30***)	(4.33***)	(3.49***)
NN5L	0.74	0.84	0.86	0.80	0.82
	(4.61***)	(5.22***)	(4.71***)	(4.52***)	(3.79***)
Ens_linear	1.11	1.17	1.19	1.07	1.16
	(5.45***)	(5.59***)	(5.27***)	(5.09***)	(4.60***)
Ens_Net	1.50	1.70	1.74	1.64	1.75
	(5.20***)	(6.04***)	(5.47***)	(5.52***)	(4.79***)
Ens_total	1.63	1.81	1.83	1.69	1.83
	(5.02***)	(5.55***)	(5.10***)	(5.12***)	(4.52***)

Table 18: Risk-Adjusted Spread Portfolio Returns over Expansion and Recession Periods (Predictors: 120 Lagged Return) (Cont'd)

the best-predicted quintile and selling the worst-predicted quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Recession Period (Predictors: 120 lagged returns)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.75	1.82	1.83	1.79	1.36
	(2.77**)	(2.81***)	(2.86***)	(3.53***)	(2.26**)
PCR	1.24	1.27	1.36	1.35	1.09
	(2.31**)	(2.42**)	(2.73**)	(2.94***)	(2.36**)
PLS	1.71	1.77	1.82	1.80	1.43
	(2.75**)	(3.00***)	(3.18***)	(3.56***)	(2.66**)
Ridge	1.75	1.82	1.83	1.79	1.36
	(2.77**)	(2.81***)	(2.86***)	(3.53***)	(2.26**)
Lasso	1.79	1.86	1.90	1.86	1.38
	(2.83***)	(2.93***)	(3.09***)	(3.67***)	(2.36**)
ENet	1.79	1.86	1.90	1.86	1.38
	(2.83***)	(2.93***)	(3.09***)	(3.67***)	(2.36**)
NN1L	2.02	2.06	2.10	2.07	1.66
	(2.94***)	(2.94***)	(3.05***)	(3.63***)	(2.52**)
NN2L	2.01	2.04	2.13	2.12	1.64
	(3.31***)	(3.32***)	(3.42***)	(3.90***)	(2.70**)
NN3L	1.84	1.85	1.92	1.92	1.48
	(2.82***)	(2.75**)	(2.69**)	(2.85***)	(2.03**)
NN4L	1.78	1.78	1.85	1.84	1.54
	(2.62**)	(2.49**)	(2.48**)	(2.84***)	(2.02**)
NN5L	1.69	1.70	1.79	1.78	1.45
	(2.85***)	(2.74**)	(2.81***)	(3.05***)	(2.28**)
Ens_linear	2.26	2.34	2.35	2.30	1.94
	(2.69**)	(2.77**)	(2.95***)	(3.67***)	(2.86***)
Ens_Net	2.32	2.40	2.42	2.39	1.90
	(2.00*)	(2.08**)	(2.12**)	(2.29**)	(1.96*)
Ens_total	2.59	2.69	2.73	2.67	2.20
	(2.18**)	(2.28**)	(2.33**)	(2.65**)	(2.21**)

Table 19: Risk-Adjusted Spread Portfolio Returns over Expansion and Recession Periods (Predictors: MA Signals)

This table reports risk-adjusted spread portfolio returns (alpha in %) for expansion and recession periods where 60 price and 60 trading volume moving average signals are used for predictors. A given month is labeled expansion or recession based on the business cycle definition on National Bureau of of Economic Research (NBER) website. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted return from each model and

Panel A: Expansion Period (Predictors: MA Signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	0.57	0.66	0.66	0.55	0.55
	(3.88***)	(4.62***)	(4.10***)	(3.01***)	(2.61**)
PCR	0.66	0.76	0.79	0.62	0.67
	(3.36***)	(3.87***)	(3.58***)	(2.60**)	(2.47**)
PLS	0.65	0.76	0.80	0.62	0.69
	(3.12***)	(3.66***)	(3.46***)	(2.50**)	(2.44**)
Ridge	0.58	0.66	0.66	0.55	0.55
	(3.88***)	(4.64***)	(4.10***)	(3.03***)	(2.62**)
Lasso	0.77	0.87	0.90	0.76	0.78
	(3.84***)	(4.55***)	(4.11***)	(3.08***)	(2.82***)
ENet	0.77	0.87	0.90	0.76	0.79
	(3.85***)	(4.57***)	(4.13***)	(3.09***)	(2.84***)
NN1L	0.93	1.06	1.07	0.96	0.96
	(4.56***)	(5.49***)	(4.70***)	(3.71***)	(3.40***)
NN2L	0.94	1.08	1.11	0.95	0.99
	(4.15***)	(4.99***)	(4.40***)	(3.36***)	(3.11***)
NN3L	0.88	1.01	1.06	0.90	0.95
	(3.73***)	(4.45***)	(4.05***)	(3.11***)	(2.89***)
NN4L	0.78	0.90	0.94	0.77	0.82
	(3.33***)	(3.85***)	(3.59***)	(2.69**)	(2.57**)
NN5L	0.77	0.91	0.95	0.80	0.85
	(3.40***)	(4.18***)	(3.83***)	(2.84***)	(2.68**)
Ens_linear	0.94	1.05	1.12	0.92	1.00
	(3.82***)	(4.35***)	(4.11***)	(3.06***)	(3.00***)
Ens_Net	1.34	1.49	1.55	1.36	1.44
	(4.77***)	(5.47***)	(5.03***)	(4.05***)	(3.74***)
Ens_total	1.72	1.85	1.94	1.74	1.90
	(5.32***)	(5.93***)	(5.52***)	(4.66***)	(4.43***)

Table 19: Risk-Adjusted Spread Portfolio Returns over Expansion and Recession Periods (Predictors: MA Signals) (Cont'd)

construct spread portfolios by buying the best-predicted quintile and selling the worst-predicted quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Recession Period (Predictors: MA Signals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.51	1.53	1.54	1.53	0.85
	(2.27**)	(2.09**)	(2.24**)	(2.72**)	(1.25)
PCR	1.57	1.59	1.53	1.51	0.60
	(1.72*)	(1.57)	(1.59)	(1.89*)	(0.63)
PLS	1.54	1.56	1.49	1.48	0.56
	(1.79*)	(1.64)	(1.63)	(1.88*)	(0.64)
Ridge	1.50	1.52	1.53	1.51	0.84
	(2.24**)	(2.06**)	(2.21**)	(2.69**)	(1.22)
Lasso	1.59	1.61	1.59	1.57	0.67
	(1.88*)	(1.72*)	(1.80*)	(2.17**)	(0.78)
ENet	1.59	1.61	1.58	1.56	0.67
	(1.87*)	(1.71*)	(1.79*)	(2.15**)	(0.77)
NN1L	1.85	1.86	1.87	1.85	1.00
	(2.39**)	(2.20**)	(2.21**)	(2.72**)	(1.19)
NN2L	1.96	1.95	1.89	1.88	0.98
	(2.38**)	(2.24**)	(2.19**)	(2.67**)	(1.19)
NN3L	1.85	1.83	1.81	1.79	0.84
	(2.27**)	(2.14**)	(2.01**)	(2.39**)	(0.98)
NN4L	1.69	1.68	1.62	1.60	0.69
	(2.17**)	(2.04**)	(1.92*)	(2.31**)	(0.88)
NN5L	1.43	1.43	1.39	1.37	0.47
	(1.96*)	(1.80*)	(1.74*)	(2.13**)	(0.62)
Ens_linear	2.22	2.24	2.16	2.15	1.09
	(2.33**)	(2.14**)	(2.19**)	(2.56**)	(1.16)
Ens_Net	2.57	2.55	2.46	2.44	1.42
	(2.59**)	(2.45**)	(2.31**)	(2.69**)	(1.44)
Ens_total	3.17	3.19	3.05	3.04	1.84
	(3.13***)	(3.00***)	(2.88***)	(3.20***)	(1.86*)

Table 20: Risk-Adjusted Spread Portfolio Returns over Expansion and Recession Periods (Predictors: 79 firm characteristics)

This table reports risk-adjusted spread portfolio returns (alpha in %) for expansion and recession periods where 79 firm characteristics are used for predictors. A given month is labeled expansion or recession based on the business cycle definition on National Bureau of of Economic Research (NBER) website. Newey-West t-statistics with 12 lags are reported in parentheses. At the beginning of each month, I independently sort stocks into quintile portfolios on predicted return from each model and construct spread portfolios by buying

Panel A: Expansion Period (Predictors: 79 firm fundamentals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	1.16	1.35	1.40	1.24	1.33
	(4.22***)	(4.81***)	(4.32***)	(4.19***)	(3.60***)
PCR	1.26	1.46	1.50	1.35	1.46
	(4.56***)	(5.23***)	(4.60***)	(4.59***)	(3.92***)
PLS	1.18	1.37	1.41	1.25	1.37
	(4.33***)	(4.96***)	(4.39***)	(4.19***)	(3.71***)
Ridge	1.15	1.34	1.39	1.23	1.33
	(4.20***)	(4.80***)	(4.30***)	(4.18***)	(3.59***)
Lasso	1.23	1.44	1.50	1.34	1.43
	(4.23***)	(4.91***)	(4.38***)	(4.29***)	(3.71***)
ENet	1.23	1.44	1.50	1.34	1.43
	(4.23***)	(4.91***)	(4.38***)	(4.29***)	(3.71***)
NN1L	1.22	1.43	1.49	1.29	1.43
	(4.23***)	(4.90***)	(4.40***)	(4.29***)	(3.67***)
NN2L	1.26	1.46	1.53	1.34	1.47
	(4.22***)	(4.87***)	(4.35***)	(4.27***)	(3.66***)
NN3L	1.25	1.46	1.53	1.36	1.48
	(4.21***)	(4.91***)	(4.36***)	(4.36***)	(3.70***)
NN4L	1.20	1.42	1.46	1.33	1.43
	(4.12***)	(4.88***)	(4.25***)	(4.21***)	(3.62***)
NN5L	1.20	1.39	1.45	1.30	1.41
	(4.14***)	(4.73***)	(4.16***)	(4.15***)	(3.57***)
Ens_linear	1.47	1.68	1.75	1.56	1.70
	(4.77***)	(5.47***)	(4.88***)	(4.72***)	(4.16***)
Ens_Net	1.62	1.86	1.93	1.72	1.92
	(4.61***)	(5.38***)	(4.72***)	(4.66***)	(4.11***)
Ens_total	1.81	2.06	2.14	1.92	2.11
	(4.99***)	(5.75***)	(5.06***)	(5.00***)	(4.38***)

Table 20: Risk-Adjusted Spread Portfolio Returns over Expansion and Recession Periods (Predictors: 79 firm characteristics) (Cont'd)

the best-predicted quintile and selling the worst-predicted quintile. For a given family of models, ensemble portfolios are constructed by buying the winner quintile agreed-upon unanimously by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes all 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel B: Recession Period (Predictors: 79 firm fundamentals)

Model Name	Mean	CAPM	FF3	Carhart4	FF5
OLS	2.35	2.23	2.31	2.31	1.59
	(2.81**)	(2.57**)	(2.39**)	(2.55**)	(1.84*)
PCR	2.16	2.04	2.14	2.15	1.47
	(2.71**)	(2.45**)	(2.31**)	(2.39**)	(1.70*)
PLS	2.47	2.35	2.41	2.41	1.68
	(3.03***)	(2.78**)	(2.57**)	(2.71**)	(1.91*)
Ridge	2.36	2.24	2.33	2.32	1.61
	(2.84***)	(2.60**)	(2.42**)	(2.58**)	(1.87*)
Lasso	2.46	2.33	2.41	2.42	1.68
	(2.98***)	(2.71**)	(2.52**)	(2.66**)	(1.90*)
ENet	2.46	2.34	2.42	2.42	1.68
	(2.98***)	(2.71**)	(2.52**)	(2.66**)	(1.91*)
NN1L	2.21	2.07	2.16	2.17	1.47
	(2.55**)	(2.32**)	(2.19**)	(2.26**)	(1.60)
NN2L	2.11	1.96	2.03	2.05	1.43
	(2.41**)	(2.22**)	(2.05**)	(2.07**)	(1.52)
NN3L	2.08	1.92	2.01	2.02	1.39
	(2.51**)	(2.28**)	(2.15**)	(2.19**)	(1.57)
NN4L	2.11	1.97	2.07	2.09	1.45
	(2.59**)	(2.37**)	(2.22**)	(2.21**)	(1.59)
NN5L	2.08	1.93	2.05	2.07	1.40
	(2.86***)	(2.64**)	(2.45**)	(2.43**)	(1.71*)
Ens_linear	2.72	2.60	2.66	2.65	1.91
	(2.79**)	(2.56**)	(2.35**)	(2.50**)	(1.89*)
Ens_Net	2.79	2.61	2.71	2.73	1.92
	(2.66**)	(2.48**)	(2.32**)	(2.35**)	(1.84*)
Ens_total	3.11	2.95	3.05	3.07	2.22
	(2.83***)	(2.63**)	(2.43**)	(2.48**)	(2.02**)

Table 21: Summary Statistics of Combination of Predictors - Conventional Forecasts (1978:01 - 2017:12)

This table presents summary statistics of spread portfolio return where the three sets of predictors are combined in different ways. Summary statistics include average return (in % per month), t-statistics, volatility (standard deviation), monthly Sharpe ratio, skewness, kurtosis, proportion of positive returns, minimum monthly return, and maximum drawdown. In Panel A, predictors include 120 monthly lagged stock returns plus 120 moving average trading signals; in Panel B predictors include 120 monthly lagged stock returns plus 79 firm fundamentals; in Panel C predictors include 120 moving average trading signals plus 86 firm fundamentals; in panel D predictors include 120 monthly lagged stock returns, 120 moving average trading signals, and 86 firm fundamentals.

Panel A: Lagged Return and Moving Average Trading Signals

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.54	4.13***	2.53	0.21	-0.40	5.76	0.63	-0.14	-0.38
PCR	0.88	5.31***	4.04	0.22	0.90	16.04	0.60	-0.22	-0.43
PLS	0.98	5.07***	4.23	0.23	1.21	16.56	0.62	-0.22	-0.41
Ridge	0.61	4.77***	2.43	0.25	0.02	4.79	0.64	-0.12	-0.37
Lasso	0.84	5.40***	3.16	0.27	0.59	15.78	0.64	-0.19	-0.40
ENet	0.84	5.40***	3.16	0.27	0.58	15.80	0.64	-0.19	-0.40
NN1L	0.90	5.70***	3.27	0.27	0.86	17.58	0.67	-0.19	-0.39
NN2L	1.11	5.56***	4.56	0.24	1.02	12.12	0.64	-0.23	-0.42
NN3L	1.03	4.87***	5.03	0.20	1.27	13.23	0.60	-0.26	-0.45
NN4L	1.07	4.97***	5.16	0.21	1.51	13.62	0.62	-0.24	-0.43
NN5L	0.96	4.64***	4.94	0.19	1.08	10.95	0.60	-0.24	-0.43
Ens_Linear	1.36	5.02***	5.65	0.24	2.06	21.46	0.64	-0.27	-0.53
Ens_Net	1.66	5.93***	6.61	0.25	1.27	12.62	0.64	-0.30	-0.50
Ens_Total	1.78	4.83***	7.99	0.22	1.17	11.91	0.60	-0.36	-0.57

Panel B: Lagged Return and Firm Characteristics

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.94	5.33***	3.87	0.24	0.39	8.93	0.65	-0.20	-0.40
PCR	1.26	5.89***	4.99	0.25	1.65	16.54	0.65	-0.23	-0.38
PLS	1.30	5.85***	5.37	0.24	1.69	15.29	0.65	-0.22	-0.39
Ridge	0.94	5.35***	3.87	0.24	0.39	8.89	0.65	-0.20	-0.40
Lasso	1.09	5.78***	4.20	0.26	0.56	10.49	0.66	-0.23	-0.41
ENet	1.09	5.78***	4.20	0.26	0.56	10.49	0.66	-0.23	-0.41
NN1L	1.04	5.92***	3.85	0.27	0.26	11.10	0.68	-0.21	-0.41
NN2L	1.23	5.57***	5.10	0.24	0.82	10.67	0.64	-0.26	-0.45
NN3L	1.24	5.19***	5.57	0.22	0.97	11.08	0.64	-0.28	-0.51
NN4L	1.14	4.72***	5.69	0.20	1.13	12.80	0.63	-0.29	-0.50
NN5L	1.06	4.71***	5.48	0.19	1.11	13.54	0.61	-0.29	-0.52
Ens_Linear	1.72	6.00***	6.56	0.26	1.93	18.24	0.66	-0.28	-0.46
Ens_Net	1.95	5.81***	7.62	0.26	0.86	9.92	0.65	-0.35	-0.59
Ens_Total	2.03	5.68***	8.18	0.25	1.03	10.28	0.64	-0.37	-0.61

Table 21: Summary Statistics of Combination of Predictors - Conventional Forecasts (1988:01 - 2017:12) - Cont'd

At the beginning of each month, I independently sort stocks into quintile portfolios on predicted return from each model and dataset and construct spread portfolios by buying the best-predicted quintile and selling the worst-predicted quintile. For a given family of models, ensemble portfolios are constructed by buying the unanimous winner quintile by all models in the family and selling the unanimous loser quintile. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. NYSE 10% market value breakpoint is applied and stocks are equal-weighted within quintile. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: Moving Average Trading Signals and Firm Characteristics

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	1.27	5.84***	5.21	0.24	0.80	11.59	0.64	-0.24	-0.38
PCR	1.36	5.01***	6.42	0.21	0.74	9.39	0.62	-0.30	-0.50
PLS	1.43	5.13***	6.51	0.22	0.87	9.62	0.62	-0.30	-0.45
Ridge	1.28	5.84***	5.22	0.24	0.80	11.52	0.64	-0.24	-0.38
Lasso	1.49	5.58***	6.35	0.24	0.69	9.79	0.65	-0.30	-0.49
ENet	1.50	5.58***	6.35	0.24	0.69	9.79	0.65	-0.30	-0.49
NN1L	1.40	5.11***	6.37	0.22	0.89	11.85	0.64	-0.31	-0.47
NN2L	1.36	4.71***	6.75	0.20	0.81	10.90	0.63	-0.32	-0.55
NN3L	1.42	4.72***	6.96	0.20	0.86	11.65	0.61	-0.36	-0.57
NN4L	1.34	4.46***	6.93	0.19	1.00	12.13	0.61	-0.36	-0.58
NN5L	1.28	4.41***	6.80	0.19	0.80	10.41	0.61	-0.35	-0.56
Ens_Linear	1.87	5.56***	7.68	0.24	0.89	9.03	0.64	-0.32	-0.53
Ens_Net	1.82	5.05***	8.24	0.22	0.75	9.79	0.62	-0.42	-0.65
Ens_Total	2.19	5.44***	8.96	0.24	0.79	8.58	0.63	-0.39	-0.64

Panel D: Lagged Return, MA Trading Signals, and Firm Characteristics

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.72	5.43***	3.27	0.22	-0.10	11.34	0.63	-0.22	-0.39
PCR	1.25	5.77***	4.99	0.25	0.92	10.90	0.64	-0.23	-0.41
PLS	1.23	5.17***	5.31	0.23	1.12	11.28	0.62	-0.25	-0.41
Ridge	0.80	5.51***	3.29	0.24	0.20	11.86	0.66	-0.21	-0.39
Lasso	1.09	5.67***	4.20	0.26	0.43	9.86	0.66	-0.23	-0.41
ENet	1.09	5.67***	4.20	0.26	0.43	9.86	0.66	-0.23	-0.41
NN1L	1.04	5.58***	4.05	0.26	0.38	9.97	0.66	-0.23	-0.39
NN2L	1.25	5.43***	5.21	0.24	0.86	10.93	0.65	-0.27	-0.45
NN3L	1.22	5.01***	5.74	0.21	1.27	13.12	0.62	-0.28	-0.46
NN4L	1.12	4.61***	5.80	0.19	1.44	14.09	0.60	-0.28	-0.49
NN5L	1.12	4.61***	5.71	0.20	1.18	12.76	0.63	-0.28	-0.50
Ens_Linear	1.73	5.15***	6.98	0.25	2.03	19.08	0.62	-0.31	-0.51
Ens_Net	1.92	5.58***	7.40	0.26	1.10	11.32	0.63	-0.36	-0.58
Ens_Total	2.20	5.19***	8.58	0.26	1.35	13.79	0.63	-0.40	-0.65

Table 22: Summary Statistics of Spread Portfolio - P-Lasso Forecasts (1988:01 - 2017:12)

This table presents portfolio performance statistics for P-Lasso allocation whose weights are linear combination of spread portfolio weights computed from models using 120 month lagged return, 120 moving average signals, and 79 fundamentals as predictors independently. I use the global minimum variance portfolio weights computed from data through month t-1 to determine the linear combination for month t. For each month, companies that appear in at least one of the three samples are considered. Weights for companies that does not appear in all three samples for a given month are filled with zeros. Summary statistics include average return (in % per month), t-statistics, volatility (standard deviation), monthly Sharpe ratio, skewness, kurtosis, proportion of positive returns, minimum monthly return, and maximum drawdown. In Panel A, predictors include 120 monthly lagged stock returns plus 120 moving average trading signals; in Panel B predictors include 120 monthly lagged stock returns plus 79 firm fundamentals; in Panel C

Panel A: Lagged Return and Moving Average Trading Signals

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.74	5.01***	3.00	0.25	0.67	19.31	0.65	-0.20	-0.37
PCR	0.78	5.87***	3.10	0.25	0.65	14.82	0.63	-0.19	-0.31
PLS	1.00	6.69***	3.36	0.30	1.11	15.77	0.66	-0.20	-0.32
Ridge	0.74	5.01***	3.00	0.25	0.66	19.31	0.65	-0.20	-0.37
Lasso	0.81	5.23***	3.02	0.27	0.55	15.40	0.65	-0.19	-0.37
ENet	0.81	5.23***	3.02	0.27	0.55	15.40	0.65	-0.19	-0.37
NN1L	0.89	5.52***	3.21	0.28	0.36	9.77	0.65	-0.18	-0.41
NN2L	1.01	5.87***	3.83	0.27	0.86	13.23	0.65	-0.21	-0.39
NN3L	0.99	5.82***	4.05	0.24	0.92	16.83	0.66	-0.24	-0.44
NN4L	0.87	4.63***	4.33	0.20	1.10	19.93	0.63	-0.28	-0.46
NN5L	0.85	5.21***	4.04	0.21	1.28	21.34	0.62	-0.26	-0.45
Ens_Linear	1.17	5.65***	4.39	0.27	1.36	19.72	0.66	-0.24	-0.42
Ens_Net	1.50	5.55***	6.35	0.24	1.02	14.12	0.65	-0.35	-0.50
Ens_Total	1.73	5.79***	6.92	0.25	1.36	15.88	0.62	-0.35	-0.49

Panel B: Lagged Return and Firm Characteristics

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.82	5.20***	3.01	0.27	0.82	17.56	0.65	-0.19	-0.36
PCR	0.77	5.54***	3.06	0.25	0.91	12.42	0.62	-0.18	-0.30
PLS	1.02	6.54***	3.29	0.31	1.30	14.23	0.66	-0.19	-0.32
Ridge	0.82	5.20***	3.01	0.27	0.82	17.56	0.65	-0.19	-0.36
Lasso	0.84	5.28***	3.07	0.27	0.67	16.48	0.64	-0.19	-0.38
ENet	0.84	5.28***	3.07	0.27	0.67	16.48	0.64	-0.19	-0.38
NN1L	0.90	5.26***	3.26	0.27	0.50	10.92	0.64	-0.18	-0.42
NN2L	0.96	5.38***	3.77	0.26	0.91	10.85	0.62	-0.20	-0.39
NN3L	0.87	4.90***	3.82	0.23	0.64	11.31	0.62	-0.21	-0.47
NN4L	0.76	3.81***	4.03	0.19	1.23	17.69	0.56	-0.26	-0.46
NN5L	0.73	4.63***	3.78	0.19	1.27	19.53	0.60	-0.23	-0.43
Ens_Linear	1.28	5.85***	4.37	0.29	1.61	20.51	0.66	-0.24	-0.41
Ens_Net	1.64	5.52***	6.44	0.25	1.23	13.72	0.63	-0.34	-0.52
Ens_Total	1.85	5.78***	7.01	0.26	1.44	14.46	0.63	-0.35	-0.53

Table 22: Summary Statistics of Spread Portfolio - P-Lasso Forecasts (1979:01 - 2017:12) - Cont'd

predictors include 120 moving average trading signals plus 86 firm fundamentals; in panel D predictors include 120 monthly lagged stock returns, 120 moving average trading signals, and 86 firm fundamentals. At the beginning of each month, I calculate the variance and covariance from previous spread portfolio returns and use them to calculate global minimum variance portfolio weights. For ensemble portfolios, I first construct the unanimously agreed top and bottom quintile for each set of predictors independently. Then I calculate the average weights of the two quintiles across three sets of predictors in the same manner as for individual models. Linear family includes OLS, PCR, PLS, Ridge, Lasso, and Elastic net. Network family includes 5 neural networks. Total ensemble includes both linear and network families, totalling 11 models. NYSE 10% market value breakpoint is applied and stocks are equal-weighted within quintile. *,**,*** represent statistical significance at 1%, 5%, and 10% levels.

Panel C: Moving Average Trading Signals and Firm Characteristics

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.65	3.77***	4.18	0.15	0.02	14.96	0.60	-0.26	-0.46
PCR	0.88	4.10***	5.46	0.16	0.42	13.04	0.62	-0.32	-0.51
PLS	0.84	3.71***	5.53	0.15	0.54	11.71	0.58	-0.30	-0.51
Ridge	0.64	3.76***	4.18	0.15	0.01	14.93	0.61	-0.26	-0.46
Lasso	0.95	4.34***	5.55	0.17	0.01	13.02	0.62	-0.33	-0.58
ENet	0.96	4.36***	5.54	0.17	0.03	12.97	0.62	-0.33	-0.57
NN1L	1.04	4.62***	5.61	0.19	0.14	13.87	0.63	-0.34	-0.54
NN2L	1.05	4.37***	6.07	0.17	0.07	14.34	0.61	-0.41	-0.59
NN3L	1.00	4.04***	6.11	0.16	0.41	14.70	0.60	-0.40	-0.59
NN4L	0.90	3.69***	5.96	0.15	0.58	14.18	0.59	-0.37	-0.56
NN5L	0.88	3.78***	5.94	0.15	0.32	15.05	0.59	-0.39	-0.59
Ens_Linear	1.20	4.60***	6.34	0.19	0.48	11.21	0.62	-0.36	-0.57
Ens_Net	1.50	5.14***	7.25	0.21	0.61	11.49	0.61	-0.41	-0.60
Ens_Total	1.91	6.06***	7.87	0.24	0.70	10.47	0.65	-0.42	-0.59

Panel D: Lagged Return, MA Trading Signals, and Firm Characteristics

Model Name	Mean	t-stat	Vol	Sharpe	Skew	Kurt	Pr(R>0)	Min	MDD
OLS	0.71	4.68***	3.00	0.24	0.73	20.67	0.65	-0.20	-0.38
PCR	0.72	5.23***	3.09	0.23	0.63	11.97	0.61	-0.19	-0.32
PLS	0.93	6.10***	3.35	0.28	1.15	13.98	0.65	-0.20	-0.34
Ridge	0.71	4.68***	3.00	0.24	0.72	20.65	0.65	-0.20	-0.38
Lasso	0.82	5.23***	3.04	0.27	0.52	15.30	0.65	-0.20	-0.38
ENet	0.82	5.23***	3.04	0.27	0.52	15.29	0.65	-0.20	-0.38
NN1L	0.86	5.10***	3.26	0.26	0.29	11.19	0.65	-0.18	-0.43
NN2L	0.95	5.27***	3.81	0.25	0.77	11.87	0.63	-0.22	-0.40
NN3L	0.86	4.83***	3.84	0.22	0.54	11.89	0.61	-0.22	-0.47
NN4L	0.72	3.61***	4.07	0.18	1.12	17.91	0.56	-0.27	-0.45
NN5L	0.69	4.29***	3.84	0.18	0.88	18.24	0.59	-0.26	-0.45
Ens_Linear	1.18	5.55***	4.43	0.27	1.44	20.42	0.66	-0.25	-0.43
Ens_Net	1.51	5.45***	6.39	0.24	1.09	14.64	0.63	-0.36	-0.50
Ens_Total	1.78	5.85***	6.97	0.26	1.39	15.92	0.62	-0.35	-0.49

References

Gopal K. Basak, Ravi Jagannathan, Tongshu Ma, Jackknife Estimator for Tracking Error Variance of Optimal Portfolios. Management Science 2009 55:6, 990-1002

Bryan Kelly, Seth Pruitt, The three-pass regression filter: A new approach to forecasting using many predictors. Journal of Econometrics, Volume 186, Issue number 2, Pages 294-316, June 1 201

Bryan Kelly, Seth Pruitt, Market Expectations in the CrossSection of Present Values. Journal of Finance, Volume 68, Issue 5 October 2013 Pages 1721-1756

Butaru, Florentin and Chen, Qingqing and Clark, Brian J. and Das, Sanmay and Lo, Andrew W. and Siddique, Akhtar R., Risk and Risk Management in the Credit Card Industry (June 12, 2016). Available at SSRN: https://ssrn.com/abstract=2618746 or http://dx.doi.org/10.2139/ssrn.2618746

Christopher J. Neely, David E. Rapach, Jun Tu and Zhou, Guofu. Forecasting the Equity Risk Premium: The Role of Technical Indicators. Management Science, 60, 2014, 1772–1791.

Cybenko, G. Approximation by superpositions of a sigmoidal function. Math. Control Signal Systems (1989) 2: 303. https://doi.org/10.1007/BF02551274

Diederik P. Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization 2014. arXiv:1412.6980v9 [cs.LG] 30 Jan 2017

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1993) 3-56. North-Holland

Eugene F. Fama and Kenneth R. French. A five-factor asset pricing model. Journal of Financial Economics Volume 116, Issue 1, April 2015, Pages 1-22

Feng, Guanhao and Giglio, Stefano and Xiu, Dacheng, Taming the Factor Zoo: A Test of New Factors (January 2, 2019). Fama-Miller Working Paper; Chicago Booth Research Paper No. 17-04. Available at SSRN: https://ssrn.com/abstract=2934020 or http://dx.doi.org/10.2139/ssrn.2934020

Francis X. Diebold. Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of DieboldMariano Tests. Journal of Business Economic Statistics. Volume 33, 2015 - Issue 1

Gu, Shihao, Kelly, Bryan T. and Xiu, Dacheng, Empirical Asset Pricing via Machine Learning. Chicago Booth Research Paper No. 18-04; 31st Australasian Finance and Banking Conference 2018. Available at SSRN: https://ssrn.com/abstract=3159577 or http://dx.doi.org/10.2139/ssrn.3159577

Gur Huberman, Shmuel Kandel. MeanVariance Spanning. The Journal of Finance. Vol. 42, No. 4 (Sep., 1987), pp. 873-888

Han, Yufeng and He, Ai and Rapach, David and Zhou, Guofu, What Firm Characteristics Drive US Stock Returns? (October 1, 2018). Available at SSRN: https://ssrn.com/abstract=3185335 or http://dx.doi.org/10.2139/ssrn.3185335

Hao Huang, Bo Chen, and Chungao Liu, Safety Monitoring of a Super-High Dam Using Optimal Kernel Partial Least Squares. Mathematical Problems in Engineering Volume 2015, Article ID 571594, 13 pages http://dx.doi.org/10.1155/2015/571594

Heaton, J.B. and Polson, Nick and Witte, Jan, Deep Learning for Finance: Deep Portfolios (September 5, 2016). Available at SSRN: https://ssrn.com/abstract=2838013 or http://dx.doi.org/10.2139/ssrn.2838013

Huang, Dashan and Zhang, Huacheng and Zhou, Guofu and Zhu, Yingzi, Twin Momentum: Fundamental Trends Matter (January 7, 2019). Available at SSRN: https://ssrn.com/abstract=2894068 or http://dx.doi.org/10.2139/ssrn.2894068

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative Adversarial Networks. arXiv:1406.2661

Jeremiah Green John R. M. Hand X. Frank Zhang The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. The Review of Financial Studies, Volume 30, Issue 12, December 2017, Pages 43894436

Joachim Freyberger, Andreas Neuhierl, Michael Weber. Dissecting Characteristics Nonparametrically. Review of Financial Studies (accepted for publication)

Joseph Granville. Book Granville's New Key to Stock Market Profits, 1963.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. Microsoft Research arXiv:1502.01852

Kurt Hornik, Maxwell Stinchcombe, Halbert White. Multilayer feedforward networks are universal approximators. Neural Networks Volume 2, Issue 5, 1989, Pages 359-366

LeCun, Y.A., L. Bottou, G. B. Orr, and K.-R. Muller (2012). Efficient Backprop, Neural Networks: Tricks of the Trade, pp. 9-48.

Light, Nathaniel and Maslov, Denys and Rytchkov, Oleg, Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach (January 21, 2016). Review of Financial Studies (RFS), Vol. 30, No. 4, pp. 1339-1381, 2017. Available at SSRN: https://ssrn.com/abstract=2307563 or http://dx.doi.org/10.2139/ssrn.2307563

Malcolm Baker and Jeffrey Wurgler. Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance. Vol LXI, NO.4, August 2006

Mark M. Carhart. On Persistence in Mutual Fund Performance. The Journal of Finance. Vol. LII, NO.1. March 1997

Masters, Timothy, 1993, Practical Neural Network Recipes in C++ (Academic Press).

Rajneesh Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15 (2014) 1929-1958

Rapach, David E, Jack K Strauss, and Guofu Zhou, 2013, International stock return predictability: what is the role of the united states?, The Journal of Finance 68, 1633-1662.

Rapach, David and Strauss, Jack and Tu, Jun and Zhou, Guofu, Industry Return Predictability: A Machine Learning Approach (October 8, 2018). Available at SSRN: https://ssrn.com/abstract=3120110 or http://dx.doi.org/10.2139/ssrn.3120110

Rapach, David and Zhou, Guofu, Sparse Macro Factors (January 31, 2019). Available at SSRN: https://ssrn.com/abstract=3259447 or http://dx.doi.org/10.2139/ssrn.3259447

Raymond Kan, Guofu Zhou. MeanVariance Spanning. Annals of Economics and Finance. 13-1, 145193 (2012)

R. David Mclean, Jeffery Pontiff Does Academic Research Destroy Stock Return Predictability? Journal of Finance 2015

Sergey Ioffe, Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:448-456, 2015.

Sirignano, Justin and Sadhwani, Apaar and Giesecke, Kay, Deep Learning for Mortgage Risk (November 20, 2018). Available at SSRN: https://ssrn.com/abstract=2799443 or http://dx.doi.org/10.2139/ssrn.2799443

Stambaugh, Robert F. Yu, Jianfeng Yuan, Yu, 2012. "The short of it: Investor sentiment and anomalies," Journal of Financial Economics, Elsevier, vol. 104(2), pages 288-302.

Tarun Chordia, Avanidhar Subrahmanyam, Qing Tong. Have Capital Market Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity?. Journal of Accounting Economics (JAE), Vol. 58, No. 1, 2014

Xavier Glorot, Yoshua Bengio. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR 9:249-256, 2010.

Yufeng Han, Guofu Zhou, Yingzi Zhu. A trend factor: Any economic gains from using information over investment horizons? Journal of Financial Economics Volume 122, Issue 2, November 2016, Pages 352-375

Appendices

Algorithm 1 NIPALS for Principal Component Analysis

Require: De-meaned data matrix X, convergence threshold δ

Select the i^{th} column vector of X, $u = x_i$
while $d > \delta$ **do**
    Project the matrix X onto u and calculate loading $v = \dfrac{X^T u}{u^T u}$
    Normalize the loading vector v to norm one, $p = v/|v|$
    Copy vector $u_{old} = u$
    Project matrix X onto p to find corresponding new score vector $u = \dfrac{Xp}{p^T p}$
    Measure the convergence $d = u_{old} - u$
end while
Orthogonalize X with respect to the component $X = X - up^T$

Algorithm 2 Stochastic Gradient Descent (SGD) with momentum

Require: Learning rate ϵ, momentum parameter α

Require: Initial parameter θ, initial velocity v

while stopping criterion not met **do**
    Sample a mini-batch of m examples from the training set $\{\mathbf{x^{(1)}}, ..., \mathbf{x^{(m)}}\}$
    with corresponding targets $\mathbf{y^{(i)}}$
    Compute gradient estimate: $\mathbf{g} \leftarrow \frac{1}{m} \nabla_\theta \sum_i L(f(x^{(i)}; \theta), y^{(i)})$.
    Compute velocity update: $v \leftarrow \alpha v - \epsilon g$
    Apply update: $\theta \leftarrow \theta + v$
    End while
end while

Algorithm 3 Backward computation for the deep neural network of the forward algorithm, which uses, in addition to the input $\mathbf{x}$, a target $\mathbf{y}$. This computation yields the gradient on the activation $\mathbf{a}^{(k)}$ for each layer k, starting from the output layer and going backwards to the first hidden layer. From these gradients, which can be interpreted as an indication of how each layer's output should change to reduce error, one can obtain the gradient on the parameters of each layer. The gradients on weights and biases can be immediately used as part of a stochastic gradient update (performing the update right after the gradients have been computed) or used with other gradient-based optimization methods

After the forward computation, compute the gradient on the output layer:
$\mathbf{g} \leftarrow \bigtriangledown_{\hat{\mathbf{y}}} J = \bigtriangledown_{\hat{\mathbf{y}}} L(\hat{\mathbf{y}}, \mathbf{y})$
for $k = l, l-1 ..., 1$ **do**
    Convert the gradient on the layer's output into a gradient into the pre-nonlinearity activation (element-wise multiplication if f is element-wise):
    $\mathbf{g} \leftarrow \bigtriangledown_{\alpha^{(\mathbf{k})}} J = \mathbf{g} \odot f'(\alpha^{(\mathbf{k})})$
    Compute gradients on weights and biases (including the regularization terms, where needed):
    $\bigtriangledown_{b^{(k)}} J = \mathbf{g} + \lambda \bigtriangledown_{\mathbf{b}^{(k)}} \Omega(\theta)$
    $\bigtriangledown_{\mathbf{W}^{(k)}} J = \mathbf{g}\mathbf{h}^{(\mathbf{k-1})^T} + \lambda \bigtriangledown_{\mathbf{W}^{(k)}} \Omega(\theta)$ Propagate the gradients w.r.t. the next lower-level hidden layer's activations:
    $\mathbf{g} \leftarrow \bigtriangledown_{\mathbf{h}^{(\mathbf{k-1})}} J = \mathbf{W}^{\mathbf{k}^T} \mathbf{g}$
end for

Algorithm 4 The Adam algorithm

Require: Step size ϵ (Suggested default: 0.001)

Require: Exponential decay rates for moment estimates, ρ_1 and ρ_2 in [0,1). (Suggested defaults: 0.9 and 0.999 respectively)

Require: Small constant δ used for numerical stabilization (Suggested default:10^{-8})

Require: Initial parameters θ

Initialize 1st and 2nd moment variables $\mathbf{s} = \mathbf{0}, \mathbf{r} = \mathbf{0}$

Initialize time step t=0

while stopping criterion not met **do**

  Sample a mini-batch of m examples from the training set $\{\mathbf{x^{(1)}}, ..., \mathbf{x^{(m)}}\}$

  with corresponding targets $\mathbf{y^{(i)}}$

  Compute gradient estimate: $\mathbf{g} \leftarrow \frac{1}{m} \nabla_\theta \sum_i L(f(x^{(i)}; \theta), y^{(i)})$.

  $t \leftarrow t + 1$

  Update biased first moment estimate: $s \leftarrow \rho_1 s + (1 - \rho_1) g$

  Update biased second moment estimate: $r \leftarrow \rho_2 r + (1 - \rho_2) g \odot g$

  Correct bias in first moment: $\hat{s} \leftarrow \frac{s}{1 - \rho_1^t}$

  Correct bias in second moment: $\hat{r} \leftarrow \frac{r}{1 - \rho_2^t}$

  Compute update: $\delta\theta = -\epsilon \frac{\hat{s}}{\sqrt{\hat{r}} + \delta}$ (operations applied element-wise)

  Apply update: $\theta \leftarrow \theta + \delta\theta$

  End while

end while

Algorithm 5 Forward propagation through a typical deep neural network and the computation of the cost function. The loss $L(\hat{y}, y)$ depends on the output $\hat{y}$. To obtain the total cost J, the loss may be added to a regularizer $\Omega(\theta)$, where θ contains all the parameters. For simplicity, this algorithm uses only a single input example x. Practical applications should use a minibatch, i.e., of 32 examples as in our study.

Require: Network depth l
Require: $\mathbf{W}^{(i)}, i \in \{1, ..., l\}$, the weight matrices of the model
Require: $\mathbf{b}^{(i)}, i \in \{1, ..., l\}$, the bias parameters of the model
Require: $\mathbf{x}$, the input to process
Require: $\mathbf{y}$, the target output

```
h^(0) = x
for k = 1, ..., l do
    a^(k) = b^(k) + W^(k) h^(k-1)
    h^(k) = f(a^(k))
end for
ŷ = h^(l)
J = L(ŷ, y) + λΩ(θ)
```

$\mathbf{h}^{(0)} = x$
for $k = 1, ..., l$ **do**
  $\mathbf{a}^{(k)} = \mathbf{b}^{(k)} + \mathbf{W}^{(k)}\mathbf{h}^{(k-1)}$
  $\mathbf{h}^{(k)} = f(\mathbf{a}^{(k)})$
end for
$\hat{y} = \mathbf{h}^{(l)}$
$J = L(\hat{y}, y) + \lambda\Omega(\theta)$

Algorithm 6 The early stopping meta-algorithm for determining the best amount of time to train. This meta-algorithm is a general strategy that works well with a variety of training algorithms and ways of quantifying error on the validation set.

Require: Let n be the number of steps between evaluations
Require: Let p be the "patience," the number of times to observe worsening validation set error before giving up
Require: Let $\theta_{\mathbf{0}}$ be the initial parameters

$\theta \leftarrow \theta_{\mathbf{0}}; i \leftarrow 0; j \leftarrow 0; v \leftarrow \infty; \theta^* \leftarrow \theta; i^* \leftarrow i$
while j ¡ p **do**
  Update θ by running the training algorithm for n steps.
  $i \leftarrow i + n$
  $v' \leftarrow ValidationSetError(\theta)$
  if $v' < v$ **then**
    $j \leftarrow 0; \theta^* \leftarrow \theta; i^* \leftarrow i; v \leftarrow v'$
  else$j \leftarrow j + 1$

    Best parameters are θ^*, best number of training steps is i^*