

Article

# Predicting the Profitability of Directional Changes Using Machine Learning: Evidence from European Countries

Nicholas D. Belesis , Georgios A. Papanastasopoulos \* and Antonios M. Vasilatos 

Department of Business Administration, University of Piraeus, 18534 Piraeus, Greece; nmpeleisis@unipi.gr (N.D.B.); vasilatos@unipi.gr (A.M.V.)

\* Correspondence: papanast@unipi.gr

**Abstract:** In this paper, we follow the suggestions of past literature to further explore the prediction of the profitability direction by employing different machine learning algorithms, extending the research in the European setting and examining the effect of profits mean reversion for the prediction of profitability. We provide evidence that simple algorithms like LDA can outperform classification trees if the data used are preprocessed correctly. Moreover, we use nested cross-validation and show that sample predictions can be obtained without using the classic train–test split. Overall, our prediction results are in line with previous studies, and we also found that cash flow-based measures like Free Cash Flow and Operating Cash Flow can be predicted more accurately compared to accrual-based measures like return on assets or return on equity.

**Keywords:** profitability; directional changes; machine learning; mean reversion



**Citation:** Belesis, Nicholas D., Georgios A. Papanastasopoulos, and Antonios M. Vasilatos. 2023. Predicting the Profitability of Directional Changes Using Machine Learning: Evidence from European Countries. *Journal of Risk and Financial Management* 16: 520. <https://doi.org/10.3390/jrfm16120520>

Academic Editor: Marius Sikveland

Received: 13 November 2023

Revised: 12 December 2023

Accepted: 14 December 2023

Published: 18 December 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The forecasting of earnings has long been an important topic in accounting research because of the correlation between earnings and market return (Beaver 1968). Implementing a fundamental analysis requires three steps. The initial and most important phase is an accurate forecast of a company's future profitability across the forecast horizon. Without the correct forecasts, the second step in estimating the intrinsic value is inaccurate, and the third-step comparison of the market value to the intrinsic value likely results in the wrong investment decision. In addition to accurate projections, fundamental analysis requires out-of-sample-based forecasts of profitability because investors require such projections. Profitability and growth projections are necessary for the valuation of a company. Previous research on financial statement analysis (Freeman et al. 1982; Fairfield et al. 1996; Nissim and Penman 2001) has investigated methods for enhancing profitability and growth forecasts.

According to Chen et al. (2022), examining the direction of future profits rather than their level is preferable for various reasons. First, it is difficult to estimate how much future earnings will change, as studies have demonstrated that the projection of earnings based on business characteristics is not much more accurate than those based on the random-walk model (Gerakos and Gramacy 2013; Li and Mohanram 2014).

Second, Freeman et al. suggest that the variance in the increase in earnings is too great to be compared to the variance in projected earnings and changes contingent on explanatory variables. They propose reducing the changes in earnings variance by translating the amount into the direction of earnings changes, anticipating which is more realizable.

Third, from an economic standpoint, anticipating the direction of change in profits is significant and beneficial since portfolios are constructed depending on the direction of changes in earnings (Ou and Penman 1989; Wahlen and Wieland 2011).

We were motivated by the call of Anand et al. (2019) to further explore the profitability of directional change predictions using other machine learning algorithms apart from the random forest they used previously. We also followed their suggestion on handling

categorical variables properly. We performed this by hot encoding categorical variables. Taking into account their choice to use a time-series cross-validation, we suggested the use of a nested time-series cross-validation, which gave us the benefit of using the entire dataset to train the algorithm rather than portioning the data in a train–test ratio. We provide evidence that with pre-processed data, algorithms that are easier to implement compared to random forest perform better. Furthermore, we extended the analysis by Anand et al. on the effect of profitability’s mean reversion on the predictability of profitability by examining mean reversion within each country and within the entire universe of firms. Our results appear similar to their approach, so we cannot provide solid evidence that the mean reversion improves predictability.

This paper is structured as follows: In Section 2, we review the phenomenon of the mean reversion of profits and present studies from the past literature relative to the use of machine learning in predicting profitability and the key differences and advantages compared to traditional regression methods. In Section 3, we explain the motivation and benefits of our approach, and we present the detailed steps we followed to establish our models. Moreover, we explain the data analysis and the preprocessing we performed. In Section 4, we present the empirical results, and Section 5 contains our concluding remarks and proposals for future research.

## 2. Literature Review

### 2.1. Mean Reversion of Profitability

As enterprises operate in a competitive environment, the economic theory predicts that their profitability rates revert to the economy-wide mean and converge. Despite the fact that businesses strive to maintain any competitive advantage and avoid mean reversion (Porter 1985), they are often governed by the economic laws of competition (Aghion et al. 2001; Aghion 2002). New entrants provide competition for firms with higher performance, hence cutting future rents. Similarly, underperforming enterprises either survive by increasing their profitability or fail, resulting in a reversion to the mean of profitability. Accounting return rates exhibit signs of mean reversion, according to an extensive body of accounting research. Beaver (1970) found that firms with high ROEs face a subsequent fall in ROE, but firms with low ROEs experience subsequent growth, albeit at a slower rate. Penman (1991) discovered that although ROA reverts to the mean, it also has a persistent component that enables enterprises with a high ROA to continue to outperform in the future.

The existence of a mean reversion in profitability and growth (Freeman et al. 1982; Nissim and Penman 2001) towards the economy-wide mean is the key finding of prior research. However, the optimal way to model such mean reversion remains unanswered. Most of the research is predicated on the assumption that all enterprises in the economy exhibit the same level of mean reversion and return to a standard.

Numerous subsequent studies have endeavored to comprehend the drivers of business profitability time-series features—see Kothari (2001) for a review. According to these studies, the reversion of the mean of accounting returns is influenced by both firm- and industry-level variables. Firm-level drivers include firm size (Lev 1983), future investment opportunities (Nissim and Penman 2001), and measurement errors in accounting (Penman and Zhang 2002).

Given the significance of membership in this sector, Fairfield et al. (2009) investigated whether a mean-reverting model at the industry level enhances the precision of profitability and growth estimates. While they found that industry-level analyses are incrementally informative for forecasting growth, forecasts of profitability are not improved by industry-level analyses.

Healy et al. (2014) investigated the effect of cross-country disparities in product, capital, and labor market competition, as well as profit management, on the mean reversion of accounting return on assets. Using a sample of 48,465 unique enterprises from 49 nations from 1997 to 2008, they discovered, as predicted by the economic theory, that accounting for the mean of returns reverts faster in countries with higher product and capital market

competition. In contrast, countries with more competitive labor markets experience a delayed return to the mean. The relationship between variations in country in labor market competition, earnings management, and mean reversion in accounting returns varies with business performance. When unexpected returns are favorable, country labor competition promotes a mean reversion, but when unexpected returns are negative, it decreases it. Accounting returns in nations with a higher mean earnings management reverts more slowly for profitable organizations and more quickly for loss-making businesses. Thus, incentives for earnings management = to slow or speed up the mean reversion of accounting returns are amplified in nations with a high earnings management tendency. Overall, these results indicate that national factors explain the mean reversion of accounting returns and are, therefore, relevant to firm valuation.

## 2.2. Machine Learning

During the past thirty years since [Ou and Penman \(1989\)](#) reported their findings, computing power, and machine learning techniques have advanced dramatically, allowing researchers to examine whether additional independent variables and more computer-intensive methodologies are useful for predicting future earnings. Over the past two decades, statisticians have developed dozens of forecasting methods, each with its own advantages and disadvantages. The conventional stepwise logistic model ([Ou and Penman 1989](#)), elastic net ([Zou and Hastie 2005](#)), and random forest model are these strategies ([Breiman 2001](#)).

The evolution of profit forecasting reflects not only the growth of accounting studies but also the development of statistics and computer science. Early studies predicted future earnings using random walk and time-series models ([Ball and Watts 1972](#)). Previous studies included more fundamental data in linear regression ([Deschamps and Mehta 1980](#)) or logistic regression-based prediction models ([Ou and Penman 1989](#)).

Numerous studies, in light of the rapid growth of computer science, have revealed the remarkable potential of machine learning models, which might significantly improve the accuracy of a firm's earnings forecasts and subsequently generate abnormal profits.

The forecasting performance of regression models is contingent upon several variables, including the functional form, the choice of predictors, the choice of estimator, and the behavior of the error term. Perhaps a few of these limiting variables are responsible for the inferior performance of regression models compared to the random walk model in out-of-sample predictions. For financial analysts, a vast body of research, as summarized by [Ramnath et al. \(2008\)](#), demonstrates that analysts are systematically optimistic in their projections, either due to incentive-driven strategic reporting or innate cognitive bias.

Machine learning (ML) approaches are more flexible than regression methods because they do not rely on limiting statistical and economic assumptions and are not influenced by cognitive biases. Instead, they utilize past data patterns and trends to generate forecasts. ML focuses on maximizing the accuracy of prediction ([Mullainathan and Spiess 2017](#)). While the term machine learning is widely used, several machine learning methods have been more adept at handling econometric difficulties in the data, such as multicollinearity and nonlinearity, than regression-based methods. In addition, many ML approaches do not require the user to provide a functional form beforehand, resulting in additional flexibility to discover a functional form that best matches the data. As a compromise, many ML approaches are more opaque than conventional regression. Thus, a researcher who selects ML over regression often improves the accuracy of out-of-sample forecasts but sacrifices interpretability. Given the issue at hand, which is enhancing the accuracy of out-of-sample profit estimates, we believe this is a worthwhile tradeoff.

Recent applications of random forests and stochastic gradient boosting have shown surprising results ([Zhou 2012](#); [Mullainathan and Spiess 2017](#)). Both methods, which are based on ensemble learning, combine a large number of decision tree estimators. A key aspect of these methods, as opposed to regressions, is their capacity to estimate models with a greater number of predictors than observations. In addition, the theoretical literature

provides little direction for the selection of crucial financial variables and functional forms in financial statement analysis. High-dimensional predictor sets may enter in a nonlinear manner with several interactions.

By contrast, machine learning algorithms are specifically built to tolerate complicated relationships and cast a wide net in their specification search. They give a strong out-of-sample predictive performance by employing “regularization” (e.g., tweaking a parameter like the number of decision trees in random forests) for model selection and overfitting mitigation.

Decision trees are a prominent method for incorporating nonlinearities and interactions in statistical learning. Contrary to regressions, trees are designed to group observations with comparable predictors and are constructed nonparametrically. The forecast is the mean of the outcome variable within each group.

Two approaches are used to regularize decision trees in random forests. Initially, in the bootstrap aggregation technique, also known as “bagging” (Breiman 2001), a tree is constructed based on each  $n$  distinct data bootstrap samples. There are  $n$  projections for a particular observation, and the final forecast is the simple average of  $n$  predictions. The tendency of trees to overfit individual bootstrap samples renders their forecasts inefficient.

Anand et al. (2019) investigated whether classification trees, as a machine learning technique, can produce out-of-sample profitability estimates that are superior to random walk forecasts. Out-of-sample forecasts of directional changes (either increases or decreases) are generated for the following five profitability measures: return on equity (ROE), return on assets (ROA), return on net operating assets (RNOA), cash flow from operations (CFO), and Free Cash Flow (FCF). Based on a minimal set of independent variables, their methodology obtains classification accuracies ranging from 57 to 64% as opposed to 50% for the random walk, and the proportional differences between ML and random walk are highly significant. In addition, they found that the predictive ability of their ML approach is unaffected by a five-year forecast horizon. Furthermore, they found that the two cash-flow measures (CFO and FCF), particularly when accruals are included in the prediction of cash flows, have a greater classification accuracy than the three earnings-based measures of profitability (ROE, ROA, and RNOA). Overall, their findings suggest that ML approaches have the potential to be utilized for predicting profitability.

Xinyue et al. (2020) exhaustively evaluated the feasibility and suitability of adopting the machine learning model for forecasting company earnings and compared their results with analysts’ consensus estimates and traditional statistical models with logistic regression. They discovered that their approach only surpasses logistic regression methods and cannot surpass analysts.

In their work, Kureljusic and Reisch (2022) used publicly available information for European enterprises and new machine-learning algorithms to estimate future revenues in an IFRS environment, investigating the advantages of predictive analytics for both the preparers and users of financial predictions. Their empirical findings, based on 3000 firm-year data from 2010 to 2019, show that machine learning gives revenue estimates that are as accurate or more accurate than those of financial experts. Yet, their sample size is very small to employ machine-learning algorithms and generalize their results.

We build upon the existing literature in several ways. We were motivated by the growing literature on machine learning and earnings prediction. We introduce in the literature the concept of the nested cross-validation of algorithms. In the relevant literature, we spotted the following gaps. Both Anand et al. (2019), Bao et al. (2020), and Chen et al. (2022) made arbitrary splits in the training and test samples. These splits affect the reported performance of their algorithms (Rácz et al. 2021). With the nested cross-validation, we tackle this drawback, and we report more robust results. In addition to this, relevant studies use “random walk”<sup>1</sup> as a benchmark; this comparison is “unfair” as all complex machine learning algorithms can beat the random walk. We create a benchmark, starting with a simple algorithm, and then employ more complex algorithms to examine whether we can beat this benchmark. Furthermore, we expand the study of Anand et al. (2019)

by examining the mean reversion of earnings not only within industries but also within countries and the entire sample of firms.

### 3. Data and Methodology

#### 3.1. Research Design

Previous researchers in the field followed the same approach in order to build a model and make predictions. They selected their features (variables) from theory, and based on the properties and nature of their dataset, they chose an appropriate algorithm, made predictions, and then reported a performance metric such as accuracy or the ROC-AUC score. By comparing their performance metric to similar past works of research using the same methodology and similar features, they evaluated their model.

Although this approach is the gold standard in the literature thus far, it comes with some drawbacks. Every dataset in each study is different, and these data have different properties. One of the first design decisions the researcher makes when predicting with machine learning is the train–test split ratio of the data. The rule of thumb is to split the data in a ratio of 70–30 or 80–20. What happens if we vary this ratio and make predictions using different splits, such as 50–50 or 65–35? Our results vary. Which model generalizes better with unseen data? Is there an optimal train–test split? Which results should we report?

The next major design choice a researcher makes is the selection of the algorithm. Not all algorithms perform well with all datasets. Some algorithms perform better in linear problems and underperform in nonlinear problems; other more complex, general-purpose algorithms perform well in all kinds of problems (such as random forests or decision trees). Does a simple-to-implement algorithm give us better or similar prediction results than a more complex algorithm? Is a slight improvement in prediction accuracy using a complex estimation method worthwhile compared to a simpler one, or does it justify the use of a more complex algorithm?

These questions stemming from two simple design choices have motivated us to differentiate our approach. We did not split our data in a certain ratio, but we used nested time-series cross-validation in the entire dataset to tune our algorithms, validate our models, and make sample predictions. In this way, we used the entire dataset to train our algorithms, and thus, we did not waste data. By cross-validating our data, we trained, validated, and tested several models and data splits. In this way, we reported more robust results. Moreover, we did not use just a single algorithm to make predictions. We used a simple algorithm in linear problems, the Linear Discriminant Analysis (LDA), as a benchmark. When we established the best-performing model estimated with the LDA, we used more complex algorithms such as the K-Nearest Neighbors (KNN), decision trees (DT), and random forest (RF), and we attempted to beat the benchmark model we established. We built our benchmark by sequentially adding features to our model until we could report no further improvement in our prediction performance.

##### 3.1.1. Effect of Sequentially Adding Features

In the first step of our analysis, we established a base model using only the profitability measure, industry, country, and year as features to predict the target with the LDA algorithm. Our target variable, if there was an increase in profitability ( $\Delta ROA/\Delta ROE/\Delta RNOA/\Delta FCF/\Delta CFO$ ), took the value of one, or if there was a decrease,  $-1$ .

- (a) Analysis 1.1-a: Measure (ROA/ROE/RNOA/FCF/CFO), Industry, Country, Year;
- (b) Analysis 1.1-b: Measure, 1st Lag Measure, Industry, Country, Year;
- (c) Analysis 1.1-c: Measure, 1st Lag Measure, 2nd Lag Measure, Industry, Country, Year;
- (d) Analysis 1.1-d: Measure, 1st Lag Measure, 1st Diff. Measure, Industry, Country, Year.

In Analysis 1.2, we selected the best performing model (Benchmark) from Analysis 1.1 estimated with LDA, and we used more algorithms, trying to improve upon LDA.

1. Analysis 1.2-a: Estimated Benchmark with KNN;
2. Analysis 1.2-b: Estimated Benchmark with DT;



### 3. Analysis 1.2-c: Estimated Benchmark with RF.

#### 3.1.2. Mean Reversion and Profitability Prediction

To examine the effect of the profitability of mean reversion in our predictions, we followed the approach of Zhou (2012). In year 0, we created 10 portfolios by ranking the profitability measures in quintiles. We performed the ranking in 2000, 2005, 2010, and 2015. For these years and the following five years, we calculated the medians of the profitability measure. We used, as proxies for mean reversion, the mean/median/standard deviation of the profitability measure for the industry/country/the whole universe of firms; the same method was used by Anand et al. (2019). We expected that if there was a mean reversion of profits within the industries, countries, or the entire sample, the features that were used as proxies of the mean reversion of profits would eventually increase their prediction performance.

#### 3.2. Nested Cross-Validation

It is well documented in the literature that standard cross-validation, such as k-fold cross-validation, is not suitable for accounting data, as accounting data are time-dependent (Chen et al. 2022; Zhu et al. 2022; Bao et al. 2020). For this reason, we followed a similar approach to Anand et al. (2019). However, even with a simple time-series cross-validation, we still kept a percentage of our data as a test set. The algorithm's parameters had to be hyper-tuned during cross-validation, causing a leak from the test data to the training data. Instead, we chose to perform nested cross-validation using the entire dataset. With nested cross-validation, we performed two cross-validations. One inner cross-validation was performed in each training fold to hyper-tune the algorithm's parameters and validate the model, and one outer cross-validation was performed to evaluate the model in the unseen test folders.

Since each training fold was further cross-validated, the outer cross-validation provided unbiased results (Varma and Simon 2006). The outer loop of the cross-validation can be described in the following way. The data are partitioned into the following nine blocks:

- Block 1: Train (2000–2001)-Test (2003–2004);
- Block 2: Train (2000–2003)-Test (2005–2006);
- Block 3: Train (2000–2005)-Test (2007–2008);
- .
- .
- .
- Block 9: Train (2000–2017)-Test (2019–2020).

In each training fold, the inner loop cross-validation is repeated in the same way to tune each algorithm's hyperparameters. Because we used lagged variables in the model, we preferred to leave a gap between the train and test folds. For example, in Block 1, we did not train the model in 2002. Otherwise, there would be a data leakage from the test fold to the training fold because if, for example, the test fold included ROA2003 and the first lag of ROA2003 (which is ROA2002), then, at the same time, we trained and tested using the first lag of ROA2003 (ROA2002).

The performance of our algorithms was evaluated on the unseen test folds. For every test fold, we measured performance with a weighted F1 score. The F1 score is more suitable than accuracy as a performance metric. Accuracy is the standard classification performance metric, and we could obtain it using the following formula:

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

where  $TP$  = true positive and is the number of firm years correctly classified as an "increase in profitability" (e.g., when  $\Delta ROA = 1$  is correctly classified),  $TN$  = true negative and is the number of firm years correctly classified as a "decrease in profitability" (e.g.,  $\Delta ROA = -1$ ).

FN (false negative) is the number of firm years misclassified as an “increase in profitability”, and FP (false-positive) is the number of firm years misclassified as a “decrease in profitability”. This performance metric clearly failed when our data were imbalanced and provided us with overconfident performance metrics. If most of the firm years are classified as an “increase in profitability”, then the accuracy metric can fluctuate because only 20% of our sample is classified as a “decrease in profitability”. On the other hand, the F1 score is given by the following formula:

$$F1 = 2 \times \frac{\textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}} \tag{2}$$

or alternatively in terms of TP and TN as

$$F1 = \frac{TP}{TP + \frac{1}{2} \times (FP + FN)} \tag{3}$$

To derive the weighted F1 score, we simply multiplied the F1 score for each class (1, -1) with each class’s “support proportion”, resulting in the sum of firm years for each class to the total firm years. The out-of-sample performance of the model was calculated by averaging the F1 score of all test folds (k).

$$\textit{Out of Sample Performance} = \frac{1}{10} \times \sum_{k=1}^{10} F1 \textit{ Score}_k \tag{4}$$

### 3.3. Sample

Our sample consisted of public firms in 15 European Union countries (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Switzerland, United Kingdom) plus Switzerland. Financials were excluded because these countries follow different accounting rules. The countries are classified as advanced economies and follow the same accounting rules (IFRS since 2005) (Artikis et al. 2022). The period examined ranged from 2000 to 2020. All data were downloaded from the Thomson Reuters WorldScope-Datastream Database. The sample size was 29,589 firm years. Data are summarized in Table 1.

**Table 1.** Descriptive statistics.

Variables	Count	Mean	std	min	25%	50%	75%	max
ROA	29589	−0.00029	3.853063	−18.571	0.012029	0.038413	0.065984	9.430038
ROE	29589	5.607812	6.694	−9.8982	0.079385	0.74055	2.545368	4.3362
FCF	29589	−0.02018	4.947218	−7.714	−0.06354	0.024818	0.103388	4.925294
RNOA	29589	0.169964	5.725635	−12.427	0.030032	0.106526	0.205016	11.9359
CFO	29589	0.035151	4.278692	−14.143	0.037866	0.076719	0.118143	12.17509

Note: Variables were winsorized by 2.5% on each side of the distribution.

By examining the targets, it was clear that the two classes were quite balanced. Nonetheless, even with a small class imbalance, as in the case of DFCF, where there were 2500 more firm years that exhibited an increase in profitability, the accuracy score increased in favor of the positive class (1), making the F1 score a more appropriate choice. Target balance is presented in Table 2.

**Table 2.** Class balance (in firm years).

Target	DROA	DROE	DRNOA	DFCF	DCFO
Increase in profitability (1)	14,726	15,342	14,587	16,063	14,483
Decrease in profitability (−1)	14,863	14,247	15,002	13,526	15,105
Class balance (%)	49.8–50.2	51.9–48.1	49.3–50.7	54.3–45.7	48.9–51.1

### 3.4. Data Analysis

During the nested cross-validation, in each fold, with the use of a pipeline, we performed data preprocessing. Before we fed data into an algorithm, we had to ensure that some conditions and statistical properties of these data were not violated. There are algorithms, such as random forest, that work equally well with raw data, but others, such as LDA and KNN, require us to preprocess our data. First, we dropped all rows with missing variables in our five target variables ( $\Delta ROA$ ,  $\Delta ROE$ ,  $\Delta RNOA$ ,  $\Delta FCF$ ,  $\Delta CFO$ ). Then, we win-sorized our dataset by 2.5% on each side of the distribution to remove outliers. Finally, into the pipeline, in each cross-validation fold, the following processes took place sequentially.

- Missing values were imputed with a K-Nearest Neighbor algorithm. Missing value imputation is a more sophisticated technique than mean imputation (Zhang 2012);
- We standardized our data by removing the mean and scaling to unit variance. The result of this was to transform each feature’s distribution closer to Gaussian. Standardized data improved the prediction performance and led to more stable models (Shanker et al. 1996).

The target variables were encoded as follows. If there was an increase in the profitability measurement, we coded this increase with 1 and  $-1$  otherwise. The time, country, and industry variables were one-hot encoded. We noticed that when we one-hot-encoded a variable, such as the country variable, each value became a new variable. If we used label encoding, we could imply that there was an ordinal relationship between the variable values, e.g., France > Germany > Greece, which was not true in our case. One-hot encoding leads to an increase in the dimensionality of the dataset, as in this case, instead of one country categorical variable, we had fifteen country variables; however, this is not an issue in our study since during estimation, we used PCA to reduce the dataset’s dimension.

### 4. Empirical Results

We started our analysis by establishing our benchmark. For each profitability measure, we trained several models by sequentially adding features. In Analysis 1.1, we used only the profitability measure and country, year, and industry variables. Then, we added the first lag of the measure, the second lag, and the first difference sequentially. We used the weighted F1 score to evaluate our results. In contrast to accuracy, the weighted F1 score took into consideration a possible class imbalance in our target. The results for Analysis 1.1 are presented in Table 3. The derivation of measures of profitability is presented in Appendix A (Table A1).

**Table 3.** Analysis 1.1 results.

Target	$\Delta ROA$	$\Delta ROE$	$\Delta RNOA$	$\Delta FCF$	$\Delta CFO$
(a) Measure	0.58	0.53	0.57	0.67	0.56
(b) Measure, Measure-1st Lag	0.87	0.62	0.77	0.93	0.89
(c) Measure, Measure-1st Lag, Measure-2nd Lag	0.87	0.62	0.77	0.93	0.89
(d) Measure, Measure-1st Lag, Measure-1st Diff.	0.92	0.7	0.85	0.96	0.95

Note: We added features sequentially, as per Analysis 1.1. Country, year, and industry dummies are included in all steps of the analysis. Measure is the target variable ( $\Delta ROA$ ,  $\Delta ROE$ ,  $\Delta RNOA$ ,  $\Delta FCF$ ,  $\Delta CFO$ ). We used the F1 score to measure prediction performance.

By sequentially adding variables, we found that adding a second lag to the measure in Model (b) did not improve our results, so we removed it and added the first difference. Model (d) is the best-performing model estimated by the LDA for all the targets, so this is the benchmark we try to beat by using more complex algorithms. To further enhance our results, we plotted the decision surface of the algorithm in Figure 1 and identified how it separated data into classes. Yellow color represents firm-years with increase in profitability, while purple color represents firm-years with decrease in profitability.



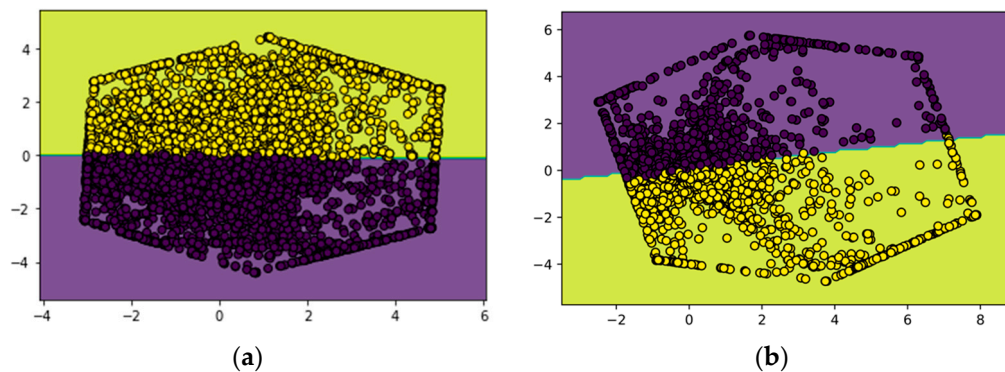


Figure 1. Decision surface plot. (a) Decision surface for  $\Delta ROA$ ; (b) Decision surface for  $\Delta ROE$ .

We chose two models, one predicting  $\Delta ROA$  and the other predicting  $\Delta ROE$ , and the plots showed how, in  $\Delta ROA$ , the algorithm can adequately separate the two classes of this target, while in  $\Delta ROE$ , several data points are misclassified. In the case of  $\Delta ROE$ , we expected more complex algorithms to outperform LDA. In Analysis 1.2, we predicted the best model from Analysis 1.1 using the KNN, DT, and RF algorithms. Results are presented in Table 4.

Table 4. Analysis 1.2 results.

Algorithm	$\Delta ROA$	$\Delta ROE$	$\Delta RNOA$	$\Delta FCF$	$\Delta CFO$
(a) KNN	0.89	0.8 *	0.86 *	0.94	0.91
(b) DT	0.86	0.79 *	0.8	0.9	0.82
(c) RF	0.79	0.63	0.75	0.9	0.82
Benchmark	0.92	0.7	0.85	0.96	0.95

Note: We took the best model, Analysis 1.1-d (from now on Benchmark), estimated with LDA, and we re-estimated it using more complex algorithms. We denote F1 scores with (\*) when they beat the benchmark.

In two cases ( $\Delta ROE$  and  $\Delta RNOA$ ), more complex algorithms managed to outperform our benchmark. Especially in the case of  $\Delta ROE$ , we observed a 10% increase in the F1 score. In Figure 2, we plotted the decision surfaces of  $\Delta ROE$  using the best-performing algorithms (KNN and DT) and saw how they compared to the previous plot made with the LDA.

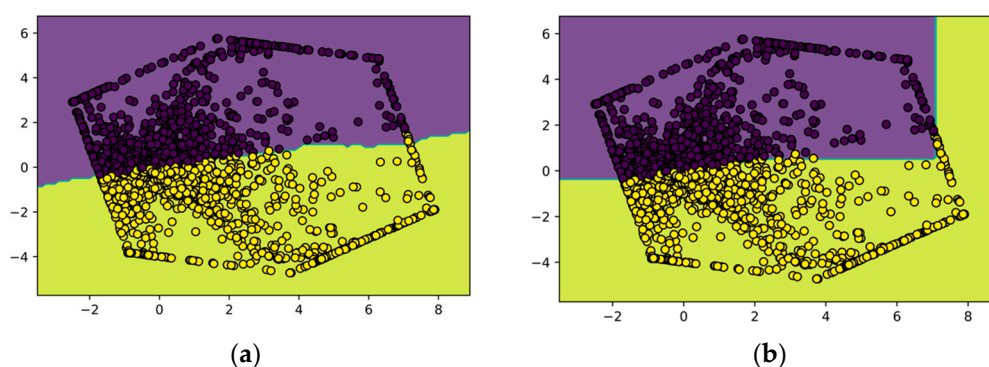


Figure 2. Decision surface plot. (a) Decision surface for  $\Delta ROE$  using KNN; (b) Decision surface for  $\Delta ROE$  using DT.

It appears that, in the case of  $\Delta ROE$ , the straight line used by LDA was not sufficient to correctly classify the data points. KNN and DT, which are suited for both linear and nonlinear classification problems, overperformed LDA.

In the second part of our analysis, we explored whether profitability’s mean reversion increased our prediction score. Nissim and Penman proved that profitability results in means reverting in the entire universe of firms. Anand et al. supported the fact that

profitability is mean reverting within industries and that it increases prediction accuracy. We extended their analysis and tested whether the mean reversion of profitability increases the predictability of the profitability of directional changes when profits are mean reverting (a) within the entire universe of firms, (b) within industries, and (c) within countries.

We took our benchmark model from Analysis 1.2 and added features related to the mean reversion and see whether they improved our prediction scores in a similar manner to Anand et al. For each profitability measure, we calculated the mean, median, and standard deviation per industry, country, and portfolio as proxies for the mean reversion within the industry, country, and the entire universe of firms, respectively. Furthermore, we created two more features, including the measure minus measure mean industry/country/portfolio and the measure minus measure median industry/country/portfolio. As in Analysis 1.2, we sequentially added these features to the benchmark model. For all the targets, we found that the best-performing model was the benchmark plus the profitability measure mean industry/country/portfolio and the measure minus measure mean industry/country/portfolio. Our results are presented in the tables below.

In the tables describing the empirical results, we examine the effect of adding mean reversion proxies to the prediction performance. In Table 5, we added three more features to the benchmark model in the form of the industry mean, median, and standard deviation of the profitability measure. We noticed that the predictability of Free Cash Flows remained unimproved. In all other measures, we noticed an improvement at a maximum of 1%. Similarly, in Table 6, we added the country mean, median, and standard deviation of the profitability measure to the benchmark. Again, we noticed that there was, in some cases, a 1% improvement in the prediction performance, and the predictability of Free Cash Flows remained once again unaffected. Lastly, in Table 7, we added the portfolio (see Section 3.1.2) mean, median, and standard deviation of the profitability measure. Again, Free Cash Flows remained unaffected in terms of the predictability of improvement, and we also saw a noticeable improvement of 3% in the predictability of Operating Cash Flows. The return on equity predictability performance did not improve as well. All other measures' predictability improved by 1%.

**Table 5.** Mean reversion within industry model.

	$\Delta ROA$		$\Delta ROE$		$\Delta RNOA$		$\Delta FCF$		$\Delta CFO$	
	Bench.	Ind.	Bench.	Ind.	Bench.	Ind.	Bench.	Ind.	Bench.	Ind.
LDA	0.92	0.93 *	0.7	0.7	0.85	0.86 *	0.96	0.96	0.95	0.95
KNN	0.89	0.9 *	0.8	0.8	0.86	0.85	0.94	0.94	0.91	0.91
DT	0.86	0.82	0.79	0.76	0.8	0.81 *	0.9	0.86	0.82	0.83 *
RF	0.79	0.78	0.63	0.62	0.75	0.73	0.9	0.87	0.82	0.83 *

Note: We estimated the model of Analysis 1.1-d and augmented with the industry's mean, median and standard deviation of the measure (as proxies of the profitability mean reversion). We denote F1 scores with (\*) when they beat the benchmark.

**Table 6.** Mean reversion within country model.

	$\Delta ROA$		$\Delta ROE$		$\Delta RNOA$		$\Delta FCF$		$\Delta CFO$	
	Bench.	Cntr.	Bench.	Cntr.	Bench.	Cntr.	Bench.	Cntr.	Bench.	Cntr.
LDA	0.92	0.92	0.7	0.7	0.85	0.86 *	0.96	0.96	0.95	0.95
KNN	0.89	0.9 *	0.8	0.8	0.86	0.86	0.94	0.94	0.91	0.91
DT	0.86	0.84	0.79	0.72	0.8	0.79	0.9	0.9	0.82	0.87 *
RF	0.79	0.78	0.63	0.64 *	0.75	0.75	0.9	0.9	0.82	0.82

Note: We estimated the model of Analysis 1.1-d and augmented with the country's mean, median and standard deviation of the measure (as proxies of profitability mean reversion). We denote F1 scores with (\*) when they beat the benchmark.

**Table 7.** Mean reversion within universe of firms model.

	$\Delta ROA$		$\Delta ROE$		$\Delta RNOA$		$\Delta FCF$		$\Delta CFO$	
	Bench.	Univ.	Bench.	Univ.	Bench.	Univ.	Bench.	Univ.	Bench.	Univ.
LDA	0.92	0.93 *	0.7	0.7	0.85	0.86 *	0.96	0.96	0.95	0.95
KNN	0.89	0.9 *	0.8	0.75	0.86	0.84	0.94	0.94	0.91	0.92 *
DT	0.86	0.86	0.79	0.76	0.8	0.74	0.9	0.9	0.82	0.85 *
RF	0.79	0.77	0.63	0.63	0.75	0.73	0.9	0.9	0.82	0.82

Note: We estimated the model of Analysis 1.1-d and augmented with the whole sample’s mean, median and standard deviation of the measure (as proxies of profitability mean reversion). We denote *F1* scores with (\*) when they beat the benchmark.

### 5. Discussion

In this research, we predict the profitability of directional changes using five measures of profitability, ROA, ROE, RNOA, FCF, and CFO. We propose an approach in which we estimate a simple model with a simple algorithm and sequentially add features until a final benchmark model is established. Then, the benchmark model is estimated with more complex algorithms. In only two cases ( $\Delta ROE$  and  $\Delta RNOA$ ), the more complex algorithms beat the benchmark. This approach suggests that a complex algorithm such as random forest, which has many parameters that need hyper-tuning and requires much computing power and time to make predictions, compared to LDA, which requires seconds to run and, in some cases, provides us with better prediction scores, is not a panacea.

Changes in cash flow-based measures of profitability ( $\Delta FCF$ ,  $\Delta CFO$ ) are predicted more accurately than accrual-based measures ( $\Delta ROA$ ,  $\Delta ROE$ ,  $\Delta RNOA$ ). This is a very interesting result and in line with the research of Foerster et al. (2017), who found that cash flow measures exhibit higher predictability compared to accrual measures. Adding variables for the mean reversion in all types of mean reversion that we examined (within the industry, country, and universe of firms) provided little to no improvement in predictability. Only in the case of DROA, DRNOA, and DCFO did we note an improvement of 1%-5% compared to the benchmark model, but overall, evidence that profitability’s mean reversion improves predictions cannot be supported.

In general, our results are in line with Anand et al. (2019). Anand et al. also found that cash-flow-based profitability measures can be forecasted with higher accuracy. On the contrary, we cannot support the idea that proxies for mean reversion improve the forecasting accuracy significantly in the European setting. Chen et al. (2022), who predicted profitability directional changes, found that machine learning outperformed traditional models; however, the comparison between econometric methods and machine learning was “unfair”. Chen et al. used two complex algorithms, random forests, and stochastic gradient boosting. On the other hand, relating to the novelty of our research, there are occasions where even simple algorithms, like LDA and KNN, can outperform more complex algorithms.

Our study contributes to the relevant literature by proposing a novel approach to how to use machine learning to predict profitability directional changes. In our approach, we began by estimating a simple model with a simple algorithm (LDA), and we sequentially added features until we established the best-performing model, which acted as the benchmark. Next, we used the benchmark model for predictions by employing more complex algorithms (KNN, random forest) and proved that a complex algorithm does not always guarantee improved forecasting performance. Our findings contribute to the growing body of literature that supports the idea that cash-flow-based measures of profitability are superior compared to accrual-based measures. Furthermore, we challenge the assumption that the mean reversion of profits enhances the forecasting accuracy, at least in the European setting.

Our research creates practical implications for both academic researchers and practitioners. The choice between simple and complex algorithms should be made by not only relying on predictive performance but also on computational efficiency. In addition to

that, practitioners (analysts, managers) should rely more on cash-flow measures when they assess profitability.

While this research offers some interesting insights, we acknowledge several limitations. First, we do not have available data for private companies. According to the literature, there is a difference in the quality of earnings between private and public firms (Ball and Shivakumar 2005). This difference can affect our results significantly. Second, there are further profitability models, like the DuPont analysis, which have not been tested in this research.

In this section, we propose potential directions for future research. It is interesting to investigate profitability prediction inter-regionally<sup>2</sup>. The different economic characteristics of groups of countries, like Europe, Africa, or Asia and the Pacific, certainly produce interesting implications and allow for useful comparisons. Finally, future researchers may opt to test forecasting profitability using raw accounting numbers from financial statements instead of ratios or combine raw numbers and ratios to create hybrid models.

**Author Contributions:** Conceptualization, N.D.B. and A.M.V.; methodology, A.M.V.; software, A.M.V.; validation, G.A.P. and A.M.V.; formal analysis, N.D.B.; investigation, N.D.B.; resources, N.D.B.; data curation, A.M.V.; writing—original draft preparation, N.D.B.; writing—review and editing, G.A.P.; visualization, A.M.V.; supervision, G.A.P.; project administration, G.A.P.; funding acquisition, G.A.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) under the 4th Call for HFRI PhD Fellowships (Fellowship Number: 009116) And the APC was funded by the University of Piraeus Research Center.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author, only with the permission of LSEG (Datastream).

**Acknowledgments:** The publication of this paper is partly supported by the University of Piraeus Research Center.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Description of variables.

Variable	Description
$ROA_{i,t}$	Income before extraordinary items minus net financial expense/Average total assets
$ROE_{i,t}$	Income before extraordinary items minus net financial expense/Average common equity
$RNOA_{i,t}$	Operating income before depreciation and amortization minus depreciation/Average net operating assets
$FCF_{i,t}$	Net income/Average total assets minus accruals
$CFO_{i,t}$	Net cash flow from operating activities/Average total assets
$\Delta ROA_{i,t}$	Dummy variable, 1 if $ROA_{i,t} > 0$ , otherwise 0
$\Delta ROE_{i,t}$	Dummy variable, 1 if $ROE_{i,t} > 0$ , otherwise 0
$\Delta RNOA_{i,t}$	Dummy variable, 1 if $RNOA_{i,t} > 0$ , otherwise 0
$\Delta FCF_{i,t}$	Dummy variable, 1 if $FCF_{i,t} > 0$ , otherwise 0
$\Delta CFO_{i,t}$	Dummy variable, 1 if $CFO_{i,t} > 0$ , otherwise 0

## Notes

- <sup>1</sup> Random walk is a random process where the probability of a firm to exhibit increased (decreased) profitability is equal to 0.5.
- <sup>2</sup> We thank the anonymous referee for raising this point.

## References

- Aghion, Philippe. 2002. Schumpeterian Growth Theory and the Dynamics of Income Inequality. *Econometrica* 70: 855–82. [CrossRef]
- Aghion, Philippe, Christopher Harris, Peter Howitt, and John Vickers. 2001. Competition, Imitation and Growth with Step-by-Step Innovation. *The Review of Economic Studies* 68: 467–92. [CrossRef]

- Anand, Vic, Robert Brunner, Kelechi Ikegwu, and Theodore Sougiannis. 2019. Predicting Profitability Using Machine Learning. SSRN 3466478. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3466478](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3466478) (accessed on 1 September 2022).
- Artikis, Panagiotis G., Lydia Diamantopoulou, Georgios A. Papanastasopoulos, and John N. Sorros. 2022. Asset growth and stock returns in European equity markets: Implications of investment and accounting distortions. *Journal of Corporate Finance* 73: 102193. [CrossRef]
- Ball, Ray, and Lakshmanan Shivakumar. 2005. Earnings quality in UK private firms: Comparative loss recognition timeliness. *Journal of Accounting and Economics* 39: 83–128. [CrossRef]
- Ball, Ray, and Ross Watts. 1972. Some time series properties of accounting income. *The Journal of Finance* 27: 663–81. [CrossRef]
- Bao, Yang, Bin Ke, Bin Li, Y. Julia Yu, and Jie Zhang. 2020. Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research* 58: 199–235. [CrossRef]
- Beaver, William H. 1968. The Information Content of Annual Earnings Announcements. *Journal of Accounting Research* 6: 67–92. [CrossRef]
- Beaver, William H. 1970. The Time Series Behavior of Earnings. *Journal of Accounting Research* 8: 62–99. [CrossRef]
- Breiman, Leo. 2001. Random Forests. *Machine Learning* 45: 5–32. [CrossRef]
- Chen, Xi, Yang Ha (Tony) Cho, Yiwey Dou, and Baruch Lev. 2022. Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data. *Journal of Accounting Research* 60: 467–515. [CrossRef]
- Deschamps, Benoit, and Dileep R. Mehta. 1980. Predictive ability and descriptive validity of earnings forecasting models. *The Journal of Finance* 35: 933–49. [CrossRef]
- Fairfield, Patricia M., Richard J. Sweeney, and Teri Lombardi Yohn. 1996. Accounting Classification and the Predictive Content of Earnings. *The Accounting Review* 71: 337–55.
- Fairfield, Patricia M., Sundaresh Ramnath, and Teri Lombardi Yohn. 2009. Do industry—level analyses improve forecasts of financial performance? *Journal of Accounting Research* 47: 147–78. [CrossRef]
- Foerster, Stephen, John Tsagarelis, and Grant Wang. 2017. Are Cash Flows Better Stock Return Predictors Than Profits? *Financial Analysts Journal* 73: 73–99. [CrossRef]
- Freeman, Robert N., James A. Ohlson, and Stephen H. Penman. 1982. Book Rate-of-Return and Prediction of Earnings Changes: An Empirical Investigation. *Journal of Accounting Research* 20: 639–53. [CrossRef]
- Gerakos, Joseph, and Robert Gramacy. 2013. *Regression-Based Earnings Forecasts*. Chicago Booth Research Paper. Amsterdam: SSRN.
- Healy, Paul, George Serafeim, Suraj Srinivasan, and Gwen Yu. 2014. Market competition, Earnings Management, and persistence in accounting profitability around the world. *Review of Accounting Studies* 19: 1281–308. [CrossRef]
- Kothari, S. 2001. Capital markets research in accounting. *Journal of Accounting and Economics* 31: 105–231. [CrossRef]
- Kureljusic, Marko, and Lucas Reisch. 2022. Revenue forecasting for European capital market-oriented firms: A comparative prediction study between financial analysts and machine learning models. *Corporate Ownership & Control* 19: 159–78.
- Lev, Baruch. 1983. Some economic determinants of time-series properties of earnings. *Journal of Accounting and Economics* 5: 31–48. [CrossRef]
- Li, Kevin K., and Partha Mohanram. 2014. Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies* 19: 1152–85. [CrossRef]
- Mullainathan, Sendhil, and Jann Spiess. 2017. Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives* 31: 87–106. [CrossRef]
- Nissim, Doron, and Stephen H. Penman. 2001. Ratio Analysis and Equity Valuation: From Research to Practice. *Review of Accounting Studies* 6: 109–54. [CrossRef]
- Ou, Jane A., and Stephen H. Penman. 1989. Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics* 11: 295–329. [CrossRef]
- Penman, Stephen H. 1991. An evaluation of accounting rate-of-return. *Journal of Accounting, Auditing & Finance* 6: 233–55.
- Penman, Stephen H., and Xiao-Jun Zhang. 2002. Accounting conservatism, the quality of earnings, and stock returns. *The Accounting Review* 77: 237–64. [CrossRef]
- Porter, Michael E. 1985. Technology and Competitive Advantage. *Journal of Business Strategy* 5: 60–78. [CrossRef]
- Rácz, Anita, Dávid Bajusz, and Károly Héberger. 2021. Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification. *Molecules* 26: 1111. [CrossRef]
- Ramnath, Sundaresh, Steve Rock, and Philip Shane. 2008. The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting* 24: 34–75. [CrossRef]
- Shanker, Murali, Michael Y. Hu, and Mingshing S. Hung. 1996. Effect of data standardization on neural network training. *Omega* 24: 385–97. [CrossRef]
- Varma, Sudhir, and Richard Simon. 2006. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7: 91. [CrossRef]
- Wahlen, James M., and Matthew M. Wieland. 2011. Can financial statement analysis beat consensus analysts' recommendations? *Review of Accounting Studies* 16: 89–115. [CrossRef]
- Xinyue, Chi, Xu Zhaoyu, and Zhou Yue. 2020. Using Machine Learning to Forecast Future Earnings. *Atlantic Economic Journal* 48: 543–45. [CrossRef]
- Zhang, Shichao. 2012. Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software* 85: 2541–52. [CrossRef]



---

Zhou, Zhi-Hua. 2012. *Ensemble Methods: Foundations and Algorithms*. Boca Raton: CRC Press.

Zhu, Weidong, Tianjiao Zhang, Yong Wu, Shaorong Li, and Zhimin Li. 2022. Research on optimization of an enterprise financial risk early warning method based on the DS-RF model. *International Review of Financial Analysis* 81: 102140. [[CrossRef](#)]

Zou, Hui, and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67: 301–20. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.