Machine Learning, Earnings Forecasting, and Implied Cost of Capital - US and International Evidence

Akash Chattopadhyay
University of Toronto

Bingxu Fang
Sinagpore Management University

Partha Mohanram*
University of Toronto

March 2022

Preliminary Draft

(Please do not cite or distribute without permission)

Abstract

Keywords: Machine Learning; Forecasting; Valuation; Implied Cost of Capital

JEL: G18, G34, G38, G41, L51, M14, M52

*First draft: March 2022. Chattopadhyay (akash.chattopadhyay@rotman.utoronto.ca) is an Assistant Professor of Accounting at the University of Toronto. Fang (bxfang@smu.edu.sg) is an Assistant Professor of Accounting at the Singapore Management University. Mohanram (partha.mohanram@rotman.utoronto.ca) is the John Watson Chair in Value Investing and Professor of Accounting at the University of Toronto. Comments are welcome.

1. Introduction

Forecasts of future earnings are an integral input for research in valuation in finance and accounting. Researchers have used these forecasts in a variety of settings. For instance, Frankel and Lee (1998) generates estimates of intrinsic value to identify mispriced stocks. Forecasts are also the building blocks in estimating implied cost of capital (ICC), a metric of risk and expected returns that finance and accounting researchers have used to answer several important questions.

Until recently, the only forecast source was from sell-side analysts, usually obtained from data sources such as I/B/E/S or First call. There were two fundamental problems with analyst forecasts. One - analysts typically follow only a subset of the universe of firms, usually the larger firms with high levels of institutional investment. Two, prior research has shown that analyst forecasts are often optimistically biased and inaccurate. Previous research has shown the poor quality of analyst estimates to be responsible for the relatively weak performance of the ICC metrics in terms of their correlation with future realized returns (e.g., Easton and Monahan, 2005; Mohanram and Gode, 2013).

The past decade has seen the emergence of cross-sectional regression-based models that provide forecasts of future earnings. The most common models are the HVZ model from Hou, van Dijk, and Zhang (2012) and the earnings persistence (EP) and residual income (RI) models from Li and Mohanram (2014). They address two problems – lack of coverage, especially for smaller firms in weaker information environments, and the bias/inaccuracy of analyst forecasts. They, however, do have the problem of being inaccurate (though generally unbiased), with some models even underperforming a naïve Random Walk model (Gerakos and Gramacy, 2013).

The last few years have seen the emergence of models using machine learning (ML). ML models are potentially attractive for a few reasons. First, ML models can accommodate a large number of predictors, thus allowing researchers to use a broader information set than simple cross-sectional models. Secondly, several ML models also allow for non-linear and non-parametric relationships between the underlying data and the outcome variable (future earnings). In this paper, we evaluate two classes of ML models — penalized linear models (Lasso and Ridge) and decision-tree based models (Random Forest and Gradient Boosting Regression) — in their earnings forecasting efficacy and the predictive ability of implied cost of capital (ICC) estimates based on those forecasts.

We begin by testing various ML models' performance compared to the extant cross-sectional models in a sample of US firms. The ML models generally perform better than the cross-sectional models and significantly better than the naïve random walk model. Among the cross-sectional models, the EP and RI models from Li and Mohanram (2014) generally perform the best. Among the ML models, we find that the Gradient Boosting Regression (GBR) and Random Forest (RF) models perform the best, not just among ML models but among all models. However, the improvement of the best ML model over the best cross-sectional model is relatively modest, with a 3-6% reduction in mean absolute forecast errors (MAFE).

ML models are computationally intensive relative to the parsimonious traditional forecasting model (HVZ, EP, and RI), which raises the question of whether the modest improvement in MAFE is worth the additional effort. To better understand the relative performance of these models, we partition the sample based on several characteristics related to the information environment — firm size, analyst following, and earnings volatility. For firms with volatile earnings or small firms in the early stages of their life-cycle, it is plausible that the parsimonious linear cross-sectional forecasting models might not effectively mimic the underlying earnings-generating process. ML models accommodating a more comprehensive set of predictors and non-linearities in the data might be better suited to forecast the earnings of such firms. Consistent with our intuition, we find that the ML models perform significantly better in firms with relatively weaker information environments. The improvement in forecast accuracy is over 7% percent for small firms and firms with high earnings volatility. Interestingly, we find that ML models add to forecasting accuracy both for firms with and without analyst following

Having demonstrated the efficacy of the ML models in the US, especially for firms with arguably "hard-to-forecast" earnings, the natural question to ask is how these models would perform in an international setting. Ex-ante, one can think of why such models might perform better or worse. ML models might perform better because the information environment tends to be weaker in international settings because of poorer accounting standards, weaker enforcement, and weaker governance. Moreover, given the variation in accounting standards and economic forces across countries, it is likely that a simple linear model which works in the US might not be adequate outside the US. However, ML models with their extensive data and estimation requirements face specific challenges outside the US. Estimating ML models is likely to require pooling of data across countries which can differ substantially in their countries and institutional characteristics. Moreover, there is likely to be more variation in data quality outside the US. Consequently, ML models with a large number of predictors are likely to be more affected by noise in the data. Therefore, the relative efficacy of the ML models outside the US is an empirical question.

We examine this question using an extensive global sample from 61 countries. The extant cross-sectional models perform poorly in an international context, producing estimates with more error than the naive random walk model. The ML models, in general, outperform the extant models in a global setting. Consistent with non-parameterization and non-linearities being more important for the diverse international sample, we find that the GBR model performs the best outside the US. The improvement in forecasting accuracy for the GBR model is substantial, often more than 10%. The improvements from the GBR model in the international sample hold for firms with strong or weak information environments.

In addition to forecast accuracy, researchers might also care about the average level of bias in the generated forecasts. This is especially the case if one is interested in questions that look at aggregate information, e.g. what is the average market premium and how is it changing across time? Additional tests indicate that the best ML models (GBR and RF) perform as well as, if not better than, the cross-sectional models in terms of providing unbiased forecasts, both in the US and in the international samples.

Given that forecasts are vital inputs in valuation models, we test the efficacy of the various forecasting models by examining the predictive ability of implied cost of capital (ICC) metrics that can be calculated using the forecasts. We find that the ICCs estimated from the GBR model perform the best in terms of correlation with realized future returns in both the US and international samples.

This paper makes the following important contributions to the literature in finance and accounting. First, ML models (specifically the GBR model) generate substantially better forecasts of earnings, both in terms of lower forecast error and less bias. Second, the GBR model performs particularly well in the international context, where the extant cross-sectional models perform poorly. Third, using forecasts from the GBR model generates ICCs that perform the best as measures of expected returns, both in the US as well as internationally. We recommend that future research that requires earnings forecasts, whether they are used as measures of expected earnings or they are used to compute ICCs, use the GBR model both for the US and for international settings.

The rest of the paper is organized as follows. Section 2 positions our paper in the related literature on forecasting, the use of machine learning models, implied cost of capital, and differences between the US and international context. Section 3 outlines our research methodology — model description, estimation, and validation. Section 4 presents the results of our estimation and compares the performance of the different models in the US as well as international settings. Section 5 concludes.

2. Relation to Literature

Forecasting profitability constitutes a rich literature in accounting. This literature has often intersected productively with another body of methodological work in accounting — estimating expected returns using the so-called Implied Cost of Capital (ICC). However, researchers in this area are only beginning to explore innovations in machine learning for forecasting. Further, while a large body of work uses outputs of forecasting models and ICCs in research in an international context, methodological work remains scarce outside the US context.

2.1. Cross-sectional Models in Forecasting

Researchers in the area of valuation have long relied on forecasts of future earnings as crucial inputs into their analysis. The forecasts have been used in a variety of contexts - e.g. Frankel and Lee (1998) who used the forecasts to come up with measures of intrinsic value (V) and estimate a Value to Price or V/P ratio to identify undervalued and overvalued stocks, or the entire literature on implied cost of capital (e.g., Gebhardt, Lee, and Swaminathan, 2001; Gode and Mohanram, 2003; Easton, 2004; Botosan and Plumlee, 2002). For a long time, the only source of forecasts was analyst forecasts from sources such as I/B/E/S or First Call.

Using analyst forecasts presents researchers with two major problems. First, analyst forecasts are generally optimistic and not very accurate. Easton and Monahan (2005), Easton and Sommers (2007) , and Mohanram and Gode (2013) are among the papers that show that inaccurate and optimistic forecasts are the leading reasons for why ICC models do not perform well in predicting future returns. Second, analyst forecasts are not available for all firms, as analysts tend to follow larger firms in strong information environments. This skewed lack of coverage essentially means that researchers are unable to answer interesting questions on topics such as information quality and disclosure in the subset of firms where the answers to such questions would be particularly insightful. Researchers have also tried to use time-series models to come up with forecasts. However, the lengthy firm-specific time-series of data required to estimate such models renders them ineffective in the subset of younger firms that do not have a lengthy history.

Cross-sectional forecasting, a technique that has emerged in the last decade, attempts to address both these shortcomings. First, it uses the cross-section of data without imposing any firm-specific data limitation - i.e. a firm does not need to have existed in the entire estimation period. This essentially allows one to estimate the forecasts for nearly the entire universe of firms. Secondly, the models are not subject to the behavioral biases that plague analyst forecasts and generally produce unbiased forecasts.

The paper that pioneered the of cross-sectional forecasts is Hou et al. (2012), who build on models in Fama and French (2000, 2006) and regress future earnings on total assets, dividends, earnings and accruals. However, Gerakos and Gramacy (2013) show that in terms of forecast accuracy, the HVZ model underperforms a naïve random walk model that simply sets future earnings to equal past earnings. Further, the forecast errors of the HVZ model is rather high – the mean absolute error (scaled by price) for one-year-ahead earnings is 0.084 for firms with analyst coverage (Table 3 of Hou et al. (2012), page 9). If one assumes an average P/E ratio of 12, this represents an absolute error that is on average equal to the estimate of earnings itself. More importantly, the HVZ model generates larger forecast errors for firms without analyst coverage where the need for a forecasting model is crucial.

Li and Mohanram (2014) attempt to improve on the HVZ model with a model that is motivated by the literature in accounting (e.g., Dechow, 1994) which has generally shown that accrual based measures like earnings show greater persistence and predictability than the cash flow based measures that the HVZ model relies on. They present two alternative models - the earnings persistence model (EP) and the residual income model (RI) - and show that both models outperform the HVZ model as well as the random walk model. Further, the EP and RI models perform particularly well in the subset of small firms and firms without analyst following, where the utility of these models is the most salient. However, in absolute terms, the average level of forecast error reported in Li and Mohanram (2014) is still rather high. For instance, the mean absolute forecast error for one-year ahead earnings for the EP and RI models is 0.073 - significantly better than the HVZ and RW models, but still high. In this paper, we will attempt to see if the use of machine learning based approaches can generate forecasts with significantly greater accuracy.

2.2. Application of Cross-sectional Forecasts: Implied Cost of Capital

Cost of equity plays a central role in valuation, portfolio selection, and capital budgeting. Therefore, measuring and validating cost of equity metrics has been the subject of much research. Inferring cost of equity ex-post from realized returns is problematic because the correlation between expected returns and realized returns is weak (Elton, 1999). Prior research has often documented a weak or even non-existent relation between conventional measures of risk (e.g., beta and realized returns (Fama and French, 1992). This has led to the use of implied cost of capital (ICC), which is the discount rate that equates current stock price to the present value of expected future dividends.

Prior literature has taken different approaches towards measuring ICC. Gebhardt et al. (2001) and Claus and Thomas (2001) use variants of the residual income model to solve for the discount rate that equates price to the sum of book value and the present value of future abnormal earnings. Gode and Mohanram (2003) and Easton (2004) develop proxies based on the abnormal earnings growth model of Ohlson and Juettner-Nauroth (2005).

ICCs are widely used by researchers in accounting and finance in a variety of contexts. Most commonly, researchers have used ICC to see if a firm's information environment has changed in response to changes in voluntary disclosure (Dhaliwal, Li, Tsang, and Yang, 2011), accounting standards (Daske, Hail, Leuz, and Verdi, 2008; Li, 2010), securities regulation (Hail and Leuz, 2006), tax laws (Dhaliwal, Krull, Li, and Moser, 2005), etc.. Typically, researchers use the average of the four commonly used ICC metrics mentioned above as their measure of ICC. Mohanram and Gode (2013) validate this approach by showing that such a composite ICC metric has lower measurement error than any of the individual ICC metrics.

Earnings forecasts are a crucial input in the estimation of ICC. Usually, the models require both a short term forecast (i.e. forecast of one-year-ahead EPS) as well as long-term forecasts (either forecasts over longer horizons or forecasts of long-term growth). Most of the early papers relied on earnings forecasts to generate ICCs, despite the problems of limited coverage and bias in the forecasts. Easton and Monahan (2005) show that ICCs perform rather poorly in terms of predicting future returns, while Easton and Sommers (2007)) demonstrate that most ICC models generate measures that are biased upwards because analyst forecasts are optimistic. Mohanram and Gode (2013) show that when analyst forecasts are adjusted for predictable error and bias, the ICCs generated from these forecasts perform better.

With the emergence of cross-sectional forecasts, most researchers use these models, either the HVZ model or the EP or RI model from Li and Mohanram (2014) to estimate ICCs. The ICC paradigm is the most direct application of forecasting. In addition, these papers use ICC as a tool to validate the forecasts - i.e. for the forecasts to be any good, the ICCs that one can generate from these forecasts should also perform better. In this paper as well, we will analyze the performance of the ICCs that are generated from the ML based forecasting models and test whether they outperform the ICCs generated from the extant models.

2.3. Emergence of Machine Learning in Forecasting

Machine Learning (ML) as a tool to solve the two canonical prediction problems in accounting and finance research — predicting profitability and returns — is receiving considerable attention because of the ability of ML models to handle correlated, high-dimensional data and unspecified non-linearities within the data.

This growing literature has primarily focused on predicting asset returns (e.g., Freyberger, Neuhierl, and Weber, 2020; Gu, Kelly, and Xiu, 2020). A key motivation behind this work has been to solve the dimensionality problem created by the proliferation of return-predictive signals. This literature suggests that machine learning models successfully identify the return predictors with independent information and generate significant improvement over existing models in terms of the quality of predictions. Moreover, this literature suggests that the advantage of machine learning models is realized in methods that allow for nonlinear predictor interactions.

The literature focusing on the utility of ML in forecasting the other key input of equity valuation — profitability — is relatively nascent (e.g., Cao and You, 2020; van Binsbergen, Han, and Lopez-Lira, 2020; De Silva and Thesmar, 2021). While the overarching goal of this literature has been to evaluate the efficacy of ML models in forecasting future firm profitability, the key focus of the individual papers differs. van Binsbergen et al. (2020) and De Silva and Thesmar (2021) focus on using ML to create an optimal earnings benchmark to precisely identify expectation errors by analysts. In work related to ours, Cao and You (2020) focus on the entire cross-section of US firms to identify the optimal forecast of future profitability. Consistent with the findings from the literature using ML to predict returns, ML models incorporating non-linearities yield the best results in forecasting profitability. Cao and You (2020) find that the nonlinear ML models (Random Forest, Gradient Boosting Regressions, and Artificial Neural Networks) yield predictions that outperform those from a naive benchmark from the RW model while those from the extant linear models (HVZ, RI, and EP) do not. To our knowledge, we are the first to examine the efficacy of these models outside the US.

2.4. US vs International Evidence on Forecasting and ICCs

The body of methodological work described above almost exclusively focuses on the US context. There is little evidence on the efficacy of the various linear cross-sectional models in forecasting profitability and serving as inputs to ICC models outside the US. The performance of the extant models outside the US is not obvious ex-ante. First, at a general level, the US and other global markets vary on economic, political, legal, and institutional dimensions (Ball, 2016) which should motivate examination of economic models outside the US. Secondly, empirical evidence suggests that institutional and economic differences across countries also affect the earnings process, the critical outcome of interest in this literature (Healy, Serafeim, Srinivasan, and Yu, 2014). Consequently, it is plausible that findings based on linear parameterized earnings forecasting models estimated on US data might not extend internationally.

The question of identifying earnings forecasting models that perform robustly internationally is an important one because a burgeoning body of work in accounting and finance uses ICCs to study the influence of various policies on firms' cost of capital in an international context (Chattopadhyay, Lyle, and Wang, 2022). As Chattopadhyay et al. (2022) document, the performance of ICCs using forecasts from the extant cross-sectional models is relatively inconsistent outside the US, while the availability of analyst forecasts is sparse. Moreover, as Fang, Hope, Huang, and Moldovan (2020) document, the availability of analyst forecasts in Europe is declining following the enactment of MiFID II.

We fill this gap and build on the literature described above by examining the efficacy of the ML models outside the US context, as well as the validity of the ICC estimates generated using the outputs of these models. The latter analysis differs from the literature on using ML to predict returns in one key way. Instead of fitting a wide set of accounting characteristics to a noisy outcome variable (returns), we use ML to generate optimal forecasts of a less noisy accounting variable (earnings). We map these earnings to returns by using the theoretically motivated present-value approach to estimate ICCs.

3. Research Methodology

In this section, we briefly discuss the conceptual underpinnings of the various forecasting models and ICC measures we consider and detail their estimation processes. We also discuss our empirical framework for evaluating the forecasting models and the ICCs.

3.1. Forecasting Profitability

We evaluate four candidate machine learning models — two from the class of linear penalized models (Lasso and Ridge) and two tree-based models incorporating non-linearities (Random Forest and Gradient Boosting Regression) — against the extant models described in Section *2.1.*.

3.1..1 Traditional Models

The three traditional cross-sectional forecasting models we consider are the HVZ model developed by Hou et al. (2012) and the RI and EP models are based on the work by Li and Mohanram (2014). We use a naive random-walk model (RW) prediction as a benchmark. All three cross-sectional models produce earnings forecasts by estimating the following general model:

$$\mathbb{E}[E_{i,t+\tau}] \quad = \beta_0 + \beta_1 X_{i,t} + \epsilon_{i,t} \tag{1}$$

where $\mathbb{E}[E_{i,t+\tau}]$ represents expectation of earnings τ periods away and X_i represents firm-level characteristics measured at time t. The Appendix describes each model in terms of the firm characteristics involved. One difference between the HVZ model and the RI and EP models is that HVZ estimates earnings while RI and EP estimate earnings per share. We follow the extant literature in estimating the HVZ, EP,and RI models (Hou et al., 2012; Li and Mohanram, 2014). Specifically, for each year t in our sample, we use the previous 10 years' of observations ($t-1$ to $t-10$) as the training sample to estimate the parameters of each model, and then we use the parameters and the financial information in year t to generate the forecasts for year $t+1$ to year $t+3$.[1] When estimating the models for the international setting, we use a pooled sample with observations across all countries in our sample to increase the size of the training sample. [2]

3.1..2 ML Models: Background and Estimation

We briefly discuss the machine learning models we evaluate and detail our estimation process. Readers should refer to Hastie, Tibshirani, and Friedman (2009) for a significantly more technical description of these models.

The first class of models we evaluate is the so-called penalized class of models. The essential advantage of these models over linear regression is their lesser susceptibility to overfitting as the

[1]To mitigate look-ahead bias, we follow Li and Mohanram (2014) and assume that firms with fiscal year ending in April to June do not have their financial information available by end of June. We only include firms with fiscal year ending in April of year t-1 to March of year t when estimating the models for year t.

[2]There is a trade-off in the international setting between the size and the relevance of the training sample. Using a pooled sample across countries increases the size of the training sample, while countries with heterogeneous business environments are designed to have the same model parameters. On the other hand, estimating the models by country will allow each country to have different parameters, but it decreases the training sample sizes. In untabulated analyses, we estimate the models by country in the international setting and our inferences regarding the superior performance of machine learning models remain unchanged. However, the performance of each model generally becomes worse because of the smaller training samples.

number of parameters to be estimated increases. OLS estimates parameters to minimize a standard least squares objective function :

$$\beta^{OLS} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{2}$$

where we are estimating a linear model with p predictors and N observations. With an increase in the number of parameters to be estimated, OLS is prone to overfitting the model in-sample, leading to poor predictive performance out-of-sample. Moreover, coefficient estimates in a linear regression can be poorly determined and harder to interpret with many correlated predictors. Penalized models, also referred to as *Shrinkage* methods, are constrained to place the greatest weight on the subset of predictors with the highest predictive content. Penalized models thus allow for bias in the parameter estimates to minimize expected prediction error. By shrinking coefficients, penalized models also avoid the problem of overfitting for high-dimensional models. We examine two popular candidates from this class of estimators, Ridge Regression and Lasso. Ridge Regression minimizes a penalized sum of squares:

$$\beta^{Ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\} \tag{3}$$

where λ is the shrinkage parameter which scales coefficient values lower. Lasso minimizes the following penalized sum of squares:

$$\beta^{Lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{4}$$

where λ is the corresponding shrinkage parameter for Lasso. We can describe the key difference between Ridge and Lasso in a simplified manner as saying that Ridge shrinks all coefficients proportionally while Lasso shifts coefficients by a constant, truncating at zero. Thus, in Ridge, all predictors will technically get a non-zero coefficient, while Lasso will altogether discard some predictors. Moreover, while the Ridge estimate has a closed-form solution, Lasso requires numerical estimation.

We estimate all machine learning models with 56 predictors in total, including 28 financial statement line items obtained from Compustat and their changes relative to the previous year.[3] Similar to the estimation process for the traditional models, we use previous 10 years' of data when estimating machine learning models and we use the pooled sample across all countries when estimating the models in the international setting to increase the size of training sample. To reduce the subjectivity of researchers and search for the optimal choice of hyperparameters of machine learning models, we use a five-fold cross-validation process to search among a group of candidate values and identify the optimal hyperparameter values for each year (Hastie et al. (2009)).[4]

Finally, we consider two non-parametric decision-tree-based models — Random Forest (RF) and Gradient Boosting Regression (GBR). Both RF and GBR are ensemble learning models, i.e., they are a collection of individual models. The fundamental model in each ensemble is a regression tree. A regression tree partitions the data into a set of regions where the predicted value in each region is a constant that minimizes a squared-error loss function. So, for a dataset partitioned into M regions R_1, R_2, ..., R_m, a decision tree can be represented as:

$$f(x) = \sum_{m=1}^{M} c_m \mathbb{1}(x \in R_m) \tag{5}$$

Since the loss function is a sum of squares, the predicted value in each region is simply the average of the outcome variable for that region. A decision tree is modeled using a greedy algorithm

[3]Detailed definitions of the predictors are available in the Appendix. The number of predictors become 54 for the international sample because XAD is unavailable in Compustat Global.

[4]The K-fold cross-validation process is a popular and useful way to reduce potential over-fitting of the estimated models. The process starts with randomly dividing the training sample into K groups without replacement and then uses one of the groups as testing data and the other K-1 groups as training data. This process is repeated K times so that each group will be used as test data once. The average performance of a model will be calculated after using each group as test data. We use the cross-validation process to determine the shrinkage parameter for the Lasso and Ridge models and candidate values for the parameter range from 0.0001 to 0.1. with a step of a thousandth of the interval.

which starts with the entire data and splits it using a predictor j and a split point s into regions R_1 and R_2 that solves:

$$\min_{j,s} \left[\sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \tag{6}$$

For any j and s, the minimization is solved by the average of the outcome in each region. Thus, a splitting point s can be found for each predictor. As the first split, the algorithm chooses the predictor which produces the lowest value of the sum of squared errors as described in Equation 6. The algorithm repeats the process for each subsequent resulting region. Since a few iterations of this process are likely to result in a very complex tree, one of the critical parameters for designing a regression tree is the depth of the tree. The tree depth is usually chosen by pruning a large tree. The pruned tree is selected by minimizing a loss function which includes a penalty for the number of terminal nodes in the smaller tree. The key advantage of regression trees is their conceptual simplicity and flexibility to accommodate non-linearities and interactions in the data. However, the flexibility of an individual tree is undermined by its high variance. A small change in the data can lead to a tree of a very different structure. Therefore, an ensemble approach is preferred to produce more robust models.

RF and GBR are two popular ensemble approaches to arrive at more robust models based on regression trees. The RF approach uses bootstrapped data samples to create multiple regression trees. The final output for the random forest is the average of the outputs of the individual trees. Thus the critical parameters of a random forest are the number of trees and the depth of each tree. On the other hand, GBR uses an iterative approach instead of aggregating independent trees. GBR starts from a simple tree that produces errors that are not significantly lower than using the sample average as the predictor. GBR fits a regression tree to the residuals from the first tree in the next iteration. It keeps repeating this process for a set number of iterations, with the final output being an addition of the individual regression trees.

The estimation process of the RF and GBR models are similar to that for the Lasso and Ridge models. We use the same set of 56 predictors and we use previous 10 years' of data to the estimate the models for each year. In addition, we use the five-fold cross-validation process to search for the optimal choice of hyperparameters of the two models for each year.[5].

3.2. Proxying for the Information Environment

While it is necessary and important to focus on improving earnings forecasting accuracy for the entire population of firms, this task becomes particularly important when firms are in weak information environments. This is because forecasts from analysts are often of poor quality or even non-existent. This means that model-based forecasts are even more important.

We consider three proxies for information environment. Prior research has long used firm size as a proxy for the quality of the information environment (e.g., Brown, Richardson, and Schwager, 1987; Wiedman, 1996) because small firms have poorer disclosure quality, weaker auditing, less coverage in media, less institutional investment, and poorer analyst following. Our first proxy is hence firm size, measured by market capitalization.

Second, we consider whether a firm has analyst following or not. Prior research would simply delete such firms in their analysis till the emergence of cross-sectional forecasts. Cross-sectional forecasting allows researchers to generate estimates of future earnings. However, given the lack of other alternatives, forecast accuracy is particularly important for this subgroup.

Our third proxy is the volatility of earnings. Volatile earnings make forecasting more difficult. Prior research (e.g., Bhushan, 1989; O'Brien and Bhushan, 1990) shows that analysts often shy away from providing forecasts for firms with volatile earnings. In addition, the volatility also makes cross-sectional forecasting potentially less effective, as extrapolating from the past to the

[5]For the RF model, we set the number of trees in the forest to be 500. The candidate values for the depth of the tree range from 20 to 35 with a step of 5, and the candidate values for the minimum number of samples required to be at a leaf node ranges from 15 to 50 with a step of 5. For the GBR model, we use the huber loss function and set the number of trees to 500.The candidate values for the depth of the tree are 1, 3, and 5, and the candidate values for the minimum number of samples required to be at a leaf node ranges from 75 to 150 with a step of 25. The choices of the parameters are following those in Cao and You (2020)

present is more difficult. We measure earnings volatility as the standard deviation of the firm's quarterly return on assets (ROA) in the previous 5 years.

3.3. Estimating Implied Cost of Capital (ICC)

We evaluate the various earnings forecasting models by also examining the predictive content of ICCs calculated using their ouputs. We compute four ICC variants commonly used in the literature. Two are based on the residual income valuation model from Ohlson (1995) and Feltham and Ohlson (1995) - the GLS model from Gebhardt et al. (2001) and the CT model from Claus and Thomas (2001). Two are based on the abnormal earnings growth model from Ohlson and Juettner-Nauroth (2005) - the OJ model from Gode and Mohanram (2003) and a simplifed PEG model from Easton (2004). We estimate annual ICCs at the end of June of each year and winsorize each ICC estimate at the 1% and 99% levels for each cross-section. To further mitigate the effect of outliers, following Mohanram and Gode (2013), we calculate a composite ICC as the average of the ICCs from the four approaches mentioned above. If one or more of the four individual ICCs are unavailable, we follow Hou et al. (2012) and compute the composite ICC as the average of those available. Finally, we convert the composite annual ICCs into monthly ICCs for our regression-based validation tests described in the following section. Our monthly composite ICC estimate is therefore given by:

$$ICC_{monthly} = (1 + ICC_{annual})^{(1/12)} - 1 \tag{7}$$

3.4. Validating Forecasts and ICCs

Following prior research (Li and Mohanram, 2014), we evaluate the earnings forecasting models by examining the mean absolute forecast error (MAFE) of the output of each model. More specifically, we compute the cross-sectional average of the MAFE from each model and then examine the time series of these averages. We perform this exercise for forecasts up to three years ahead for both the US and our international sample. We also compare the relative efficacy of the ML models over the traditional models by examining the difference in the cross-sectional averages of various models.

We also follow prior research (e.g., Chattopadhyay et al., 2022; Li and Mohanram, 2014; Lewellen, 2015) in evaluating the ICCs computed using the outputs of the various forecasting models. We assess the ICCs by examining their association with future realized returns. We use both a regression-based approach and a non-parametric approach using portfolio sorts to validate that the cross-sectional differences in ICCs are directionally consistent with the differences in realized returns. For the regression-based method, we estimate Fama-Macbeth (FM) regressions of one-month-ahead realized returns on the various composite ICC estimates:

$$R_{i,t+1} = \beta_0 + \beta_1 ICC_{i,t+1} + \epsilon_{i,t+1} \tag{8}$$

A positive and significant β_1 would validate the predictive ability of an ICC estimate. Ideally, we would expect β_1 to be 1 for an accurate measure of expected returns. Consequently, we also evaluate ICCs on whether we can statistically distinguish β_1 from 1. As Chattopadhyay et al. (2022) discuss, this is a minimally sufficient criterion for evaluating a proxy of expected returns. We also examine the equal-weight predicted and realized returns of monthly decile portfolios based on each ICC measure.

3.5. Data Sources

We obtain annual financial information for US firms from Compustat and stock returns from CRSP, and we collect financial and stock prices information for the international sample from Compustat Global. Analyst coverage information is obtained from I/B/E/S. Our US sample covers the period from 1969 to 2017 and we require each firm to have common shares listed on the NYSE, AMEX, and NASDAQ, and the stock price at the end of June to be higher than 1 USD. We exclude financial and utilities firms from our sample. After further dropping observations with missing values for our predictors, we have 90,405 firm-year observations in the final sample. Our international sample is constructed with similar data requirements and we convert all variables to US dollars before applying the filters. We further require each country in the sample to have at least 100 observations, which leaves us with a final sample of 120,296 firm-year observations from 61 countries.

4. Empirical Findings

4.1. Model Performance in the US: Absolute Forecast Error

4.1..1 Overall Sample

We begin our analysis by analyzing the forecast errors in the sample of US firms from the following models. We consider the following cross-sectional models. HVZ is from Hou et al. (2012). RW is a naive random walk model, which is used as a benchmark to compare the performance of the cross-sectional models. EP and RI are the earnings persistence and residual income models respectively from Li and Mohanram (2014). We also consider the following four machine-learning based models, Lasso, Ridge, Gradient Boosting Regression (GBR) and Random Forest (RF). The metric we focus on is Mean Absolute Forecast Error (henceforth MAFE), calculated as the mean of the absolute value of the difference between the estimated earnings per share (estimated Net Income in dollars for HVZ) and the actual realized earnings per share (realized Net Income for HVZ), scaled by price per share (market capitalization for HVZ). We calculate MAFE for one-year-ahead, two-year-ahead and three-year-ahead horizons. The results are presented in Table 1.

The first set of columns present the results for one-year-ahead forecasts. Among the cross-sectional models, the EP and RI models have the lowest MAFE at 0.057. This is significantly better than 0.064 for RW and 0.078 for HVZ. Among the ML models, the GBR and RF models perform the best, with MAFE of 0.054. This is significantly better than the best of the cross-sectional forecasting models. The last four rows of the table present a pair wise comparison of the MAFE across models. For brevity, we only compare the two best ML models (GBR and RF) with the benchmark of the random walk model (RW) as well as the two best cross-sectional models (EP and RI). The results indicate that the ML models produce MAFE that is significantly lower by around 0.003-0.004, which represents an improvement of around 5-7% compared to the average error of the cross-sectional models.

The next two sets of columns repeat the analysis for the two-year-ahead and three-year-ahead forecast horizons. The results follow generally similar trends, with some noticeable differences. As expected, the MAFE increases for all models as the horizon becomes more distant. Among the cross-sectional models, the RI model is clearly the most accurate. Among the ML models, the Lasso and Ridge models actually perform worse than the RI model. Consistent with our earlier results, the GBR and RF models perform the best among all the ML models and indeed among all models. Both GBR and RF produce errors that are significantly lower than the best cross-sectional model (RI). The improvement continues to be around 0.003, which while significant, only represents an improvement of around 4%.

What is the source of the superiority of the two ML models over the more parsimonious models? Is it the superiority of the ML technique, or could be driven by the greater information being used as an input to these models? Stated otherwise, do we unintentionally hamstring the cross-sectional model by using a far more parsimonious set of predictors? To answer this question, we estimate a conventional regression based cross-sectional model using the entire set of variables used in the ML models. The results are not tabulated for brevity, but discussed below. We find that adding additional variables in such an augmented cross-sectional worsens rather than improving the performance of cross-sectional models. The one-year-ahead MAFE of such an augmented model is 0.058, or slightly worse than the 0.057 reported for both EP and RI. The augmented model's performance worsens for the two-year-ahead horizon (0.089 vs 0.081 for EP and 0.079 for RI) as well as the three-year-ahead horizon (0.136 vs 0.101 for EP and 0.089 for RI). This suggests that the improvement seen in the ML models is not driven by the additional explanatory variables, but rather by the ML models themselves.

The analysis of forecasting accuracy presented above suggests that the best of the ML models (GBR and RF) perform better than the best of the cross-sectional forecasting models (EP and RI). However, the extent of the improvement is modest, especially given that the ML models are both data and computationally intensive. What might make these models worthwhile to pursue is if they show superior performance in settings where forecasting is difficult. We examine this in the next sub-section.

4.1..2 Partitions based on the Information Environment

In the prior sub-section, the results indicated that the best of the ML models (GBR, RF) generated forecasts that were on average significantly more accurate than the best of the cross-sectional models (EP, RI). We next examine the performance of the cross-sectional models and the ML models across partitions based on the information environment to see where we see the most significant improvement. For researcher to deem the ML models to be worth the additional investment of data and computational intensity, we should ideally find that the improvements in forecast accuracy are greater in settings where forecasting is more difficult, i.e. firms in weaker information environment. We consider the following three partitions - firm size, analyst following and earnings volatility. The results are presented in Table 2.

Panel A presents the results partitioned by size (Market Capitalization). For each year, we split the sample into two equal groups (labelled Small Firms and Large Firms). The MAFE for each group is presented pooled across the entire time period. The first set of columns present the results for small firms. As the results indicate, we see a significant improvement in forecasting accuracy with the ML models. For a one-year horizon, GBF and RF both see an improvement of 0.005 over the errors of EP and RI at 0.078, which represents an improvement of over 6%. The improvement persists for longer horizons. For two-year-ahead forecasts as well, we see that MAFE reduces from 0.106 for RI to 0.100, and for GBR and RF, a reduction in error of 0.006 or almost 6%. Finally, for three-year-ahead forecasts, the MAFE reduces from 0.131 for RI to 0.125 for GBR and RF.

The next set of columns present the results for larger firms. At the outset, it is easy to see that the forecast error is much smaller for larger firms. For instance the MAFE for the best cross-sectional model EP is only 0.036 which is less than half the error for small firms (EP and RI at 0.078). Interestingly, the improvement with ML models, while statistically significant for one-year-ahead forecasts is minuscule with both GBR and RF having MAFE of 0.035. For longer horizons, there is literally no improvement. For two-year-ahead forecasts, the MAFE for the best cross-sectional model (EP and RI at 0.051) is identical to that for the best ML model (GBR and RF at 0.051). Similarly, for three-year-ahead forecasts, the MAFE for the best cross-sectional model (EP and RI at 0.064) is identical to that for the best ML model (GBR and RF at 0.064). Thus the results of the first partition suggest that the improved forecasting accuracy of ML models is concentrated in the crucial subgroup of small firms.

Panel B presents the results partitioned by analyst following. We consider two groups - firms with and without analyst following, which obviously do not have to be of the same size. The first set of columns present the results for the crucial subgroup of firms without analyst following. For this partition, there is no alternative of analyst forecasts, so researchers have to rely on model-based forecasts. Here, we see a significant improvement in forecasting accuracy with the ML models. For a one-year horizon, GBR and RF both see an improvement of 0.005 over the errors of EP and RI at 0.071, which represents an improvement of almost 7%. The improvement persists for longer horizons. For two-year-ahead forecasts as well, we see that MAFE reduces from 0.096 for RI to 0.091 for GBR and RF, a reduction of error of 0.005 or almost 6%. Finally, for three-year-ahead forecasts, the MAFE reduces from 0.120 for RI to 0.114 for GBR and RF, an improvement of 0.006.

The next set of columns present the results for firms with analyst following. At was the case for large firms, the forecast error is smaller for followed firms. For instance the MAFE for the best cross-sectional model (EP and RI) is only 0.048 as opposed to 0.071 for firms without following. Interestingly, we find that ML models perform better even in the subset with analyst following. For one-year-ahead forecasts, the best ML model (GBR) has a MAFE of 0.043, significantly better than the 0.048 for EP and RI. We find similar improvements for the two and three-year-ahead forecast horizons.

Most regression based earnings models extrapolate from current performance into the future, a task that becomes more challenging when earnings are volatile. Our final partition is that of earnings volatility, defined as the standard deviation of net income scaled by total assets over the past N years for a given firm. We partition our sample into two equal groups each year and present the results pooled across time in Panel C. The results mirror those for the size partition. We find that the ML models significantly outperform the cross-sectional models for the high volatility subsample. For one-year-ahead forecasts, the best cross-sectional model (RI) generates MAFE of 0.075, while the best ML model (GBR) generates MAFE of 0.068, an improvement of 0.007 or almost 10%. GBR continues to be the best performing ML model for two and three-year-ahead horizons as well, generating improvements of 0.007 and 0.005 respectively.

The next set of columns present the results for the low volatility subsample. Consistent with our results for large firms, we find either insignificant or marginally significant improvements for the best ML model (GBR or RI) as compared to the best cross-sectional model (EP or RI).

Overall, the results from the partition analysis validate the usefulness of machine learning models. The superior performance of the ML models is concentrated in subsamples of small firms and firms with the most volatile earnings, where the extant cross-sectional models tend to have high levels of forecast error. This suggests that the machine learning models are particularly useful in settings with weaker information environments.

4.2. Model Performance Internationally: Absolute Forecast Error

Our analysis thus far has only focused on US firms. Whether the findings that machine learning models generate significant improvements over extant cross-sectional models would translate to an international setting is unclear. Our partition analysis suggests that ML models are particularly useful in settings with weaker information environments. Prior research has generally concluded that informational environment is the strongest in the US. For instance, a vast literature on cross-listing suggests that the information environment of international firms improves after they cross-list in the US (e.g, Lang, Lins, and Miller, 2003; Bailey, Karolyi, and Salva, 2006). If indeed it is the case that international firms operate in weaker information environments, it is possible that the ML models might show significant improvements in forecasting ability as compared to the extant cross-sectional models. Conversely, the performance of cross-sectional models or machine learning models for that matter has not been examined in an international setting. Because of smaller sample sizes, researchers have to either run country-specific estimations that are noisier, or pool disparate observations from different countries into a single estimation. In this subsection, we examine the performance of the cross-sectional models and machine learning models using a sample of international firms. To ensure the maximum sample coverage, we pool observations across all countries (other than the US) in our estimation, both for the cross-sectional as well as the ML models.

4.2..1 Overall Sample

We begin by examining the mean absolute forecast errors for the models using the international sample. The results are presented in Table 3. The first set of columns present the MAFE for one-year-ahead forecasts. Consider the cross-sectional forecasts first. None of the three regression-based models (HVZ, EP or RI) perform as well as even the naive random walk model. The MAFE is 0.126 for HVZ, 0.098 for EP and 0.092 for RI all of which is worse than 0.078 for the RW model. Consider the ML models next. While GBR and RF perform better than the cross-sectional models, only the GBR model outperforms the RW model. The MAFE for GBR is 0.073, which is 0.005 lower than the RW model (an improvement of around 7%). The next set of columns presents the MAFE for two-year-ahead forecasts. Again, with the exception of the GBR model, none of the other models perform better than RW (MAFE=0.099). The GBR model performs the best with MAFE of 0.090 which is approximately 9% better than the RW model. We find similar results for three-year-ahead results, with the GBR model (MAFE=0.105) performing the best and delivering a significant improvement over all other models.[6]

The results in Table 3 highlight an important contribution of our paper. Cross-sectional forecasts, which tend to be error prone in the US context (see Gerakos and Gramacy (2013)), perform even worse in an international context. The EP and RI models from Li and Mohanram (2014) perform well in the US context, but these models also fare worse than RW internationally. Given this, the GBR model offers promise as a model to use in the international context.

4.2..2 Partitions based on the Information Environment

We next consider the performance of the models in the international sample partitioned on the basis of our proxies for information environment - size, analyst following and earnings volatility. The results are presented in Table 4.

Panel A presents the results partitioned by size (market capitalization). The first set of columns present the results for small firms. Consistent with the results for the overall sample, the cross-sectional models all perform poorly. The MAFE for the HVZ (0.213), EP (0.133) and RI (0.125) are all worse than the naive benchmark of RW (0.117). Among the ML models, the Lasso and Ridge models both perform poorly. The RF model performs marginally better than the RW model with a MAFE of 0.116. The GBR model, on the other hand, is the best performing overall with a MAFE of 0.106, which is almost 10% better than RW. For the two-year-ahead horizon, both the RF and GBR models outperform the RW model, with MAFE of 0.141 and 0.126 respectively as opposed to 0.147 for RW. Again, the best performing model is GBR, with an error reduction of 0.020 (around 14%). For the three-year-ahead horizon, again the best performing model is GBR with a MAFE of 0.146, which is substantially better than RW with a MAFE of 0.168. The error reduction with GBR is 0.022 (around 13%).

[6]As we do for the US sample, we examined an augmented cross-sectional regression model that uses all the explanatory variables used in the ML models. Such a model performs the worst of all models in all horizons, suggesting that the improved forecasting accuracy arises not from the additional variables alone, but from the ML modelling technique.

The next set of columns present the results for larger firms. While the average errors are substantially lower, none of the models perform as well as the naive RW model, which has the lowest MAFE for all three horizons (0.039, 0.052 and 0.062 for 1,2 and 3-year-ahead horizons respectively). Among the cross-sectional models, the HVZ model performs well with MAFE of 0.040, 0.052 and 0.065 for 1,2 and 3-year-ahead horizons respectively. Among the ML models, the only model to perform reasonably well is GBR with MAFE of 0.040, 0.054 and 0.065 for 1,2 and 3-year-ahead horizons respectively. Thus the results of the first partition using international data suggest that the improved forecasting accuracy of ML models is concentrated in the crucial subgroup of small firms. For large firms, while the errors are lower, no model beats the RW benchmark.

Panel B presents the results partitioned by analyst following. The first set of columns present the results for the important subgroup of firms without analyst following. Here too, we find that none of the cross-sectional forecasts are able to improve on the naive RW benchmark for all three horizons. Among the ML models, again it is the GBR model which alone performs strongly, outperforming all models across all horizons. For one-year-ahead forecasts, the MAFE of GBR is 0.089 as opposed to 0.094 for RW (improvement of 0.005 or around 6%). For two-year-ahead forecasts, the MAFE of GBR is 0.109 as opposed to 0.121 for RW (improvement of 0.012 or around 10%). Finally, for three-year-ahead forecasts, the MAFE of GBR is 0.130 as opposed to 0.142 for RW (improvement of 0.012 or over 8%).

The next set of columns present the results for firms with analyst following. At was the case for large firms, the forecast error is smaller for followed firms. Interestingly, we find similar patterns in the subset with analyst following. None of the cross-sectional models are able to outperform the naive model. Among the ML models, the GBR model performs the strongest, significantly improving on the RW model in all three horizons. For one-year-ahead forecasts, GBR has a MAFE of 0.064, significantly better than the 0.068 for RW. We find similar improvements for the two and three-year-ahead forecast horizons.

Panel C presents the results partitioned by earnings volatility. The results largely mirror that for the size partition. For firms with volatile earnings, only the GBR model delivers earnings forecasts with significantly lower error compared to the RW benchmark. All the other cross-sectional and ML models fare worse in all three horizons. For one-year-ahead forecasts, GBR generates MAFE of 0.081, which is around 10% better than the MAFE of 0.089 for the RW model. We see similar strong reduction in MAFE for GBR over the two-year-ahead (0.097 for GBR vs 0.111 for RW) and three-year-ahead (0.111 for GBR vs 0.127 for RW) forecast horizons.

The next set of columns present the results for the low volatility subsample. Again, while the errors are lower for all models are expected, the only model to perform as well as or better than the RW model is the GBR model. However, the improvements are statistically insignificant.

Overall, our partition analysis in the international setting also validates the usefulness of machine learning models in the key subsamples of small firms and firms with the most volatile earnings. There are however two major differences in the international context. One - all the extant cross-sectional models perform poorly and fail to beat the benchmark of the RW model, and perform particularly poorly in the crucial subsamples of small firms, firms without analyst following and firms with volatile earnings. Two - only the GBR model consistently produces earnings forecasts with the lowest absolute forecast error in almost every subsample.

4.3. Model Bias

Our tests thus far have focused on forecasting accuracy measured by unsigned forecast error (MAFE). We now turn our attention to forecast bias. Bias and forecast accuracy are fundamentally different concepts - bias is a measure of signed error, while forecast accuracy is a measure of unsigned error. It is possible for a model with higher MAFE to have lower bias, if the errors "cancel out". Conversely, a model can have a high degree of accuracy and yet be biased. Why might researchers care about bias? Bias may not be that important if the focus is on the firm-level. However, if the focus is on the aggregate, e.g. estimating the aggregate market premium as in Claus and Thomas (2001) or estimating aggregate implied cost of capital as in Li, Ng, and Swaminathan (2013), then it is important to have unbiased forecasts. Easton and Sommers (2007)) show that ICC estimates derived from analyst forecasts are systematically biased upwards because the forecasts used to generate them are optimistically biased.

In our next set of tests, we examine the bias of the forecasts generated by the cross-sectional as well as the ML models. We define bias as the difference between the predicted earnings forecasts minus actual earnings , scaled by either price per share (for models that forecast EPS) or market capitalization (for models that forecast unscaled total earnings). A positive bias indicates that the forecast is higher than the actual,i.e. optimistic forecasts. Conversely, a negative bias on the other hand indicates that forecasts are pessimistic. The results are presented in Table 5.

Panel A presents the results for the international sample. Unsurprisingly, the RW model performs poorly, especially as the horizon gets longer. The mean bias increases in magnitude from -0.018 for one-year-ahead to -0.034 for two-year-ahead to -0.049 for three-year-ahead forecasts. This is because the static RW model is inherently pessimistic, as it does not incorporate any growth into its forecasts. Among the cross-sectional forecasts, the HVZ shows a high level of optimistic bias - i.e. the actual earnings are considerably less than the forecasted earnings (0.030, 0.059 and 0.088 for 1,2 and 3-year-ahead forecasts respectively). The EP model also produces optimistically biased forecasts, but the bias is far less than that of the HVZ model (0.007, 0.010, 0.013). The cross-sectional model that performs the best is the RI model for which the mean bias across the three-years are insignificantly different from zero (0.002, 0.000, and 0.002 for 1,2 and 3-year-ahead forecasts respectively). Among the ML models, the Lasso and Ridge models produce unbiased forecasts across all horizons. GBR and RF produce slightly biased forecasts, but the level of bias is relatively low, especially for GBR. The mean bias for GBR is 0.001, -0.007, and -0.014, while the mean bias for RF is -0.004, -0.010 and -0.016 for 1,2 and 3-year-ahead forecasts respectively. What the bias results indicate is that researchers may have a trade-off to make with model selection in the US context. The models that produce the least biased forecasts are not the same as the models that produce the most accurate forecasts. Among the cross-sectional models, the RI model produces forecasts that are reasonably accurate (Tables 1 and 2) as well as unbiased. Among the ML models, the GBR model produces forecasts that are the most accurate (Tables 1 and 2) and have, what might be considered, an acceptable level of bias.

Panel B presents the results for the US sample. Here too, the static RW model performs poorly, especially as the horizon gets longer. The mean bias increases from -0.015 to -0.029 to -0.039 for 1,2 and 3-year-ahead forecasts respectively. Among the cross-sectional forecasts, the RI model dominates with unbiased forecasts for all three horizons. Among the ML models, the bias results are very different compared to the US sample. The Lasso and Ridge models produce extremely biased (optimistic) forecasts. The GBR model performs the best, with unbiased forecasts for all horizons. The RF model also performs reasonably well with unbiased forecasts for 1 and 2-year-ahead forecasts, but biased forecasts for 3-year-ahead forecasts. Combining these bias results with the earlier results for forecast accuracy produces a clear winner in the international context. The GBR model produces the most accurate forecasts (Tables 3 and 4) which are also unbiased. Among the cross-sectional models, the RI model produces unbiased forecasts, but the forecast error can often be higher than that of the naive RW model.

4.4. Performance of Model-Based ICCs

Our final set of tests uses the Implied cost of capital (ICC) paradigm to test the quality of the forecasts. As Mohanram and Gode (2013) show, the poor performance of ICC metrics can be largely attributed to the poor quality of the forecasts. When forecasts are accurate, the ICCs generated from them "perform well" - i.e. they are a reliable measure of expected returns. Li and Mohanram (2014) test their proposed cross-sectional models (EP and RI) and show that these models generate better ICCs than those from the HVZ model.

Using forecasts from each of the cross-sectional (HVZ, EP and RI) as well as ML models (Lasso, Ridge, GBR and RF), we generate measures of ICC, which is defined as the average of the ICC from the GLS,CT, PEG and OJ models. Note that the naive RW model cannot be used to compute

ICC, as it does not provide any estimate of earnings growth which is a prerequisite for two of the ICC models (PEG and OJ). We test the performance of the ICCs generated from the different forecasting models using two sets of tests. First, we run univariate regressions of realized returns on the measure of ICC and test how close the coefficient on the ICC is to the theoretical benchmark of "1". Second, we create portfolios based on the level of ICC and examine the pattern of returns to see if there is a monotonic relationship between ICC and future returns. For both sets of tests, we express ICC in terms of monthly returns and use monthly returns for the one-year-ahead horizon.[7]

4.4..1 US results

Table 6 presents the results for the sample of US firms. Panel A presents the results of the univariate regression of ICC on realized returns. Each of the seven models produces ICCs that are significantly positively correlated with future returns. Among the cross-sectional models, HVZ performs the worst with the smallest coefficient of 0.346, EP the next best with a coefficient of 0.497, while RI performs the best with a coefficient of 0.593. However, all these coefficients are significantly below the theoretical benchmark of 1. All the ML models have higher coefficients, 0.665 for Lasso, 0.704 for Ridge, 0.912 for GBR and 0.992 for RF. However, only the GBR and RF have coefficients that are statistically indistinguishable from the theoretical benchmark of 1.

Panel B of Table 6 presents the portfolio results for the US sample across deciles of ICC generated using each of the seven measures. In this table, a model can be deemed to perform well if we find a significant spread in returns across extreme ICC deciles, and if the spread in realized returns is comparable to the spread in the ICC. For all seven models, we find significant return spreads between the lowest and highest ICC deciles. The cross-sectional models have the lowest spread, with HVZ (0.68%), EP (0.68%) and RI (0.70%) all showing return spreads between the extreme deciles that are far less than the spread in ICCs. The Lasso and Ridge models perform better with

[7] We use monthly returns for this analysis because annual returns are more likely to be affected by delisting bias; delisting-adjusted returns are not available for the international sample. In untabulated analysis, we find that using unadjusted annual returns produces qualitatively and statistically similar results.

return spreads of 0.80% and 0.77% respectively. However, the models that perform the best by far are GBR (0.96%) and RF (0.89%). For these two models, the spread in returns is almost as large as the spread in ICC (1.10% for GBR, 1.06% for RF).

The results from Table 6 highlight the importance of forecast quality for the estimation of ICCs. It is not surprising that the models that perform the best, GBR and RF among the ML models and the RI model among the cross-sectional models, are also the models that produces the most accurate and least biased forecasts.

4.4..2 International results

Table 7 presents the results for the international sample. Panel A presents the results of the univariate regression of ICC on realized returns. Among the cross-sectional models, the HVZ model performs poorly and actually shows a statistically insignificant correlation with realized returns (coefficient = 0.074, t-stat = 1.01). The EP and RI perform moderately well with significant coefficients of 0.529 and 0.649 respectively. Among the ML models, the Lasso and Ridge models perform poorly, with coefficients of 0.204 and 0.217 respectively. The RF model performs better with a significant coefficient of 0.639. However, the model that stands out is the GBR model with a coefficient of 1.013 which is indistinguishable from the theoretical benchmark of 1.

Panel B of Table 7 presents the portfolio results for the international sample across deciles of ICC generated using each of the seven measures. These results also mirror the regression results. Among the cross-sectional model, the HVZ model performs poorly with return spread of 0.42%, far less than the ICC spread of 3.15%. The EP and RI models perform moderately with return spreads of 1.08% and 1.14% as compared to ICC spreads of 1.65% and 1.39%. Among the ML models, both the Lasso and Ridge models perform poorly, with return spreads under 1%, far less than the ICC spreads that are in excess of 2%. By far, the best performing model is GBR, with return spreads of 1.22% that is comparable to the ICC spread of 1.04%. The RF model also performs moderately with return spreads of 1.08% as compared to the ICC spread of 1.49%.

The results from Table 7 highlight an important contribution of this paper. One particular model, the GBR model, dominates all other models in its ability to generate the most accurate and least biased forecasts, in both US and international samples. Unsurprisingly, it also generates ICCs that perform the best.

5. Conclusion

In this paper, we test whether recently developed machine learning (ML) techniques can help researchers seeking to generate accurate and unbiased forecasts of future earnings, and whether these forecasts can lead to better estimates of implied cost of capital (ICC). We examine these questions, not just in US firms like most prior research, but also in an international sample. We consider four ML models - Lasso regression, Ridge regression, Gradient boosting regression (GBR) and Random Forest (RF). We benchmark the performance of these models against both a naive random walk (RW) model as well as extant cross-sectional models of forecasting, specifically the HVZ model from Hou et al. (2012) and the earnings persistence (EP) and residual income (RI) models from Li and Mohanram (2014).

Within the US sample, we find that the GBR and RF models both perform well, generating forecasts with the greatest ex-post accuracy. The EP and RI models also perform reasonably well, while the HVZ, Lasso and Ridge models perform poorly, often underperforming even the naive RW model. The improvements generated by the GBR and RF model, while not dramatic, are concentrated in the important subgroups of small firms and firms with volatile earnings. However, it is in the international sample where one ML model, the GBR model, really shines in its ability to generate forecasts with dramatically better forecasting accuracy. The results from ICC tests mirror the forecast accuracy tests - with the GBR model performing the strongest, especially for international firms.

The results of this paper have important methodological contributions for researchers in finance and accounting, striving to generate accurate earnings forecasts and reliable measures of expected risk. We recommend that future research use the GBR model to generate estimates of future earnings as well as ICC. This recommendation is particularly important in the subset of international earnings for two reasons. First, our results show that the cross-sectional models that perform moderately well in the US sample, do not fare as well internationally. Second, the problem of scarce coverage and volatile earnings is likely to more severe in international settings, and these are some of the subsamples in which the GBR model does extremely well.

We must mention that ours is only a first attempt at showing that ML models can add a lot of value in both forecasting space as well as the estimation of ICC. In fact, one can view our results as a lower bound of what ML models can do. We have used a simple and static (though reasonably exhaustive) set of potential explanatory variables in our estimation models. Using a wider set of variables, including non-financial variables as well as market-based signals, might also increase the accuracy of the forecasts and the performance of the ICCs from these forecasts. We leave this question for future research to examine.

References

Bailey, W., G. A. Karolyi, and C. Salva (2006). The economic consequences of increased disclosure: Evidence from international cross-listings. *Journal of Financial Economics 81*, 175–213.

Ball, R. (2016). Why we do international accounting research. *Journal of International Accounting Research 15*(2), 1–6.

Bhushan, R. (1989). Firms characteristics and analyst following. *Journal of Accounting and Economics 11*, 255–274.

Botosan, C. A. and M. A. Plumlee (2002). A Re-examination of Disclosure Level and the Expected Cost of Equity Capital. *Journal of Accounting Research 40*(1), 21–40.

Brown, L. D., G. D. Richardson, and S. J. Schwager (1987). An information interpretation of financial analyst superiority in forecasting earnings. *Journal of Accounting Research*, 49–67.

Cao, K. and H. You (2020). Fundamental analysis via machine learning. *HKUST Business School Research Paper* (2020-009).

Chattopadhyay, A., M. R. Lyle, and C. C. Wang (2022). Expected stock returns worldwide: A log-linear present-value approach. *The Accounting Review 97*(2), 107–133.

Claus, J. and J. Thomas (2001). Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets. *Journal of Finance 56*(5), 1629–1666.

Daske, H., L. Hail, C. Leuz, and R. Verdi (2008). Mandatory IFRS Reporting around the World: Early Evidence on the Economic Consequences. *Journal of Accounting Research 46*(5), 1085–1142.

De Silva, T. and D. Thesmar (2021). Noise in expectations: Evidence from analyst forecasts. Technical report, National Bureau of Economic Research.

Dechow, P. M. (1994). Accounting & Economics The role of accounting accruals. *Journal of Accounting and Economics 18*, 3–42.

Dhaliwal, D., L. Krull, O. Z. Li, and W. Moser (2005). Dividend Taxes and Implied Cost of Equity Capital. *Journal of Accounting Research 43*(5), 675–708.

Dhaliwal, D. S., O. Z. Li, A. Tsang, and Y. G. Yang (2011). Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting. *The Accounting Review 86*(1), 59–100.

Easton, P. and S. Monahan (2005). An evaluation of accounting-based measures of expected returns. *The Accounting Review 80*(2), 501–538.

Easton, P. D. (2004). PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital. *The Accounting Review 79*(1), 73–95.

Easton, P. D. and G. A. Sommers (2007). Effect of Analysts' Optimism on Estimates of the Expected Rate of Return Implied by Earnings Forecasts. *Journal of Accounting Research 45*(5), 983–1015.

Elton, E. J. (1999). Expected Return, Realized Return, and Asset Pricing Tests. *Journal of Finance 54*(4), 1199–1220.

Fama, E. F. and K. French (2000). Forecasting profitability and earnings. *The Journal of Business 73*, 161–175.

Fama, E. F. and K. R. French (1992). The Cross-Section of Expected Stock. *Journal of Finance 47*(2), 427–465.

Fama, E. F. and K. R. French (2006). Profitability, investment and average returns. *Journal of Financial Economics 82*(3), 491–518.

Fang, B., O.-K. Hope, Z. Huang, and R. Moldovan (2020). The effects of mifid ii on sell-side analysts, buy-side analysts, and firms. *Review of Accounting Studies 25*(3), 855–902.

Feltham, G. a. and J. a. Ohlson (1995). Valuation and Clean Surplus Accounting for Operating and Financial Activities. *Contemporary Accounting Research 11*(2), 689–731.

Frankel, R. M. and C. M. C. Lee (1998). Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics 25*, 283–319.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically.

Gebhardt, W. R., C. M. C. Lee, and B. Swaminathan (2001). Toward an Implied Cost of Capital. *Journal of Accounting Research 39*(1), 135–176.

Gerakos, J. and R. B. Gramacy (2013). Regression-based earnings forecasts.

Gode, D. and P. Mohanram (2003). Inferring the cost of capital using the Ohlson–Juettner Model. *Review of Accounting Studies 8*, 399–431.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Hail, L. and C. Leuz (2006). International Differences in the Cost of Equity Capital: Do Legal Institutions and Securities Regulation Matter? *Journal of Accounting Research 44*(3), 485–531.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Second ed.). Springer Science & Business Media.

Healy, P., G. Serafeim, S. Srinivasan, and G. Yu (2014). Market competition, earnings management, and persistence in accounting profitability around the world. *Review of Accounting Studies 19*, 1281–1308.

Hou, K., M. A. van Dijk, and Y. Zhang (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics 53*, 504–526.

Lang, M. H., K. V. Lins, and D. P. Miller (2003). ADRs, analysts, and accuracy: Does cross listing in the US improve a firm's information environment and increase market value. *Journal of Accounting Research 41*, 317–345.

Lewellen, J. (2015). The cross-section of expected stock returns. *Critical Finance Review 4*(1), 1–44.

Li, K. K. and P. Mohanram (2014). Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies 19*(3), 1152–1185.

Li, S. (2010). Does Mandatory Adoption of International Financial Reporting Standards in the European Union Reduce the Cost of Equity Capital ? *The Accounting Review 85*(2), 607–636.

Li, Y., D. T. Ng, and B. Swaminathan (2013). Predicting market returns using aggregate implied cost of capital. *Journal of Financial Economics 110*(2), 419–436.

Mohanram, P. and D. Gode (2013). Removing predictable analyst forecast errors to improve implied cost of equity estimates. *Review of Accounting Studies 18*, 443–478.

O'Brien, P. and R. Bhushan (1990). Analyst following and institutional ownership. *Journal of Accounting Research 28*, 55–82.

Ohlson, J. A. (1995). Earnings, Book Values and Dividends in Equity Valuation. *Contemporary Accounting Research 11*(2), 661–687.

Ohlson, J. A. and B. E. Juettner-Nauroth (2005). Expected EPS and EPS growth as determinants of value. *Review of Accounting Studies 10*, 349–365.

van Binsbergen, J. H., X. Han, and A. Lopez-Lira (2020). Man vs. machine learning: The term structure of earnings expectations and conditional biases. Technical report, National Bureau of Economic Research.

Wiedman, C. I. (1996). The relevance of characteristics of the information environment in the selection of a proxy for the market's expectations for earnings: An extension of brown, richardson, and schwager [1987]. *Journal of Accounting Research 34*(2), 313–324.

Appendix: Description of Models and Variables

This table describes the models used for generating earnings forecasts and computations of ICCS, and defines the variables used in earnings forecast models, including traditional models (e.g., HVZ, EP, and RI) and machine learning models. All variables are obtained from Compustat North America or Compustat Global; non-U.S. fundamental data are converted to U.S. dollars using Compustat's exchange-rate file.

Variable	Description	Computation
A1. Earnings Forecasting Models		
	Models Estimated	
HVZ	Generate earnings forecasts using the model of Hou et al. (2012)	$\mathbb{E}[E_{i,t+\tau}] = \beta_0 + \beta_1 A_{i,t} + \beta_2 D_{i,t} + \beta_3 DD_{i,t} + \beta_4 E_{i,t} + \beta_5 Neg\_E_{i,t} + \beta_6 ACCRUAL_{i,t} + \epsilon_{i,t}$
EP	Generate earnings forecasts using the model of Li and Mohanram (2014)	$\mathbb{E}[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t} + \beta_2 Neg\_E_{i,t} + \beta_3 Neg\_E_{i,t} \times E_{i,t} + \epsilon_{i,t}$
RI	Generate earnings forecasts using the model of Li and Mohanram (2014)	$\mathbb{E}[E_{i,t+\tau}] = \beta_0 + \beta_1 E_{i,t} + \beta_2 Neg\_E_{i,t} + \beta_3 Neg\_E_{i,t} \times E_{i,t} + \beta_4 B_{i,t} + \beta_5 TACC_{i,t} + \epsilon_{i,t}$
A2. Computation of ICCS		
	Valuation Models	
ICC_{GLS}	R_e computed using the model in Gebhardt et al. (2001)	$P_{i,t} = B_{i,t} + \sum_{\tau=1}^{11} \frac{\mathbb{E}_t[E_{i,t+\tau}] - (R_e - 1) \times \mathbb{E}_t[B_{i,t+\tau-1}]}{(R_e)^\tau} + \frac{\mathbb{E}_t[E_{i,t+12}] - (R_e - 1) \times \mathbb{E}_t[B_{i,t+11}]}{(R_e - 1)(R_e)^{11}}$
ICC_{CT}	R_e computed using the model in Claus and Thomas (2001)	$P_{i,t} = B_{i,t} + \sum_{\tau=1}^{3} \frac{\mathbb{E}_t[E_{i,t+\tau}] - (R_e - 1) \times \mathbb{E}_t[B_{i,t+\tau-1}]}{(R_e)^\tau} + \frac{\mathbb{E}_t[E_{i,t+3}] - (R_e - 1) \times \mathbb{E}_t[B_{i,t+2}]}{((R_e - 1) - g)(R_e)^3}(1+g)$
ICC_{PEG}	R_e computed using the "PEG" model in Easton (2004)	$R_e = 1 + \sqrt{\frac{E_{i,t+2} - E_{i,t+1}}{P_{i,t}}}$
ICC_{OJ}	R_e computed using the model in Ohlson and Juettner-Nauroth (2005)	$P_t = \frac{E_{t+1}}{(R_e - 1)} + \frac{E_{t+1}(E_{t+2} + (R_e - 1)D_{t+1} - (R_e)E_{t+1})}{(R_e - 1)(R_e - 1) - \frac{E_{t+3} + (R_e - 1)D_{t+2} - R_e I_{t+2}}{E_{t+2} + (R_e - 1)D_{t+1} - R_e E_{t+1}})}$
A3. Definitions of Variables in the HVZ Model		
$E_{i,t+\tau}$	Earnings in year t+τ	ib-spi
A_t	Total assets in year t	at

Variable	Description	Computation
D_t	Dividend payment in year t	dvc
DD_t	Dividend Payer Indicator	An indicator variable that equals 1 if dividend is higher than 0
Neg_E_t	Negative earnings indicator	An indicator variable that equals 1 for firms with negative earnings
$Accruals$	Accruals	Change in non-cash current assets (act - che) minus change in current liabilities excluding short-term debt and taxes payable (lct - dlc - txp) minus depreciation and amortization (dp)
A4. Definitions of Variables in *EP* and *RI* Models		
$E_{i,t+\tau}$	Earnings per share in year t+τ	((ib-spi)/csho)
Neg_E_t	Negative earnings indicator	An indicator variable that equals 1 for firms with negative earnings
B	Book value of equity per share	ceq/csho
$TACC$	Total accruals	Sum of the change in WC ((act - che) - (lct - dlc)), change in NCO ((at - act - ivao) - (lt - lct - dltt)), and change in FIN ((ivst + ivao) - (dltt + dlc + pstk))
A5. Definitions of Variables in Machine Learning Models		
$Sale$	Total sales	sale/csho
$COGS$	Cost of goods sold	cogs/csho
$XSGA$	Selling, general, and administrative expenses	xsga/csho
XAD	Advertising expense	xad/csho
XRD	Research and development expense	xrd/csho
DP	Depreciation and amortization	dp/csho
$XINT$	Interest and related expense	xint/csho
$NOPIO$	Non-operating income *expense*	nopio/csho
TXT	Income taxes	txt/csho
$XIDO$	Extraordinary items and discontinued operations	xido/csho
E	Earnings	(ib - spi)/csho
DVC	Common dividend	dvc/csho
CHE	Cash and short-term investments	che/csho
$INVT$	Inventories	invt/csho
$RECT$	Receivables	rect/csho
ACT	Total current assets	act/csho

Variable	Description	Computation
$PPENT$	Property, plant, and equipment (Net)	ppent/csho
$IVAO$	Investments and advances	ivao/csho
$INTAN$	Intangible assets	intan/csho
AT	Total assets	at/csho
AP	Accounts payable	ap/csho
DLC	Debt in current liabilities	dlc/csho
TXP	Income taxes payable	txp/csho
LCT	Total current liabilities	lct/csho
$DLTT$	Long-term debt	dltt/csho
LT	Total liabilities	lt/csho
CEQ	CommonOrdinary equity	ceq/csho
CFO	Cash flow from operating activities	(oancf - xidoc)/csho

Table 1. Forecasting Performance for the US Sample

This table presents the time-series average of the mean absolute forecasting errors (MAFE) using the US sample for both traditional models and machine learning models. Forecasting error for the HVZ model is calculated as the absolute value of the difference between forecast earnings and actual earnings, scaled by market value of equity at the fiscal year end. Forecasting error for all other models (RW, EP, RI, Lasso, Ridge, GBR, and RF) is calculated as the absolute value of the difference between forecast earnings per share and actual earnings per share, scaled by the stock price at the prior fiscal year end. The t-statistics are reported in the parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively

	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)
Model	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat
HVZ	0.078***	[31.18]	0.124***	[31.24]	0.165***	[29.62]
RW	0.064***	[30.27]	0.093***	[28.04]	0.117***	[25.72]
EP	0.057***	[26.97]	0.081***	[31.95]	0.101***	[31.38]
RI	0.057***	[27.11]	0.079***	[31.73]	0.098***	[29.84]
Lasso	0.056***	[28.55]	0.079***	[30.71]	0.100***	[29.13]
Ridge	0.056***	[28.52]	0.079***	[30.53]	0.100***	[29.16]
GBR	0.054***	[26.77]	0.075***	[28.59]	0.095***	[26.36]
RF	0.054***	[27.18]	0.075***	[28.69]	0.094***	[26.44]
Comparison						
GBR - RW	-0.010***	[-6.76]	-0.018***	[-8.66]	-0.022***	[-8.72]
GBR - EP	-0.004***	[-7.01]	-0.005***	[-6.48]	-0.007***	[-6.04]
GBR - RI	-0.003***	[-7.66]	-0.003***	[-5.24]	-0.003***	[-3.98]
RF - RW	-0.009***	[-6.63]	-0.018***	[-8.78]	-0.022***	[-8.84]
RF - EP	-0.003***	[-5.75]	-0.005***	[-6.73]	-0.007***	[-6.80]
RF - RI	-0.003***	[-6.25]	-0.003***	[-5.39]	-0.003***	[-4.93]

Table 2. Forecasting Performance in the US: Sub-sample Analyses

This table presents the time-series average of the mean absolute forecasting errors of the traditional and machine learning models for sub-samples within the US. Panels A, B, and reports results of the sample partitioned by firm size, analyst coverage, and earnings volatility, respectively. Forecasting error for the HVZ model is calculated as the absolute value of the difference between forecast earnings and actual earnings, scaled by market value of equity at the fiscal year end. Forecasting error for all other models (RW, EP, RI, Lasso, Ridge, GBR, and RF) is calculated as the absolute value of the difference between forecast earnings per share and actual earnings per share, scaled by the stock price at the fiscal year end. The t-statistics are reported in the parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively

Panel A: Partition Analyses by Firm Size

	Small Firms						Large Firms					
	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*		*One-Year Ahead*		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat
HVZ	0.118***	[29.97]	0.191***	[28.37]	0.259***	[26.30]	0.038***	[25.63]	0.056***	[25.32]	0.071***	[24.09]
RW	0.089***	[30.24]	0.128***	[28.47]	0.161***	[26.13]	0.038***	[26.46]	0.058***	[24.40]	0.073***	[22.63]
EP	0.078***	[28.26]	0.110***	[32.57]	0.139***	[32.72]	0.036***	[22.44]	0.051***	[26.98]	0.064***	[25.80]
RI	0.078***	[28.16]	0.106***	[32.33]	0.131***	[31.19]	0.037***	[23.00]	0.051***	[26.88]	0.064***	[24.72]
Lasso	0.075***	[29.46]	0.106***	[31.64]	0.133***	[30.54]	0.036***	[24.39]	0.053***	[25.98]	0.066***	[24.22]
Ridge	0.076***	[29.45]	0.106***	[31.44]	0.133***	[30.62]	0.036***	[24.36]	0.053***	[25.90]	0.066***	[24.16]
GBR	0.073***	[27.54]	0.100***	[29.14]	0.125***	[27.34]	0.035***	[22.96]	0.051***	[24.86]	0.064***	[22.77]
RF	0.073***	[28.01]	0.100***	[29.37]	0.125***	[27.61]	0.035***	[23.32]	0.051***	[24.70]	0.064***	[22.58]
Comparison												
GBR - RW	-0.016***	[-7.98]	-0.029***	[-10.20]	-0.035***	[-9.95]	-0.004***	[-3.74]	-0.007***	[-4.75]	-0.008***	[-5.10]
GBR - EP	-0.006***	[-8.28]	-0.010***	[-8.63]	-0.014***	[-8.77]	-0.001***	[-3.82]	-0.000	[-0.79]	0.001	[0.73]
GBR - RI	-0.005***	[-8.60]	-0.006***	[-6.84]	-0.006***	[-6.18]	-0.002***	[-5.30]	-0.001	[-1.41]	0.000	[0.63]
RF - RW	-0.016***	[-7.96]	-0.029***	[-10.25]	-0.036***	[-10.05]	-0.003***	[-3.33]	-0.007***	[-4.87]	-0.009***	[-5.22]
RF - EP	-0.005***	[-7.13]	-0.010***	[-9.22]	-0.014***	[-9.90]	-0.001**	[-2.44]	-0.000	[-0.60]	0.000	[0.46]
RF - RI	-0.005***	[-7.36]	-0.006***	[-7.45]	-0.007***	[-7.66]	-0.001***	[-3.56]	-0.000	[-1.10]	0.000	[0.23]

Table 2. (Continued)

Panel B: Partition Analyses by Analyst Coverage

	No Coverage						With Coverage					
	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*		*One-Year Ahead*		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat
HVZ	0.107***	[26.71]	0.177***	[25.10]	0.241***	[24.51]	0.053***	[25.17]	0.082***	[12.55]	0.101***	[16.52]
RW	0.077***	[31.84]	0.112***	[33.49]	0.140***	[31.75]	0.052***	[15.46]	0.078***	[10.12]	0.096***	[9.66]
EP	0.071***	[27.46]	0.100***	[35.09]	0.127***	[37.95]	0.048***	[12.26]	0.063***	[22.06]	0.080***	[14.93]
RI	0.071***	[27.61]	0.096***	[34.95]	0.120***	[37.69]	0.048***	[12.97]	0.061***	[30.99]	0.076***	[20.14]
Lasso	0.068***	[30.92]	0.096***	[36.33]	0.122***	[38.25]	0.046***	[17.16]	0.061***	[37.05]	0.077***	[22.93]
Ridge	0.069***	[30.81]	0.096***	[36.12]	0.122***	[38.12]	0.046***	[16.63]	0.061***	[35.98]	0.076***	[26.51]
GBR	0.066***	[29.09]	0.091***	[34.57]	0.114***	[35.21]	0.044***	[17.44]	0.057***	[37.17]	0.071***	[32.38]
RF	0.066***	[29.70]	0.091***	[34.76]	0.114***	[35.32]	0.045***	[15.06]	0.057***	[37.02]	0.071***	[30.20]
Comparison												
GBR - RW	-0.011***	[-6.67]	-0.021***	[-8.78]	-0.026***	[-8.37]	-0.009***	[-4.98]	-0.020***	[-2.75]	-0.025***	[-2.84]
GBR - EP	-0.005***	[-7.67]	-0.009***	[-7.74]	-0.013***	[-7.69]	-0.005***	[-3.15]	-0.006***	[-2.85]	-0.008**	[-2.11]
GBR - RI	-0.005***	[-8.04]	-0.006***	[-6.51]	-0.006***	[-5.65]	-0.005***	[-3.55]	-0.004***	[-4.62]	-0.005**	[-2.24]
RF - RW	-0.011***	[-6.63]	-0.021***	[-8.96]	-0.026***	[-8.43]	-0.008***	[-5.20]	-0.020***	[-2.73]	-0.025***	[-2.91]
RF - EP	-0.005***	[-6.66]	-0.009***	[-7.92]	-0.013***	[-8.31]	-0.004***	[-3.66]	-0.006***	[-2.81]	-0.008**	[-2.23]
RF - RI	-0.005***	[-6.92]	-0.006***	[-6.52]	-0.006***	[-6.50]	-0.004***	[-4.37]	-0.004***	[-4.47]	-0.005**	[-2.47]

Table 2. (Continued)

Panel C: Partition Analyses by Earnings Volatility

	High Volatility						Low Volatility					
	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*		*One-Year Ahead*		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat
HVZ	0.096***	[26.63]	0.153***	[27.36]	0.201***	[25.88]	0.053***	[27.66]	0.083***	[26.92]	0.110***	[24.13]
RW	0.079***	[27.12]	0.114***	[28.85]	0.139***	[28.16]	0.045***	[24.56]	0.068***	[21.35]	0.088***	[18.20]
EP	0.075***	[24.27]	0.103***	[35.02]	0.126***	[36.12]	0.040***	[23.29]	0.057***	[27.40]	0.072***	[23.27]
RI	0.075***	[24.46]	0.100***	[35.12]	0.120***	[35.82]	0.040***	[23.26]	0.057***	[26.62]	0.072***	[22.16]
Lasso	0.070***	[30.87]	0.099***	[37.60]	0.121***	[34.60]	0.040***	[24.32]	0.058***	[25.00]	0.074***	[21.47]
Ridge	0.070***	[29.79]	0.098***	[35.18]	0.121***	[31.96]	0.040***	[24.47]	0.058***	[24.89]	0.074***	[21.31]
GBR	0.068***	[27.30]	0.093***	[32.93]	0.115***	[32.45]	0.038***	[22.98]	0.056***	[24.04]	0.072***	[20.48]
RF	0.070***	[20.05]	0.094***	[29.55]	0.115***	[32.79]	0.038***	[23.80]	0.056***	[24.25]	0.071***	[20.58]
Comparison												
GBR - RW	-0.011***	[-5.13]	-0.020***	[-6.04]	-0.024***	[-6.14]	-0.007***	[-6.12]	-0.012***	[-6.68]	-0.016***	[-6.92]
GBR - EP	-0.007***	[-5.53]	-0.010***	[-9.45]	-0.012***	[-5.80]	-0.001***	[-3.53]	-0.001	[-1.19]	-0.001	[-0.78]
GBR - RI	-0.007***	[-5.28]	-0.007***	[-8.22]	-0.005***	[-3.38]	-0.002***	[-4.42]	-0.001*	[-1.99]	-0.000	[-0.87]
RF - RW	-0.009**	[-2.50]	-0.019***	[-4.81]	-0.024***	[-5.70]	-0.007***	[-6.03]	-0.012***	[-6.78]	-0.016***	[-7.05]
RF - EP	-0.005***	[-4.23]	-0.009***	[-6.34]	-0.011***	[-4.99]	-0.001***	[-2.84]	-0.001	[-0.89]	-0.001	[-1.19]
RF - RI	-0.005***	[-4.53]	-0.006***	[-4.95]	-0.005**	[-2.67]	-0.001***	[-3.60]	-0.001	[-1.50]	-0.001	[-1.48]

Table 3. Forecasting Performance for the International Sample

This table presents the time-series average of the MAFE in our international sample for both traditional models and machine learning models. Forecasting error for the HVZ model is calculated as the absolute value of the difference between forecast earnings and actual earnings, scaled by market value of equity at the fiscal year end. Forecasting error for all other models (RW, EP, RI, Lasso, Ridge, GBR, and RF) is calculated as the absolute value of the difference between forecast earnings per share and actual earnings per share, scaled by the stock price at the fiscal year end. The t-statistics are reported in the parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively

	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)
Model	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat
HVZ	0.126***	[16.74]	0.188***	[18.75]	0.246***	[17.11]
RW	0.078***	[12.19]	0.099***	[12.94]	0.115***	[13.29]
EP	0.098***	[13.74]	0.130***	[12.73]	0.153***	[12.61]
RI	0.092***	[12.25]	0.121***	[11.45]	0.139***	[11.73]
Lasso	0.119***	[12.59]	0.165***	[11.61]	0.199***	[13.41]
Ridge	0.120***	[13.22]	0.165***	[11.68]	0.199***	[13.42]
GBR	0.073***	[13.35]	0.090***	[15.17]	0.105***	[15.28]
RF	0.082***	[11.82]	0.103***	[12.67]	0.128***	[11.35]
Comparison						
GBR - RW	-0.005***	[-4.09]	-0.009***	[-3.88]	-0.010***	[-3.98]
GBR - EP	-0.025***	[-9.26]	-0.040***	[-7.26]	-0.047***	[-6.90]
GBR - RI	-0.019***	[-6.36]	-0.031***	[-5.58]	-0.034***	[-5.25]
RF - RW	0.005***	[2.99]	0.004	[1.25]	0.013**	[2.63]
RF - EP	-0.016***	[-4.91]	-0.028***	[-5.01]	-0.024***	[-4.10]
RF - RI	-0.010**	[-2.83]	-0.018***	[-3.20]	-0.011	[-1.61]

Table 4. Forecasting Performance Internationally: Sub-sample Analyses

This table presents the time-series average of the MAFE of the traditional and machine learning models for sub-samples within the international data. Panels A, B, and C reports results of the sample partitioned by firm size, analyst coverage, and earnings volatility, respectively. Forecasting error for the HVZ model is calculated as the absolute value of the difference between forecast earnings and actual earnings, scaled by market value of equity at the fiscal year end. Forecasting error for all other models (RW, EP, RI, Lasso, Ridge, GBR, and RF) is calculated as the absolute value of the difference between forecast earnings per share and actual earnings per share, scaled by the stock price at the fiscal year end. The t-statistics are reported in the parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively

Panel A: Partition Analyses by Firm Size

	Small Firms						Large Firms					
	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*		*One-Year Ahead*		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat
HVZ	0.213***	[16.45]	0.324***	[18.26]	0.426***	[14.99]	0.040***	[13.15]	0.052***	[12.99]	0.065***	[14.67]
RW	0.117***	[11.92]	0.147***	[12.89]	0.168***	[13.00]	0.039***	[12.23]	0.052***	[12.21]	0.062***	[12.57]
EP	0.133***	[13.32]	0.173***	[12.90]	0.202***	[12.49]	0.064***	[13.28]	0.088***	[11.60]	0.103***	[11.75]
RI	0.125***	[12.30]	0.161***	[12.14]	0.184***	[12.42]	0.059***	[11.14]	0.081***	[9.74]	0.095***	[9.73]
Lasso	0.158***	[12.20]	0.213***	[11.79]	0.253***	[12.97]	0.080***	[12.49]	0.116***	[10.66]	0.145***	[13.27]
Ridge	0.161***	[12.82]	0.214***	[11.81]	0.254***	[13.07]	0.080***	[13.06]	0.117***	[10.80]	0.144***	[13.11]
GBR	0.106***	[12.66]	0.126***	[14.83]	0.146***	[15.12]	0.040***	[14.31]	0.054***	[14.34]	0.065***	[13.86]
RF	0.116***	[11.54]	0.141***	[12.85]	0.172***	[11.76]	0.049***	[11.89]	0.065***	[12.03]	0.084***	[10.18]
Comparison												
GBR - RW	-0.011***	[-6.32]	-0.020***	[-5.67]	-0.022***	[-5.58]	0.002**	[2.27]	0.002	[1.32]	0.003*	[2.08]
GBR - EP	-0.027***	[-9.08]	-0.047***	[-7.15]	-0.056***	[-6.74]	-0.023***	[-8.43]	-0.034***	[-6.80]	-0.038***	[-6.52]
GBR - RI	-0.020***	[-6.38]	-0.034***	[-5.57]	-0.038***	[-5.33]	-0.019***	[-6.08]	-0.028***	[-5.32]	-0.030***	[-4.88]
RF - RW	-0.000	[-0.21]	-0.006*	[-1.77]	0.004	[0.76]	0.010***	[5.52]	0.013***	[4.36]	0.022***	[4.43]
RF - EP	-0.016***	[-5.10]	-0.032***	[-5.42]	-0.030***	[-4.48]	-0.015***	[-4.25]	-0.023***	[-4.09]	-0.019***	[-3.17]
RF - RI	-0.009**	[-2.73]	-0.020***	[-3.38]	-0.011	[-1.54]	-0.010**	[-2.78]	-0.017**	[-2.85]	-0.011	[-1.61]

Table 4. (Continued)

Panel B: Partition Analyses by Analyst Coverage

	No Coverage						With Coverage					
	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*		*One-Year Ahead*		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat
HVZ	0.188***	[17.35]	0.289***	[18.29]	0.386***	[15.90]	0.089***	[12.77]	0.128***	[14.35]	0.162***	[14.52]
RW	0.094***	[14.42]	0.121***	[14.34]	0.142***	[13.76]	0.068***	[10.25]	0.087***	[11.44]	0.099***	[12.02]
EP	0.115***	[15.58]	0.151***	[14.03]	0.180***	[13.46]	0.088***	[11.99]	0.118***	[11.20]	0.136***	[11.19]
RI	0.109***	[14.18]	0.142***	[12.49]	0.167***	[12.02]	0.082***	[10.66]	0.109***	[10.37]	0.123***	[11.02]
Lasso	0.131***	[17.08]	0.178***	[20.67]	0.218***	[23.47]	0.112***	[9.65]	0.158***	[8.40]	0.189***	[9.44]
Ridge	0.134***	[18.60]	0.180***	[19.42]	0.220***	[24.44]	0.113***	[9.96]	0.158***	[8.51]	0.188***	[9.36]
GBR	0.089***	[15.01]	0.109***	[15.66]	0.130***	[14.73]	0.064***	[11.55]	0.079***	[13.78]	0.091***	[14.14]
RF	0.102***	[13.09]	0.126***	[13.21]	0.160***	[11.33]	0.071***	[10.33]	0.089***	[11.58]	0.110***	[10.54]
Comparison												
GBR - RW	-0.005***	[-4.65]	-0.012***	[-5.86]	-0.013***	[-5.69]	-0.004***	[-3.03]	-0.008**	[-2.69]	-0.008**	[-2.83]
GBR - EP	-0.026***	[-8.11]	-0.042***	[-7.43]	-0.051***	[-7.27]	-0.025***	[-9.06]	-0.040***	[-6.53]	-0.045***	[-6.18]
GBR - RI	-0.020***	[-5.94]	-0.033***	[-5.47]	-0.038***	[-4.92]	-0.019***	[-6.29]	-0.030***	[-5.32]	-0.032***	[-5.33]
RF - RW	0.008***	[3.00]	0.005	[1.64]	0.018***	[3.19]	0.003**	[2.39]	0.003	[0.91]	0.011*	[1.90]
RF - EP	-0.013***	[-2.91]	-0.025***	[-3.71]	-0.021**	[-2.82]	-0.017***	[-6.37]	-0.029***	[-5.50]	-0.026***	[-4.68]
RF - RI	-0.007	[-1.56]	-0.016**	[-2.31]	-0.008	[-0.94]	-0.011***	[-3.76]	-0.019***	[-3.73]	-0.013*	[-1.98]

Table 4. (Continued)

Panel C: Partition Analyses by Earnings Volatility

	High Volatility						Low Volatility					
	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*		*One-Year Ahead*		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Model	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat	Mean Coeff	*t*-stat
HVZ	0.155***	[18.38]	0.229***	[17.50]	0.301***	[16.82]	0.088***	[14.89]	0.128***	[16.18]	0.167***	[16.67]
RW	0.089***	[12.85]	0.111***	[12.66]	0.127***	[12.38]	0.061***	[11.41]	0.079***	[12.14]	0.093***	[13.30]
EP	0.109***	[14.05]	0.143***	[12.22]	0.168***	[12.29]	0.083***	[12.72]	0.111***	[11.31]	0.129***	[11.55]
RI	0.102***	[12.70]	0.132***	[10.90]	0.152***	[11.27]	0.078***	[11.57]	0.104***	[10.71]	0.119***	[11.16]
Lasso	0.125***	[14.16]	0.167***	[11.38]	0.204***	[13.27]	0.106***	[9.76]	0.150***	[9.08]	0.181***	[10.52]
Ridge	0.128***	[14.79]	0.169***	[11.62]	0.205***	[13.28]	0.106***	[9.98]	0.150***	[9.14]	0.180***	[10.32]
GBR	0.081***	[14.09]	0.097***	[14.69]	0.111***	[14.03]	0.060***	[12.68]	0.076***	[14.60]	0.090***	[14.49]
RF	0.093***	[12.69]	0.113***	[11.64]	0.140***	[10.09]	0.067***	[11.50]	0.085***	[12.47]	0.106***	[11.16]
Comparison												
GBR - RW	-0.007***	[-5.16]	-0.014***	[-5.56]	-0.016***	[-5.44]	-0.001	[-1.28]	-0.003	[-1.09]	-0.003	[-1.12]
GBR - EP	-0.028***	[-8.98]	-0.046***	[-7.40]	-0.056***	[-7.55]	-0.023***	[-7.89]	-0.034***	[-5.81]	-0.039***	[-5.85]
GBR - RI	-0.021***	[-6.42]	-0.035***	[-5.53]	-0.041***	[-5.49]	-0.018***	[-5.83]	-0.027***	[-4.92]	-0.029***	[-4.75]
RF - RW	0.004*	[1.91]	0.002	[0.41]	0.013**	[2.20]	0.006***	[4.96]	0.006**	[2.30]	0.013**	[2.51]
RF - EP	-0.016***	[-4.05]	-0.030***	[-4.63]	-0.028***	[-4.07]	-0.016***	[-5.21]	-0.026***	[-4.81]	-0.023***	[-4.23]
RF - RI	-0.009**	[-2.30]	-0.020***	[-2.91]	-0.012	[-1.57]	-0.011***	[-3.44]	-0.019***	[-3.56]	-0.013*	[-2.06]

Table 5. Forecasting Bias

This table presents the time-series average of the mean forecasting bias for both the traditional and machine learning models. Panel A reports results for the US sample while Panel B reports results for the international sample. Forecasting bias for the HVZ model is calculated as the difference between forecast and actual earnings, scaled by market value of equity at the fiscal year end. Forecasting bias for all other models (RW, EP, RI, Lasso, Ridge, GBR, and RF) is calculated as the difference between forecast earnings per share and actual earnings per share, scaled by the stock price at the fiscal year end. The t-statistics are reported in the parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively

Panel A: Forecasting Bias for the US Sample

	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)
Model	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat
HVZ	0.030***	[11.19]	0.059***	[12.98]	0.088***	[15.30]
RW	-0.018***	[-8.19]	-0.034***	[-9.27]	-0.049***	[-9.21]
EP	0.007***	[3.40]	0.010***	[3.42]	0.013***	[3.19]
RI	0.002	[0.79]	0.000	[0.02]	-0.002	[-0.60]
Lasso	0.002	[0.97]	-0.001	[-0.35]	-0.003	[-0.79]
Ridge	0.002	[1.15]	-0.001	[-0.32]	-0.004	[-0.86]
GBR	0.001	[0.33]	-0.007**	[-2.54]	-0.014***	[-3.40]
RF	-0.004**	[-2.11]	-0.010***	[-3.35]	-0.016***	[-3.65]

Panel B: Forecasting Bias for the International Sample

	One-Year Ahead		*Two-Year Ahead*		*Three-Year Ahead*	
	(1)	(2)	(3)	(4)	(5)	(6)
Model	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat
HVZ	0.065***	[11.95]	0.119***	[11.15]	0.171***	[9.93]
RW	-0.015***	[-4.19]	-0.029***	[-4.28]	-0.039***	[-4.90]
EP	0.023**	[2.46]	0.037**	[2.62]	0.042**	[2.48]
RI	0.009	[1.01]	0.014	[0.98]	0.010	[0.64]
Lasso	0.061***	[6.44]	0.094***	[6.06]	0.124***	[8.31]
Ridge	0.060***	[6.36]	0.090***	[5.68]	0.120***	[7.91]
GBR	-0.001	[-0.26]	-0.007	[-1.13]	-0.011	[-1.37]
RF	0.007	[1.53]	0.012	[1.47]	0.025**	[2.45]

Table 6. Performance of Model-Based ICCs for the US Sample

This table presents the performance of model-based ICCs for the US sample. The implied cost of capital is computed as the average value based on four models, GLS, CT, PEG, and OJ. Panel A presents the univariate Fama-MacBeth regression results, with one-month-ahead realized return as the dependent variable and the model-based ICC as the independent variable. Panel B presents the results of firms sorted into deciles by the model-based ICCs. The odd columns report the equal-weighted mean ICC of the portfolios while the even columns report the equal-weighted mean realized returns of the portfolios. The last row of Panel B reports results of the spread between the highest and lowest decile of firms.The t-statistics are reported in the parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

Panel A: Regression Analyses

	(1)	(2)	(3)	(4)	(5)	(6)
Model	Slope Coeff	t-stat	Intercept	t-stat	R^2	F-test for Slope = 1
HVZ	0.346***	[4.19]	0.010***	[3.96]	0.022	62.80***
EP	0.497***	[2.86]	0.010***	[3.94]	0.021	8.37***
RI	0.593***	[3.40]	0.010***	[3.79]	0.020	5.44**
Lasso	0.665***	[3.60]	0.010***	[3.98]	0.020	3.29*
Ridge	0.704***	[3.93]	0.009***	[3.89]	0.020	2.73*
GBR	0.912***	[5.02]	0.008***	[3.21]	0.020	0.23
RF	0.992***	[4.66]	0.008***	[3.07]	0.020	0.00

Table 6. (Continued)

Panel B: Portfolio Analyses

	HVZ		*RI*		*EP*		*Lasso*		*Ridge*		*GBR*		*RF*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Decile	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns
1	0.34	1.18	0.33	1.00	0.26	1.09	0.30	1.01	0.30	1.02	0.26	0.88	0.26	0.96
2	0.52	1.21	0.44	1.31	0.39	1.26	0.40	1.20	0.40	1.16	0.38	1.28	0.38	1.12
3	0.64	1.25	0.51	1.32	0.46	1.32	0.46	1.29	0.46	1.26	0.44	1.29	0.44	1.31
4	0.75	1.25	0.57	1.44	0.52	1.46	0.51	1.40	0.51	1.38	0.50	1.21	0.50	1.37
5	0.87	1.46	0.62	1.38	0.58	1.32	0.57	1.38	0.57	1.42	0.55	1.53	0.55	1.34
6	1.00	1.42	0.68	1.50	0.64	1.32	0.63	1.38	0.63	1.38	0.61	1.41	0.60	1.50
7	1.16	1.31	0.76	1.38	0.72	1.42	0.71	1.43	0.71	1.44	0.68	1.43	0.66	1.43
8	1.38	1.58	0.86	1.49	0.82	1.52	0.81	1.54	0.81	1.57	0.78	1.53	0.75	1.52
9	1.75	1.58	1.03	1.51	0.96	1.57	0.98	1.59	0.98	1.63	0.92	1.58	0.88	1.57
10	2.63	1.86	1.41	1.70	1.28	1.77	1.32	1.82	1.32	1.79	1.22	1.98	1.15	2.01
Spread	2.29***	0.68***	1.08***	0.70***	1.01***	0.68***	1.02***	0.80***	1.02***	0.77***	0.96***	1.10***	0.89***	1.06***

Table 7. Performance of Model-Based ICCs for the International Sample

Panel A: Regression Analyses

This table presents the performance of model-based ICCs for the international sample. The implied cost of capital is computed as the average value based on four models, GLS, CT, PEG, and OJ. Panel A presents the univariate Fama-MacBeth regression results, with one-month-ahead realized return as the dependent variable and the model-based ICC as the independent variable. Panel B presents the results of firms sorted into deciles by the model-based ICCs. The odd columns report the equal-weighted mean ICC of the portfolios while the even columns report the equal-weighted mean realized returns of the portfolios. The last row of Panel B reports results of the spread between the highest and lowest decile of firms.The t-statistics are reported in the parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Model	Slope Coeff	t-stat	Intercept	t-stat	R^2	F-test for Slope = 1
HVZ	0.074	[1.01]	0.006**	[2.26]	0.016	161.09***
EP	0.529***	[3.20]	0.002	[0.52]	0.016	8.12***
RI	0.649***	[3.34]	0.001	[0.37]	0.017	3.27*
Lasso	0.204**	[2.09]	0.004	[1.20]	0.018	66.75***
Ridge	0.217**	[2.22]	0.003	[1.17]	0.017	63.66***
GBR	1.013***	[5.71]	0.001	[0.25]	0.014	0.01
RF	0.639***	[4.18]	0.002	[0.75]	0.013	5.58**

Table 7. (Continued)

Panel B: Portfolio Analyses

	HVZ		*RI*		*EP*		*Lasso*		*Ridge*		*GBR*		*RF*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Decile	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns	ICC	Realized Returns
1	0.23	0.09	0.18	0.03	0.15	-0.07	0.22	0.11	0.22	0.15	0.16	-0.10	0.18	-0.11
2	0.41	0.44	0.33	0.32	0.29	0.32	0.41	0.37	0.40	0.32	0.26	0.09	0.29	0.37
3	0.54	0.44	0.44	0.24	0.38	0.32	0.58	0.53	0.57	0.56	0.32	0.62	0.37	0.46
4	0.66	0.59	0.54	0.60	0.47	0.35	0.77	0.72	0.75	0.68	0.38	0.78	0.44	0.61
5	0.78	0.88	0.64	0.43	0.56	0.86	0.96	0.67	0.94	0.54	0.45	0.67	0.51	0.71
6	0.94	0.78	0.75	0.97	0.65	0.70	1.17	0.72	1.15	0.82	0.52	0.72	0.60	0.82
7	1.14	1.05	0.88	0.87	0.77	0.89	1.40	0.82	1.38	0.99	0.61	0.81	0.70	0.98
8	1.42	0.86	1.05	1.00	0.90	1.01	1.67	0.95	1.64	0.84	0.72	0.69	0.85	0.67
9	1.94	0.86	1.27	0.96	1.10	1.09	2.01	0.71	1.98	0.66	0.87	1.13	1.06	0.97
10	3.38	0.51	1.83	1.11	1.55	1.07	2.54	0.93	2.52	0.96	1.20	1.12	1.68	0.97
Spread	3.15***	0.42*	1.65***	1.08***	1.39***	1.14***	2.32***	0.82***	2.30***	0.81***	1.04***	1.22***	1.49***	1.08***