Two Tales of Dividends and Future Earnings: Evidence from Machine Learning Models

Ebenezer Asem[1], Elnaz Davoodi[2], and Hamed Ghanbari[1 1]

[1]Dhillon School of Business, University of Lethbridge

[2]Google DeepMind

July 31, 2024

Two Tales of Dividends and Future Earnings: Evidence from Machine Learning Models

Ebenezer Asem[1], Elnaz Davoodi[2], and Hamed Ghanbari[†1]

[1]Dhillon School of Business, University of Lethbridge

[2]Google DeepMind

July 31, 2024

Abstract

Studies of the information content of dividends show that dividend changes convey information about earnings in the next four quarters. We examine the extent to which dividend changes contribute to predicting earnings in the subsequent four quarters and find marginal contributions based on both in-sample and out-of-sample tests. Further, we use state-of-the-art machining learning models to examine the contributions of dividend changes. These models deliver improved predictions of future earnings, but the contribution of dividend changes remains marginal that is robust across a variety of machine learning specifications. Our results suggest that the predictive power of the dividend changes is marginal even when these dividend changes convey information about future earnings.

Keywords: Dividends, Earning expectations, Machine Learning Models, Gradient Boosting Regression Trees, Random Forest Classifiers, Neural Networks

[*]Corresponding author email: hamed.ghanbari@uleth.ca

[†]Corresponding author email: hamed.ghanbari@uleth.ca

1 Introduction

One of the most debated topics in corporate finance is whether a firm's dividend change conveys information about its future earnings. While both theory and surveys point to dividend changes conveying information about future earnings (e.g., Miller and Modigliani (1961) and Brav et al. (2005)), the empirical evidence has been mixed. Studies such as Brickley (1983), Aharony and Dotan (1994), and Nissim and Ziv (2001) all find that future earnings performance is directly related to dividend changes. However, these studies provide contradicting results about the persistence and the economic significance of future earnings response to such dividend announcements. By contrast, Watts (1973), Penman (1983), Benartzi et al. (1997), and Grullon et al. (2005) find no relation between future earnings and dividend changes. These studies propose that market reaction to changes in dividend should be realized elsewhere or even explained through behavioral theories.

More recently, Ham et al. (2020) revisit the empirical debate and document a positive relation between future earnings performance and dividend changes. They note that most of the previous disagreements about this empirical relation can be resolved by focusing on quarterly data and using an event-window approach to compute future earnings versus fiscal year approach previously used to compute future earnings. We continue this line of research, reexamine this relationship and then focus on the predictive power of the dividend changes. In particular, we investigate the extent to which the relation between future earnings and dividend changes help better predict future earnings. Historically, the former relationship was evaluated using hand-crafted functional forms between dividend changes and future earning changes including a set of control variables by applying conventional econometrics techniques.

In this study, we depart from the realm of linear regression analysis by relying on the strength of ensemble Machine Learning (ML) algorithms in automatic discovery of interactions between the relevant variables without any prior assumptions about the nature of this relationship. ML algorithms offer spectrum choices from linear models to boosted trees and neural networks designed to approximate complex non-linearity. Some of these algorithms are also robust to outliers and multicollinearity among the independent variables which make them ideal candidates to explore the intricate relationships such as ours. In addition, feature ablation studies, feature importance measurement, parameter penalization and out-of-sample evaluations help avoid overfitting biases and false discoveries.

We build, train and test several state-of-the-art ensemble and boosting classification and regression Machine Learning (ML) algorithms to evaluate the modeling and predictive power of dividend changes. See, for instance, Dietterich (2000), Breiman (2017), Zhou (2012), Schapire (2003), Friedman (2001), Natekin and Knoll (2013) and references therein. Following Ham et al. (2020), we adopt an event-window approach to distinguish the earnings after the dividend change announcements versus the fiscal year approach. We find that the dividend changes contain some information about the earnings changes in upcoming years post dividend announcements. Nonetheless, the incremental information content is marginal and the feature importance of dividend changes is less than 2%. This finding is robust across different algorithms examined in this study. In addition, we find that the predictive power of dividend changes is very low in predicting unexpected changes in future earnings. Ablation studies, feature importance scorers, and SHAP values point to the marginal contribution of dividend changes to the overall R-squared in all ML models applied in this study. In addition, in a restricted model without any control variables, the R-squared of ML regression models is negligible and the accuracy of ML classifiers are as low as the ZeroR model. All of which points to the relatively low marginal contribution of dividend changes to modeling the evolution of unexpected future earnings. Accordingly, we conclude that although future earnings are related to dividend changes, the dividend changes have low[marginal] modeling and predictive power.

This study makes at least three key contributions to the literature on the relation between future earnings and dividend changes. *First*, we confirm that dividend changes convey marginal information about future earnings. *Second*, given that future earnings are related to dividend changes, we investigate whether this knowledge is beneficial in predicting future earnings and we find that this is not the case. *Third*, to our knowledge, this is the first study that employs state-of-the-art Machine Learning (ML) algorithms, which do not impose any functional or distributional assumptions on the models used to evaluate the relation between future earnings and dividend changes. We find 50% higher out-of-sample predictive R-squared relative to preceding literature that is robust across a variety of machine learning specifications. Nonetheless, the prediction power of dividend changes remains limited. Thus, we conclude that machine learning algorithms have the potential to improve our empirical understanding of this highly debated topic over the past 60 years.

While dividend increases (cuts) are associated with positive (negative) abnormal returns, the drivers of these abnormal returns have confounded researchers for a long time. Although there are several theories to explain this observation, a lot of the literature has focused on earnings as a possible driver and we follow this line of research.[1] Lintner (1956) is the first to document a connection between earnings and dividend changes, reporting that managers increase dividends only when they believe that the current earnings increases are permanent. After this, Miller and Modigliani (1961) model posits that dividend changes convey managerial information about future earnings, predicting a positive relation between the two variables. Nonetheless, the empirical evidence has been mixed for several decades.

The empirical evidence on the relation between future earnings and dividend changes is extensive and we cannot provide detailed discussions of this literature here. Rather we focus on the more recent result from Ham et al. (2020). They note that most of the previous disagreements about the empirical relation can be resolved by focusing on quarterly data and using an event-window approach to compute future earnings. They report a positive relation between dividend changes and earnings in the next four quarters and that the increase in earnings persists for the subsequent two years. This is consistent with Aharony and Dotan (1994) finding that dividend increases convey information about earnings in the subsequent four quarters. Thus, the evidence on the relation between dividend

[1]For instance, the wealth redistribution hypothesis suggests that the positive abnormal return associated with dividend increases is due to redistribution wealth away from bondholders to shareholders and the negative price reaction to dividend cuts capture redistribution of wealth in the opposite direction (e.g., Handjinicolaou and Kalay (1984)).

changes and earnings performance in the subsequent four quarters appears clear: there is a positive relation. Given that dividend changes convey information about the next four quarters' earnings, we ask a simple question. While earnings in the four quarters after a dividend change are related to the dividend change, it is unclear whether this knowledge translates into a better prediction of earnings in the four quarters after a dividend change.

The rest of the paper is organized as follows. Section 2 presents the dataset and its descriptive statistics. Section 3 describes a benchmark model adopted from previous literature. Section 4 introduces machine learning algorithms used in our analysis, presents empirical findings and discusses the robustness of results. Section 5 concludes.

2 Data

We use the same dataset as in Ham et al. (2020), which contains the data for the ordinary common stocks (nonfinancial firms only) listed on NYSE, Amex, and Nasdaq exchanges over the period of 1971-2016. We record observations in which the firm made a previous quarterly dividend declaration in the past 180 days. Dividend declarations data are from the Center for Research in Securities Prices (CRSP) events database.[2] Table 1 presents summary statistics of the dataset. Overall, the sample has 165,558 observations where 14.22% (1.12%) of dividend declarations exhibit increase (decrease) in the level of dividends compared to the prior quarter with average dividend increase (decrease) of 18.38% (43.46%). The table shows that dividend change announcements tend to be preceded by return of the same sign as dividend change across different horizons up to one year before the announcement date. Earnings changes are computed as the difference between the sum of the four consecutive quarterly earnings announced after the dividend change and earnings for the four quarters before the dividend change, where earnings is income before extraordinary items. Dividend Changes is the percentage dividend changes in a quarter relative to the dividends in the previous quarter.

3 Benchmark Models from Prior Literature

As a benchmark for comparison, we follow Ham et al. (2020) and measure the information content of dividend changes (ΔDIV) about future earnings. We reproduce some of its results that are summarized in this section. We estimate model 1 by regressing unexpected earnings changes over one year (four quarters) after dividend change announcements on the percentage dividend changes and a set of control variables. These control variables include past returns from one-month up to one-year, level and change of the past quarterly earnings up to four quarters before dividend change announcements, and six nonlinear controls for the level and change of the corresponding dependent variable for a year before the dividend announcements.

[2]See, Ham et al. (2020) for further details about the dataset.

$$\Delta E_{it+n} = \beta_0 + \beta_1 \Delta DIV_{it} + \sum_{j=2}^{j=19} \beta_j Controls + \varepsilon_{it+n} \tag{1}$$

The regression analysis in model 1 is repeated using earnings changes over the second and the third year after dividend change announcement to document the persistence of changes in earnings. As we are interested in the relative contribution of the dividend changes to the model, we conduct a similar regression in the absence of dividend changes as a covariate, as shown in model 2. The left (right) Panel in Table 2 presents the results of regression analysis in 1 (with ΔDIV) and 2 (without ΔDIV), including the coefficients, t-stats, and adjusted R-squared over the next three years post dividend change announcements. We also conduct a similar regression analysis as in model 3 that only contains dividend changes without any control variables. The results of this analysis are reported in Table 3.

$$\Delta E_{it+n} = \beta_0 + \sum_{j=1}^{j=18} \beta_j Controls + \varepsilon_{it+n} \tag{2}$$

$$\Delta E_{it+n} = \beta_0 + \beta_1 \Delta DIV_{it} + \varepsilon_{it+n} \tag{3}$$

In addition, we examine the persistence of earning changes following dividend change announcements using alternative measures of dividend news. As in Ham et al. (2020), we use percentile rank of dividend changes to control for the skewness in the distribution of dividend changes. These results are reported in Table 4. Last, we measure the robustness of results by using a different measure of net income. Table 5 shows the coefficients and R-squared when all models are estimated using gross profit as a measure on net income, both as dependent variables and as some of control variables.

The left panel in Table 2 shows that the dividend changes coefficient ($\beta_1 = 0.0255$ and $t\text{-stat} = 5.13$) is highly significant in the first year after dividend change announcements and remains significant in the second (third) year post announcements with $\beta_1 = 0.0182$ and $t\text{-stat} = 3.18$ ($\beta_1 = 0.0183$ and $t\text{-stat} = 2.67$). In the absence of all control variables, Table 3 shows that dividend changes coefficient is marginally higher and remains statistically significant over the three years post dividend change announcements. The coefficients are respectively 0.0309 (4.8521), 0.0199 (2.6698), and 0.0177 (2.0712) with t-stats in parentheses. Comparing predictive power of the models 1-3, all three models show that including ΔDIV produces a very marginal gain in modeling future earnings. For instance, comparing the R-squared in Column I to that in Column IV of Table 2 shows ΔDIV increases the R-squared by only 1.19% ($\frac{0.1866}{0.1844} - 1$). The similar results can be inferred by comparing Akaike and Bayesian Information Criterion across two models. Model 3 also confirms this finding as the reported R-squared statistics in 3 are very close to zero across all three years post dividend announcements. Nonetheless, the point estimates of the ΔDIV is statistically significant. The results in the first three columns in Tables 4-5 reproduced those in Panels A and B of Table 5 in Ham et al. (2020). The last three columns in each table show that the model's goodness-of-fit remains the same in the absence of ΔDIV.

4 Machine Learning Models

We use a set of candidate ML models for classification and regression analysis. This includes linear classifiers such as Naive Bayes classifier and logistic regression (with different parameter penalizations such as lasso, and elasticnet), and non-linear classifiers such as Support Vector Machines, Random Forests, Decision Trees, and Neural Networks. In addition, we conduct regression analysis using Gradient Boosting Trees and Random Forests. We aim to provide a detailed description of the models that make it self-explanatory in particular for readers with limited prior knowledge of these models. In general, the purpose of machine learning models is to recognize patterns from data in order to make predictions or decisions without hand-crafting or imposing explicit relationships ex-ante. For the entire analysis, we focus on supervised learning models where the data labels identify covariates and independent variables.

Our analysis relies on two categories of models, regression models and classification models. Regression analysis aims to examine the information content and predictive power of dividend changes together with a set of control variables similar to those used in the benchmark OLS analysis. We explore the extent to which dividend changes is useful in predicting future unexpected earnings. The supervised classification models aim to predict whether positive/negative change in future unexpected earnings is related to positive/negative change in dividends, while controlling for several other covariates. For both categories, we evaluate the extent to which the predictive power/classification is persistent over three years post dividend change announcements. In addition, we measure the robustness of results using different measures of earnings change and dividend change.

4.1 Gradient Boosting Trees

Gradient Boosting Trees (GBT) are nonparametric machine learning models that recursively and sequentially partition the space of independent variables into smaller regions with similar observations, fitting the trees sequentially given the residuals from the previous tree to minimize the residual errors, and then combine (boost) forecasts from these simplified decision trees into a single forecast. See, Freund (1995), and Friedman (2001) among others. The boosting theory suggests that the combination of several smaller trees, as an ensemble, increases stability and helps prevent overfitting. GBT minimizes a loss function by gradient descent (i.e., mean squared errors of regression) to fit subsequent trees. The process is repeated for a specified number of iterations or until a stopping criterion is met.

GBT often incorporates regularization techniques to find the right balance between training accuracy and generalizability to prevent overfitting. Training a GBT model requires tuning hyperparameters of the model, including number of trees, learning rate, and depth of the tree. Increasing the number of trees in the ensemble could increase the performance but also increases complexity and computational time. Learning rate, also known as the shrinkage parameter, controls the contribution of each tree to the final prediction. A lower learning rate requires more trees to achieve the same performance but can improve generalization and prevent overfitting. Deeper trees in the ensemble could capture more complex relationships in the data but are more prone to overfitting. Limiting the depth of a tree helps prevent overfitting and improves generalization.

4.2 Random Forests

Random Forest Models are another nonparametric supervised learning algorithm that combines the prediction of a collection of decision trees to predict the outcome. See, for instance, Breiman (2001), Louppe (2014), and Biau (2012). They belong to a general class of bootstrap aggregation procedure (also known as bagging) that randomly samples the training data with replacement to create multiple bootstrap samples. Each tree in the ensemble is trained on a different bootstrap sample drawn by bagging procedure, and then the average of the forecasts is used for final prediction to ensure diversity among the trees, reducing variations, and stabilizing the trees' predictive performance. Splitting criteria at each node of the decision tree (also known as dropout) lets the model select a random subset of variables to reduce the correlation between trees, further improve the variance reduction, and ensure the robustness of the ensemble given the noise in the dataset.

Training a random forests model requires tuning hyperparameters, including number of trees, maximum depth of the tree, and the minimum sample split, among others. Although deeper decision trees in the forests could capture more complex relationships in the data, such models are more prone to overfitting. Limiting depth of trees helps prevent overfitting and improves generalization. It is important to find the minimum number of samples required to split an internal node. Increasing this value can prevent the trees from splitting too early, leading to a simpler model and reducing overfitting. To ensure the reproducibility of results, a random seed should be set ex-ante. The optimal set of hyperparameters can be obtained through various techniques such as grid search, random search, and Bayesian optimization. Model performance can be measured using mean squared errors, mean absolute errors, or R-squared.

4.3 Regression Analysis with Machine Learning

Linear regression models similar to those applied historically in this literature can be thought of as the first order approximation of the data generating process. These models may restrict our capability to capture the true relations embedded in the observed data. Generalized linear models allow for more flexibility and could capture nonlinear relations to some extent. Nonetheless, without priori assumptions about nonlinear interactions between independent variables, these models become computationally infeasible when the set of independent variables becomes larger. Machine learning algorithms are alternative non-parametric approaches that accommodate nonlinearity among predictors. Unlike conventional regressions these models do not require a priori as they consider all possible interactions across predictors while remaining computationally feasible. In this study, we use particular implementations of the Gradient Boosting Regression Trees (GBRT), Random Forests, and Neural Networks.

In our analysis, we use the same set of variables as those in Ham et al. (2020). That includes dividend changes, past returns, past earnings (both level and changes) and a set of interaction variables. We start with the same set of variables to make sure that a consistent comparison is feasible between our findings and those reported in previous studies. We also repeat our analysis by dropping all interaction variables and rely on the capacity of machine learning models to detect and capture nonlinear relations between predictor variables. We evaluate each model using a 10-fold cross validation procedure. In this procedure, the dataset is divided into 10 equal subsets (fold). The training is done on nine of the 10 folds and validation (performance measurement) is done on the remaining fold. The process is repeated 10 times using a fixed random seed, with each fold serving only once as the validation set. The validation results across all 10 folds are then aggregated to generate a performance statistic such as R-squared, implying that this is an out-of-sample validation technique.

Table 6 reports predictive R-squared for Gradient Boosting Trees (GBT) and Random Forests (RF), and R-squared for Ordinary Least Squares (OLS) for models 1-2 in three years post dividend announcements. The results in this table show that GBT and RF outperform OLS with and without ΔDIV as a covariate. Panel A in Table 6 shows that in-sample R-squared of GBT and OLS models with (without) ΔDIV are 0.5161 (0.5188) and 0.1866 (0.1844) respectively. Using a 10-fold cross-validation procedure, the out-of-sample R-squared of the GBT model reduces to 0.2702 (0.2662) with (without) ΔDIV, which is still 45% (44%) better than in-sample R-squared of the OLS model. Evaluating the performance of the models in capturing future earnings changes two and three years after the dividend changes, we note that GBT and RF models retain more than 80% (64%) of their performance in-sample (out-of-sample) while this is only about 50% for the OLS model in-sample.

Next, we drop all interaction (non-linear) variables and repeat regression analysis. Although feature engineering in machine learning is a critical step in the model development pipeline and can have a significant impact on the performance of the final model, we evaluate the robustness of our results in this analysis. This new set of variables allows us to solely rely on the capacity of machine learning models to detect and capture nonlinear relations between predictor variables. Entries in Panel B of Table 6 suggest that GBT remains the best performing model in terms of goodness-of-fit statistics in the absence of all interaction variables. The results hold over the second and third year with and without ΔDIV in both in-sample and out-of-sample analysis. Although it was expected, we note that the OLS model loses around 35% of its R-squared in the absence of interaction variables whereas the GBT model only loses around 4% of its R-squared in capturing earnings changes in the first year after announcements.

We quantify the marginal contribution of each variable (feature) in ML regression models to the overall goodness-of-fit statistics. We identify the influential variables according to a notion of feature importance, where features with higher importance scores are considered more important and the scores sum to one. In linear regression models, the magnitude of the coefficient associated with each independent variable indicates its importance where the larger coefficients suggest stronger relationships with the dependent variable. There are different approaches to measure feature importance in regression-based ML models. In our tree-based models, feature importance score is based on how frequently a feature is used to split the data across all decision trees in the ensemble. Figures 1-4 present the feature importance scores based on the model performance in the first, second, and third year following the dividend announcements. In all figures, the left (right) panel shows scores with (without) ΔDIV as an independent variable in the respective model. Figures 1 and 2 report scores when the models are evaluated using a 10-fold cross-validation process whereas 1 and 2 show in-sample scores measured over the entire sample and thus more consistent with OLS results. Figures 2 and 4 report scores when all interaction variables are dropped from the model.

The left panel in Figure 1 shows that ΔDIV is ranked as the least important variable in OOS of GBT in all three years post earning announcement. In the absence of all interaction variables, the left panel in Figure 2 confirms this finding as ΔDIV has consistently the lowest score. Figures 3-4 also show that ΔDIV is the least favorable variable in explaining earnings changes when evaluating the GBT model in sample in all three years post earning announcement. Figures 1-4 show that in the absence of ΔDIV as a covariate, the relative feature importance of all variables in the right panels remain comparable to those in the left panels across all three years. Comparing the GBT's importance scores in Figure 1 and the regression coefficients in Table 2, we observe that the relative contributions of variables to the performance of the GBT model is relatively comparable to the regression coefficients in 2.

4.4 Classification Analysis with Machine Learning

A classification model aims to assign inputs into one of available classes (categories) based on a series of common features and patterns that characterize the decision boundary to separate different classes. The question at hand can also be formulated as a supervised learning classification problem, in which the goal is to predict whether positive/negative changes in future unexpected earnings is related to positive/negative change in dividends, while controlling for several other covariates. We train different types of classifiers to examine the impact of positive versus negative dividends changes in direction (positive versus negative) of the unexpected changes in future earnings.

Supervised machine learning classification models require labeled training dataset. Labeled dataset refers to data in the form of pairs of input (independent variables) and the corresponding correct label (dependent variable) for each observation (data point or entry). The representation of inputs plays an important role in the accuracy (or success rate) of a supervised machine learning model. Typically, the input is transformed into a feature vector, which contains several features (independent variables) that describe the observation. The number of features should not be too large, because of the curse of dimensionality, but should contain enough information to accurately predict the output/label.

We quantify the performance of each classifier by measuring its Accuracy, Area Under the Receiver Operating Characteristics (AUROC), Precision, Recall, and F1-score. Accuracy measures the overall proportion of the correct predictions. Mathematically, it is $\frac{TP+TN}{TP+TN+FP+FN}$ where TP (FN) is the total number of true positive (false negative) instances predicted by the model. Accuracy shows how often the model made a correct prediction across the entire dataset. This is a reliable measure if the dataset is class-balanced with a relatively similar number of observations from each class. However, if the data set is imbalanced and the majority of rows is one or zero, the model can obtain high accuracy in predicting the majority class but low accuracy in predicting the minority class. To define a benchmark accuracy score a baseline model (also known as ZeroR classifier) is required that simply predicts the majority category (class) for all instances in the dataset. In other words, it trivially predicts the most frequent class as the predicted class for every observation. Any model must achieve an accuracy score better than the baseline value to be considered useful in distinguishing the two classes.

The Area Under the Receiver Operating Characteristics (AUROC), also known as AUC, is another performance evaluation metric used in our analysis. It measures the ability of the model to distinguish between the two classes, positive class (class 1) and negative class (class 0), across different threshold values. It quantifies the overall performance of the classification model by calculating the area under the ROC curve. It is defined based on the ROC curve that plots the true positive (TP) rate versus the false positive (FP) rate at different classification thresholds. The thresholds are different probability cutoffs that separate the two classes in binary classifications. The AUC score ranges from zero to one. The AUC scores closer to one implies that the model has a higher discriminatory power and could distinguish between two classes more effectively. However, AUC scores closer to 0.5 indicates that the model performs no better than random guessing. Overall, models with higher AUC scores are preferred all else equal.

Precision is another evaluation metric that quantifies the extent to which the classification model can identify the positive class while minimizing false positives. It defines the proportion of observations that are correctly classified as the positive class. Mathematically it is defined as $\frac{TP}{TP+FP}$. A higher precision value implies that the model has a lower rate of false positives, and thus it is more precise in identifying positive instances. Overall, precision score helps measure the reliability and accuracy of the model.

Precision is often used together with recall (sensitivity) to provide a balanced assessment of a classification model. Recall measures the model's ability to correctly identify all positive instances. Mathematically, it is defined as $\frac{TP}{TP+FN}$. A higher recall score indicates that the model is better at identifying positive instances and has a lower rate of false negatives. In other words, the model is more sensitive to positive instances and has a higher probability of correctly detecting them. Finally, F1-Score combines Precision and Recall statistics using their harmonic mean, $2 \times \frac{\text{Prisicion} \times \text{Recall}}{\text{Prisicion} + \text{Recall}}$ to evaluate the balance between the model's ability to make accurate positive predictions and its ability to capture all positive instances.

We start our analysis by pre-processing the dataset to make it suitable for classification analysis. We convert the dependent variable to a categorical variable with three classes (categories), reflecting the direction of changes in future earnings. Positive (negative) class is assigned to each observation if earnings changes are positive (negative) post dividend change announcements. The class of zero is assigned to each observation if unexpected changes in the future earnings are zero. Since there are only a few observations belonging to class zero, we drop this class and train all models using only positive and negative classes with assigned labels of 1 and 0 respectively. We train each model by finding an optimal set of hyperparameters that gives the best performance.

We evaluate the performance of each model given the metrics such as accuracy, AUC, Precision, Recall, and F-score. In addition, we use feature importance scores to quantify the relative significance of each independent variable (feature) in classifying/predicting the sign of unexpected earning changes. We identify variables that have the most influence/influence on the model's predictions. We also evaluate the importance of each independent variable and its contribution to the overall performance of the model by systematically removing each variable from the model and observing the performance of the model in the absence of that variable. This approach, also known as ablation study, helps identify key components that contribute to performance of the model.

To obtain a more robust estimate of the model's performance and reduce any possible bias in model evaluation we train and evaluate models using K-fold cross-validation technique. It involves partitioning the original dataset into 10 equal-sized subsets (folds) and then training and evaluating the model 10 times, each time using one different fold as the validation set and the remaining nine folds as the training set. This technique gives a more accurate estimate of how well the model will generalize to unseen data. In particular, it could provide a more statistically significant estimate of model performance compared to a single train-test split, especially when dealing with small or imbalanced datasets.

Table 7 reports the classification results. The left panel presents the results for the Gradient Boosting Trees (XGBoost) and the right panel presents results for the Random Forests classifier adopted in this study. The performance metrics in each panel are the average scores in 10-fold cross-validation analysis. As in the regression analysis, we conduct ablation studies to measure the importance of each covariates. Panel A reports accuracy for the baseline benchmark model in which the classifiers assign the majority class label in the dataset to all observations. Panel B shows that if dividend change is the only independent variable in the model in absence of all other control variables, the accuracy of both classifiers are as low as those of the benchmark model. In other words, none of the classifiers make any better classification than the base case classifier, implying that dividend change contains no incremental information to improve the predictive power of these classifiers.

We also conduct the classification analysis with all covariates as in Ham et al. (2020). The results are reported in Panel C, including dividend change, and Panel D in the absence of dividend change. The reported statistics in Panel C shows that the accuracy of both classifiers improved extensively from 0.59 (0.64) to 0.94 (0.97) for XGBoost and RF respectively, when all covariates are present. Comparing statistics in Panels C and D, we observe that the contributions of dividend change in the accuracy of the classifiers are marginal. Table 8 reports variable (feature) importance score for each covariates with and without dividends in both classifiers. As can be seen in this table, dividend change is among the least important features in both classifiers with importance scores of 1.4%. These results are consistent with those obtained from GBRT models.

We also conduct the classification analysis with all covariates as in Ham et al. (2020). The performance of these classification analyses are reported in Panel C, including dividend change as an independent variable, and in Panel D without dividend change. Adding all control variables to the classifiers, Panel C shows that the performance of both classifiers improved extensively compared to the classifiers that only had dividend change as an independent variable. For instance, the accuracy score increased from 0.59 to 0.94 for XGBoost and from 0.60 to 0.97 for RF when all covariates are present. Comparing statistics in Panels C and D, we observe that the contributions of dividend change in the accuracy of the classifiers are marginal. The performance improvement is consistent across all different measures. We also measure the importance score for all variables in classification models to rank variables based on their influence on the model's prediction. Table 8 reports feature importance scores for each covariates with and without dividend change in both classifiers. Looking across all covariates, we observe that dividend change is among the least important features with the importance score of 1.4% in both classifiers. These results are consistent with those obtained from regression models, confirming our finding about the marginal predictive power of dividend changes on the unexpected earnings changes.

Re-write: 9 and ?? compares contribution of features wrt. feature importance and shap value. Div change seems to consistently have one of the lowest feature importance and shap values.

5 Conclusion

Miller and Modigliani (1961)' proposition that dividend changes convey managerial information about future earnings sparked active research in empirically investigating the relation between these two variables. Unfortunately, the evidence has been mixed, particularly when annual dividends and earnings are used. Quarterly dividends and earnings provide more consistent results, pointing to a positive relation between dividend changes and earnings in the next four quarters (e.g., Aharony and Dotan (1994) and Ham et al. (2020)). We extend this line of research to investigate the contribution of dividend changes to modeling and predicting earnings in the next four quarters.

Using standard regression models, we find that the contribution of dividend changes to the models of earnings in the next four quarters is marginal, increasing the R-squared by only 1.19% at best. Consistent with this, the dividend changes have minimal predictive power both in-sample and out-of-sample. Next, we use state-of-the-art machine learning models to investigate both the modeling and predictive power of the dividend changes. We build and train gradient boosting regression trees, random forest regressions, and a variety of deep neural network regressions and find that the results are no different from those from the standard regressions: dividend changes have marginal modeling and predictive power. We also confirm these results by building and training several supervised classification models. Collectively, dividend changes convey information about earnings in the subsequent four quarters, but the information is marginal in modeling or predicting these earnings.

References

Aharony, J. and Dotan, A. (1994). Regular dividend announcements and future unexpected earnings: An empirical analysis. *Financial Review*, 29(1):125–151.

Benartzi, S., Michaely, R., and Thaler, R. (1997). Do changes in dividends signal the future or the past? *The Journal of finance*, 52(3):1007–1034.

Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, 13(1):1063–1095.

Brav, A., Graham, J. R., Harvey, C. R., and Michaely, R. (2005). Payout policy in the 21st century. *Journal of financial economics*, 77(3):483–527.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Brickley, J. A. (1983). Shareholder wealth, information signaling and the specially designated dividend: An empirical study. *Journal of Financial Economics*, 12(2):187–209.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer.

Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information and computation*, 121(2):256–285.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Grullon, G., Michaely, R., Benartzi, S., and Thaler, R. H. (2005). Dividend changes do not signal changes in future profitability. *The Journal of Business*, 78(5):1659–1682.

Ham, C. G., Kaplan, Z. R., and Leary, M. T. (2020). Do dividends convey information about future earnings? *Journal of Financial Economics*, 136(2):547–570.

Handjinicolaou, G. and Kalay, A. (1984). Wealth redistributions or changes in firm value: An analysis of returns to bondholders and stockholders around dividend announcements. *Journal of Financial Economics*, 13(1):35–63.

Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. *The American Economic Review*, 46(2):97–113.

Louppe, G. (2014). Understanding random forests: From theory to practice. *arXiv preprint arXiv:1407.7502*.

Miller, M. H. and Modigliani, F. (1961). Dividend policy, growth, and the valuation of shares. *the Journal of Business*, 34(4):411–433.

Natekin, A. and Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7:21.

Nissim, D. and Ziv, A. (2001). Dividend changes and future profitability. *The Journal of Finance*, 56(6):2111–2133.

Penman, S. H. (1983). The predictive content of earnings forecasts and dividends. *The Journal of finance*, 38(4):1181–1199.

Schapire, R. E. (2003). The boosting approach to machine learning: An overview. *Nonlinear estimation and classification*, pages 149–171.

Watts, R. (1973). The information content of dividends. *The Journal of Business*, 46(2):191–211.

Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*. CRC press.

Tables and Figures

Table 1: Summary Statistics

	$\Delta DIV = 0$		$\Delta DIV < 0$		$\Delta DIV > 0$	
	Mean	Median	Mean	Median	Mean	Median
ΔDIV	0.0000	0.0000	-0.4346	-0.4974	0.1838	0.1111
$Ret_{(-2,-20)}$	-0.0002	-0.0043	-0.0173	-0.0189	0.0065	0.0017
$Ret_{(-21,-40)}$	0.0001	-0.0035	-0.0137	-0.0172	0.0043	0.0009
$Ret_{(-41,-60)}$	0.0029	-0.0010	-0.0100	-0.0106	0.0069	0.0019
$Ret_{(-61,-120)}$	0.0032	-0.0050	-0.0349	-0.0418	0.0144	0.0046
$Ret_{(-121,-240)}$	0.0093	-0.0093	-0.0239	-0.0442	0.0389	0.0153
N	140,042		1,975		23,541	

Note: The table reports mean and median values for dividend changes, the earnings, and the return realizations around earning announcements separately for positive, negative, and no dividends change. ΔDIV is current quarterly dividend less the prior quarterly dividend divided by the prior quarterly dividend. $Ret_{(-2,-20)}$ is daily compounded returns from twenty to two trading days before the dividend declaration less the daily compounded return to the value-weighted market portfolio over the same period. $E_{(y-1)}$ is the sum of the four quarterly earnings before dividend announcements.

Table 2: Dividend Changes and Future Earning Changes

	With ΔDIV			Without ΔDIV		
	$\Delta E_{(Y+1)}$	$\Delta E_{(Y+2)}$	$\Delta E_{(Y+3)}$	$\Delta E_{(Y+1)}$	$\Delta E_{(Y+2)}$	$\Delta E_{(Y+3)}$
ΔDIV	0.0255***	0.0182**	0.0183*			
	(5.13)	(3.18)	(2.67)			
$Ret_{(-2,-20)}$	0.0861***	0.1001***	0.0952***	0.0874***	0.1011***	0.0961***
	(15.30)	(10.81)	(9.62)	(15.46)	(10.83)	(9.61)
$Ret_{(-21,-40)}$	0.0778***	0.0907***	0.0940***	0.0787***	0.0913***	0.0947***
	(11.45)	(10.51)	(9.14)	(11.53)	(10.54)	(9.17)
$Ret_{(-41,-60)}$	0.0744***	0.0841***	0.0792***	0.0752***	0.0847***	0.0797***
	(11.62)	(9.36)	(8.26)	(11.64)	(9.35)	(8.25)
$Ret_{(-61,-120)}$	0.0626***	0.0697***	0.0714***	0.0632***	0.0701***	0.0719***
	(11.20)	(10.06)	(9.51)	(11.21)	(10.07)	(9.49)
$Ret_{(-121,-240)}$	0.0328***	0.0373***	0.0404***	0.0332***	0.0376***	0.0407***
	(8.26)	(8.16)	(8.90)	(8.25)	(8.16)	(8.90)
$E_{(q-1)}$	0.4121***	0.3415***	0.3324***	0.4189***	0.3464***	0.3373***
	(13.16)	(4.74)	(3.73)	(13.38)	(4.80)	(3.76)
$E_{(q-2)}$	0.0664	0.0789	0.1278	0.0669	0.0792	0.1282
	(1.78)	(1.10)	(1.36)	(1.79)	(1.10)	(1.37)
$E_{(q-3)}$	0.0214	0.0301	0.0603	0.0209	0.0299	0.0602
	(0.56)	(0.41)	(0.64)	(0.55)	(0.41)	(0.64)
$E_{(q-4)}$	-0.0987*	-0.0611	0.0072	-0.1007*	-0.0626	0.0056
	(-2.05)	(-0.81)	(0.07)	(-2.08)	(-0.83)	(0.05)
$\Delta E_{(q-1)}$	0.4521***	0.2981***	0.2567**	0.4570***	0.3016***	0.2599**
	(8.53)	(4.27)	(3.22)	(8.45)	(4.28)	(3.24)
$\Delta E_{(q-2)}$	0.1613***	0.0781	0.0392	0.1644***	0.0803	0.0410
	(4.39)	(1.19)	(0.49)	(4.43)	(1.22)	(0.51)
$\Delta E_{(q-3)}$	0.0745	0.0175	-0.0469	0.0761	0.0185	-0.0459
	(1.56)	(0.23)	(-0.56)	(1.57)	(0.24)	(-0.55)
$\Delta E_{(q-4)}$	0.0355	-0.0346	-0.1593	0.0394	-0.0318	-0.1568
	(0.67)	(-0.42)	(-1.36)	(0.73)	(-0.38)	(-1.34)
$E_{(y-1)}1_{-}$	-0.9134***	-1.2856***	-1.1340***	-0.9093***	-1.2831***	-1.1304***
	(-7.95)	(-6.75)	(-4.91)	(-7.92)	(-6.74)	(-4.91)
$E^2_{(y-1)}1_{-}$	0.4281	-0.4457	2.7144	0.4259	-0.4506	2.7180
	(0.46)	(-0.32)	(1.56)	(0.46)	(-0.32)	(1.56)
$E^2_{(y-1)}1_{+}$	-0.4650***	-0.3625	-0.3088	-0.4650***	-0.3625	-0.3086
	(-4.21)	(-1.97)	(-0.96)	(-4.18)	(-1.97)	(-0.96)
$\Delta E_{(y-1)}1_{-}$	-0.3942***	-0.3125***	-0.3976***	-0.3875***	-0.3075***	-0.3919***
	(-7.55)	(-4.17)	(-4.31)	(-7.37)	(-4.10)	(-4.25)
$\Delta E^2_{(y-1)}1_{-}$	-0.3989	-0.1843	-0.5319	-0.3666	-0.1604	-0.5053
	(-1.65)	(-0.71)	(-1.80)	(-1.51)	(-0.62)	(-1.71)
$\Delta E^2_{(y-1)}1_{+}$	-0.9171***	-0.9624***	-0.8431***	-0.9202***	-0.9637***	-0.8435***
	(-5.73)	(-5.44)	(-4.65)	(-5.68)	(-5.41)	(-4.63)
Constant	-0.0068**	0.0000	0.0042	-0.0063*	0.0003	0.0045
	(-2.74)	(0.00)	(0.82)	(-2.55)	(0.08)	(0.88)
Observations	165558	154945	145042	165558	154945	145042
R^2	0.1866	0.1137	0.0928	0.1844	0.1132	0.0924
AIC	-439292	-287935	-206212	-438850	-287843	-206159
BIC	-439081	-287726	-206005	-438650	-287644	-205962

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Dividend Changes and Future Earning Changes

	$\Delta E_{(Y+1)}$	$\Delta E_{(Y+2)}$	$\Delta E_{(Y+3)}$
ΔDIV	0.0309***	0.0199*	0.0177*
	(4.8521)	(2.6698)	(2.0712)
Constant	0.0054*	0.0126***	0.0217***
	(2.6461)	(3.5530)	(4.4272)
Observations	165558	154945	145042
R^2	0.0033	0.0007	0.0003
Adjusted R^2	0.0033	0.0006	0.0003
AIC	-405678	-269371	-192174
BIC	-405657	-269351	-192154

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Percentile Rank Dividend Changes and Future Earning Changes

	With Div			Without Div		
	$\Delta E_{(Y+1)}$	$\Delta E_{(Y+2)}$	$\Delta E_{(Y+3)}$	$\Delta E_{(Y+1)}$	$\Delta E_{(Y+2)}$	$\Delta E_{(Y+3)}$
rank($\Delta DIV > 0$)	0.0182***	0.0172***	0.0168***			
	(10.13)	(5.77)	(4.25)			
rank($\Delta DIV < 0$)	-0.0431***	-0.0181*	-0.0205*			
	(-5.46)	(-2.26)	(-2.21)			
$Ret_{(-2,-20)}$	0.0854***	0.0996***	0.0946***	0.0874***	0.1011***	0.0961***
	(15.21)	(10.75)	(9.57)	(15.46)	(10.83)	(9.61)
$Ret_{(-21,-40)}$	0.0773***	0.0903***	0.0936***	0.0787***	0.0913***	0.0947***
	(11.41)	(10.48)	(9.12)	(11.53)	(10.54)	(9.17)
$Ret_{(-41,-60)}$	0.0738***	0.0837***	0.0788***	0.0752***	0.0847***	0.0797***
	(11.61)	(9.34)	(8.26)	(11.64)	(9.35)	(8.25)
$Ret_{(-61,-120)}$	0.0622***	0.0694***	0.0711***	0.0632***	0.0701***	0.0719***
	(11.23)	(10.08)	(9.52)	(11.21)	(10.07)	(9.49)
$Ret_{(-121,-240)}$	0.0326***	0.0371***	0.0402***	0.0332***	0.0376***	0.0407***
	(8.28)	(8.17)	(8.89)	(8.25)	(8.16)	(8.90)
$E_{(q-1)}$	0.4043***	0.3359***	0.3265***	0.4189***	0.3464***	0.3373***
	(12.88)	(4.67)	(3.68)	(13.38)	(4.80)	(3.76)
$E_{(q-2)}$	0.0622	0.0760	0.1247	0.0669	0.0792	0.1282
	(1.67)	(1.06)	(1.33)	(1.79)	(1.10)	(1.37)
$E_{(q-3)}$	0.0193	0.0301	0.0599	0.0209	0.0299	0.0602
	(0.51)	(0.41)	(0.64)	(0.55)	(0.41)	(0.64)
$E_{(q-4)}$	-0.1006*	-0.0619	0.0063	-0.1007*	-0.0626	0.0056
	(-2.09)	(-0.82)	(0.06)	(-2.08)	(-0.83)	(0.05)
$\Delta E_{(q-1)}$	0.4512***	0.2968***	0.2552**	0.4570***	0.3016***	0.2599**
	(8.57)	(4.28)	(3.21)	(8.45)	(4.28)	(3.24)
$\Delta E_{(q-2)}$	0.1620***	0.0775	0.0386	0.1644***	0.0803	0.0410
	(4.39)	(1.18)	(0.48)	(4.43)	(1.22)	(0.51)
$\Delta E_{(q-3)}$	0.0742	0.0159	-0.0483	0.0761	0.0185	-0.0459
	(1.54)	(0.21)	(-0.58)	(1.57)	(0.24)	(-0.55)
$\Delta E_{(q-4)}$	0.0358	-0.0354	-0.1600	0.0394	-0.0318	-0.1568
	(0.67)	(-0.43)	(-1.36)	(0.73)	(-0.38)	(-1.34)
$E_{(y-1)}1_{-}$	-0.9121***	-1.2838***	-1.1319***	-0.9093***	-1.2831***	-1.1304***
	(-7.97)	(-6.77)	(-4.92)	(-7.92)	(-6.74)	(-4.91)
$E^2_{(y-1)}1_{-}$	0.4390	-0.4321	2.7291	0.4259	-0.4506	2.7180
	(0.47)	(-0.31)	(1.57)	(0.46)	(-0.32)	(1.56)
$E^2_{(y-1)}1_{+}$	-0.4620***	-0.3640	-0.3093	-0.4650***	-0.3625	-0.3086
	(-4.18)	(-1.98)	(-0.97)	(-4.18)	(-1.97)	(-0.96)
$\Delta E_{(y-1)}1_{-}$	-0.4067***	-0.3222***	-0.4071***	-0.3875***	-0.3075***	-0.3919***
	(-7.78)	(-4.30)	(-4.43)	(-7.37)	(-4.10)	(-4.25)
$\Delta E^2_{(y-1)}1_{-}$	-0.4427	-0.2257	-0.5717	-0.3666	-0.1604	-0.5053
	(-1.83)	(-0.87)	(-1.93)	(-1.51)	(-0.62)	(-1.71)
$\Delta E^2_{(y-1)}1_{+}$	-0.9023***	-0.9491***	-0.8299***	-0.9202***	-0.9637***	-0.8435***
	(-5.63)	(-5.36)	(-4.56)	(-5.68)	(-5.41)	(-4.63)
Constant	-0.0071**	-0.0006	0.0037	-0.0063*	0.0003	0.0045
	(-2.84)	(-0.14)	(0.71)	(-2.55)	(0.08)	(0.88)
Observations	165558	154945	145042	165558	154945	145042
R^2	0.1885	0.1145	0.0933	0.1844	0.1132	0.0924
AIC	-439676	-288066	-206289	-438850	-287843	-206159
BIC	-439456	-287847	-206071	-438650	-287644	-205962

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Percentile Rank Dividend Changes and Future Gross Profit Changes

	With Div			Without Div		
	$\Delta GP_{(Y+1)}$	$\Delta GP_{(Y+2)}$	$\Delta GP_{(Y+3)}$	$\Delta GP_{(Y+1)}$	$\Delta GP_{(Y+2)}$	$\Delta GP_{(Y+3)}$
rank($\Delta DIV > 0$)	0.0132***	0.0203***	0.0244***			
	(5.82)	(4.60)	(3.59)			
rank($\Delta DIV < 0$)	-0.0204***	-0.0306***	-0.0422**			
	(-3.58)	(-3.67)	(-3.42)			
$Ret_{(-2,-20)}$	0.0687***	0.0973***	0.1247***	0.0699***	0.0993***	0.1271***
	(11.08)	(7.53)	(8.26)	(11.34)	(7.73)	(8.47)
$Ret_{(-21,-40)}$	0.0685***	0.1112***	0.1396***	0.0696***	0.1128***	0.1415***
	(10.10)	(9.43)	(9.74)	(10.29)	(9.57)	(9.89)
$Ret_{(-41,-60)}$	0.0561***	0.0916***	0.1113***	0.0571***	0.0929***	0.1130***
	(6.87)	(5.90)	(6.08)	(6.97)	(5.96)	(6.11)
$Ret_{(-61,-120)}$	0.0460***	0.0735***	0.1019***	0.0468***	0.0746***	0.1033***
	(7.64)	(7.08)	(8.00)	(7.72)	(7.15)	(8.04)
$Ret_{(-121,-240)}$	0.0175***	0.0335***	0.0569***	0.0181***	0.0343***	0.0579***
	(3.80)	(4.89)	(6.20)	(3.89)	(4.98)	(6.28)
$GP_{(q-1)}$	0.2814***	0.3623***	0.4806***	0.2870***	0.3712***	0.4917***
	(17.71)	(8.72)	(8.34)	(17.81)	(8.80)	(8.50)
$GP_{(q-2)}$	-0.0244	-0.0221	0.0271	-0.0227	-0.0192	0.0312
	(-1.54)	(-0.95)	(0.65)	(-1.43)	(-0.82)	(0.74)
$GP_{(q-3)}$	-0.0440*	-0.0339	-0.0192	-0.0455*	-0.0364	-0.0222
	(-2.29)	(-0.99)	(-0.43)	(-2.34)	(-1.05)	(-0.50)
$GP_{(q-4)}$	-0.1785***	-0.1640***	-0.1690***	-0.1780***	-0.1636***	-0.1693***
	(-12.37)	(-5.34)	(-3.75)	(-12.25)	(-5.29)	(-3.73)
$\Delta GP_{(q-1)}$	1.3415***	1.4852***	1.7118***	1.3463***	1.4920***	1.7202***
	(37.30)	(21.99)	(16.68)	(36.79)	(21.89)	(16.66)
$\Delta GP_{(q-2)}$	0.0587	0.2217***	0.2972**	0.0599	0.2237***	0.2995**
	(1.58)	(3.63)	(3.37)	(1.61)	(3.66)	(3.39)
$\Delta GP_{(q-3)}$	-0.2176***	-0.0787	0.0304	-0.2148***	-0.0740	0.0357
	(-5.95)	(-1.13)	(0.27)	(-5.85)	(-1.06)	(0.32)
$\Delta GP_{(q-4)}$	0.1762***	0.4717***	0.7095***	0.1811***	0.4793***	0.7186***
	(5.07)	(6.85)	(7.39)	(5.16)	(6.88)	(7.41)
$GP_{(y-1)}1_-$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(.)	(.)	(.)	(.)	(.)	(.)
$GP^2_{(y-1)}1_-$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(.)	(.)	(.)	(.)	(.)	(.)
$GP^2_{(y-1)}1_+$	0.0357***	0.0713***	0.0989***	0.0346***	0.0695***	0.0968***
	(5.20)	(4.65)	(4.13)	(5.07)	(4.53)	(4.04)
$\Delta GP_{(y-1)}1_-$	-0.2081***	-0.4587***	-0.7280***	-0.1953***	-0.4398***	-0.7034***
	(-4.27)	(-4.22)	(-4.78)	(-3.88)	(-4.03)	(-4.63)
$\Delta GP^2_{(y-1)}1_-$	1.8078***	2.7158***	2.9671***	1.8882***	2.8379***	3.1195***
	(4.22)	(4.19)	(3.56)	(4.33)	(4.36)	(3.76)
$\Delta GP^2_{(y-1)}1_+$	-0.1736**	-0.4113***	-0.5642**	-0.1782**	-0.4183***	-0.5725**
	(-3.15)	(-3.89)	(-2.94)	(-3.20)	(-3.93)	(-2.97)
Constant	0.0050***	0.0105***	0.0150***	0.0053***	0.0110***	0.0156***
	(4.28)	(4.17)	(4.01)	(4.56)	(4.35)	(4.09)
Observations	134439	124683	115986	134439	124683	115986
R^2	0.2902	0.1994	0.1945	0.2890	0.1985	0.1938
AIC	-331422	-148327	-55913	-331189	-148187	-55818
BIC	-331226	-148132	-55720	-331012	-148012	-55644

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Machine Learning Regressions vs OLS Regression

	With ΔDIV			Without ΔDIV		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
Panel A: All Variables						
GBT (OOS)	0.2702	0.2093	0.1723	0.2662	0.2110	0.1750
RF (OOS)	0.1808	0.1808	0.1471	0.1808	0.1590	0.1471
GBT (IS)	0.5161	0.4526	0.4263	0.5188	0.5188	0.4232
RF (IS)	0.1971	0.1741	0.1645	0.1971	0.1741	0.1645
OLS (IS)	0.1866	0.1137	0.0928	0.1844	0.1132	0.0924
Panel B: Without Interaction Variables						
GBT (OOS)	0.2441	0.1826	0.1463	0.2406	0.1874	0.1491
RF (OOS)	0.1266	0.1111	0.1044	0.1267	0.1112	0.1044
GBT (IS)	0.4976	0.4335	0.4038	0.4980	0.4335	0.4044
RF (IS)	0.1466	0.1276	0.1217	0.1466	0.1276	0.1217
OLS (IS)	0.1215	0.0682	0.0518	0.1198	0.0679	0.0515

Note: The table reports the R-squared for Gradient Boosting Trees (GBT) regressions, Random Forests (RF) regressions, and Ordinary Least Squares (OLS) regressions. The regressions evaluate the relation between dividend changes (ΔDIV) and unexpected changes in future earnings in three years post dividend announcements. Panel A shows regression results with all variables as in Ham et al. (2020). Panel B shows regression results without interaction variables. The left (right) panel reports results for model 1 (2) with (without) ΔDIV as a covariate. Out-of-sample (OOS) R-squared refers to the average R-squared using 10-fold cross-validation.

Table 7: Evaluation Results Classification Models

	XGBoost			Random Forests		
	Class 0	Class 1	W-Avg	Class 0	Class 1	W-Avg
Panel A: Baseline Majority Class (ZeroR Classifier)						
Accuracy		0.5933			0.6398	
Panel B: Dividends Only - One Feature						
Accuracy		0.6007			0.6007	
AUC		0.54 +/-0.01			0.54 +/-0.01	
Precision	0.71	0.6	0.64	0.69	0.6	0.64
Recall	0.03	0.99	0.60	0.03	0.99	0.60
F1 Score	0.06	0.75	0.47	0.06	0.75	0.47
Support	24,890	36,315	61,205	24,890	36,315	61,205
Panle C: Full Model with Ablation Study						
Accuracy		0.9384			0.9688	
AUC		0.96 +/-0.11			0.96 +/-0.11	
Precision	0.93	0.94	0.94	0.97	0.97	0.97
Recall	0.91	0.96	0.94	0.95	0.98	0.97
F1 Score	0.92	0.95	0.94	0.96	0.97	0.97
Support	24,890	36,315	61,205	24,890	36,315	61,205
Panel D: Full Model without Ablation Study						
Accuracy		0.9373			0.9690	
AUC		0.96 +/-0.11			0.96 +/-0.11	
Percision	0.94	0.94	0.94	0.97	0.97	0.97
Recall	0.91	0.96	0.94	0.95	0.98	0.97
F1 Score	0.92	0.95	0.94	0.96	0.97	0.97
Support	24,890	36,315	61,205	24,890	36,315	61,205

Note: The table reports the performance of Gradient Boosting Tree classifier (XGBoost) and Random Forests Classifier. Panel A reports the accuracy of ZeroR classifier, in which the model assign the majority category to each observation. Panel B reports model performance using Accuracy, Precision, Recall, and F1-Score for the model that only contains ΔDIV. Panel C (D) reports similar statistics for the full model including all covariates with (without) ablation study.

Table 8: Feature Importance Classification Models

XGBoost				Random Forest			
W/ Ablation		W/out Ablation		W/ Ablation		W/out Ablation	
$\Delta E_{(q-1)}$	0.113	$\Delta E_{(q-1)}$	0.110	$\Delta E_{(q-1)}$	0.080	$\Delta E_{(q-1)}$	0.082
$E^2_{(y-1)}1_+$	0.089	$E^2_{(y-1)}1_+$	0.098	$ret_{(-61,-120)}$	0.068	$ret_{(-61,-120)}$	0.069
$E_{(q-1)}$	0.068	$ret_{(-61,-120)}$	0.065	$ret_{(-121,-240)}$	0.067	$ret_{(-121,-240)}$	0.068
$ret_{(-2,-20)}$	0.068	$E_{(q-4)}$	0.065	$ret_{(-2,-20)}$	0.065	$ret_{(-2,-20)}$	0.066
$ret_{(-61,-120)}$	0.066	$ret_{(-2,-20)}$	0.065	$E_{(q-4)}$	0.065	$E_{(q-4)}$	0.066
$E_{(q-4)}$	0.064	$ret_{(-121,-240)}$	0.065	$ret_{(-21,-40)}$	0.064	$ret_{(-21,-40)}$	0.065
$ret_{(-121,-240)}$	0.063	$E_{(q-1)}$	0.064	$E_{(q-1)}$	0.063	$E_{(q-1)}$	0.064
$E_{(q-2)}$	0.061	$\Delta E_{(q-2)}$	0.059	$E^2_{(y-1)}1_+$	0.063	$E^2_{(y-1)}1_+$	0.064
$\Delta E_{(q-2)}$	0.058	$E_{(q-2)}$	0.059	$ret_{(-41,-60)}$	0.062	$ret_{(-41,-60)}$	0.063
$ret_{(-41,-60)}$	0.056	$\Delta E_{(q-4)}$	0.059	$E_{(q-3)}$	0.060	$E_{(q-3)}$	0.061
$\Delta E_{(q-4)}$	0.056	$\Delta E_{(q-4)}$	0.058	$\Delta E_{(q-2)}$	0.059	$\Delta E_{(q-2)}$	0.060
$E_{(q-3)}$	0.054	$ret_{(-41,-60)}$	0.055	$E_{(q-2)}$	0.059	$E_{(q-2)}$	0.060
$\Delta E_{(q-4)}$	0.053	$ret_{(-21,-40)}$	0.055	$\Delta E_{(q-4)}$	0.059	$\Delta E_{(q-4)}$	0.059
$ret_{(-21,-40)}$	0.053	$E_{(q-3)}$	0.054	$\Delta E_{(q-4)}$	0.057	$\Delta E_{(q-4)}$	0.058
$\Delta E^2_{(y-1)}1_+$	0.032	$\Delta E^2_{(y-1)}1_+$	0.036	$\Delta E^2_{(y-1)}1_+$	0.038	$\Delta E^2_{(y-1)}1_+$	0.038
ΔDiv	0.014	$\Delta E^2_{(y-1)}1_-$	0.015	$\Delta E_{(y-1)}1_-$	0.022	$\Delta E^2_{(y-1)}1_-$	0.022
$\Delta E^2_{(y-1)}1_-$	0.012	$\Delta E_{(y-1)}1_-$	0.008	$E^2_{(y-1)}1_-$	0.022	$\Delta E_{(y-1)}1_-$	0.021
$\Delta E_{(y-1)}1_-$	0.009	$E^2_{(y-1)}1_-$	0.008	ΔDiv	0.014	$E_{(y-1)}1_-$	0.007
$E^2_{(y-1)}1_-$	0.007	$E_{(y-1)}1_-$	0.003	$E^2_{(y-1)}1_-$	0.006	$E^2_{(y-1)}1_-$	0.006
$E_{(y-1)}1_-$	0.004			$E_{(y-1)}1_-$	0.006		

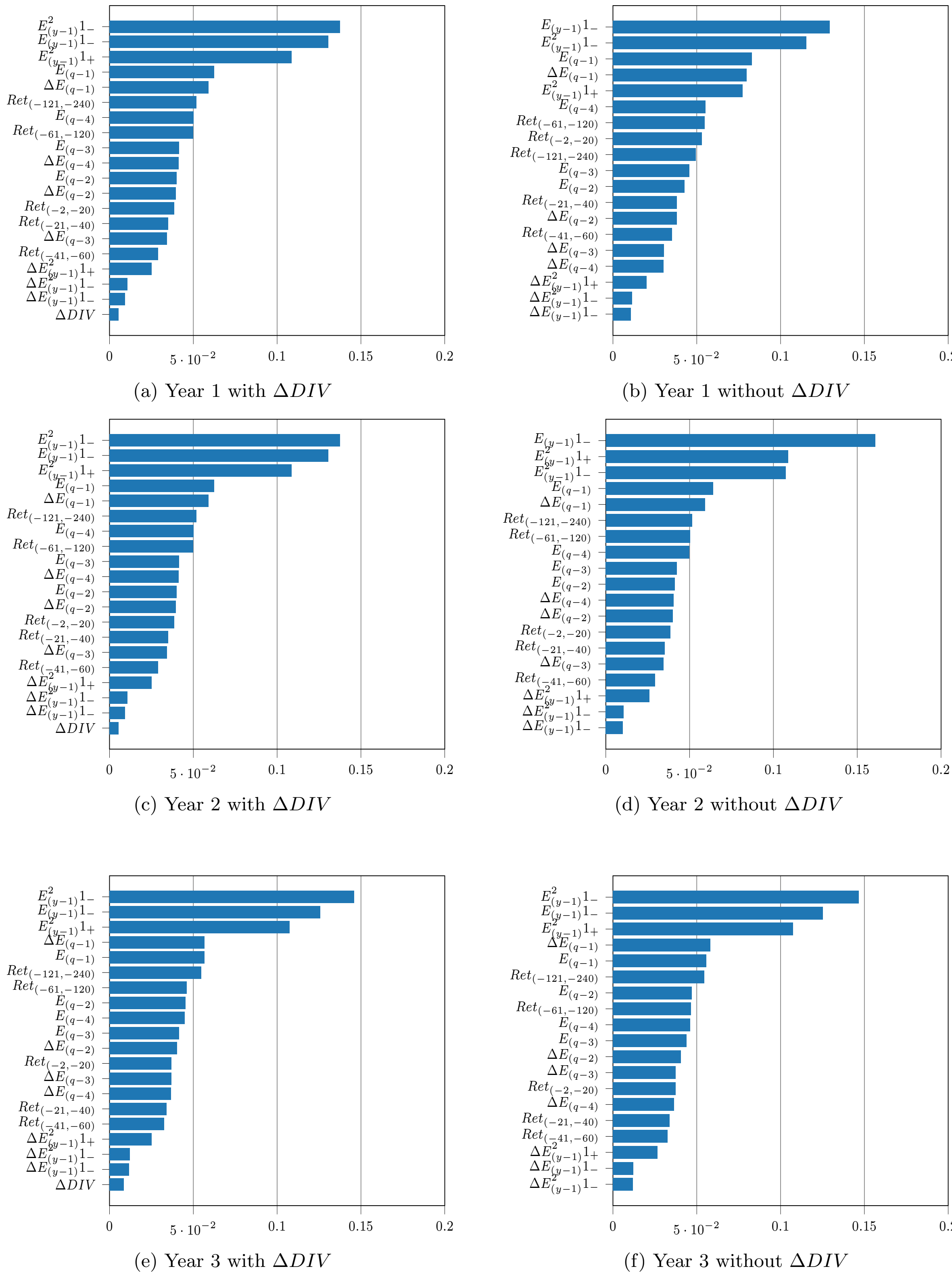
Note: The table reports feature (variable) importance scores for a Gradient Boosting Trees (XGBoost) and a Random Forest (RF) classifiers. Both models are trained with all the covariates as in Ham et al. (2020).

Table 9: VIF in Earnings Prediction

	With ΔDIV			Without ΔDIV		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
$E_{(y-1)}1_{-}$	19.77	19.76	19.45	19.77	19.76	19.45
$E^2_{(y-1)}1_{-}$	17.29	17.27	17.01	17.29	17.27	17.01
$\Delta E_{(y-1)}1_{-}$	14.14	14.08	14.04	14.13	14.07	14.04
$\Delta E^2_{(y-1)}1_{-}$	9.51	9.48	9.47	9.51	9.47	9.46
$E^2_{(y-1)}1_{+}$	7.71	7.85	7.95	7.71	7.85	7.95
$E_{(q-1)}$	3.16	3.18	3.18	3.16	3.17	3.18
$E_{(q-2)}$	3.11	3.13	3.15	3.11	3.13	3.15
$\Delta E^2_{(y-1)}1_{+}$	2.93	2.95	2.96	2.93	2.95	2.96
$E_{(q-3)}$	2.88	2.9	2.91	2.88	2.9	2.91
$E_{(q-4)}$	2.76	2.78	2.79	2.76	2.78	2.79
$\Delta E_{(q-2)}$	2.62	2.64	2.66	2.62	2.64	2.66
$\Delta E_{(q-1)}$	2.58	2.59	2.6	2.58	2.58	2.6
$\Delta E_{(q-3)}$	2.54	2.56	2.58	2.54	2.56	2.58
$\Delta E_{(q-4)}$	2.4	2.42	2.42	2.4	2.42	2.42
$Ret_{(-121,-240)}$	1.15	1.15	1.15	1.15	1.15	1.15
$Ret_{(-61,-120)}$	1.06	1.06	1.06	1.06	1.06	1.06
$Ret_{(-41,-60)}$	1.02	1.02	1.02	1.02	1.02	1.02
ΔDIV	1.02	1.02	1.02			
$Ret_{(-21,-40)}$	1.02	1.02	1.02	1.02	1.02	1.02
$Ret_{(-2,-20)}$	1.02	1.02	1.02	1.02	1.02	1.02
Mean	4.98	4.99	4.97	5.19	5.2	5.18

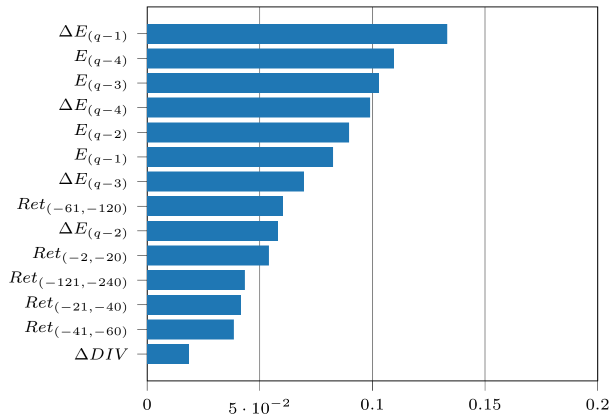
Note: The table presents Variance Inflation Factor (VIF) in regression of earning changes in the upcoming three years post earning announcement on several independent variables. The initial OLS regression is reported in Table 2. VIF quantify the extent to which the variance (or standard error) of the estimated regression coefficient is inflated due to collinearity. VIF_i for each independent variable is computed as $VIF_i = 1/(1 - R_i^2)$, where R_i^2 is the unadjusted coefficient of determination for regressing the i$^{\text{th}}$ independent variable on the remaining ones. If R^2 is equal to 0, the variance of the remaining independent variables cannot be predicted from the i$^{\text{th}}$ independent variable. Thus, when VIF is equal to 1, the i$^{\text{th}}$ independent variable is not correlated to the remaining ones, which means multicollinearity does not exist in this regression model.

Figure 1: Feature Importance in Gradient Boosting Regression Trees OOS

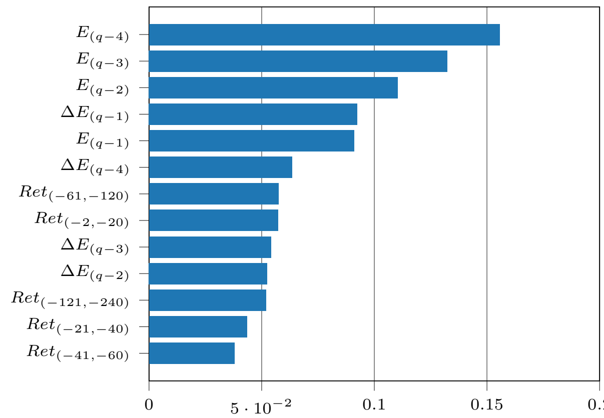


Note: The figure reports the out-of-sample feature (variable) importance for all the covariates in gradient boosting regression trees. Feature importance are sorted based on the most to the least important variables in predicting one-year, two-year and three-year ahead earning changes with (the left panel) and without (the right panel) dividend changes as a covariate. Feature importance sums to one. The overall performance of the model is measured by R^2.

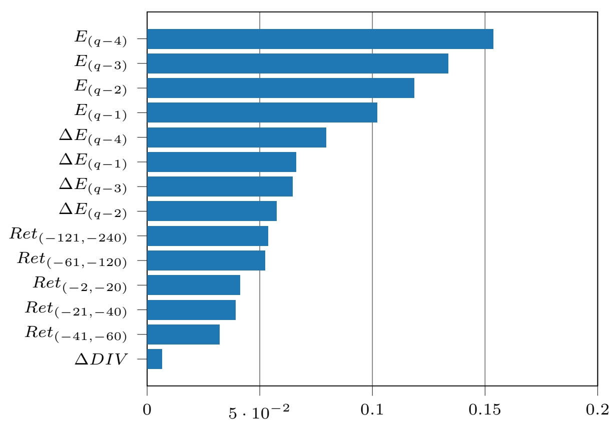
Figure 2: Feature Importance in Gradient Boosting Regression Trees without Interaction Variables OOS



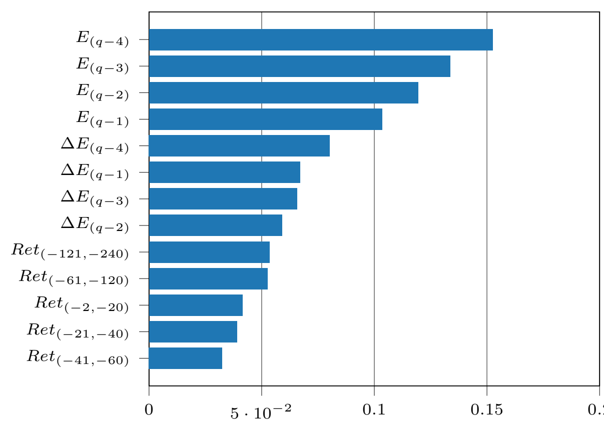
(a) Year 1 with ΔDIV



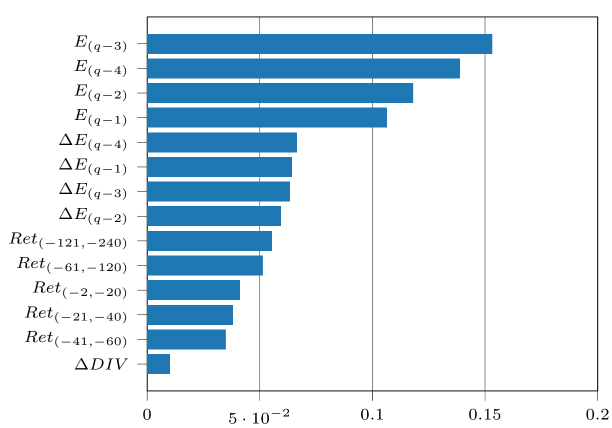
(b) Year 1 without ΔDIV



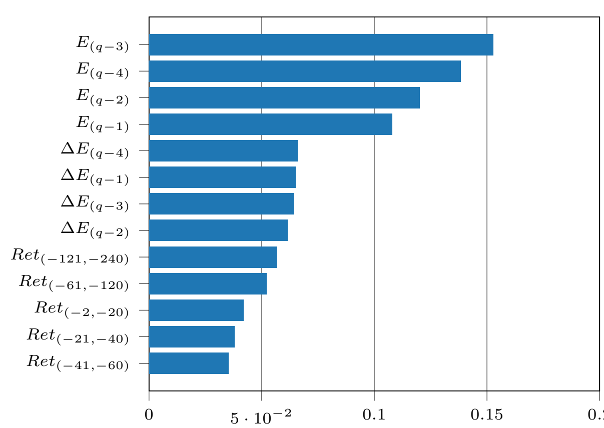
(c) Year 2 with ΔDIV



(d) Year 2 without ΔDIV



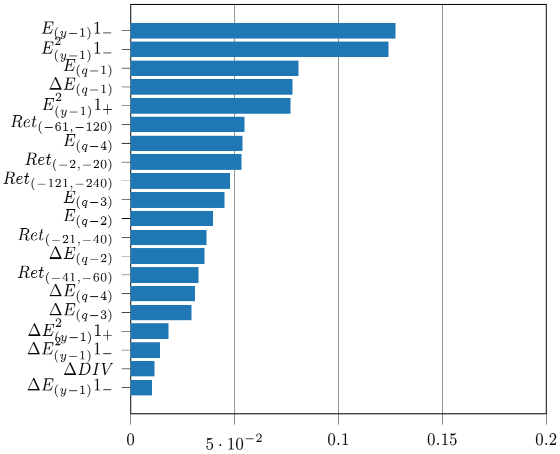
(e) Year 3 with ΔDIV



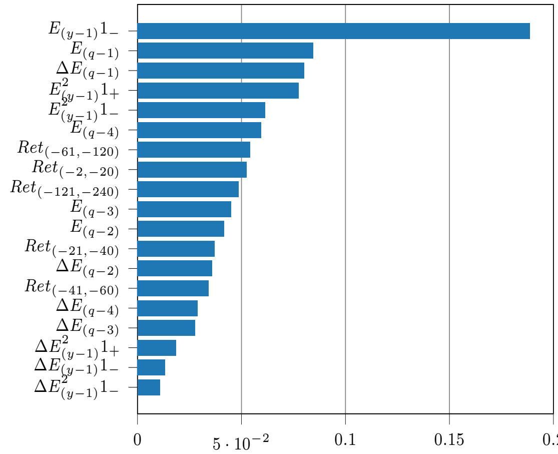
(f) Year 3 without ΔDIV

Note: The figure reports the out-of-sample feature (variable) importance for all but the control covariates in gradient boosting regression trees. Feature importance are sorted based on the most to the least important variables in predicting one-year, two-year and three-year ahead earnings with (the left panel) and without (the right panel) dividend changes as a covariate. Feature importance sums to one. The overall performance of the model is measured by R^2.

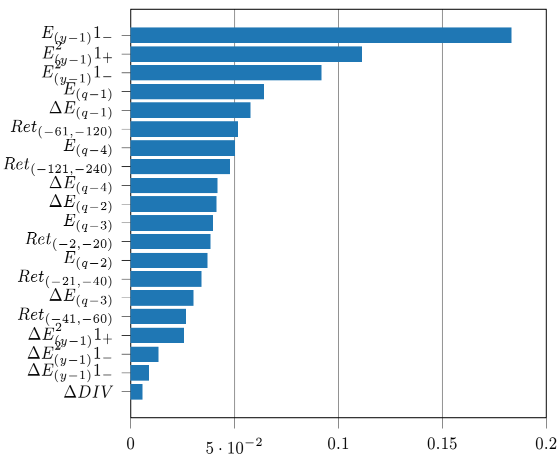
Figure 3: Feature Importance in Gradient Boosting Regression Trees IS



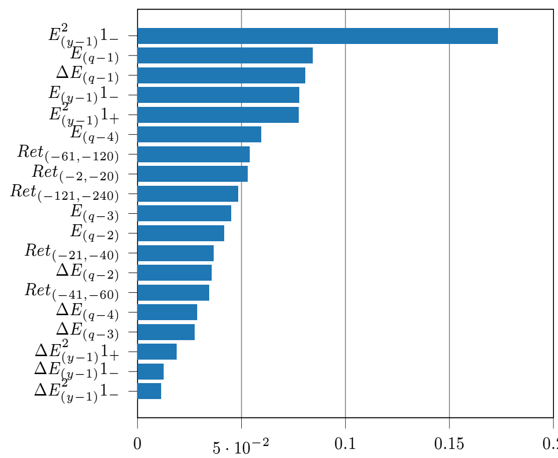
(a) Year 1 with ΔDIV



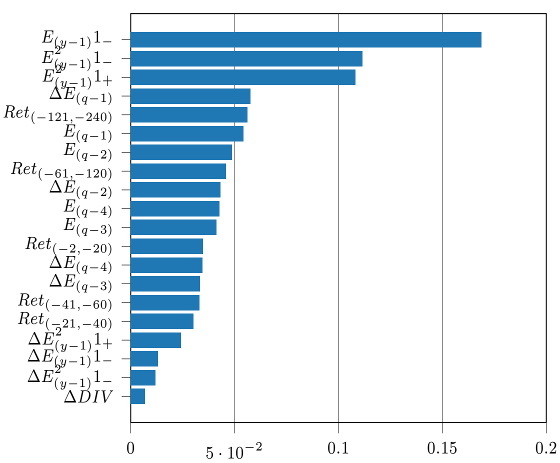
(b) Year 1 without ΔDIV



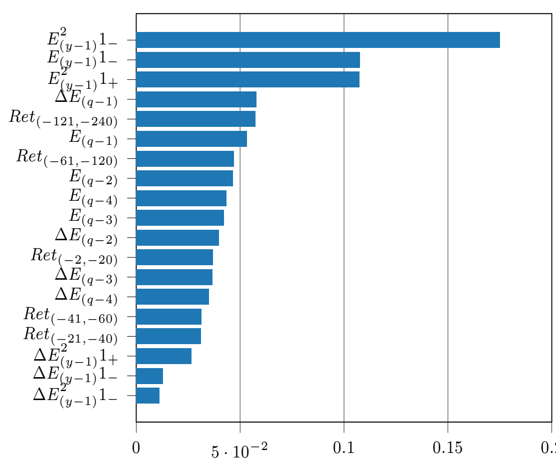
(c) Year 2 with ΔDIV



(d) Year 2 without ΔDIV



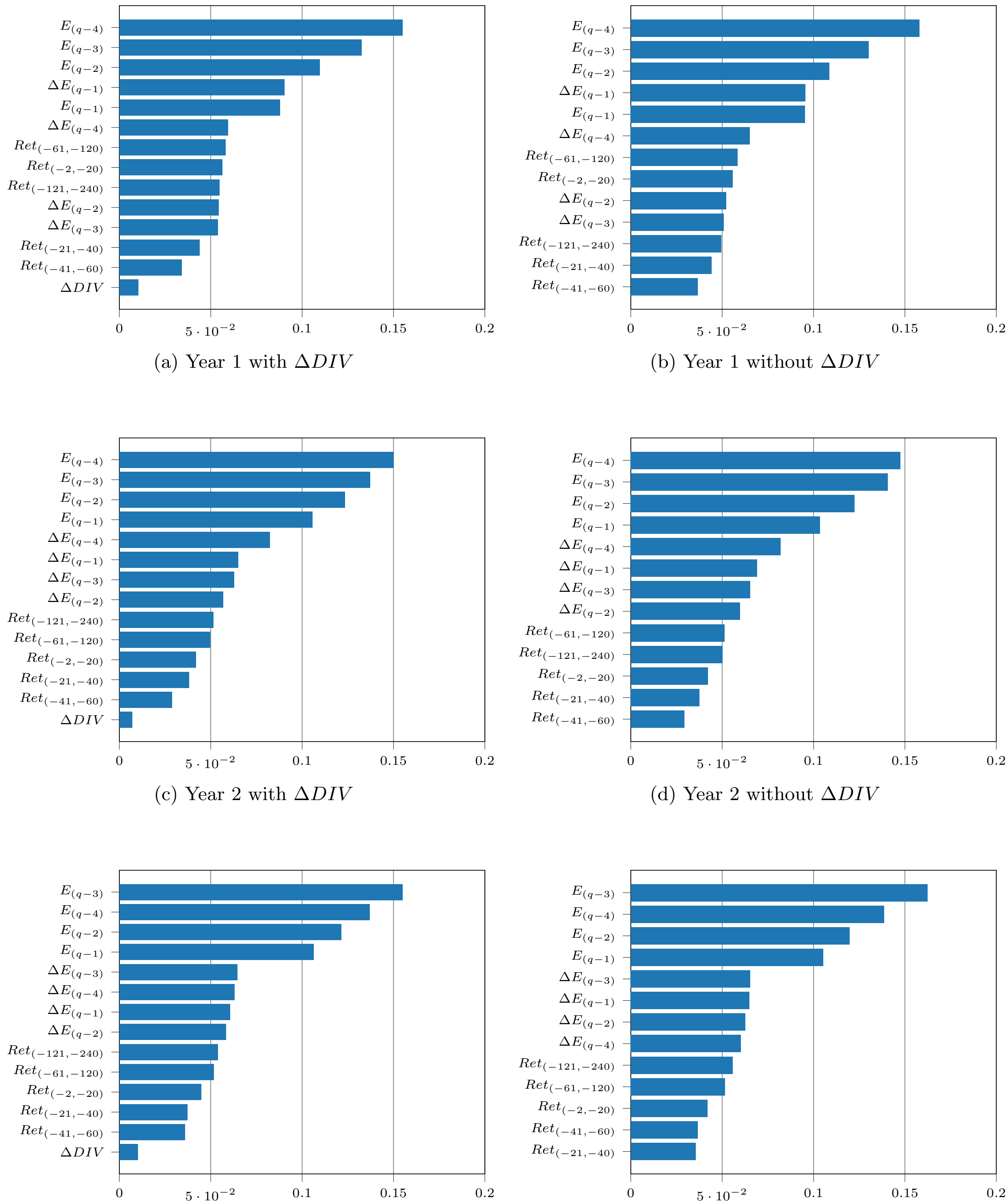
(e) Year 3 with ΔDIV



(f) Year 3 without ΔDIV

Note: The figure reports the in-sample feature (variable) importance for all the covariates in gradient boosting regression trees. Feature importance are sorted based on the most to the least important variables in predicting one-year, two-year and three-year ahead earning changes with (the left panel) and without (the right panel) dividend changes as a covariate. Feature importance sums to one. The overall performance of the model is measured by R^2.

Figure 4: Feature Importance in Gradient Boosting Regression Trees without Interaction Variables IS



(a) Year 1 with ΔDIV

(b) Year 1 without ΔDIV

(c) Year 2 with ΔDIV

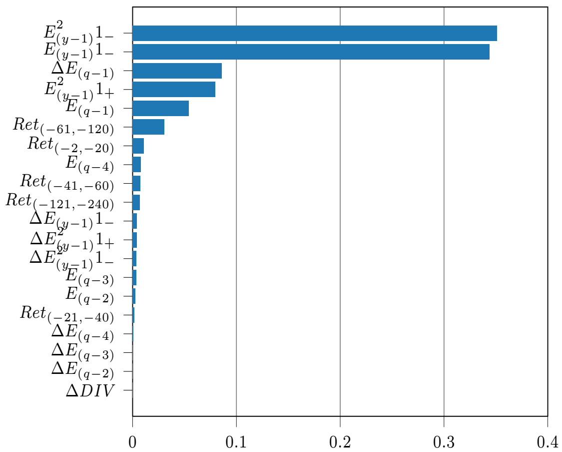
(d) Year 2 without ΔDIV

(e) Year 3 with ΔDIV

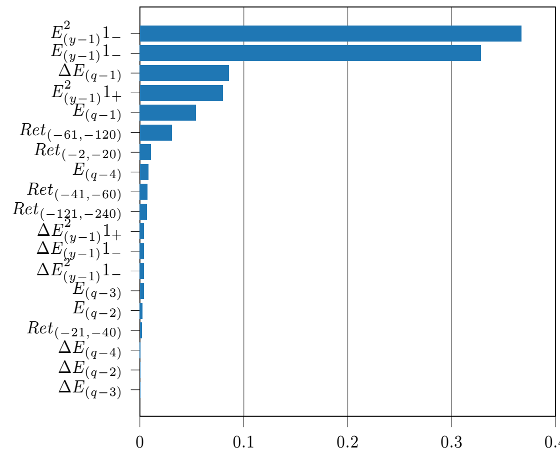
(f) Year 3 without ΔDIV

Note: The figure reports the in-sample feature (variable) importance for all but the control covariates in gradient boosting regression trees. Feature importance are sorted based on the most to the least important variables in predicting one-year, two-year and three-year ahead earnings with (the left panel) and without (the right panel) dividend changes as a covariate. Feature importance sums to one. The overall performance of the model is measured by R^2.

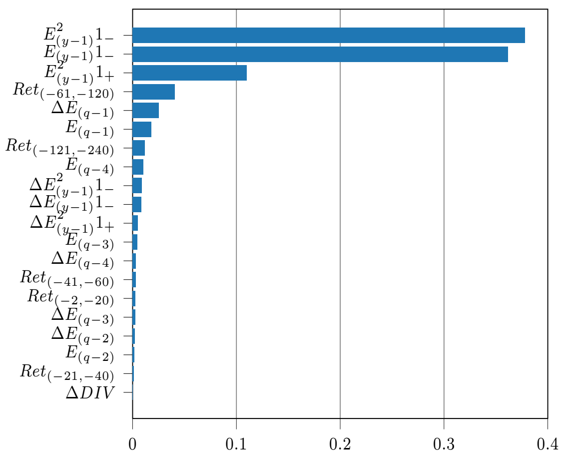
Figure 5: Feature Importance in Random Forest IS



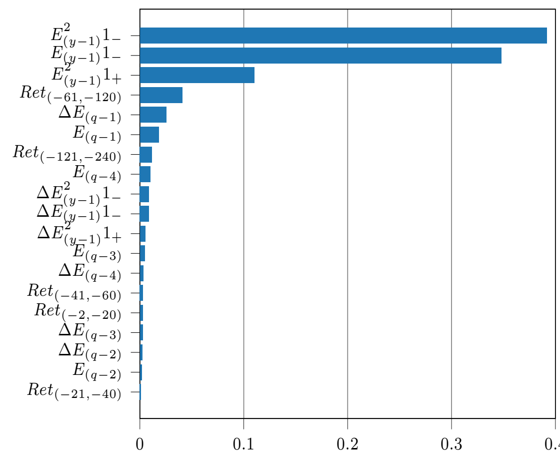
(a) Year 1 with ΔDIV



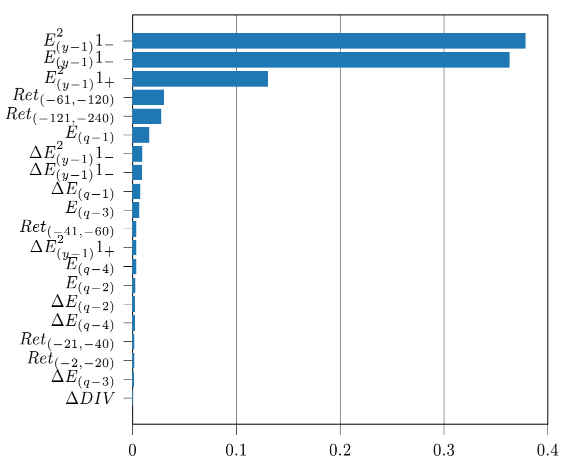
(b) Year 1 without ΔDIV



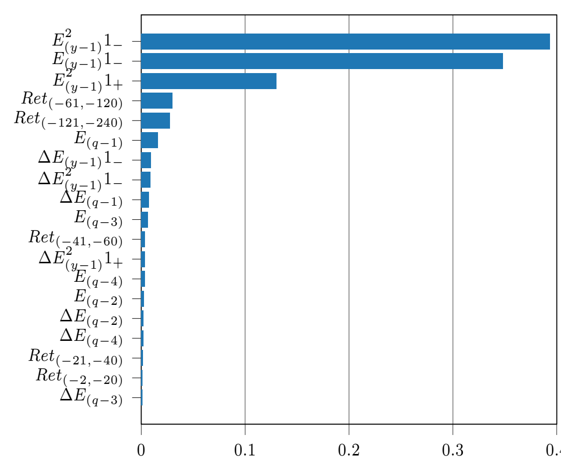
(c) Year 2 with ΔDIV



(d) Year 2 without ΔDIV



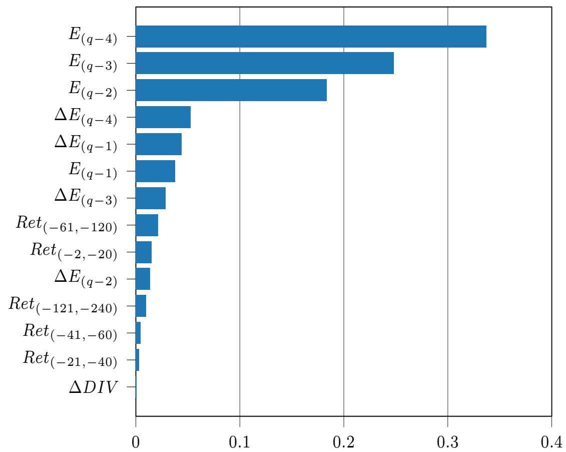
(e) Year 3 with ΔDIV



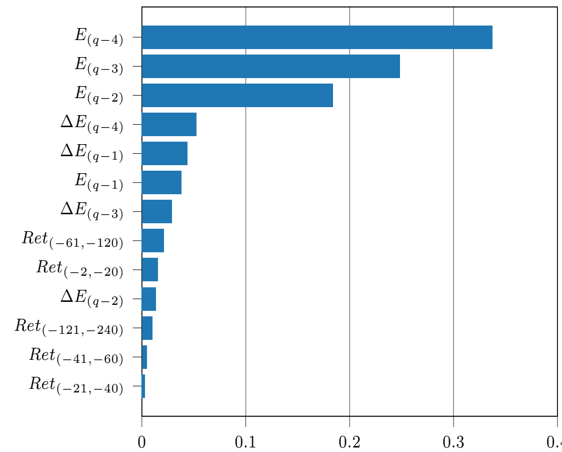
(f) Year 3 without ΔDIV

Note: The figure reports the in-sample feature (variable) importance for all the covariates in random forest regression trees. Feature importance are sorted based on the most to the least important variables in predicting one-year, two-year and three-year ahead earning changes with (the left panel) and without (the right panel) dividend changes as a covariate. Feature importance sums to one. The overall performance of the model is measured by R^2.

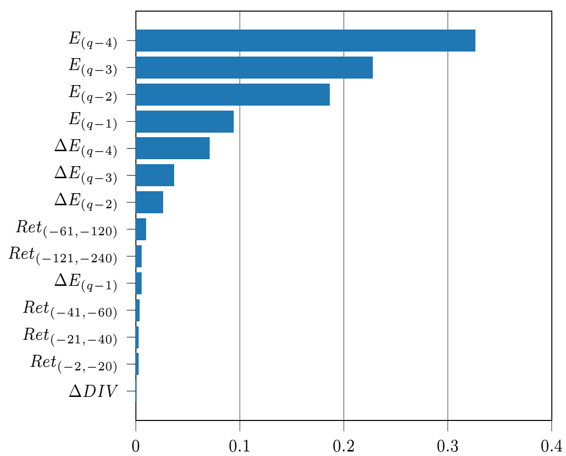
Figure 6: Feature Importance in Random Forest without Interaction Variables IS



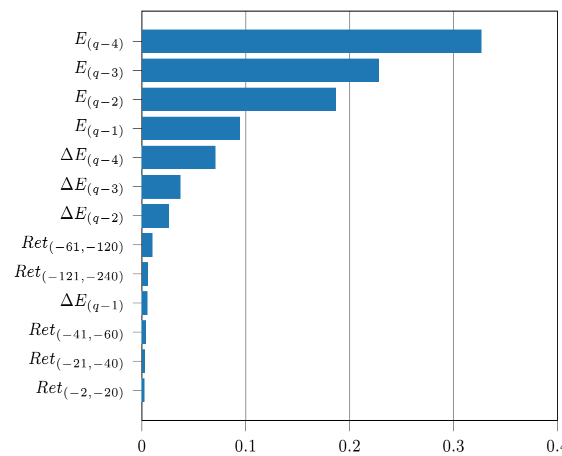
(a) Year 1 with ΔDIV



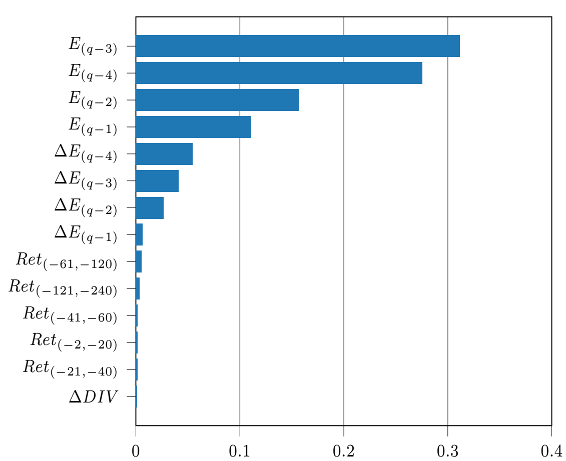
(b) Year 1 without ΔDIV



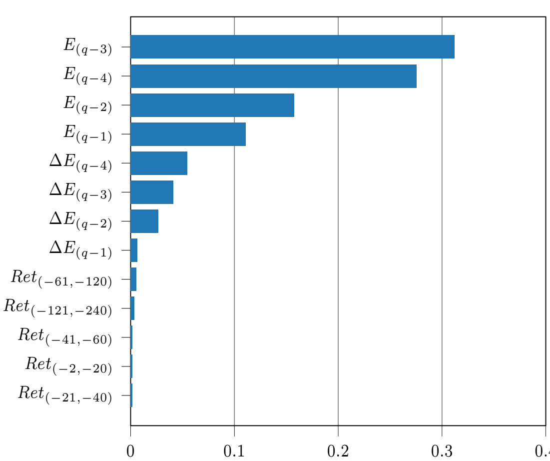
(c) Year 2 with ΔDIV



(d) Year 2 without ΔDIV



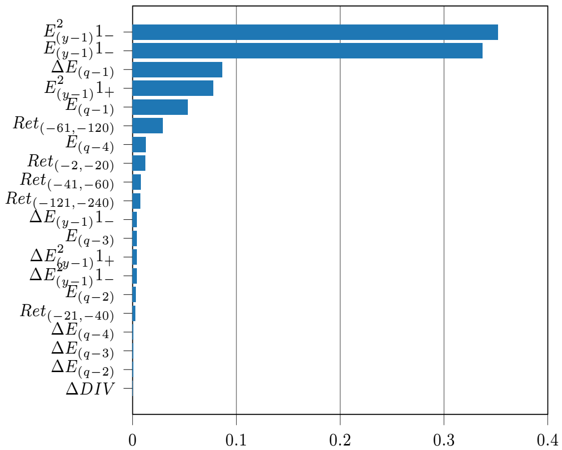
(e) Year 3 with ΔDIV



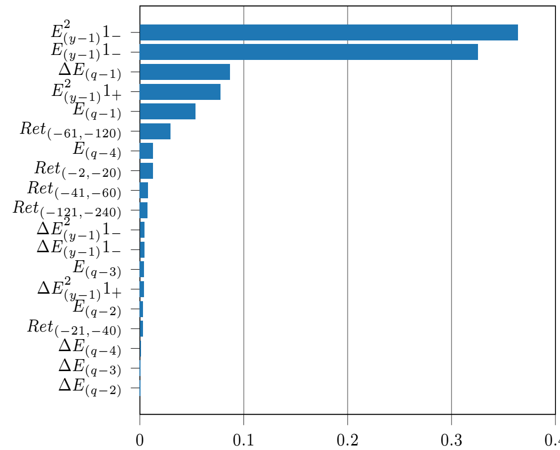
(f) Year 3 without ΔDIV

Note: The figure reports the in-sample feature (variable) importance for all the covariates in random forest regression trees. Feature importance are sorted based on the most to the least important variables in predicting one-year, two-year and three-year ahead earning changes with (the left panel) and without (the right panel) dividend changes as a covariate. Feature importance sums to one. The overall performance of the model is measured by R^2.

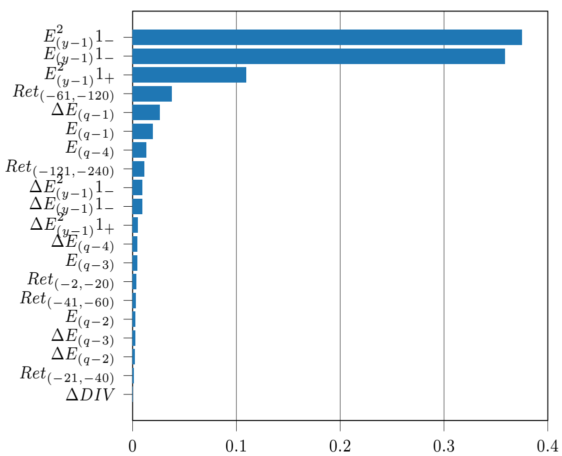
Figure 7: Feature Importance in Random Forest OOS



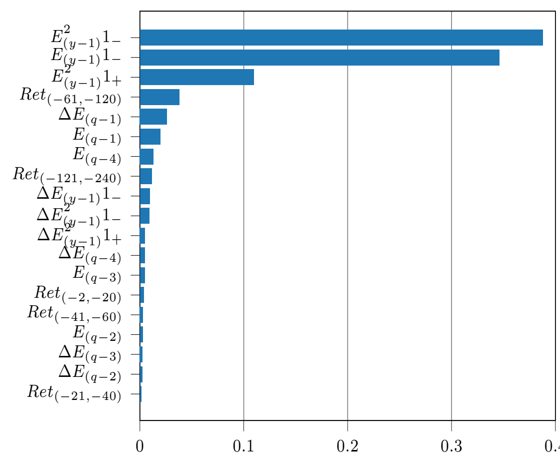
(a) Year 1 with ΔDIV



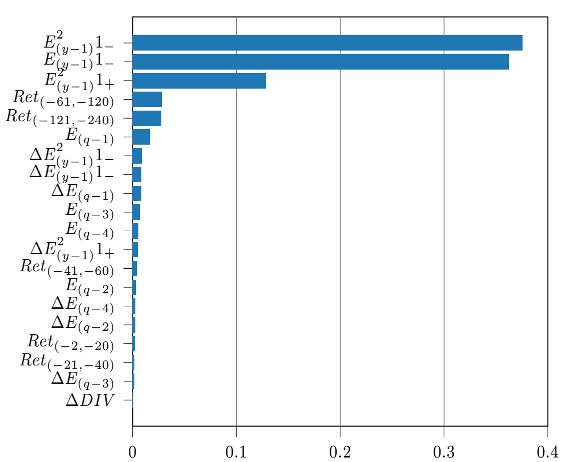
(b) Year 1 without ΔDIV



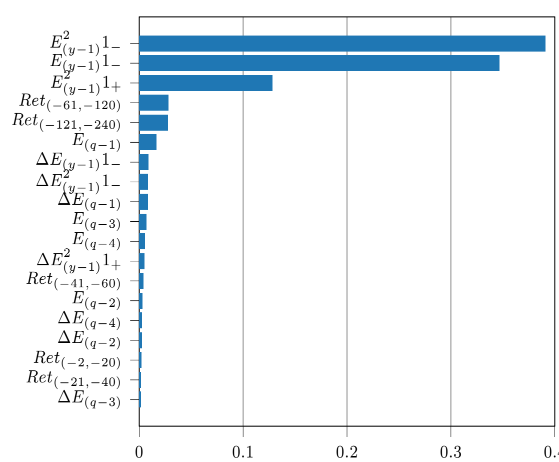
(c) Year 2 with ΔDIV



(d) Year 2 without ΔDIV



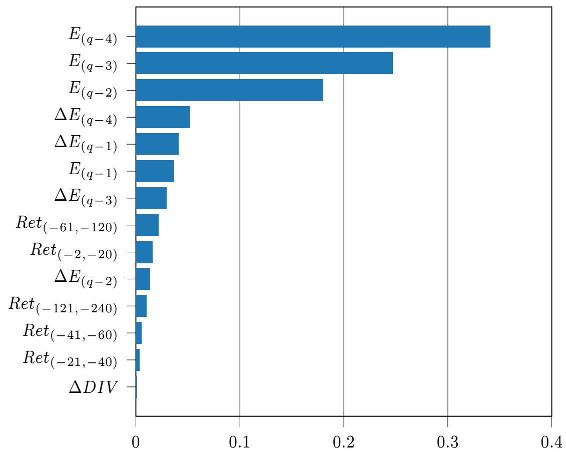
(e) Year 3 with ΔDIV



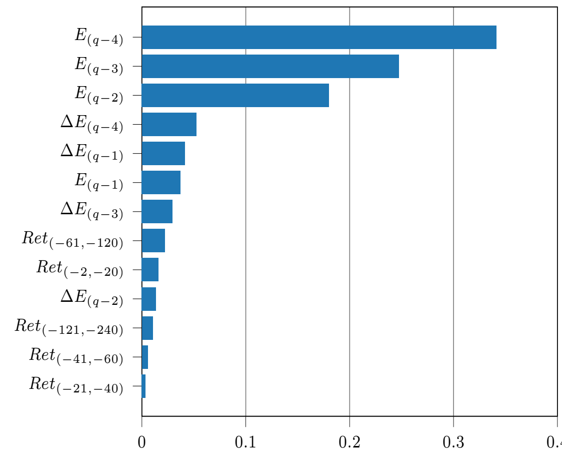
(f) Year 3 without ΔDIV

Note: The figure reports the in-sample feature (variable) importance for all the covariates in random forest regression trees. Feature importance are sorted based on the most to the least important variables in predicting one-year, two-year and three-year ahead earning changes with (the left panel) and without (the right panel) dividend changes as a covariate. Feature importance sums to one. The overall performance of the model is measured by R^2.

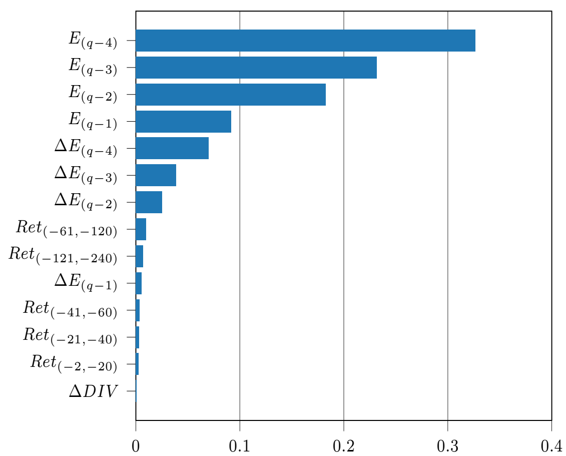
Figure 8: Feature Importance in Random Forest without Interaction Variables OOS



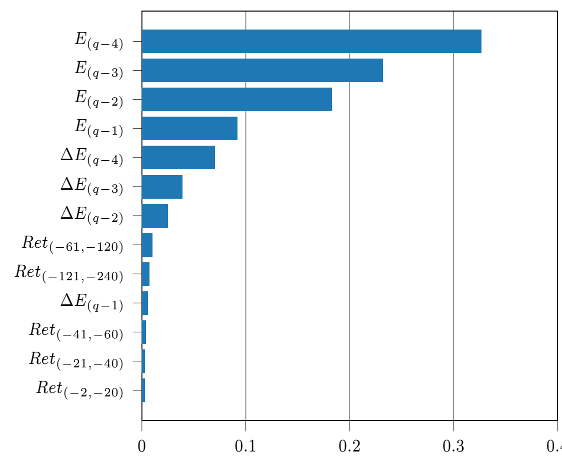
(a) Year 1 with ΔDIV



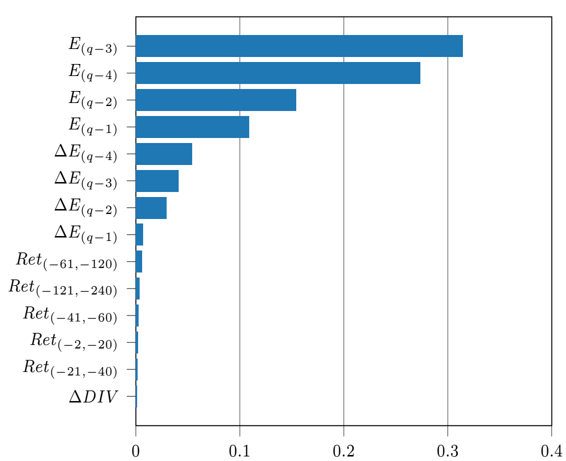
(b) Year 1 without ΔDIV



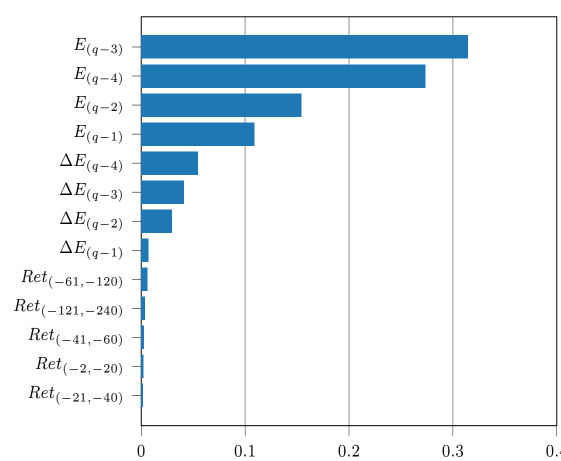
(c) Year 2 with ΔDIV



(d) Year 2 without ΔDIV



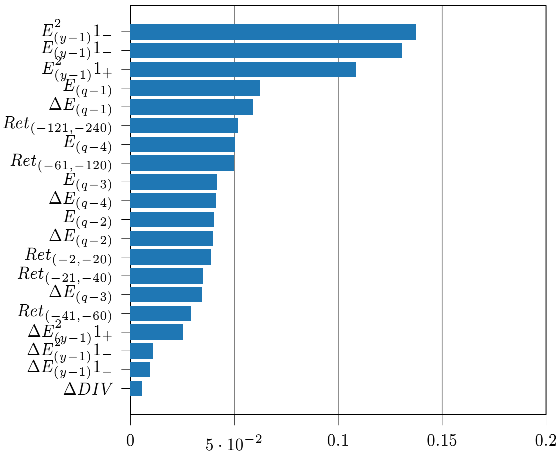
(e) Year 3 with ΔDIV



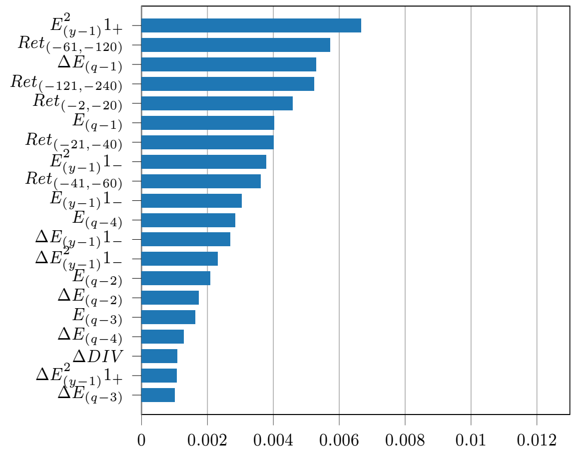
(f) Year 3 without ΔDIV

Note: The figure reports the in-sample feature (variable) importance for all the covariates in random forest regression trees. Feature importance are sorted based on the most to the least important variables in predicting one-year, two-year and three-year ahead earning changes with (the left panel) and without (the right panel) dividend changes as a covariate. Feature importance sums to one. The overall performance of the model is measured by R^2.

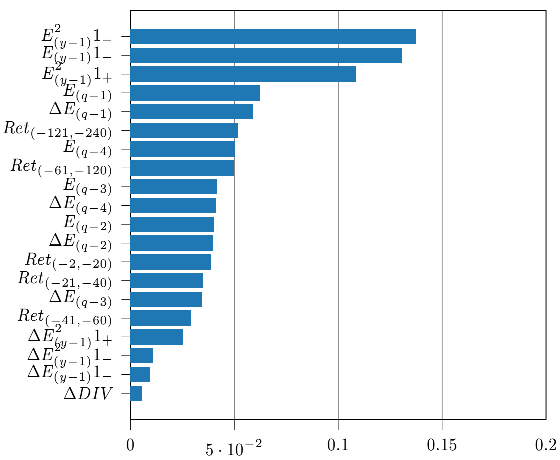
Figure 9: Feature Importance vs Mean SHAP Value in Gradient Boosting Trees OOS



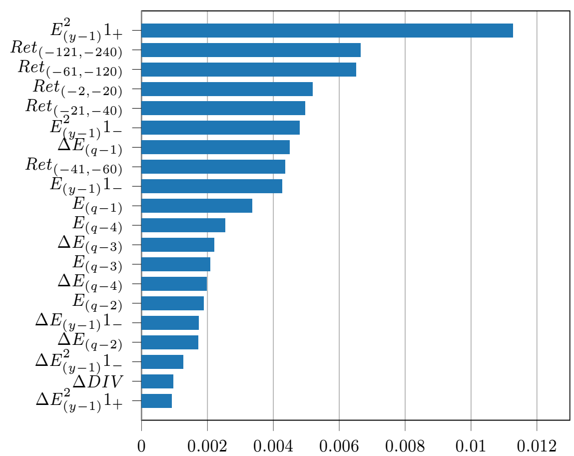
(a) Year 1



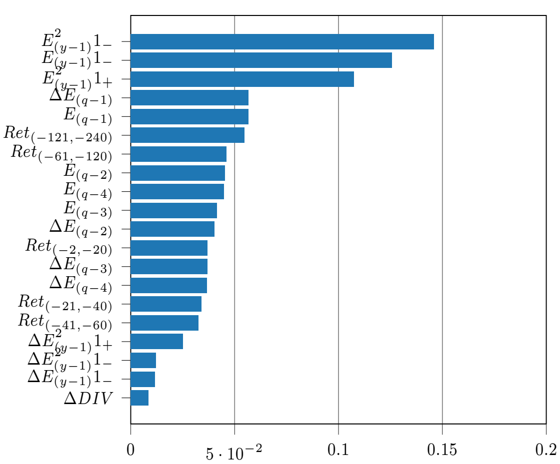
(b) Year 1



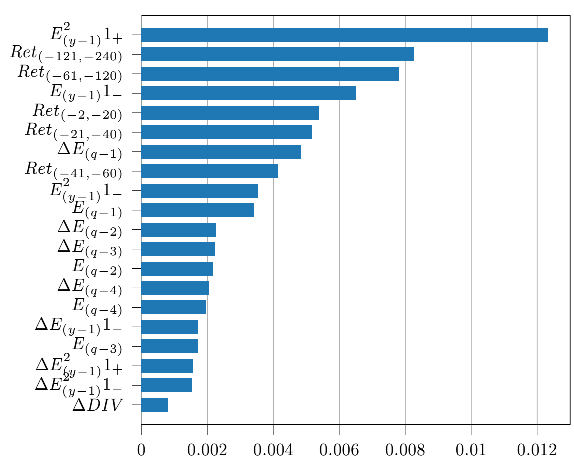
(c) Year 2



(d) Year 2



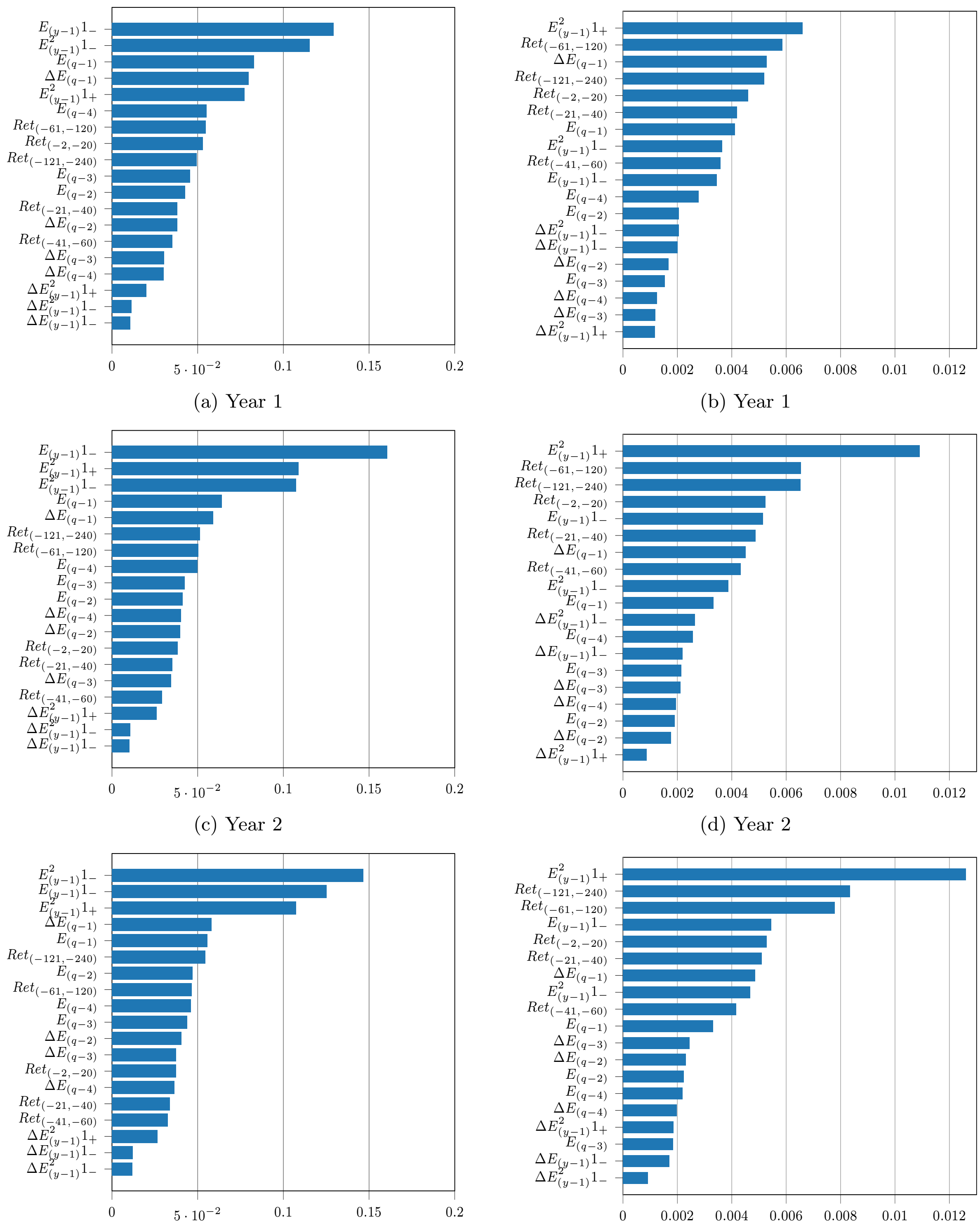
(e) Year 3



(f) Year 3

Note: feature importance (left), mean shap value (right), GBR, OOS

Figure 10: Feature Importance vs Mean SHAP Value in Gradient Boosting Trees without ΔDiv OOS



(a) Year 1 (b) Year 1

(c) Year 2 (d) Year 2

(e) Year 3 (f) Year 3

Note: feature importance (left), mean shap value (right), GBR, OOS