Improving Earnings Predictions with Machine Learning

Joshua O. S. Hunt
Mississippi State University

James N. Myers
University of Tennessee

Linda A. Myers
University of Tennessee

December 2019

This paper is based on Joshua Hunt’s dissertation. He would like to thank his dissertation committee – Vernon Richardson (chair), Cory Cassell, David Douglas, and James Myers – for their helpful comments and guidance. We also thank Zhan Gao, Steve Young, and other workshop participants at the University of Arkansas and at Lancaster University for helpful comments and suggestions. James Myers gratefully acknowledges financial support from the Dennis Hendrix Distinguished Professorship at the University of Tennessee, and Linda Myers gratefully acknowledges financial support from the Haslam Chair of Business and the Vallett Family Outstanding Researcher Award at the University of Tennessee.
Improving Earnings Predictions with Machine Learning

Abstract: Ou and Penman (1989) use stepwise logit regression to predict the sign of future earnings changes. Using these predictions, they form a profitable hedge portfolio long in firms predicted to have an increase in earnings and short in firms predicted to have a decrease in earnings. Dramatic increases in computing power and recent advances in machine learning allow us to extend Ou and Penman (1989) using a larger dataset, more computer intensive forecasting algorithms, and modern prediction models. We find that stepwise logit continues to provide good out-of-sample predictions regarding the sign of future earnings changes but a trading strategy based on these forecasts does not generate abnormal returns on average. We next test whether a modification of stepwise logit, elastic net, improves prediction accuracy and the trading strategy but find that it does not. Next, we use a non-parametric machine learning technique, random forest. We find that this method significantly improves out-of-sample forecast accuracy and that these forecasts are useful for generating abnormal returns. Finally, we use all three of these forecasting methods to predict the direction of stock market returns. Overall, the most profitable strategy is based on a random forest model’s prediction of future earnings. Our results suggest that recent non-parametric machine learning techniques could be useful in a variety of accounting contexts where predictions of binary outcomes are needed.

Keywords: Data analytics, machine learning, stepwise logit regression, elastic net, random forest, earnings prediction, returns prediction, trading strategies
I. Introduction

Earnings expectations and earnings forecasts have been a central focus of accounting research for almost 50 years. Financial accounting research is frequently traced back to Ball and Brown (1968) and Beaver (1968). Both of these papers examine the relation between earnings innovations and stock market returns. Over the following decades, accounting researchers studied the time-series properties of earnings and learned much more about the relation between earnings innovations and stock prices. For example, Albrecht, Lookabill, and McKeown (1977) finds that although annual earnings can be well approximated using firm-specific Box Jenkins time-series models, out-of-sample, a random walk seems to provide similar forecast accuracy. Interestingly, Beaver, Lambert, and Morse (1980) shows that investors believe that the earnings process is more complex than a random walk. In subsequent work on post earnings announcement drift, Foster, Olson, and Shevlin (1984) and Bernard and Thomas (1989) show that extreme changes in quarterly earnings are more persistent than investors expect. More recently, Nichols and Wahlen (2004) extend seminal papers including Ball and Brown (1968), Kormendi and Lipe (1987), and Bernard and Thomas (1989) through 2002 and find that the inferences from these papers continue to hold. Although our understanding of the importance of earnings innovations for stock prices has improved greatly since 1968, our ability to forecast earnings has not advanced as much, and time-series models of earnings have very limited predictive power.

Beginning in the late 1980s, researchers started to use the components of earnings to improve their earnings predictions. Lipe (1986) decomposes earnings into six components and shows that these components have differing levels of persistence. Subsequent research tests whether stock market participants appear to understand these differences, at least to some extent.
For example, Sloan (1996) partitions earnings into accruals and cash flows, and finds that cash flows are more persistent, although investors appear to be naïve about the relative persistence of these earnings components. Similarly, Hayn (1995) shows that losses are less persistent than gains, and Xie (2001) shows that discretionary accruals are less persistent than non-discretionary accruals, but investors do not seem to fully understand the persistence of these items.

In much of the prior research, the authors attempt to predict the level of earnings. In contrast, in a seminal paper, Ou and Penman (1989) attempt to predict the sign of earnings changes. Their general research question is whether and to what extent standard financial ratios can be useful for financial statement analysis. Specifically, they use a two-step process. First, they predict the sign of future changes in earnings using stepwise logit regression. Because time-series models are generally not more accurate than a random walk, time-series models are not useful for predicting the sign of earnings changes.\(^1\) Therefore, Ou and Penman (1989) develop a model to estimate the historical relation between currently observable financial ratios and the sign of future earnings changes. In sample, their model correctly predicts the sign (positive or negative) of one-year-ahead earnings changes 78 percent of the time. They then use their model to make an out of sample prediction of the sign of the one-year-ahead earnings changes, and find that their model correctly predicts the sign approximately 60 percent of the time. In the second step of their analysis, Ou and Penman (1989) use this out of sample prediction to form a zero investment hedge portfolio investment strategy that is long in firms with positive predicted earnings changes and short in firms with negative predicted earnings changes. Over the

---

\(^1\) Ou and Penman (1989) estimate the sign of earnings changes after controlling for a drift. We ignore the drift term because estimating a firm-specific drift term requires a number of prior years’ data, reducing the sample size and eliminating startup firms. Overall, this makes little difference to our empirical results.
following 24 months, their zero investment hedge portfolio earns abnormal returns of more than 12 percent.

Whereas Ou and Penman (1989) uses a two-step process of first estimating the sign of earnings changes and then forming investment portfolios, Holthausen and Larcker (1992) forgo the initial step and directly predict the sign of future stock returns using stepwise logit regression. They are unable to replicate the Ou and Penman (1989) abnormal returns in their sample period, but they find that investment portfolios formed based on the predicted sign of changes in stock price (rather than earnings) earn abnormal returns.

The technique in Ou and Penman (1989) has the advantage over time-series models of using a relatively rich set of information to predict earnings changes, but their analyses was constrained by the amount of computing power available at the time. Specifically, estimating the stepwise logit regression using a large set of financial ratios was too computer intensive (Ou and Penman 1989, p. 303) so the authors used univariate predictive power to winnow down the number of included ratios to 58. In addition, rather than estimating the models annually using rolling windows of data, Ou and Penman were only able to estimate their logit regressions in two time periods.2

In the 30 years since Ou and Penman (1989) was published, computing power and machine learning techniques have advanced dramatically, allowing researchers to examine whether additional independent variables and more computer intensive methodologies might be useful for predicting future earnings. However, even though the literature has undoubtedly become more sophisticated with respect to econometric issues, the vast majority of recent papers

---

2 Specifically, they used data from 1965 through 1972 to predict the sign of earnings changes in 1973 through 1977 and data from 1973 through 1977 to predict the sign of earnings changes in 1978 through 1983.
still rely on relatively simple parametric estimation techniques such as ordinary least squares regression or traditional logit regression models.

Over the past 20 years, statisticians have devised dozens of forecasting techniques, each with its own strengths and weaknesses. In this paper, we use a data analytics approach to apply modern machine learning techniques to the Ou and Penman (1989) task of predicting the sign of future earnings changes. We then use this prediction to form investment portfolios and test whether these portfolios earn abnormal returns. Because of improvements in computer-based forecasting techniques, we are no longer constrained to parametric estimation. Thus, we compare the out-of-sample accuracy of three different estimation techniques which we use to predict the sign of future earnings changes based on observable financial statement data. These techniques are the standard stepwise logit model (Ou and Penman 1989), elastic net (Zou and Hastie 2005), and the random forest model (Breiman 2001).

We begin by comparing the out-of-sample predictive accuracy of these three models using all of the independent variables included in Ou and Penman (1989). We find that the stepwise logit regression has 62.3 percent forecast accuracy, which is similar to the findings in Ou and Penman (1989). Elastic net performs similarly to stepwise logit, and does not outperform stepwise logit in any subsample. However, random forest is significantly more accurate, with out-of-sample accuracy of up to 76.8 percent.

---

3 In untabulated tests, we also use neural networks, support vector machines, and gradient boosting machine learning algorithms. Random forest outperforms neural networks and support vector machines in all respects. Gradient boosting outperforms random forest with respect to accuracy, but not with respect to abnormal return generation.

4 All three of the forecast methods that we consider provide probability estimates of an increase in future earnings for all observations, but these forecasts tend to be unreliable when the probability of an increase is close to 50/50. That is, the predictions are much more accurate for observations with strong signals of either increasing or non-increasing earnings. When evaluating their model, Ou and Penman (1989) drop the middle 20 percent of observations. Although we tabulate the effectiveness of our models using a number of cutoffs, we focus on those samples where they are most accurate. This occurs for observations at the top and bottom five percent of the distribution. However, our inferences also hold using the Ou and Penman (1989) cutoffs.
We then test whether the models can be improved by adding the price-earnings ratio (PE) and the percent change in PE as additional independent variables. PE will be greatly influenced by the market’s expectation of future earnings growth (Penman 1996). When PE is high, investors anticipate an earnings increase, and when PE is low, investors anticipate an earnings decrease. Therefore, including PE allows us to incorporate the information incorporated in investor beliefs into the prediction model. Although the parametric models do not improve significantly with the addition of PE and the percent change in PE, the out-of-sample accuracy of random forest improves to 79.8 percent.

Next, we test whether any of the three methods can be used to form investment strategies that earn abnormal stock market returns. We find that the Ou and Penman stepwise logit model earns statistically significant abnormal returns of 3.4 percent and the elastic net model earns statistically significant abnormal returns of 4.2 percent. Moreover, the random forest model earns abnormal returns of 10.6 percent over the subsequent year. We then follow Holthausen and Larker (1992) and use the prediction models to forecast the direction of stock market returns directly (i.e., without the intermediate step of predicting the sign of earnings changes). In this setting, all three models have similar accuracy, but elastic net produce the most profitable trading strategy and the random forest model does not dominate either parametric model on any dimension. Overall, we find that the method used in Ou and Penman (1989) remains a viable method for predicting the sign of earnings changes but the forecasts can be improved with the use of modern non-parametric techniques.

Our paper should be interesting to researchers who investigate how machine learning can be used in accounting and finance applications. For example, recent research uses a variety of machine learning techniques to predict fraud (Purda and Skillicorn 2015), explain accruals
(Frankel, Jennings, and Lee 2016), predict stock returns (Gu, Kelly, and Xiu 2018), and select directors for corporate boards (Erel, Stern, Tan, and Weisbach 2019).

The remainder of this paper is organized as follows. Section II outlines each of the statistical algorithms that we examine. Section III presents the data and describes our models. Section IV presents the results, and section V concludes.

II. Statistical Algorithms

We compare the ability of three techniques to forecast the direction of the change in annual earnings. For all of the forecasts, there are two possible mutually exclusive outcomes: income is increasing or income is not increasing. Thus, we use three models that are suitable for estimating a binomial outcome: the stepwise logit model, elastic net, and random forest. We give each model the opportunity to choose the most useful subset of 60 independent variables. Each of the three models uses a different method of identifying the best subset of independent variables to be used in forecasting.

One advantage of all three models is that they provide an estimate of the probability of an earnings increase. An intuitive approach would be to partition the observations into those with an estimated probability of an earnings increase greater than 50 percent versus those with an estimated probability of less than 50 percent. However, using raw probability cutoffs yields an unbalanced sample because more than half of the observations are predicted to have increasing earnings. Furthermore, correctly predicting an increase may be less valuable than correctly predicting a decrease. Similarly, incorrect prediction of an earnings increase may be less costly than incorrect prediction of an earnings decrease. Therefore, following Holthausen and Larcker (1992), we rank the probabilities of changes in earnings in order to have a more balanced
Specifically, we split the sample based on ranked probability cutoffs of 50/50, 60/40, 70/30, 80/20, 90/10, and 95/05. That is, we first examine the ability of the top and bottom 50 percent of the sample to predict the sign of earnings changes. Next, we compare the ability of the top and bottom 40 percent of the sample to make these predictions. Using this methodology balances the sample size of the top and bottom groups, and yields a consistent number of observations across models.

Logit Regression

Logit regression is the most common binary classification algorithm in the accounting literature. It is used for explaining or predicting financial statement misstatements, auditor changes, meeting or just beating earnings expectations, etc. Logit regression is useful for predicting a dichotomous outcome when there is more than one independent variable. In our setting, logit regression estimates the conditional probability of reported earnings in year $t$ being strictly larger than reported earnings in year $t-1$. The model provides coefficient estimates and these coefficient estimates can be used to estimate the probability that earnings will increase in the following year. The first paper to use logit regression to predict the sign of earnings changes is Ou (1990). This paper examines the ability of eight financial ratios to predict the change in one-year-ahead earnings changes, and then calculates abnormal stock returns based on these predictions. In our setting, for each observation, the logit model provides an estimate of the probability that next year’s diluted earnings per share will be greater than this year’s diluted earnings per share. Whereas prior literature focused on firm-wide earnings (Ou and Penman 1989; Ou 1990; Holthausen and Larker 1992), we focus on diluted earnings per share because this is one of the earnings metrics that financial analysts and financial statement users anchor on.

---

5 Note, however, that all of our inferences remain qualitatively similar when we use raw probabilities.
The logit regression coefficients are estimated using an iterative process to find the coefficients that minimize the difference between the predicted probability of an earnings increase and the realization (1 for observations with increasing earnings, 0 otherwise). This method solves for the $\beta$ that maximize equation (1):

$$\log P(y_i|\beta, x) = \sum_{i=1}^{m} y_i \log \left( \frac{1}{1+\exp(-x_i\beta)} \right) + (1 - y_i) \log \left( \frac{\exp(-x_i\beta)}{1+\exp(-x_i\beta)} \right)$$ (1)

However, logit regression is subject to several limitations. First, the relation between the log odds of the dependent variable and the independent variables is assumed to be linear, so logit regression performs poorly in the presence of outliers. Second, multicollinearity of the independent variables also causes problems for the model. Because we use a large number of related financial ratios as independent variables, both of these limitations may be relevant. Third, for each independent variable, there must be an adequate number of observations for each category of the dependent variable. In our setting, the dependent variable is set equal to one when the earnings in year $t$ are greater than earnings in year $t-1$, and zero otherwise; where earnings are measured as split-adjusted diluted earnings per share.

Misspecification can be a serious problem for logit regression. Excluding relevant variables results in an omitted variable bias. Because of the non-linearity of logit regressions, bias in one coefficient affects all of the coefficient estimates even if the variable that is omitted is uncorrelated with the variable of interest (Gail, Wieand, and Piantadosi 1984; Wooldridge 2002, section 15.7; Mood 2010). Similarly, including extraneous variables can inflate the standard errors of the irrelevant variables as well as those of other independent variables that are correlated with them.

Whereas a non-linear relation between the log odds ratio and the independent variables may reduce logit’s out-of-sample predictive ability, the primary concerns with logit regression
relate primarily to the inclusion/exclusion of variables, especially when faced with multicollinearity and outliers. These problems are likely to be present in our setting because of the large number of variables used in the analyses. This makes it likely that irrelevant variables are included in the model. Multicollinearity is likely to occur because the majority of the variables are based on common, related financial ratios. One way to reduce this problem is to eliminate independent variables that are too highly correlated or that are not useful for predicting earnings changes. To do this, we follow prior literature and implement stepwise logit regression (Ou and Penman 1989; Holthausen and Larcker 1992; Dechow, Larson, and Sloan 2011).

**Forecasting Method 1: Stepwise Logit Regression**

The first forecasting method that we consider is stepwise logit regression. Many of the potentially useful financial variables are highly correlated with one another. Including all 60 potential independent variables in the logit regression model would cause severe multicollinearity, leading to bias and to an over fitted model which is unlikely to perform well out of sample. In addition, some of these independent variables may not be useful for predicting earnings changes at all. To form a well specified and parsimonious model, we implement the logit regression estimation using a stepwise procedure. Specifically, we begin by estimating the logit regression with the complete set of variables from Ou and Penman (1989). We then augment their analysis by adding PE and the change in PE as additional independent variables.

Stepwise logit systematically eliminates the least helpful independent variables. The logit model is estimated including all 60 independent variables. The least helpful variable is identified and removed from the regression, and then the model re-estimates and tests for a change in the overall model fit (James, Whitten, Hastie, and Tibshirani 2013). When removing the least helpful variable does not significantly reduce the model’s fit, the model permanently
removes the variable in question and repeats the procedure. This procedure continues until removing any of the remaining predictor variables significantly reduces the model’s fit. The resulting logit model is the one that we tabulate and use to compare with the other two prediction models.

**Forecasting Method 2: Elastic Net**

The second prediction model that we consider is elastic net. Elastic net is a parametric model that modifies the logit model to overcome three of stepwise logit’s weaknesses. First, the stepwise procedure performs poorly under multicollinearity (Judd and McClelland 1989; Tibshirani 1996). When the independent variables are highly correlated with one another, the stepwise procedure cannot reliably identify which of the correlated variables is least important. Therefore, the stepwise procedure can inadvertently delete important variables (Hosmer, Lemeshow, and Sturdivant 2013). Second, logit models can overfit to the sample data, resulting in inflated coefficient estimates and poor out-of-sample prediction. Third, stepwise logit regression does not guarantee the best fit because not every combination of the independent variables is tested.

The elastic net method modifies stepwise logit regression by adding two additional constraints to the logit maximum likelihood estimation. The first constraint minimizes of the sum of the squared coefficient estimates. Because logit regression coefficients are often too large for out-of-sample prediction, this constraint may improve our out-of-sample accuracy. The second constraint minimizes the sum of the absolute values of the coefficient estimates. This forces the model to eliminate the variables that have little impact on the likelihood function. This forces the method to choose parsimonious model, preventing the inclusion of unimportant

---

6 This constraint is developed in James et al. (2013). A logit model with this one additional constraint is called a ridge regression.
Comparing Equations (1) and (2) reveals how logit regression and elastic net are related. Under elastic net, we minimize the negative log likelihood with the two additional constraints:

\[
\begin{align*}
\log P(y|\beta, x) &= -\left( \sum_{i=1}^{m} y_i \log \left( \frac{1}{1+\exp(-x_i \beta)} \right) + (1 - y_i) \log \left( \frac{\exp(-x_i \beta)}{1+\exp(-x_i \beta)} \right) \right) \\
&\quad + \frac{\varphi}{2} \sum_{i=1}^{k} \beta_i^2 + \frac{\delta}{2} \sum_{i=1}^{k} |\beta_i| 
\end{align*}
\]

The penalized maximum likelihood estimation includes a tuning parameter or shrinking penalty, \( \varphi \), where higher values increase the penalty and lower values decrease the penalty. When the tuning parameter is zero, the model is a standard logit regression, but as the tuning parameter approaches infinity, all of the coefficients approach zero (James et al. 2013, pg. 215). Because elastic net shrinks the coefficients and the coefficient magnitude is dependent on their scale, the inputs must all be standardized. We use a standard z-score standardization, where the independent variables are demeaned and scaled by their standard deviation each year.

Thus, when the sample changes (for example, when we make out-of-sample forecasts), the coefficient estimates change very little. Although this attenuates the coefficient estimates to some extent, the effect is strongest for coefficients that have a small effect on the prediction. The second constraint provides a mechanism for selecting a subset of variables, similar to stepwise logit regression. This constraint effectively selects the most important variables. It contains a tuning parameter, \( \delta \), that controls the amount of shrinkage. Again, we use cross-validation to identify the best tuning parameter (as discussed in more detail in section II). These two modifications mechanically reduce the goodness of fit, but often improve out-of-sample accuracy.
Elastic net is subject to the basic assumptions of the logit regression. The main weakness is the parametric assumption present in logit regression. Elastic net also requires that the variables be standardized. The algorithm shrinks coefficients and if the coefficients do not have the same scale, then it will perform poorly.

**Cross-Validation**

Elastic net requires two tuning parameters (\( \varphi \) and \( \delta \)). To estimate these, we use cross-validation (which is a systematic method of resampling) to bootstrap optimal values for these parameters. Specifically, we use ten-fold cross-validation. Here, each training data set includes a five-year period. We draw ten random samples from each training data set and nine of the ten random samples are used to identify the optimal weights of the elastic net penalty functions. The weights on the penalty functions are randomly generated and tested on the tenth random sample so that the accuracy of each random weight can be measured. This process is completed nine more times, using a different random sample each time, but using the same initial weights. The test sample accuracy is averaged over each of the ten estimates and the random weights that produce the highest accuracy are selected. The model is then estimated on the entire training sample, using these selected weights. This yields our elastic net forecast.

**Forecasting Method 3: Random Forest**

The final forecasting method that we consider is random forest. Random forest is a nonparametric machine learning method. Both stepwise logit regression and elastic net are parametric models which assume that the independent variables are linearly related to the log odds of the dependent variable. However, this assumption may not always hold. For example, using scatter plots, Easton (1999) shows that there are the nonlinearities present in the returns-earnings relation and that the findings in Hayn (1995) and Basu (1997) can be visually confirmed
using these plots. Because of potential nonlinearities, a nonparametric forecasting model has the potential to improve accuracy.

In order to describe random forest, we begin by explaining the components of the model: decision trees and tree bagging. Decision trees are comprised of a set of nodes. Each node creates a step that represents a value of one of the independent variables. For example, if the model detects that increasing earnings are more likely for firms with large values of total assets than for firms with small values of total assets, it forms a node that has two important aspects. The first is identifying total assets as a useful partitioning variable, and the second is identifying the optimal value for partitioning the observations into the large and small total asset groups. From this step, the decision tree may create an additional node. For example, for the set of observations with small values of total assets, the model may identify the magnitude of cash flows as a second partitioning variable. This process solves for a local optimum at each node, with the goal of finding a global optimum. Specifically, it determines best alternative for the current node, but it does not consider future nodes. The process is recursive, meaning that it continues to create additional nodes. The sequence of the node creation is based on ‘purity’ or how well the node separates the data into distinct classes; every partitioning of every possible variable is considered until the partition with the highest purity is identified. This happens at each node and the process continues until additional partitions would not improve the accuracy of classification. This process will produce a decision tree that partitions observations into high and low likelihoods of a future earnings increase based on observable values of the 60 independent variables. We then use this decision tree to forecast the direction of next year’s earnings change. The out-of-sample prediction for each observation is based on the decision tree.
generated using the training sample over the previous five years applied to the currently observable values of the independent variables.

Decision trees have several strengths. First, they are nonparametric and require no distributional assumptions. Second, if the trees are small, they can be easily interpreted. Third, decision trees can accommodate a large number of independent variables and are robust to outliers and collinear variables. Overall, random forest is the most flexible of the three models that we consider, so it has the potential to improve out-of-sample prediction accuracy.

The main disadvantage of decision trees is high variability, meaning that a small change in the sample can lead to a large change in the final tree (James et al. 2013). That is, decision trees tend to overfit the training data and so they can perform poorly out-of-sample. In order to overcome this limitation, random forest identifies a large number of different decision trees using a bootstrapping technique referred to as tree bagging. Tree bagging is a method for averaging over many decision trees, where each tree is trained on a random subset of the data. This general-purpose machine learning tool is useful to reduce model variance. Tree bagging forms decision trees on bootstrapped samples (with replacement) taken from subsets of the complete training data set. Each bootstrapped sample will result in a different decision tree, and these trees are then averaged to produce the overall best fitting decision tree. Tree bagging improves classification but it loses the simple tree structure.

Tree bagging is only effective when the individual decision trees differ from one another. Random forest achieves diversity by randomly choosing a subset of input variables at each node. This is done for each tree that is grown on a bootstrapped sample. For example, if the chosen number of input variables is four, then four out of the possible 60 variables are chosen at random at each node of the decision tree. One of those four variables will provide the most purity and
will be chosen for that step in the tree. For the next step, another four random variables will be chosen from the original set of 60 variables, and the one that provides the most purity at that stage will determine the next step in the tree. With each iteration, the randomly chosen variables will differ, resulting in a unique decision tree. In the final step, to classify each observation as a predicted increase or decrease, every tree in the forest receives a vote on whether the outcome is or is not positive and the outcome based on the majority of votes.

The number of variables chosen at each step is a tuning parameter. Using a greater number of variables in each iteration produces better fit for an individual tree but results in less diversity among the trees. We use cross-validation to choose the optimal size of the subset of variables used to construct individual trees. Specifically, we try a random set of possible independent variables, limited only by the total number of variables available. We choose the number that produces the best cross-validation accuracy.7

Because random forest uses random variables at each node, the resulting trees are not highly correlated with one another by construction. To illustrate, suppose the model uses a random subset of four variables at each point in the decision tree. Because we have 60 independent variables, there is only a 4/60 chance that the variables used in the first node for one tree would be available for the first node of the next tree. Random forest benefits from the strengths of decision trees in that it performs well in the presence of outliers and highly correlated variables. It also performs well with a very large set of predictor variables. Moreover, random forest (and other tree methods) do not require any variable transformations, unlike many other machine learning algorithms (including elastic net). Finally, random forest can easily be applied to data sets with missing data, and can identify outliers and natural clusters.

---

7 We set the number of trees in the forest to 500 and let the number of variables vary from 6 to 28, with a mean of 17.
in the data.

Although random forest has many benefits, it is not the best method in all circumstances. For example, logit regression will outperform nonparametric models when the assumptions underlying logit regression are true (Shmueli, Patel, and Bruce 2010, pg. 338). However, if the parametric assumption is false, then random forest might outperform logit regression-based models.

III. Data and Methods

Each of the three models begins with the same 60 financial ratios and uses a subset of these to predict the sign of one-year-ahead earnings changes. These 60 ratios are based on the financial ratios identified in Ou and Penman (1989). In addition, we add PE and the change in PE.

To be included in a forecasting model, an observation must have a valid value for the dependent variable and for each of the independent variables. Therefore, the greater the number of independent variables that are included in the analysis, the fewer the observations that will have a full set of available data. Thus, there is an inherent tradeoff between having a large number of observations and having a large number of independent variables. Following Perols (2016), we compromise by eliminating any independent variable that is not available for at least 75 percent of the potential sample. There are 70 variables in total, 68 identified by Ou and Penman (1989) plus PE and change in PE. Ten of these variables are eliminated because they are missing for more than 25 percent of the observations. This leaves us with 60 viable input variables and 116,904 firm-year observations in our training samples. Our sample is fairly balanced, with approximately 58 percent of the sample comprised of positive changes. Each
forecast is based on only five years of training data so the average training sample size is 13,105 observations. Each of the 60 viable input variables is defined in the appendix. We make out-of-sample predictions for 108,913 observations with Compustat data from 1976 through 2015. For the returns prediction, we add the additional criteria that firms must have a December fiscal year-end, total assets of greater than $1 million, data to calculate returns available be available from the Center for Research in Security Prices, and beginning of the year share prices of at least one dollar. This results in a sample of 75,489 company-year observations for use in our returns tests.

We follow Ou and Penman (1989) and Holthausen and Larcker (1992) and use five-year rolling windows as our training sample, to predict the sign of earnings changes.\(^8\) We then use the prediction model to estimate the sign of the earnings change in the sixth year. For example, the first training sample is 1971 through 1975 and we use this sample to predict the sign of earnings changes in 1976. The out-of-sample accuracy is measured as the percentage of observations where the model trained on 1971 through 1975 data correctly predicts the direction of the earnings change in 1976. Each year, we roll the window forward, so our second training sample is 1972 through 1976, and we use this to predict the sign of earnings changes in 1977.

Each model chooses only a subset of the 60 independent variables. The included variables and the size of the coefficients differ each year. The dependent variable takes a value of one when the change in diluted earnings per share from year \(t-1\) to year \(t\) is positive, and zero otherwise.\(^9\) Each prediction model provides the estimated probability of an earnings increase for

---

\(^8\) Our methodology differs from Ou and Penman (1989) and Holshausen and Larker (1992) because those papers use five years of data to construct a prediction model. They then use this same prediction model to make forecasts for each of the next five years. In contrast, we use five years of training data to construct our prediction model, but we update the prediction model each year. Our method is more computer intensive but results in significantly improved predictions for all three models.

\(^9\) We do not adjust for the trend in earnings (as in Ou and Penman (1989) and Holthausen and Larcker (1992)) because calculating a firm-specific time trend would require a history of reported earnings for each firm, and this requirement would reduce the sample size considerably.
each observation. However, the distribution of the probability estimates differs from model to model, making comparisons based on raw probability estimates difficult. Using raw probability cutoffs yields different sample sizes and unbalanced top and bottom groups. Therefore, rather than using the directly estimated probability, we follow Holthausen and Larcker (1992) and rank the probabilities in order to have more balanced cutoffs, (i.e., we rank probabilities for each model and split the sample based on quantiles).

The 50/50 split uses the sample median to separate observations, predicting an increase in earnings for one half of the sample and a decrease in earnings for the other half of the sample. At the other extreme, the 95/05 split compares the top five percent of observations (those strongly predicted to experience an increase in earnings) with the bottom five percent of observations (those strongly predicted to experience a decrease in earnings). Using this methodology not only balances the number of observations in the top and bottom groups, but holds the number of observations consistent across models.10

IV. Results

We measure the model usefulness in two ways. First, we measure out-of-sample prediction accuracy. That is, how well the models and the calculated parameter estimates predict the direction of earnings changes in the next period. Second, we measure the profitability of a trading strategy based on these out-of-sample predictions.

Out-of-sample earnings prediction accuracy

One of the most straightforward measures of the relative usefulness of a model is its ability to predict the sign of the future earnings change. In addition to predicting whether an

10 All inferences remain qualitatively similar using raw probabilities of future changes in earnings.
observation (firm) will experience an increase in earnings, each model estimates the *probability* that the observation will experience an increase in earnings. At the extremes, confidence in the models’ predictions is very high, but at the center of the distribution, the models do not provide a strong signal in either direction. In their analysis, Ou and Penman (1989) do not make a prediction for the 20 percent of the observations in the center of the distribution. Thus, they evaluate the accuracy of the predictions and the investment returns only for observations at the top and bottom 40 percent of the sample. We calculate the accuracy using six different cutoffs. First, we partition the sample at the median level of confidence for an increase, which includes all observations in the accuracy analysis. Second, we eliminate the middle 20 percent of observations and compare the top 40 percent to the bottom 40 percent of observations. This is analogous to the Ou and Penman sample. We then repeat the procedure eliminating the middle 40, 60, 80, and 90 percent. Not surprisingly, all models become more accurate as the subsamples become more restrictive.

Table 1 presents the total number of observations per data split as well as the out-of-sample accuracy for each split and model. In Panel A, we use only the original Ou and Penman (1989) variables, and in Panel B, we add PE and percent change in PE as additional variables. The first model presented is stepwise logit regression. Overall, stepwise logit provides a statistically significant prediction of the sign of future earnings changes. However, simply partitioning the sample into the 50 percent most likely to have an increase versus the 50 percent least likely to have an earnings increase provides only modest results.

< INSERT TABLE 1>

We first analyze the results without PE and the change in PE. Random chance would lead to 50 percent of forecasts having the correct sign, and we find that the model is correct only
54 percent of the time. However, the model's accuracy increases monotonically as the subsamples become more restrictive. For the 10 percent of the sample with the strongest signals (i.e., the top and bottom five percent of the sample), the model correctly predicts the sign of the earnings change 62.3 percent of the time.

The second model is elastic net. Elastic net does not significantly outperform stepwise logit in any of the six sample splits, so we do not find any evidence that elastic net provides any benefit over stepwise logit in our setting. The fact that stepwise logit regression performs as well as elastic net suggests that the limitations of stepwise logit in this setting are not due to overfitting or multicollinearity.

The final column of Table 1 presents the prediction accuracy for the random forest model. Random forest significantly outperforms both of the other models for every split.\textsuperscript{11} When random forest is restricted to the Ou and Penman (1989) variables, the 50/50 split is correct 60.1 percent of the time, and as is the case with the other models, the accuracy of random forest is monotonic in the restrictiveness of the sample splits. For the 95/05 split, random forest is accurate almost 76.8 percent of the time, which is much more accurate than the other two models. To put the accuracy into perspective, for the 95/05 split, stepwise logit’s predictions are incorrect 37.7 percent of the time, but random forest is incorrect only 23.2 percent of the time.

In Panel B, we allow the models to utilize PE and change in PE as additional explanatory variables. Interestingly, the stepwise logit and elastic net models seem not to utilize PE, so its inclusion makes no difference to their predictions, and thus, to their accuracy. However, the inclusion of PE significantly improves the accuracy of random forest. For the 95/05 split,

\textsuperscript{11} We measure statistical significance using McNemar’s test (McNemar 1947). One limitation of this test is that it compares accuracy for only those observations that have a prediction from all of the models. For the 50/50 split, this is not problematic but for the other splits, not every observation has a forecast from all models.
including PE and change in PE improves the accuracy by approximately 3 percent, with random forest correctly predicting the sign of the earnings change almost 79.8 percent of the time.\textsuperscript{12}

In Table 2, we examine the out-of-sample accuracy for the 95/05 split across nine non-overlapping five-year periods. We find that random forest is the most accurate method in all nine periods. This holds for both the Ou and Penman variable tests in Panel A and for the full variable set tests in Panel B. Therefore, the superiority of random forest is stable over time.

\textless INSERT TABLE 2\textgreater

Next, we investigate which input variables are most important. Table 3 presents the ten input variables that are selected most often for stepwise logit regression, elastic net, and random forest, as well as the number of years in which each respective variable is selected. The largest possible number of years is 40. Random forest selects DEPR_PPE and PERATIO in every sample. Elastic net selects OPMBD in 35 of the 40 years, and similarly, the most frequently selected variables by stepwise logit regression (PCHG_AT and PCHG_CURRENT) are selected in 35 years. Elastic net selects one of the same input variables as random forest (CASH_DEBT), and stepwise logit regression and random forest select two common variables (NPM and CASH_DEBT). CASH_DEBT is the only input variable selected by all three models. Interestingly, PE and change in PE are important in 40 and 38 year respectively for random forest by not in the top ten for either of the other models. This explains why adding PE significantly improves random forest but has essentially no effect for the other two models.

\textless INSERT TABLE 3\textgreater

\textsuperscript{12} The out-of-sample area under the ROC is similar with PE. Random forest performs significantly better with explanatory power of 66.7 versus 54.9 for logistic regression.
Using the model predictions for stock picking

The original motivation behind Ou and Penman (1989) was financial statement analysis. The purpose of predicting earnings changes is to identify a trading strategy.

Trading strategy returns based on the Ou and Penman (1989) approach

For each of the models, we use the data available in year $t$ to predict the direction of the change in earnings in year $t+1$. Trading begins four months after fiscal year-end $t$ (i.e., when current-year results would be widely available for all firms). We form a portfolio that is long in the five percent of firms that the model predicts are most likely to experience an earnings increase, and short in the five percent of firms that the model predicts are most likely to experience an earnings decrease. We calculate the 12 month returns to this portfolio beginning 4 months after the first year-end.

Table 4, Panel A compares the abnormal returns to portfolios generated using each of the models restricted to the Ou and Penman (1989) variables. Over the 12 months, the stepwise logit regression-based portfolio and the elastic net portfolio earn significant abnormal returns of 3.4 and 4.2 percent, respectively. The random forest method is the clear winner, earning abnormal returns of 8.4 percent; 7 percent is generated from observations with a positive predicted change (i.e., from the long side of the portfolio) and 1.4 percent is generated from observations with a negative predicted change (i.e., from the short side of the portfolio).

< INSERT TABLE 4>

The false positive (FP) and false negative (FN) rows present the financial cost of incorrect predictions. False positives occur when the model predicts an earnings increase but the actual earnings decline, and false negatives occur when the model predicts an earnings decrease but actual earnings increase. For the random forest group, there are 582 false positive
observations with an average abnormal return equal to \(-21.1\) percent. Here, random forest was 95 percent confident that these firms would experience an earnings increase but they actually experienced an earnings decrease. In addition, there are 1,166 false negatives (i.e., observations where the model predicted an earnings decrease with 95 percent confidence but earnings actually increase); these firms earned positive abnormal returns of 19.5 percent. The superiority of random forest stems from its ability to avoid both false positives and false negatives relative to the other two models.

Adding PE to the analyses significantly improves random forest’s performance. Panel B presents the results when the prediction models use the full set of variables. Although the abnormal returns are essentially unchanged for the stepwise logit model and for elastic net, the random forest abnormal returns improve by 210 basis points. This improvement comes from both the positive and the negative sides – there are approximately 60 fewer false positives and more than 150 fewer false negatives when random forest uses PE. In untabulated results, the parametric models generate abnormal returns in the earlier portion of our sample, namely 1981-1985, whereas random forest does not generate significant abnormal returns until 1991. These results confirm the OP findings and suggest that the relation between predicted earnings changes and abnormal returns has changed over time.

**Trading strategy returns based on the Holthausen and Larker (1992) approach**

Whereas Ou and Penman (1989) uses a two-step approach to create a trading strategy, Holthausen and Larker (1992) uses stepwise logit to directly predict the sign of stock returns where the dependent variable in the prediction model is the direction of one-year-ahead stock returns. In this section, we follow the Holthausen and Larker (1992) approach and use each of the three models to directly predict the sign of stock returns. We then evaluate the effectiveness
of the models in two ways. First, we calculate the out-of-sample accuracy in terms of the frequency of correct predictions, and second, we calculate the abnormal returns for a portfolio that is generated using the predictions.

Table 5 presents the accuracy for the Holthausen and Larker (1992)-type forecasts. Using this methodology, we find that the random forest method does not dominate the other two methods. Instead, the stepwise logit regression method is most accurate for the majority of the splits. However, for the 10 percent of the sample where the confidence of the model predictions is highest, all three models predict the correct sign of the change in stock price 56.1 percent of the time. This is consistent with accounting numbers providing a modest portion of new information to the equity market (Ball 2013).

< INSERT TABLE 5>

Table 6 presents the models’ forecast accuracy by five-year periods. Stepwise logit is the most accurate in 4 of the 8 periods, and random forest is most accurate in only two periods. Perhaps not surprisingly, returns predictions are much less accurate than earnings predictions.

< INSERT TABLE 6>

Table 7 presents abnormal returns generated from the Holthausen and Larker (1992) strategy for each of the three models. Overall, the elastic net method yields the most profitable strategy. Using this approach generates abnormal returns of 7.2 percent over one year but this is approximately 300 basis points less than the 10.1 percent abnormal returns generated using the random forest method to predict the sign of changes in earnings and taking positions accordingly.

< INSERT TABLE 7>
VI. Conclusion

The goal of this paper is to test whether machine learning can improve our ability to forecast the sign of earnings changes and whether these forecasts are useful for returns prediction. We find that machine learning significantly improves the prediction of the direction of earnings changes. We do not find that elastic net improves earnings prediction relative to a traditional stepwise logit model, but we find that random forest provides better out-of-sample accuracy overall, as well as in every subsample, relative to stepwise logit regression and to net elastic. Trading strategies based on the random forest predictions also earn significantly higher abnormal returns than the other models. We also examine model performance for different time periods and find that random forest outperforms the other models in all time periods. Overall, this evidence suggests that logit regression-based models do not exploit all of the available information from the independent variables.

Although we only investigate one nonparametric method, we suggest that others would likely be useful in this setting. We use random forest because it is easily understood relative to other machine learning methods and because it does not require data preparation (that is, the variables do not need to be standardized). Future research could investigate whether alternative nonparametric methods such as artificial neural networks, support vector machines, and k-nearest neighbors yield better results in this or other settings.

Classification tasks are important in a variety of research settings. For example, binary outcomes examined in the accounting and finance literature include bankruptcy and financial distress (Ohlson 1980; Beaver, McNichols, and Rhie 2005; Campbell, Hilscher, and Szilagyi 2008; Beaver, Correia, and McNichols 2012), goodwill impairments (Francis, Hanna, and Vincent 1996; Hayn and Hughes 2006; Gu and Lev 2011; Li, Shroff, Venkataraman, and Zhang 2012).
2011; Li and Sloan 2017), write-offs (Francis et al. 1996), restructuring charges (Francis et al.
1996; Bens and Johnston 2009), initial public offerings (Friedlan 1994; Pagano, Panetta, and
Zingles 1998; Teoh, Welch, and Wong 1998; Brau, Francis, and Kohers 2003; Boehmer and
Ljungqvist 2004; Brau and Fawcett 2006), seasoned equity offerings (McLaughlin, Safieddine,
and Vasudevan 1996; Guo and Mech 2000; DeAngelo, DeAngelo, and Stulz 2009; Alti and
Sulaeman 2012; Jindra 2013; Deng, Hrnjic, and Ong 2014), and Accounting and Auditing
Enforcement Releases (Beasley 1996; Dechow, Sloan, and Sweeney 1996; Beneish 1999;
Erickson, Hanlon, Maydew 2006; Dechow et al. 2011; Feng, Ge, Luo, and Shevlin 2011; Perols
2011; Price, Sharp, and Wood 2011; Hribar, Kravet, and Wilson 2014). Therefore, the use of
methods from data analytics should receive more attention from accounting researchers.
Although we focus solely on predicting changes in earnings, the improved accuracy of these
models could benefit other binary outcomes examined in the accounting literature as well.
References


Appendix

Variables included in the earnings change analyses

POSCHG = 1 if the change in diluted earnings per share from t to t+1 is positive, 0 otherwise

PERATIO = End of fiscal year stock price / diluted earnings per share

PCHG_PERATIO = Percent change in PERATIO from t-1 to t

CURRENT = Current assets / Current liabilities

PCHG_CURRENT = Percent change in CURRENT from t-1 to t

QUICK = (Current assets – Total inventory) / Current liabilities

PCHG_QUICK = Percent change in QUICK from t-1 to t

DAYS_SALEAR = 360 / (Total sales / Total receivables)

PCHG_DAYS_SALEAR = Percent change in DAYS_SALEAR from t-1 to t

INV_TURN = Cost of goods sold / Total inventory

PCHG_INV_TURN = Percent change in INV_TURN from t-1 to t

INV_AT = Total inventory / Total assets

PCHG_INVT_AT = Percent change in INV_AT from t-1 to t

PCHG_INVT = Percent change in Total inventory from t-1 to t

PCHG_SALE = Percent change in Total sales from t-1 to t

PCHG_DEPR = Percent change in Depreciation from t-1 to t

CHG_DIVIDEND = Change in Dividend per share from t-1 to t

DEPR_PPE = Depreciation / Fixed assets

PCHG_DEPR_PPE = Percent change in DEPR_PPE from t-1 to t

ROE_OPEN = Income / (Prior year stock holders’ equity + Prior year deferred taxes – Prior year preferred stock)

PCHG_ROE_OPEN = Percent change in ROE_OPEN from t-1 to t
PCHG\_CAPXAT = Percent change in Capital expenditures / Total assets from t-1 to t

LAG1\_PCHG\_CAPXAT = Percent change in Capital expenditures / Total assets from t-2 to t-1

DEBT\_EQ = Total debt / Stockholders’ equity

PCHG\_DEBT\_EQ = Percent change in DEBT\_EQ from t-1 to t

LTDEBT\_EQ = Long-term debt / (Prior year stockholders’ equity + Prior year deferred taxes – Prior year preferred stock)

EQ\_FA = (Common equity + Preferred stock) / Fixed assets

PCHG\_EQ\_FA = Percent change EQ\_FA from t-1 to t

TIE = (Earnings + Interest + taxes) / Interest expense

PCHG\_TIE = Percent change in TIE from t-1 to t

SALE\_AT = Total sales / Total assets

PCHG\_SALE\_AT = Percent change in SALE\_AT from t-1 to t

ROA = Operating income / Total assets

ROE\_CLOSE = Income / (Stockholders’ equity + Deferred taxes – Preferred stock)

GPM = Gross profit / Total sales

PCHG\_GPM = Percent change in GPM from t-1 to t

OPMBD = Operating income / Total sales

PCHG\_OPMBD = Percent change in OPMBD from t-1 to t

PTPM = Pretax income / Total sales

PCHG\_PTPM = Percent change in PTPM from t-1 to t

NPM = Income / Total sales

PCHG\_NPM = Percent change in NPM from t-1 to t

SALE\_CASH = Total sales / Cash
SALE_RECT = Total sales / Total receivables

SALE_INV = Total sales / Total inventory

PCHG_SALE_INV = Percent change in SALE_INV from t-1 to t

SALE_NWC = Total sales / (Current assets – Current liabilities)

PCHG_SALE_NWC = Percent change in SALE_NWC from t-1 to t

SALE_FA = Total sales / Fixed assets

PCHG_COGS = Percent change in Cost of goods sold from t-1 to t

PCHG_AT = Percent change in Assets from t-1 to t

CASH_DEBT = (Income + Depreciation) / Total debt

WKCAP_AT = (Current assets – Current liabilities) / Total assets

PCHG_WKCAP_AT = Percent change in WKCAP_AT from t-1 to t

OPINC_AT = Operating income / Total assets

PCHG_OPINC_AT = Percent change in OPINC_AT from t-1 to t

REPAY_DEBT = Long-term debt reduction / Long-term debt

PCHG_DLT = Percent change in long-term debt from t-1 to t

DIV_CASH = Cash dividends / (Income + Depreciation)

PCHG_WORKINGCAP = Percent change in (Current assets – Current liabilities) from t-1 to t

NI_CF = Net income / (Income + Depreciation)
Table 1 Model accuracy forecasting the direction of earnings change over the entire sample 1976-2015

Panel A: PE not included

<table>
<thead>
<tr>
<th>Split</th>
<th>N</th>
<th>Stepwise Accuracy</th>
<th>Elastic Net Accuracy</th>
<th>Random Forest Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>50/50</td>
<td>75,489</td>
<td>0.544</td>
<td>0.544</td>
<td>0.601</td>
</tr>
<tr>
<td>60/40</td>
<td>60,404</td>
<td>0.553</td>
<td>0.552</td>
<td>0.623</td>
</tr>
<tr>
<td>70/30</td>
<td>45,302</td>
<td>0.565</td>
<td>0.564</td>
<td>0.648</td>
</tr>
<tr>
<td>80/20</td>
<td>30,206</td>
<td>0.581</td>
<td>0.579</td>
<td>0.681</td>
</tr>
<tr>
<td>90/10</td>
<td>15,106</td>
<td>0.603</td>
<td>0.599</td>
<td>0.728</td>
</tr>
<tr>
<td>95/05</td>
<td>7,550</td>
<td>0.623</td>
<td>0.623</td>
<td>0.768</td>
</tr>
</tbody>
</table>

Panel B: PE included

<table>
<thead>
<tr>
<th>Split</th>
<th>N</th>
<th>Stepwise Accuracy</th>
<th>Elastic Net Accuracy</th>
<th>Random Forest Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>50/50</td>
<td>75,489</td>
<td>0.544</td>
<td>0.544</td>
<td>0.616</td>
</tr>
<tr>
<td>60/40</td>
<td>60,404</td>
<td>0.553</td>
<td>0.553</td>
<td>0.639</td>
</tr>
<tr>
<td>70/30</td>
<td>45,302</td>
<td>0.565</td>
<td>0.564</td>
<td>0.667</td>
</tr>
<tr>
<td>80/20</td>
<td>30,206</td>
<td>0.580</td>
<td>0.579</td>
<td>0.700</td>
</tr>
<tr>
<td>90/10</td>
<td>15,106</td>
<td>0.603</td>
<td>0.598</td>
<td>0.755</td>
</tr>
<tr>
<td>95/05</td>
<td>7,550</td>
<td>0.621</td>
<td>0.620</td>
<td>0.798</td>
</tr>
</tbody>
</table>

*Table 1 presents analyses using 5-year groups along with the corresponding out-of-sample accuracy. The bolded numbers represent the largest accuracies.*
Table 2 Forecast accuracy for prediction of earnings changes - Five-year groups
95/05 Split

Panel A: PE not included

<table>
<thead>
<tr>
<th>Years</th>
<th>Stepwise Logit</th>
<th>Elastic Net</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976-1980</td>
<td>0.555</td>
<td>0.556</td>
<td>0.698</td>
</tr>
<tr>
<td>1981-1985</td>
<td>0.609</td>
<td>0.612</td>
<td>0.725</td>
</tr>
<tr>
<td>1986-1990</td>
<td>0.587</td>
<td>0.596</td>
<td>0.740</td>
</tr>
<tr>
<td>1991-1995</td>
<td>0.656</td>
<td>0.653</td>
<td>0.760</td>
</tr>
<tr>
<td>1996-2000</td>
<td>0.662</td>
<td>0.662</td>
<td>0.777</td>
</tr>
<tr>
<td>2001-2005</td>
<td>0.642</td>
<td>0.635</td>
<td>0.766</td>
</tr>
<tr>
<td>2006-2010</td>
<td>0.579</td>
<td>0.599</td>
<td>0.807</td>
</tr>
<tr>
<td>2011-2015</td>
<td>0.636</td>
<td>0.621</td>
<td>0.830</td>
</tr>
</tbody>
</table>

Panel B: PE included

<table>
<thead>
<tr>
<th>Years</th>
<th>Stepwise Logit</th>
<th>Elastic Net</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976-1980</td>
<td>0.548</td>
<td>0.542</td>
<td>0.709</td>
</tr>
<tr>
<td>1981-1985</td>
<td>0.609</td>
<td>0.609</td>
<td>0.750</td>
</tr>
<tr>
<td>1986-1990</td>
<td>0.584</td>
<td>0.597</td>
<td>0.738</td>
</tr>
<tr>
<td>1991-1995</td>
<td>0.656</td>
<td>0.653</td>
<td>0.789</td>
</tr>
<tr>
<td>1996-2000</td>
<td>0.662</td>
<td>0.660</td>
<td>0.807</td>
</tr>
<tr>
<td>2001-2005</td>
<td>0.636</td>
<td>0.634</td>
<td>0.811</td>
</tr>
<tr>
<td>2006-2010</td>
<td>0.577</td>
<td>0.585</td>
<td>0.849</td>
</tr>
<tr>
<td>2011-2015</td>
<td>0.636</td>
<td>0.624</td>
<td>0.865</td>
</tr>
</tbody>
</table>

Table 2 presents analyses using 5-year groups along with the corresponding out-of-sample accuracy. The bolded numbers represent the largest accuracies.
Table 3 Top ten most important variables for forecasts of earnings changes

<table>
<thead>
<tr>
<th>Random Forest Variables</th>
<th>Freq</th>
<th>Elastic Net Variables</th>
<th>Freq</th>
<th>Stepwise Variables</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPR_PPE</td>
<td>40</td>
<td>OPMBD</td>
<td>35</td>
<td>PCHG_AT</td>
<td>35</td>
</tr>
<tr>
<td>PERATIO</td>
<td>40</td>
<td>PTPM</td>
<td>32</td>
<td>PCHG_CURRENT</td>
<td>35</td>
</tr>
<tr>
<td>PCHG_PERATIO</td>
<td>38</td>
<td>SALE_AT</td>
<td>30</td>
<td>PTPM</td>
<td>34</td>
</tr>
<tr>
<td>NPM</td>
<td>37</td>
<td>GPM</td>
<td>29</td>
<td>NPM</td>
<td>32</td>
</tr>
<tr>
<td>TIE</td>
<td>36</td>
<td>PCHG_QUICK</td>
<td>29</td>
<td>PCHG_DEBT_EQ</td>
<td>31</td>
</tr>
<tr>
<td>CASH_DEBT</td>
<td>34</td>
<td>CASH_DEBT</td>
<td>25</td>
<td>PCHG_NPM</td>
<td>31</td>
</tr>
<tr>
<td>ROE_CLOSE</td>
<td>34</td>
<td>CURRENT</td>
<td>25</td>
<td>PCHG_PTPM</td>
<td>31</td>
</tr>
<tr>
<td>NI_CF</td>
<td>33</td>
<td>PCHG_CURRENT</td>
<td>25</td>
<td>PCHG_INVIT</td>
<td>30</td>
</tr>
<tr>
<td>PCHG_CAPXAT</td>
<td>33</td>
<td>PCHG_SALE</td>
<td>25</td>
<td>CASH_DEBT</td>
<td>29</td>
</tr>
<tr>
<td>SALE_CASH</td>
<td>30</td>
<td>QUICK</td>
<td>25</td>
<td>PCHG_WORKINGCAP</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 3 shows the top ten most commonly chosen independent variables over the sample period 1976 through 2015. The numbers represent the corresponding number of times that the variable was chosen, with 40 being the largest possible number. All variables are defined in the appendix.
Table 4 Abnormal returns based on earnings prediction strategy 1976-2015

Panel A: PE not included

<table>
<thead>
<tr>
<th></th>
<th>Stepwise Logit Regression</th>
<th>Elastic Net</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BHAR</td>
<td>PValue</td>
<td>N</td>
</tr>
<tr>
<td>Hedge</td>
<td>0.034</td>
<td>0.084</td>
<td>7550</td>
</tr>
<tr>
<td>PP</td>
<td>-0.023</td>
<td>0.125</td>
<td>3795</td>
</tr>
<tr>
<td>PN</td>
<td>-0.057</td>
<td>0.000</td>
<td>3755</td>
</tr>
<tr>
<td>TP</td>
<td>0.050</td>
<td>0.010</td>
<td>2590</td>
</tr>
<tr>
<td>TN</td>
<td>-0.176</td>
<td>0.000</td>
<td>2111</td>
</tr>
<tr>
<td>FP</td>
<td>-0.181</td>
<td>0.000</td>
<td>1205</td>
</tr>
<tr>
<td>FN</td>
<td>0.096</td>
<td>0.000</td>
<td>1644</td>
</tr>
</tbody>
</table>

Panel B: PE included

<table>
<thead>
<tr>
<th></th>
<th>Stepwise Logit Regression</th>
<th>Elastic Net</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BHAR</td>
<td>PValue</td>
<td>N</td>
</tr>
<tr>
<td>Hedge</td>
<td>0.034</td>
<td>0.076</td>
<td>7550</td>
</tr>
<tr>
<td>PP</td>
<td>-0.025</td>
<td>0.090</td>
<td>3795</td>
</tr>
<tr>
<td>PN</td>
<td>-0.060</td>
<td>0.000</td>
<td>3755</td>
</tr>
<tr>
<td>TP</td>
<td>0.049</td>
<td>0.013</td>
<td>2583</td>
</tr>
<tr>
<td>TN</td>
<td>-0.178</td>
<td>0.000</td>
<td>2104</td>
</tr>
<tr>
<td>FP</td>
<td>-0.184</td>
<td>0.000</td>
<td>1212</td>
</tr>
<tr>
<td>FN</td>
<td>0.091</td>
<td>0.000</td>
<td>1651</td>
</tr>
</tbody>
</table>

Table 4 presents abnormal returns and supplemental fit data. The confusion matrix column represents data available in a confusion matrix for the 95/05 data split. PP represents those predicted to have a positive change, PN represents those predicted to have a negative change, TP represent the true positives, TN represents the true negatives, FP represents false positives, and FN represents false negatives. BHAR represents the 12-month abnormal size adjusted returns, PValue represents the significance for the abnormal returns, and N is the number of observations.
Table 5 Holthausen and Larker (1992)-type forecasts
Prediction of the direction of returns directly - Model Accuracy 1976-2015

<table>
<thead>
<tr>
<th>Split</th>
<th>N</th>
<th>Stepwise Accuracy</th>
<th>Elastic Net Accuracy</th>
<th>Random Forest Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>50/50</td>
<td>75,650</td>
<td>0.530</td>
<td>0.527</td>
<td>0.528</td>
</tr>
<tr>
<td>60/40</td>
<td>60,530</td>
<td>0.537</td>
<td>0.532</td>
<td>0.534</td>
</tr>
<tr>
<td>70/30</td>
<td>45,394</td>
<td>0.543</td>
<td>0.539</td>
<td>0.540</td>
</tr>
<tr>
<td>80/20</td>
<td>30,268</td>
<td>0.548</td>
<td>0.544</td>
<td>0.547</td>
</tr>
<tr>
<td>90/10</td>
<td>15,134</td>
<td>0.557</td>
<td>0.552</td>
<td>0.558</td>
</tr>
<tr>
<td>95/05</td>
<td>7,566</td>
<td>0.561</td>
<td>0.561</td>
<td>0.561</td>
</tr>
</tbody>
</table>

Table 5 shows the percentile splits, corresponding sample size, and accuracy. The percentile splits are taken from ranking raw probabilities formed from each respective model. Accuracy represents how correctly each model classifies a positive change in earnings. The bold numbers represent the largest accuracies.
<table>
<thead>
<tr>
<th>Years</th>
<th>Stepwise Logit</th>
<th>Elastic Net</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976-1980</td>
<td>0.476</td>
<td>0.471</td>
<td>0.470</td>
</tr>
<tr>
<td>1981-1985</td>
<td>0.588</td>
<td>0.609</td>
<td>0.584</td>
</tr>
<tr>
<td>1986-1990</td>
<td>0.577</td>
<td>0.600</td>
<td>0.560</td>
</tr>
<tr>
<td>1991-1995</td>
<td>0.602</td>
<td>0.586</td>
<td>0.587</td>
</tr>
<tr>
<td>1996-2000</td>
<td>0.534</td>
<td>0.525</td>
<td>0.533</td>
</tr>
<tr>
<td>2001-2005</td>
<td>0.573</td>
<td>0.573</td>
<td>0.595</td>
</tr>
<tr>
<td>2006-2010</td>
<td>0.570</td>
<td>0.567</td>
<td>0.557</td>
</tr>
<tr>
<td>2011-2015</td>
<td>0.554</td>
<td>0.558</td>
<td>0.583</td>
</tr>
</tbody>
</table>

Table 6 presents analyses using 5-year groups along with the corresponding out-of-sample accuracy. The bolded numbers represent the largest accuracies.
Table 7 Holthausen and Larker (1992)-type forecasts
Predicting the direction of returns directly - abnormal returns calculations

<table>
<thead>
<tr>
<th>Panel A Stepwise Logit Regression</th>
<th>Panel B Elastic Net</th>
<th>Panel C Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confusion Matrix</td>
<td>BHAR</td>
<td>PValue</td>
</tr>
<tr>
<td>Hedge</td>
<td>0.069</td>
<td>0.001</td>
</tr>
<tr>
<td>PP</td>
<td>-0.011</td>
<td>0.289</td>
</tr>
<tr>
<td>PN</td>
<td>-0.08</td>
<td>0.000</td>
</tr>
<tr>
<td>TP</td>
<td>0.441</td>
<td>0.000</td>
</tr>
<tr>
<td>TN</td>
<td>-0.496</td>
<td>0.000</td>
</tr>
<tr>
<td>FP</td>
<td>-0.352</td>
<td>0.000</td>
</tr>
<tr>
<td>FN</td>
<td>0.859</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 7 presents abnormal returns and supplemental fit data. The confusion matrix column represents data available in a confusion matrix for the 95/05 data split. PP represents those predicted to have a positive change, PN represents those predicted to have a negative change, TP represent the true positives, TN represents the true negatives, FP represents false positives, and FN represents false negatives. BHAR represents the 12-month abnormal size adjusted returns, PValue represents the significance for the abnormal returns, and N is the number of observations.