University of Piraeus

School of Economics, Business and International Studies

Department of Business Administration

The accounting treatment of Research and Development expenditures for financial firm performance: A Machine Learning approach

Antonios M. Vasilatos

A dissertation submitted

for the degree of

Doctor of Philosophy

in the

Department of Business Administration

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)

Piraeus, 2025

Παράρτημα Β: Βεβαίωση Εκπόνησης Διδακτορικής Διατριβής

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΧΕΙΡΗΜΑΤΙΚΩΝ ΚΑΙ ΔΙΕΘΝΩΝ ΣΠΟΥΔΩΝ
ΤΜΗΜΑ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΒΕΒΑΙΩΣΗ ΕΚΠΟΝΗΣΗΣ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

(περιλαμβάνεται ως ξεχωριστή (δεύτερη) σελίδα στο σώμα της διδακτορικής διατριβής)

Δηλώνω υπεύθυνα ότι η διδακτορική διατριβή για τη λήψη του διδακτορικού τίτλου, του Τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων του Πανεπιστημίου Πειραιώς, με τίτλο

«The accounting treatment of Research and Development expenditures for financial firm performance: A Machine Learning approach (Τίτλος στα Ελληνικά: Η λογιστική αντιμετώπιση των δαπανών Έρευνας και Ανάπτυξης για την χρηματοοικονομική απόδοση των εταιρειών: Μία προσέγγιση με Μηχανική Μάθηση)»

έχει συγγραφεί από εμένα αποκλειστικά και στο σύνολό της. Δεν έχει υποβληθεί ούτε έχει εγκριθεί στο πλαίσιο κάποιου άλλου διδακτορικού, μεταπτυχιακού προγράμματος ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό, ούτε είναι εργασία ή τμήμα εργασίας ακαδημαϊκού ή επαγγελματικού χαρακτήρα.

Δηλώνω επίσης υπεύθυνα ότι οι πηγές στις οποίες ανέτρεξα για την εκπόνηση της συγκεκριμένης εργασίας αναφέρονται στο σύνολό τους, κάνοντας πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.

Υπογραφή Διδάκτορα..............................

Ονοματεπώνυμο...Αντώνιος...Βασιλάτος....

Ημερομηνία............08/01/2025..........

Examining Committee

Prof. Georgios Papanastasopoulos, Supervisor

Prof. Emeritus Ioannis Sorros, Advisory and Examining Committee

Assoc. Prof. Vasileios Zisis, Advisory and Examining Committee

Prof. Michael Sfakianakis, Examining Committee

Prof. Panagiotis Artikis, Examining Committee

Prof. Michael Bekiaris, Examining Commitee

Assist. Prof. Nikolaos Belesis, Examining Commitee

Εξεταστική Επιτροπή

Καθ. Γεώργιος Παπαναστασόπουλος, Επιβλέπων

Ομ. Καθ. Ιωάννης Σώρρος, Συμβουλευτική και Εξεταστική Επιτροπή

Αναπλ. Καθηγητής Βασίλειος Ζήσης, Συμβουλευτική και Εξεταστική Επιτροπή

Καθ. Μιχαήλ Σφακιανάκης, Εξεταστική Επιτροπή

Καθ. Παναγιώτης Αρτίκης, Εξεταστική Επιτροπή

Καθ. Μιχαήλ Μπεκιάρης, Εξεταστική Επιτροπή

Επικ. Καθ. Νικόλαος Μπέλεσης, Εξεταστική Επιτροπή

Conferences and publications list

A. Articles in Scientific Journals (Refereed)

1. "R&D accounting treatment and firm future performance: Evidence from Greece", *Theoretical Economics Letters*, 2024, vol.14, (ABS: 1) (with Georgios A. Papanastasopoulos).

2. "Predicting the Profitability of Directional Changes Using Machine Learning: Evidence from European Countries", *Journal of Risk Financial Management*, 2023, vol. 16, (ABDC: B) (with Nikolaos Belesis and Georgios A. Papanastasopoulos).

B. Working Papers

1. "Can machine learning enhance the forecasting of herding behavior in international stock markets?", *IMA Journal of Management Mathematics*, 2024, *revise and resubmit offer (minor revision)*, (ABS: 2), (with Panagiotis Artikis, Georgios Panapanastasopoulos and Polixeni Tsitsiri).

C.1. Publications in Conference Proceedings (Refereed)

1. "Benchmarking Machine Learning Algorithms to Predict Profitability Directional Changes", *Business Analytics and Decision Making in Practice. ICBAP 2024. Lecture Notes in Operations Research*, Springer, (with Panagiotis Artikis, Nikolaos Belesis and Georgios Papanastasopoulos).

2. “Can Machine Learning Enhance the Forecasting of Herding Behavior in International Stock Markets?”, *Business Analytics and Decision Making in Practice. ICBAP 2024. Lecture Notes in Operations Research*, Springer, (with Panagiotis Artikis, Georgios Papanastasopoulos and Polixeni Tsitsiri).

C.2. Conference Presentations (Refereed)

1. ”GDP Growth And R&D Accounting Treatment, *MENA-Asia FEBS Conference*, Sharjah, 2024.

2. “Benchmarking Machine Learning algorithms to predict profitability directional changes”, *International Conference on Business Analytics in Practice*, Sharjah, 2024.

3. “R&D costs capitalization and firm financial performance: Evidence from Europe”, *Southern Economic Association 93rd Annual Meeting*, New Orleans, 2023.

4. “Capitalization versus expensing of R&D costs under IAS 38: An empirical investigation using machine learning”, *European Financial Management Association 2023, Annual Meeting*, Cardiff, 2023.

5. “Forecasting herding behavior during COVID-19: A machine learning approach”, *Behavioral Finance Working Group-16th Annual Conference*, London, 2023.

6. “The Implementation of IAS 38 in Greece: which determinants drive management to capitalize R&D costs and how their choice affects future firm performance?”, *10th PhD Symposium on Contemporary Greece & Cyprus, Hellenic Observatory London School of Economics*, London, 2023.

7. “Benchmarking Machine Learning Algorithms to Classify R&D Capitalizers and Expensers Under IAS 38”, *22nd Annual Conference of the Hellenic Finance and Accounting Association (HFAA)*, Athens, 2023.

8. “Financial analysts versus accountants: Who can better predict future accounting choice? An exploration of the R&D field using machine learning”, *12th Inter-*

national Conference of the Financial Engineering and Banking Society (FEBS), Chania, 2023.

9. "Further Exploration of Profitability Directional Changes Prediction using Machine Learning: Evidence from European Countries", *21st Annual Conference of the Hellenic Finance and Accounting Association (HFAA)*, Athens, 2022.

10. "R&D Costs Accounting Treatment and Value Relevance: A Literature Review of Pre- and Post-IFRS Evidence", *20th Annual Conference of the Hellenic Finance and Accounting Association (HFAA)*, Thessaloniki, 2021.

Ευχαριστίες

Παρόλο που η συγγραφή μιας διδακτορικής διατριβής αποτελεί, κατά κύριο λόγο, ένα προσωπικό και συχνά μοναχικό ταξίδι, η ολοκλήρωση του συγκεκριμένου έργου δεν θα ήταν εφικτή χωρίς τη συμβολή πολλών ανθρώπων. Άλλοι με αμεσότητα και άλλοι με πιο έμμεσο τρόπο, προσέφεραν τη γνώση, την καθοδήγηση, την υποστήριξη και την ενθάρρυνσή τους, συμβάλλοντας καθοριστικά στην επιτυχή ολοκλήρωση αυτής της διατριβής.

Πρωτίστως, θα ήθελα να εκφράσω την ειλικρινή ευγνωμοσύνη μου στην οικογένειά μου. Χωρίς την αδιάλειπτη στήριξη, την υπομονή και την πίστη τους στις δυνατότητές μου, η ολοκλήρωση αυτής της διατριβής δεν θα ήταν εφικτή. Οι γονείς μου, ιδιαιτέρως, στάθηκαν ακούραστα στο πλευρό μου καθ' όλη τη διάρκεια της ακαδημαϊκής μου πορείας, παρέχοντάς μου τη συναισθηματική ενθάρρυνση και την υλική υποστήριξη που χρειαζόμουν για να φτάσω έως εδώ. Ιδιαίτερη μνεία αξίζει στον πατέρα μου, Μιχάλη, ο οποίος υπήρξε ο πρώτος μου «δάσκαλος» στη Λογιστική.

Στη συνέχεια θα ήθελα να ευχαριστήσω θερμά τα μέλη της τριμελούς συμβουλευτικής επιτροπής για την υποστήριξή τους και το συνεχές ενδιαφέρον τους για την πορεία της διατριβής μου. Ο επιβλέπων καθηγητής μου, Γεώργιος Παπαναστασόπουλος, στάθηκε αδιάλειπτα στο πλευρό μου σε κάθε στάδιο της εκπόνησης της διατριβής. Με την πολύτιμη καθοδήγησή του, την υπομονή και τη συνεχή ενθάρρυνσή του, με ώθησε να εμβαθύνω στις ερευνητικές μου δραστηριότητες και να αναπτύξω μια πιο ολοκληρωμένη και κριτική προσέγγιση στο αντικείμενό μου. Ιδιαίτερη αναφορά θα κάνω στον Ομότιμο Καθηγητή Ιωάννη Σώρρο, ο οποίος για μένα αποτελεί πρότυπο Πανεπιστημιακού Δασκάλου και νιώθω ιδιαίτερα τυχερός που τον είχα δίπλα μου. Τον ευχαριστώ θερμά για την βοήθεια του και το συνεχές ενδιαφέρον του για την πορεία μου. Τέλος, θα ήθελα να ευχαριστήσω τον Αναπλ. Καθηγητή Βασίλειο Ζήση για το ενδιαφέρον του ως προς την πορεία της διατριβής και για τη συνεργασία μας.

Ευχαριστώ θερμά τον Καθηγητή Παναγιώτη Αρτίκη για την υποστήριξη του καθώς και για τη συνεργασία μας στα επιστημονικά συνέδρια στα οποία συμμετείχαμε, και τον Επίκουρο Καθηγητή Νικόλαο Μπέλεση για το συνεχές του ενδιαφέρον, υποστήριξη και συνεργασία.

Τέλος, ευχαριστώ την Κα Αγάπη Φατσή για την εξαιρετική μας συνεργασία στα πλαίσια του ΠΜΣ ΛΕΕΔΟ και τη συνεχή της υποστήριξη και ενδιαφέρον. Ευχαριστώ θερμά τους συναδέλφους μου, Δρ. Αστέριο Στρουμπούλη και Κα Πολυξένη Τσιτσίρη.

Acknowledgment

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).

The accounting treatment of Research and Development expenditures for financial firm performance: A Machine Learning approach

Antonios M. Vasilatos

Abstract

This thesis analyses the predictive ability of R&D expenditures, on future firm financial performance, utilizing a European sample of publicly listed firms that reported under IFRS from 2005 to 2020. This study employs a forward-looking, out-of-sample predictive approach, contrasting with prior research that predominantly examines the relationship between R&D expenditure and subsequent profitability through in-sample regression analysis. This re-frames the enduring argument on R&D accounting treatment—whether to capitalize or expense—as a forecasting issue, coinciding with contemporary literature advocating for predictive analysis over solely explanatory methods.

Machine learning algorithms are employed, such as, logistic regression, random forests, SVM and XGBoost, to predict one-step-ahead profitability measures such as ROA, Price, Returns, and EPS. This work departs from conventional econometric modeling, which focuses on parameter estimation, and illustrates the ability of machine learning algorithms to identify non-linear correlations and produce more precise out-of-sample predictions. Furthermore, it presents a data-driven approach by comparing theoretically produced financial ratios with models constructed from raw accounting data. Remarkably, the latter frequently demonstrate comparable, if not superior, predictive ability for future firm financial performance.

To gain insights into the contribution of R&D expenditures to future firm performance, feature importance metrics and SHAP (SHapley Additive exPlanations)

values are used to improve the interpretability of ML models, offering detailed insights into the impact of capitalized vs expensed R&D expenditures on predictive accuracy. Despite previous findings that capitalization promotes earnings manipulation and diminishes predictive accuracy, the results demonstrate that capitalized R&D costs can enhance future financial performance predictions. This indicates that the existing accounting treatment under IFRS, which allows for the capitalization of development expenditures under certain conditions, may more effectively communicate the information required by investors and policymakers to forecast future firm financial performance.

This research contributes to the accounting and finance literature by providing substantial out-of-sample evidence about the predictive significance of R&D expenditures. This study offers methodological guidance for incorporating machine learning techniques into accounting research and explains the contentious topic of R&D capitalization, potentially advising standard setters, practitioners, and investors on the optimal interpretation and application of R&D information for prospective decision-making.

Η λογιστική αντιμετώπιση των δαπανών Έρευνας και Ανάπτυξης για την χρηματοοικονομική απόδοση των εταιρειών: Μια προσέγγιση με Μηχανική Μάθηση

Αντώνιος Μ. Βασιλάτος

Περίληψη

Αυτή η διατριβή αναλύει την προβλεπτική ικανότητα των δαπανών Ε&Α για την μελλοντική χρηματοοικονομική απόδοση των εταιρειών, χρησιμοποιώντας ένα ευρωπαϊκό δείγμα εισηγμένων εταιρειών που συντάσσουν τις οικονομικές τους καταστάσεις σύμφωνα με τα ΔΠΧΑ από το 2005 έως το 2020. Αυτή η μελέτη χρησιμοποιεί μια εκτός δείγματος προγνωστική προσέγγιση, σε αντίθεση με προηγούμενες έρευνες που εξετάζουν κυρίως τη σχέση μεταξύ δαπανών Ε&Α και της μελλοντικής κερδοφορίας μέσω ανάλυσης παλινδρόμησης εντός δείγματος. Αυτό επαναδιατυπώνει τη διαρκή συζήτηση σχετικά με την λογιστική αντιμετώπιση των δαπανών για Ε&Α—αν θα πρέπει να κεφαλαιοποιούνται ή να καταχωρούνται ως έξοδα—ως ένα ζήτημα πρόβλεψης, ευθυγραμμιζόμενο με τη σύγχρονη βιβλιογραφία που υποστηρίζει την προγνωστική ανάλυση έναντι των αποκλειστικά επεξηγηματικών μεθόδων.

Χρησιμοποιούνται αλγόριθμοι μηχανικής μάθησης, όπως οι logistic regression, random forest, SVM and XGBoost, για την πρόβλεψη δεικτών κερδοφορίας, όπως ο δείκτης αποδοτικότητας συνολικών κεφαλαίων, η τιμή, οι αποδόσεις και τα κέρδη ανά μετοχή. Αυτή η εργασία απομακρύνεται από την παραδοσιακή οικονομετρική ανάλυση, η οποία επικεντρώνεται στην εκτίμηση παραμέτρων, και απεικονίζει την ικανότητα των αλγορίθμων μηχανικής μάθησης να εντοπίζουν μη γραμμικές συσχετίσεις και να παράγουν πιο ακριβείς προβλέψεις εκτός δείγματος. Επιπλέον, παρουσιάζει μια προσέγγιση βασισμένη σε δεδομένα συγκρίνοντας θεωρητικούς χρηματοοικονομικούς δείκτες με μοντέλα κατασκευασμένα από ακατέργαστα λογιστικά δεδομένα. Αξιοσημείωτο είναι

ότι τα τελευταία συχνά επιδεικνύουν συγκρίσιμη, αν όχι ανώτερη, ικανότητα πρόβλεψης της μελλοντικής χρηματοοικονομικής απόδοσης της εταιρείας.

Για να αποκτήσουμε περαιτέρω πληροφόρηση σχετικά με τη συμβολή των δαπανών Έρευνας και Ανάπτυξης (Ε&Α) στην μελλοντική απόδοση των εταιρειών, χρησιμοποιούνται μέτρα «σημαντικότητας χαρακτηριστικών» και τιμές SHAP (SHapley Additive exPlanations) για τη βελτίωση της ερμηνευσιμότητας των μοντέλων μηχανικής μάθησης, προσφέροντας λεπτομερείς πληροφορίες για την επίδραση των κεφαλαιοποιημένων έναντι των «εξοδοποιημένων» δαπανών για Ε&Α στην προβλεπτική ακρίβεια. Παρά τα προηγούμενα ευρήματα που δείχνουν ότι η κεφαλαιοποίηση προάγει την χειραγώγιση των κερδών και μειώνει την προβλεπτική ακρίβεια, τα αποτελέσματα δείχνουν ότι οι κεφαλαιοποιημένες δαπάνες για Ε&Α μπορούν να βελτιώσουν τις προβλέψεις για τη μελλοντική χρηματοοικονομική απόδοση. Αυτό υποδηλώνει ότι η υπάρχουσα λογιστική αντιμετώπιση σύμφωνα με τα ΔΠΧΑ, η οποία επιτρέπει την κεφαλαιοποίηση των δαπανών ανάπτυξης υπό ορισμένες προϋποθέσεις, μπορεί να επικοινωνεί πιο αποτελεσματικά τις πληροφορίες που απαιτούνται από τους επενδυτές και τους υπεύθυνους χάραξης πολιτικής για την πρόβλεψη της μελλοντικής χρηματοοικονομικής απόδοσης της εταιρείας.

Αυτή η έρευνα συμβάλλει στη βιβλιογραφία της λογιστικής και της χρηματοοικονομικής παρέχοντας σημαντικά στοιχεία εκτός δείγματος σχετικά με την προβλεπτική ικανότητα των δαπανών Έρευνας και Ανάπτυξης (Ε&Α). Αυτή η μελέτη προσφέρει μεθοδολογική καθοδήγηση για την ενσωμάτωση τεχνικών μηχανικής μάθησης στην λογιστική έρευνα και εξηγεί το αμφιλεγόμενο θέμα της κεφαλαιοποίησης των δαπανών για Ε&Α, ενδεχομένως συμβουλεύοντας τις ρυθμιστικές αρχές, τους επαγγελματίες της λογιστικής και τους επενδυτές για την βέλτιστη ερμηνεία και εφαρμογή των πληροφοριών Ε&Α για μελλοντική λήψη αποφάσεων.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

R&D issues in economics is a heavily researched field, as innovation plays a vital role in economic growth. According to the exogenous growth model (Solow-Swan model), advancements in productivity are mostly propelled by technical progress (Solow, 1956). Romer (1990) argued that ”technological change lies at the heart of economic growth”. During the 70s and 80s, R&D research in the accounting literature was sparse. In the late 80s and the 90s, a series of seminal papers by Bublitz and Ettredge (1989), S. H. Chan et al. (1990), Lev and Sougiannis (1996), and Woolridge (1988) examined R&D relevance, investor reaction to firms' R&D announcements and analysts' forecast errors.

Since then, R&D accounting, as a research stream in the accounting and finance literature is constantly growing. Researchers are motivated by the increased importance of the intangible assets. Kaplan and Norton (2004) found that in 2002, intellectual assets consisted of nearly 70% of firm's assets. Evidence from Corrado and Hulten (2010) suggests that investment rate in intangibles has surpassed investment in tangibles. One of the most prominent academics in the intangibles field, Lev (2000), argued that the increase in intangible assets can be attributed to the increased business competition and the expansion of information technologies.

One of the issues that was examined by more recent studies was the R&D capitalization and its relationship with future profitability. The seminal paper of Aboody and Lev (1998) was the first that examined the R&D capitalization determinants. Since then, two theories have emerged. The opponents of capitalization suggest that capitalization requires managerial judgment and it allows for earnings management, therefore R&D expenses should be expensed as incurred (Burgstahler & Dichev, 1997; Cazavan-Jeny & Jeanjean, 2006; Cazavan-Jeny et al., 2011; Dinh et al., 2015; Eierle & Wencki, 2016; Osma & Young, 2009). On the other hand, the proponents of capitalization support that managers should be allowed to have the discretion to capitalize R&D costs, as capitalization can be used as a signal to the market for improved future financial performance (Market Signal Theory) (K. Ahmed & Falk, 2006; Goodwin & Ahmed, 2006; Hughes & Kao, 1991; D. R. Oswald & Zarowin, 2007).

Apart from the difference in the academic findings, standard setters also view the accounting treatment of R&D differently. Under US GAAP, R&D costs are in general expensed as occurred while under IFRS, development costs (the "D" in R&D as mentioned by Lev (2019)), under circumstances, must be capitalized. In the paper of Lev (2019), the views of Doidge et al. (2018) and Paton et al. (1940) are presented. Doidge et al. (2018) gave the example of salaries for researchers. If those salaries are not treated as investments but rather as a cost, the profitability of the firm is decreased. In this case, accounting is not informative and investors will be skeptical about firm value. In the classic book of Paton et al. (1940), accounting's main objectives are *"the periodic income determination, and the division of the stream of costs incurred between the present and the future in the process of measuring periodic income"*, where Lev (2019) paraphrased as *"the stream of (intangible) costs incurred between the present (expensed) and the future (capitalized) in the process of measuring periodic income"*.

So far in the literature, researchers have primarily focused on examining the determinants of capitalization and if R&D costs or whether capitalized development costs (for short from now on, capitalized R&D) are value-relevant and their relationship with future profitability measures. They have created theoretical models and have tested them in-sample, using regression analysis. Therefore, they quantify the relationship between the dependent variable and the independent variables. Yet, the question that so far has remained unanswered is whether capitalized R&D costs truly exhibit forecasting power and are able to forecast future firm financial performance. More specifically, in IAS 38, one of the critical attributes of the intangible assets is that future economic benefits (revenues or reduced future costs) are expected from the the recognition of the intangible asset. Furthermore, literature suggests that R&D costs are related with high uncertainty of future earnings (Kothari et al., 2002). In fact, FASB's statement regarding R&D costs, the trade-off between relevance and uncertainty of future benefits is a major consideration. Therefore, this issue motivates the first part of this thesis, meaning to approach the R&D accounting treatment and firm's future performance as a forecasting task.

Literature suggests that although explanatory analysis is the mainstream approach, often the importance of prediction problems in business and economics is neglected (Kleinberg et al., 2015; Mikko Ranta & Järvenpää, 2023; Q. Zhao et al., 2023). Moreover, it is a common belief that in-sample evidence of predictability does not mean that there will be significant out-of-sample predictability, which is often seen as a sign that in-sample evidence can be spurious. Out-of-sample forecasts can be closer to reality and simulate the challenges that real-time forecasters face (Inoue & Kilian, 2005). The call to shift from the explanatory to the predictive analysis, as well as the benefits of out-of-sample predictions further motivate this thesis to address a classic accounting issue as a forecasting exercise.

The last decade AI (artificial intelligence) is changing the world, and it is said that nowadays we live in the second digital revolution era (Brynjolfsson & McAfee, 2014). This digital revolution included big data, ML (machine learning), algorithms, which change businesses, decision making, accounting (Auvinen et al., 2018; Shrestha et al., 2019). Mikko Ranta and Järvenpää (2023) suggest that ML has revealed new opportunities in accounting research. They suggest that one of the most promising areas to use ML in accounting is to obtain better predictions. Therefore, the rapid expansion of AI and the call of researchers to use ML in accounting research have motivated the second part of this thesis to approach its forecasting tasks with the use of ML.

Why ML is more suitable than traditional econometrics? In econometrics, a function of the form $y = f(\beta, \chi)$ is specified, and parameters $\hat{\beta}$ are estimated by solving an optimization problem (Anand et al., 2019; Mullainathan & Spiess, 2017). Then, the estimated $\hat{\beta}$ are used in a different sample than the one used for estimation, to produce out-of-sample predictions (Elliott & Timmermann, 2008). On the contrary, ML produces out-of-sample predictions by finding patters in the data. ML algorithms make predictions on y by using data χ (Varian, 2014). Therefore, ML focuses on $\hat{y}$ rather than $\hat{\beta}$ (Mullainathan & Spiess, 2017), which is more suited to a forecasting task.

Taking all the above into account, this thesis aims to provide new insights on how R&D expenditures predict future financial performance for firms.

1.2 Objectives

The present thesis focuses on whether R&D costs, and more specifically capitalized R&D costs, have predictive power for firm's future performance using a European sample of listed firms during the period 2005-2020. European listed firms are chosen because they report their R&D costs under IAS 38, therefore capitalization of development costs is allowed. The sample starts in 2005 because that was the first year that IAS 38 was implemented. The purpose of this study is twofold.

First, since the R&D costs accounting treatment is an ongoing debate, this thesis seeks to approach this issue in a different way compared to the existing literature. Rather than focusing on explaining why capitalized or expensed R&D costs explain or do not explain future financial performance, this study addresses the issue by providing out-of-sample evidence on the predictive ability of R&D costs.

Secondly, because recent studies call for accounting researchers to apply ML in their studies and because ML algorithms are better suited to forecasting tasks, in this study a range of algorithms is utilized. Both simple and more complex algorithms are used in order to find which one is better suited to the specific dataset. Further, sets of raw accounting data from the financial statements are tested in order to examine their predictive power compared to theoretically specified financial ratios.

1.3 Methodology

In order to obtain one-step-ahead out-of-sample predictions, the models of Cazavan-Jeny and Jeanjean (2006) and Cazavan-Jeny et al. (2011) are used. Firm's financial performance is proxied by ROA, Price, Returns and EPS. Out-of-sample forecasts are obtained and the employed algorithm's (logistic regression, random forest, XGB) is compared in order to establish the benchmark model. Directional changes of profitability are predicted, as predictions for the level of the measures has not yielded satisfactory predictive performance (X. Chen et al., 2022).

In the next step of the analysis, a data-driven approach is performed, by using raw accounting data from the financial statements instead of theoretically specified financial ratios. This approach was inspired by the papers of Bao et al. (2020) and X. Chen et al. (2022). The raw items were transformed to have the same scale as in Bao et al. (2020). Moreover, different sets of raw items were constructed using correlation analysis as a feature selection mechanism. The out-of-sample performance of the models is evaluated by the AUC (Area Under Curve) score. Comparisons between the theoretical models and the raw accounting items suggest that the data-driven approach can perform equally well (and in some cases better).

Mikko Ranta and Järvenpää (2023) suggested that in accounting research which uses ML, focus should be given in the use of explainable AI, which helps with model interpretation. Therefore, to shed light in the "black box" of ML, coefficients magnitude, feature importance and SHAP values are used to interpret the models. In this way, the predictive power of R&D costs is examined in detail, as for each R&D feature of interest, the contribution to the forecasting of profitability is derived. SHAP values is the most "state of the art" tool to interpret ML models and it has been used as an alternative to feature importance (Bali et al., 2023).

Finally, according to Aboody and Lev (1998), analysts stated that R&D capitalization is positively related with earnings forecast errors. Typically, studies in the field reverse capitalization effect by adjusting earnings and total assets. Unadjusted variables have been used to examine the prediction performance and whether it is deteriorated because of capitalization. Prediction performance of the algorithms did not deteriorate, and in fact, there was a slight improvement for ROA, Price and EPS.

1.4 Contribution

This thesis significantly contributes to the accounting and finance literature by examining the predictive capacity of R&D spending on firm's future financial performance using an innovative methodological approach. The principal contributions are as follows:

This study shifts from the main focus of past research, which has mainly used in-sample regression analyses to explain the relationship between R&D costs and future profitability, to an out-of-sample predictive framework. The thesis offers a novel perspective on the ongoing debate surrounding the capitalization of R&D costs by framing the relationship as a forecasting task. This approach corresponds with the recommendations from Kleinberg et al. (2015) and Mikko Ranta and Järvenpää (2023) for increased focus on prediction issues in business and economics, providing a more pragmatic assessment of the significance of R&D expenditures.

This work addresses the recent call for the incorporation of artificial intelligence and machine learning (ML) into accounting research (Mikko Ranta & Järvenpää, 2023) by utilizing various ML techniques, such as logistic regression, random forests, and XGBoost, to predict future firm performance. The thesis advances accounting research methodology by illustrating the superiority of machine learning approaches compared to standard econometric models in forecasting financial outcomes. It demonstrates how machine learning can manage complex, non-linear relationships and identify patterns in data that conventional models may overlook.

Expanding upon the research of Bao et al. (2020) and X. Chen et al. (2022), the thesis employs a data-driven methodology by leveraging unprocessed accounting items from financial statements rather than depending exclusively on theoretically defined financial ratios. This strategy facilitates a more detailed examination of how specific accounting entries influence future performance forecasts. The research indicates that models utilizing raw accounting data might achieve comparable or superior performance to those relying on conventional financial ratios, implying that significant predictive insights may be contained within the raw financial statement components.

To address the frequently referenced ”black box” issue related to machine learning models, this research utilizes uses AI methodologies, including coefficient magnitude, feature importance analysis and SHAP (SHapley Additive exPlanations). This enables a deeper understanding of how different features, especially capitalized and expensed R&D costs, affect profitability prediction outcomes. This interpretation of the models enhances the transparency and usefulness of machine learning in accounting, rendering the results more accessible to practitioners and policymakers.

This research investigates the impact of capitalizing R&D expenses on the prediction efficacy of financial models. Despite the findings of prior studies that capitalization could impair predicting accuracy due to possible earnings management (Cazavan-Jeny & Jeanjean, 2006), the study reveals that predictive performance does not decline—and may even enhance—when unadjusted variables are employed. This offers empirical evidence that underscores the significance of capitalized R&D costs in predicting future financial success, so contributing to the discourse on suitable accounting approaches for R&D spending.

Out-of-sample evidence is presented about the predictive ability of capitalized R&D costs, offering significant insights for accounting standard setters and investors. The results indicate that the capitalization of development expenditures under IFRS could improve the informativeness of financial statements for future performance. This affects the current dialogues regarding the harmonization of accounting standards and the possible implementation of similar practices under US GAAP.

The study enhances the literature regarding the impact of intangible assets on firm valuation. By empirically establishing the predictive significance of R&D expenditures through sophisticated analytical methods, it underscores the critical role of intangible assets in contemporary economies, reflecting the insights of Lev (2000) and Kaplan and Norton (2004) regarding the growing importance of intellectual assets within firms' total assets.

In summary, this thesis contributes to both the theoretical and practical understanding of R&D accounting and its implications for future financial performance. It bridges the gap between traditional accounting research and modern analytical methods, providing a foundation for future studies to explore predictive modeling and machine learning applications in accounting and finance.

1.5 Structure of the thesis

The structure of the thesis is as follows. Chapter 2 presents the theoretical framework and the literature relevant for this study. Chapter 3 presents the methodology, empirical approach and data description. In Chapter 4, the empirical results are presented. Finally, in Chapter 5, the conclusions are summarized and directions for future research are given.

Chapter 2

Literature review

2.1 Introduction

This chapter reviews the relevant literature of the thesis. The first section briefly discusses the theoretical framework of accounting policies. Upon these early economic theories, researchers which examined the field of R&D costs and firm profitability have built their research hypotheses. Then, the most important papers about R&D costs and firm's financial performance are presented. Moreover, the literature about the determinants of R&D accounting treatment is presented.

2.2 The theory of the firm and accounting policies

The theory of the firm seeks to explain, based on economic theory, how firms make decisions, and how they optimize production and distribute their products to achieve their goals. One of the firm's main goal is the maximization of profits. Since Adam Smith's "An Inquiry Into the Nature and Causes of the Wealth of Nations", four distinct theories about the firm have developed over time, namely, The Neoclassical Theory, The Transactions Cost Theory, The Principal-Agent Theory and the Evolutionary Theory (Kantarelis, 2010). Since these theories though, firms have continued to grow, they have became more important for the modern economies, they have not always operated in competitive markets and they have been controlled by managers and not solely the owner. Thus, the need for new theories has emerged (Pass & Lowes, 1978)

Monsen and Downs (1965) introduced the Theory of Large Managerial Firms. According to their theory, owners desire dividends and a constant price rise of the firm's stock; yet, they do not engage in firm's, however, they do not participate in the firm's management, making them unable to pursue profit-maximizing behavior. On the other hand, managers seek to maximize their own income. The separation in ownership and management leads large firms to report large expenses, less variable earnings and be very cautious on their research programs.

Schiff (1966) explained that when managers act in their own self-interest, their decisions create outcomes that give owners the impression of profit maximization. Since owners are unaware of the alternative policies available to the firm, they have no way of knowing if profits are truly maximized. Consequently, owners often act as "satisficers" rather than "maximizers." They desire maximum profits, but due to limited information, they adopt a behaviour that differs from that of a theoretical maximizer. In corporations with diffuse ownership, if there is a conflict between the interests of managers and owners [1], the ability to select the accounting method provides managers with a significant advantage.

Schiff's study *"Accounting Tactics and the Theory of the Firm"* was one of the first that tried to validate the Theory of Large Managerial Firms and examined the case of managers who use accounting policies as instrument in their attempt to satisfy stockholders. As an example of this phenomenon, Schiff examined the accounting treatment of advertising expenditures. In 1960, the listed in the NYSE company, Chock Full O' Nuts, reported that it deferred and amortized over the period of the expected sales, the advertising and promotion costs related to the development of new markets. The increase in the net income per share ratio in the following years was attributed in the increased amount of advertising costs that have been deferred. In 1964, the company reported losses. In addition to this, in the annual report, management restated the EPS ratio of the previous years by charging back previously capitalized advertising costs.

[1] See the Principal-Agent Theory as introduced by Ross (1973) and Mitnick (1973)

The official excuse for this policy shift was that the products for which the advertisement costs were capitalized, were not expected to bring future benefits to the company. By not writing-off the capitalized advertising costs in 1964 the company, management presented an increase in EPS for 1964. Therefore, management presented a smooth increasing trend in EPS to the stockholders, compatible with the Theory of Large Managerial Firms.

In his report, Peles (1970) presented a methodology to decide upon the amortization rate of the intangible assets created by advertising expenditure. These rates may be used to determine which proportion of the advertising expenses should be capitalized and which should be expensed as incurred. Peles examined three industries; beer, cigarette and cars. The amortization rate was decided based on the effect of advertising expenditure on sales. The purpose of the study was to present an objective and uniform method for amortizing advertising expenditure and discourage accounting policy shifts, such as in the case of Chock Full O' Nuts. In his final remark, he suggests that his method may be appropriate for the accounting treatment of R&D expenditures too.

2.2.1 The Positive Accounting Theory

R. L. Watts and Zimmerman (1978) in their influential paper entitled *"Towards a Positive Theory of the Determination of Accounting Standards"* introduced the Positive Accounting Theory. Their paper makes two key assumptions. Managers seek to optimise their personal utility, which encompasses other types of wealth or benefits in addition to monetary pay. Managers make accounting decisions based on how they think they will improve their own position: by lowering corporate expenses like taxes and political risks, or by raising their incentive compensation. In terms of accounting standards, company behaviour is shaped by this self-interest. Furthermore, accounting policies are chosen by management based on what will best serve their own financial interests. This means that the potential impacts of accounting standards on cash flows, taxes, and political or regulatory scrutiny are taken into consideration in addition to their correctness or compliance. Managers might, for instance, favour accounting methods that overestimate earnings in order to get bonuses or reduce reported profits in order minimise political costs. R. L. Watts and Zimmerman (1978) formulated the following three hypotheses:

(a) *Bonus Plan Hypothesis:* In order to maximise their bonuses, managers who's compensation is based on reported earnings are inclined to use accounting methods that boost earnings. If a manager's bonus is contingent on profitability, for instance, they may opt for strategies like accelerated revenue recognition in order to boost the company's reported income.

(b) *Debt Covenant Hypothesis:* In order to remain within the restrictions of their debt covenants, firms that are about to violate them may use accounting techniques that overestimate profits or decrease liabilities. This strategy assists the company in avoiding fines, further limitations, or expenses related to covenant violations.

(c) *Political Cost Hypothesis:* In order to avoid political costs like increased regulatory scrutiny, criticism from the public, or even wealth transfers as a result of government intervention, bigger firms or those under political scrutiny may use conservative accounting techniques to minimise reported earnings. For example, oil companies may declare lower revenue during periods of heightened public and political attention in order to mitigate regulatory pressure or requests for more taxation.

R. L. Watts and Zimmerman (1979) state that although not in the way that theorists had originally intended, accounting theories have traditionally had a considerable impact on the content of financial statements. Accounting theories are frequently utilised to support pre-existing assumptions rather than offering a uniform foundation for appropriate financial reporting methods, such standard-setting bodies. The authors believe that no normative theory that aims to apply universal accounting principles can explain accounting standards. Rather, a self-interest theory is better suited to explain why accounting rules are justified, as it takes into consideration the various vested interest groups (such as legislators, investors, and managers) that prevail on various problems.

Because it is unpopular politically, this self-interest hypothesis is rarely employed to justify standards. The authors argue that there will never be a widely recognised accounting theory that can justify all standards due to the effect of vested interests and the variety of arguments for different accounting standards.

The initial empirical studies in accounting used agency costs and compensation contracts to predict accounting choice (R. L. Watts & Zimmerman, 1990). Later researchers started using the term of contracting costs (Klein, 1983). According to R. L. Watts and Zimmerman (1990), contracting costs fall into multiple categories and originate from different sources. These include transaction costs such as broking and legal fees for facilitating transactions; agency costs, which include bonding and monitoring expenses as well as losses from agents' inefficient decisions; information costs, which are incurred during the process of obtaining necessary information; renegotiation costs, which are incurred when unforeseen circumstances necessitate updating existing contracts; and bankruptcy costs, which include legal fees and costs resulting from poor decision-making that may lead to bankruptcy. The word ”contracting parties” refers to both external stakeholders like suppliers, creditors, and consumers as well as internal business members, like employees and managers.

The term ”accepted set” refers to the group of accounting practices over which managers have discretion. The parties to the contracts voluntarily established this set. Depending on the relative costs and benefits of imposing limits, different firms are expected to have different levels of managerial discretion when it comes to choosing accounting procedures (also known as the ”accepted set”). These limitations are enforced by external auditors to keep managers from employing accounting discretion in an opportunistic manner and to maintain the accepted set's ”conservative” nature (R. Watts & Zimmerman, 1986).

R. L. Watts and Zimmerman (1990) represented the accepted set of accounting methods as a Venn diagram. Their illustration is replicated in Figure 2.1. The accepted set is determined by the contracting parties in advance (ex ante) to maximize the firm's value. X1 represents the accepted set of methods available to managers in Firm A, and

X2 for Firm B. Managers then choose a specific method from the set (e.g., X1 for Firm A and X2 for Firm B), which may be influenced by their personal incentives. These decisions can affect the distribution of wealth among contracting parties, as managers might select methods that maximize their utility, sometimes at the expense of other parties. Separating ex ante decisions from ex post decisions (after-the-fact choices) is challenging in practice, as contracts are continuously revised and renegotiated to adapt to changing circumstances.

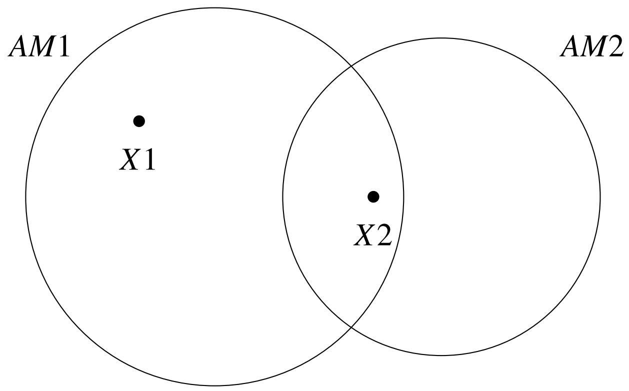

Figure 2.1: All Feasible Accounting Methods.(Replicated from R. L. Watts and Zimmerman (1990))

Note: AM1 and AM2 represent the accepted sets of accounting methods for Firms A and B, respectively. X1 and X2 denote the selected methods within each set.

2.3 R&D costs accounting treatment

So far, a brief introduction was made presenting the underlying economic theory based on which managers have motives to select certain accounting policies. In this section the two prevailing accounting policies for the treatment of R&D costs will be presented. The two major accounting standards bodies, Financial Accounting Standards Board (FASB) in the USA who overseas the US GAAP (Generally Accepted Accounting Principles) and the International Accounting Standards Board (IASB)[2] who oversea IFRS

[2]The International Accounting Standards Committee (IASC) was established in 1973. It introduced the International Accounting Standards (IAS), which were used as the primary accounting standards until 2001. In 2001, the International Accounting Standards Board (IASB) was established, replacing the IASC. The IASB took over the responsibility of developing new standards and began issuing International Financial Reporting Standards (IFRS), which succeeded the IAS. However, the IAS standards that were issued by the IASC before 2001 continue to be in effect unless they have been replaced or amended by IFRS.

(International Financial Reporting Standards), dictate different accounting treatments for R&D expenditure. In general, US GAAP state that R&D costs must be expensed as occurred. On the other hand, IFRS state that, if certain requirements are met, R&D costs must be capitalized. In Figure 2.2 the accounting treatment of R&D costs under US GAAP is illustrated as a Venn diagram. Management has no choice to make, as expensing the costs is the only available accepted method. In Figure 2.3 the managerial discretion, which is allowed by IFRS, is illustrated.

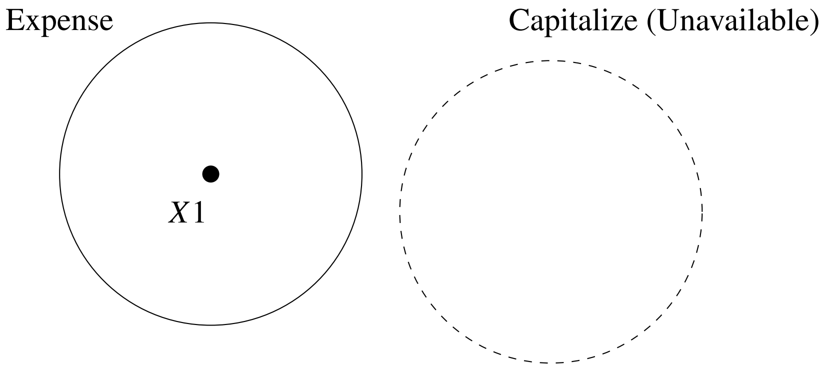

Figure 2.2: R&D costs accounting treatment under US GAAP

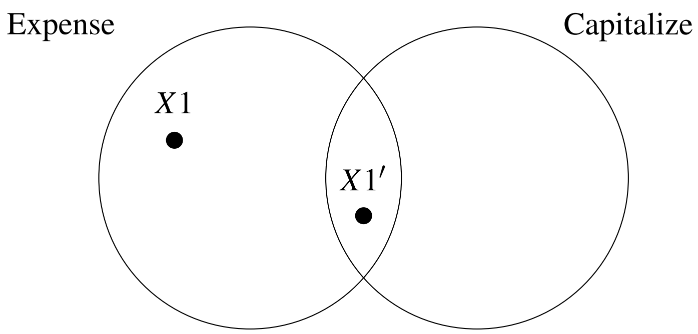

Figure 2.3: R&D costs accounting treatment under IFRS

This difference is explainable by the principles based on which the two bodies have developed their standards. Traditionally, US GAAP can be described as ”rules-based” and conservative, with specific guidelines on how to record transactions, while IFRS are described as ”principles-based” and require judgement on how to apply the standards. The financial reporting conservatism that characterizes US GAAP is seen as the solution to the agency problems described in the Theory of Large Managerial Firms (Lafond & Roychowdhury, 2008). IFRS can be characterised by conservatism too; initially they prescribe the immediate expense of R&D costs, yet they allow capitalization. Capitalization of R&D expenditure is more compatible with the matching principle[3], based on which expenses should be recorded in the same period as the revenue to which they are related.

2.3.1 R&D costs accounting treatment according to US GAAP

In 1974 FASB issued SFAS (Statement of Financial Accounting Standards) No.2, which established the standard for the reporting of R&D costs. In paragraph 8 research and development was defined as:

(a) *Research is planned search or critical investigation aimed at discovery of new knowledge with the hope that such knowledge will be useful in developing a new product or service or a new process or technique or in bringing about a significant improvement to an existing product or process*

(b) *Development is the translation of research findings or other knowledge into a plan or design for a new product or process or for a significant improvement to an existing product or process whether intended for sale or use. It includes the conceptual formulation, design, and testing of product alternatives, construction of prototypes, and operation of pilot plants. It does not include routine or periodic alterations to existing products, production lines, manufacturing processes, and other on-going operations even though those alterations may represent improvements and it does not include market research or market testing activities.*

In paragraphs 9 to 11 the standards strictly define which activities can be considered research and development. In paragraph 12 it is stated that: *"All research and development costs encompassed by this Statement shall be charged to expense when incurred"*. The interesting part of the standard are the Appendices; there the FASB explains the reasons which lead to adopt the specific policy. Initially, FASB recognizes the importance of R&D expenditures for the growth of the US economy. FASB states that there

[3] See Liao (1979) for more details in the matching principle

is a high degree of uncertainty about the future benefits of R&D projects. They mention the study of Booz-Allen and Hamilton (1982) and conclude that on average, only 2% of new product ideas and less than 15% of product development projects were commercially successful. Moreover, FASB states that according to the studies of Johnson (1967), Milburn (1971), and Newman (1968), there is a lack of causal relationship between R&D costs and future revenue in terms of future increased sales, earnings and industry sales.

In 1985, FASB issued SFAS No.86 for the accounting treatment of computer software to be sold or leased, as a response to AICPA paper *"Accounting for Costs of Software for Sale or Lease"*. The standard explains that the costs of the creation of a computer software product must be expensed as R&D until the completion of the program design, which proves the technological feasibility of the product. After this, all software production costs shall be capitalized.

More specifically, all the costs until the establishment of the technological feasibility are treated as in SFAS No.2. The technological feasibility is established when the firm has completed the planning, coding and testing of the product.

2.3.2 R&D costs accounting treatment according to IFRS

In 2001, IASB adopted IAS 38 Intangible assets which has been issued by IASC in 1998. In general, the standard defines that: *Expenditure for an intangible item is recognised as an expense, unless the item meets the definition of an intangible asset, and:*

(a) *it is probable that there will be future economic benefits from the asset; and*

(b) *the cost of the asset can be reliably measured.*

These are the general requirements for the recognition of an intangible asset. In the case of the internally generated intangible asset, the entity must classify the generation of the asset into the research phase and the development phase. IAS 38 defines research and development as:

(a) *Research is original and planned investigation undertaken with the prospect of gaining new scientific or technical knowledge and understanding. Research costs are expensed as they are incurred.*

(b) *Development is the application of research findings or other knowledge to a plan or design for the production of new or substantially improved materials, devices, products, processes, systems or services, before the start of commercial production or use. Development does not include the maintenance or enhancement of ongoing operations.*

In the research phase, an entity cannot demonstrate that an intangible asset exists and will generate probable future economic benefits, while in the development phase the entity, in some cases can identify an intangible asset which will generate future economic benefits. IAS 38 sets that an intangible asset shall be recognized if, and only if, an entity can demonstrate all of the following:

(a) *the technical feasibility of completing the intangible asset so that it will be available for use or sale.*

(b) *its intention to complete the intangible asset and use or sell it.*

(c) *its ability to use or sell the intangible asset.*

(d) *how the intangible asset will generate probable future economic benefits. Among other things, the entity can demonstrate the existence of a market for the output of the intangible asset or the intangible asset itself or, if it is to be used internally, the usefulness of the intangible asset.*

(e) *the availability of adequate technical, financial and other resources to complete the development and to use or sell the intangible asset.*

(f) *its ability to measure reliably the expenditure attributable to the intangible asset during its development.*

2.3.2.1 Accounting Standards Advisory Forum meeting-research project on Intangible Assets

IASB started a research project in April 2024 about intangible assets and the issues that stakeholders face with the accounting for intangibles. The motivation for this project is the increasing importance of intangible assets for the businesses and the need for better information for the users of financial statements about the intangible assets. One of the issues that participants agree, is that there are unrecognized internally generated intangible assets that are linked to future benefits. According to meeting participants, IASB should consider the requirement of reporting both recognized intangible assets and unrecognized intangible assets. Therefore, there are R&D projects in the development phase that meet the recognition criteria, yet management chooses not to recognize them as intangible assets.

In 2022, UKEB (UK Endorsement Board), in anticipation of the international debate on intangibles, started their own research project. Their findings are quite interesting. They conducted semi-structured interviews with 35 UK stakeholders. 85% of the respondents agreed that intangibles are very or extremely economically important. However, only 52% said that the relevant information reported in the financial statements is very or extremely useful. In addition to that, stakeholders were asked which is their preferred accounting treatment of intangible assets. In the majority of asset categories (apart from product development and software development), they indicated that they prefer the expensing of the relevant costs.

To summarise this section, the accounting treatment of R&D spending between the two standards differs in the development phase costs. While US GAAP requires the early expensing of R&D costs to emphasise financial conservatism and decrease uncertainty, IFRS allows for the capitalisation of such costs under certain situations, in accordance with the matching principle.

2.4 The value relevance of R&D expenditures: Early evidence

Academic research regarding the R&D accounting treatment and firm's future performance can be traced back in 1977. Ben-Zion (1977) showed that there is a relationship between firm's market value and R&D expenditure. Similarly, Hirschey and Weygandt (1985) found a relationship between Tobin's Q ratio and R&D to sales ratio. Sougiannis (1994) examined if extracted R&D benefits from the income statement numbers are value relevant and if investors use accounting earnings to value R&D investments. His findings suggested that one-dollar increase in R&D leads to a two-dollar increase in profitability and a five-dollar increase in market value. Hall (1993) presented findings on the opposite direction. His research examined the manufacturing sector in the US in the '80s. The evidence suggested that from 1973 to 1984 the intangible R&D assets were as value relevant as tangible capital. After 1984 though, the value relevance of R&D assets dropped dramatically. A potential explanation is that private returns to R&D have actually fallen and R&D investments create much less cash flow than traditional capital investments. Another possibility is that R&D capital is depreciating considerably faster than in the past, which could explain the drop in value. Finally, due to greater uncertainty, the stock market may be excessively discounting R&D cash flows, considering them as more risky and hence lowering their perceived value.

In their seminal paper, Lev and Sougiannis (1996) challenge the opinion of FASB, that there is not a relation between R&D expenditure and future economic benefits (based on which FASB requires the immediate expense of this expenditure). Their study explored the value-relevance of capitalized R&D and it was the first study which used firm-specific R&D capital instead of proxies for R&D investments. Lev and Sougiannis state that earnings are a function of tangible and intangible assets. This relationship is derived by the production function estimation, where sales (output) is related to labour and materials (inputs) plus stocks of physical and intangible capital (Mairesse & Sassenou, 1991). In the formulation of Lev and Sougiannis, earnings (output) minus labour and materials leaves the physical and intangible capital (assets) as their independent variables. Due to the fact that R&D capitalization is not allowed by US GAAP, the authors represented R&D capital by the lag structure of annual R&D expenditure. They used the price and returns models as suggested by Kothari and Zimmerman (1995) and examined the association of stock prices and returns with the R&D capital, where they found a strong and positive association. Thus R&D capital provides value-relevant information to investors. Yet, when the model of Fama and French (1992a) was used to examine whether investors recognise the value-relevance of capitalized R&D, they found that capitalized R&D are not fully reflected in the stock price. They attributed this to either the mispricing of the R&D firms or to that the excess returns are a compensation of the risks that are associated with the R&D projects probability of success.

Green et al. (1996) relied on the assumption that in the UK market there is excessive short-termism[4]. According to this hypothesis, investors fail to invest sufficiently in R&D. Their findings suggested that there is no evidence that the stock market totally misprices R&D expenditures. In their sample, R&D capitalization was not allowed by SSAP 13 (Statement of Standard Accounting Practice). They suggest that despite the treatment, stock market treats R&D expenses as capitalized, and adjust the reported earnings based on this.

Motivated by their findings, in a later paper, Lev and Sougiannis (1999) explored the relationship of R&D capital and excess returns. More specifically, the tried to explain the book-to-market phenomenon (BM)[5]. According to the theory, market value differs from the book value of a firm by the present value of future abnormal earnings. The abnormal earnings are frequently the result of innovation. The puzzle is that low (high) BM firms have a large (low) R&D capital. Their research addressed two main questions. Is the R&D capital proxied by BM ratio and are the returns associated with BM ratio a result of mispricing or a compensation for the risk? Their results suggest that when R&D capital is included in their model, BM is no longer associated with fu-

[4]Short-termism is when investors focus on short-term results and ignore the long-term value creation. Refer to Feigh (1994).

[5]For the BM puzzle see: Fama and French (1992b, 1995) and Lakonishok et al. (1994)

ture returns and that the relation between R&D capital and future returns is attributed to the risk factor associated with R&D investments. Al-Horani et al. (2003) provided comparable evidence for the UK market. Their results are comparable with Lev and Sougiannis (1999) as they also found that there is a relation between stock returns and R&D expenditure. Fama and French (1993, 1996) established the three factor model which explained a high proportion of portfolio returns. Al-Horani et al. (2003) showed that the three factor model modified to include R&D activity can vastly improve the explanatory power of the three factor model.

L. K. C. Chan et al. (2001) researched the stock market valuation of R&D expenditures. The immediate expense of R&D (especially in R&D intensive firms) costs may lead to a misstatement of financial ratios and because of this R&D firms may be mispriced. The authors compare the returns of firms The authors formatted portfolios based on R&D intensity, which is proxied by the ratio of R&D to sales. First the compared the returns of firms that engage in R&D activity and firms that do not do so. They found that market correctly values future benefits of R&D investments, as they found no difference in average returns between the two groups. Then, they focused in the R&D firms and created portfolios based on R&D intensity, proxied by the R&D to sales ratio. Firms with high R&D to sales behaved like glamour stocks[6], yet they did not exhibit the poor returns of glamour stocks. They found that a glamour stock with high R&D intensity earns higher returns compared to other glamour stocks. In a similar study, Amir et al. (2003) they found that analysts' forecasts of earnings compensate for R&D information, yet only for large, R&D intensive firms.

Amir, Chan, Lev, Sougiannis among others[7], were the pioneers in the research stream which documented a positive relationship between R&D intensity and future returns, and was established as the *"R&D intensity effect"*. Lev et al. (2005) further expanded the R&D literature by examining whether the expensing of R&D costs is a conservative accounting policy. The authors support that although the expensing of R&D is conservative compared to capitalization, this cannot stand throughout the

[6]See Graham and Dodd (1934) and La Porta (1996)

[7]See also Chambers et al. (2002), Eberhart et al. (2004), and Penman and Zhang (2002)

firm's life. They found that early-life (mature) firms with high (low) growth rate of R&D compared to profitability report conservatively (aggressively). The capital market implications of their findings are that the stocks of firms that report conservatively (aggressively) are undervalued (overvalued). They explain this misvaluation with the heuristic of representativeness[8], according to which investors believe that patterns in data are representative about future patterns.

Motivated by the valuation of loss firms, Darrough and Ye (2007) researched the value-relevance of R&D costs for the loss firms. The authors noted that there are firms that report losses for many years and firms with losses that exceed their BVE (Book Value of Equity) yet they do not bankrupt. They argue that these firms have invested in activities which bring future benefits, yet these activities are not reported in their book value because of the US GAAP requirements. They find that R&D activities are the main driver of this phenomenon. Firms that report losses tend to be more R&D intensive compared to profit firms. At the same time, even though these firms report losses, their R&D investments are valued by the market. The authors conclude that the increase in loss firm is associated with small firms that undertake risky R&D projects that are not profitable in the short term horizon. In a similar vein, Gu et al. (2023) find that correcting for the accounting distortion caused by expensing intangibles, the earnings of loss firms that report intangibles are equally value relevant as earnings of firms that report profits. In addition to that, the find that loss firms that report intangibles, they exhibit stronger subsequent performance compared to other firms.

Joos and Plesko (2005) researched whether investors price losses on the possibility to return to profitability according to the abandonment option hypothesis (Berger et al., 1996). They indicate that investors price differently loss firms based on whether they report R&D. In the case of firms which report consistent losses and report R&D, investors value the R&D as an asset. L. A. Franzen et al. (2007) find that healthy, R&D intensive firms are more likely to be misclassified as distressed when the O-score by Ohlson (1980) is used. This misclassification ceases to exist when adjustments are made

[8]See Grether (1992)

to the treatment of R&D. L. Franzen and Radhakrishnan (2009) investigated whether the value-relevance of R&D loss firms is extendable to profit firms. They hypothesize that profit firms that engage in R&D activities are likely to exhibit earnings that contain information about R&D productivity. They examined both profit and loss firms, and found that R&D expenditures are positively (negatively) related to stock prices for loss (profit) firms.

The early studies in the US setting examined the value relevance of R&D expenditures by simulating the capitalized outlays (see Lev and Sougiannis (1996)). In 1985 SFAS No.86 was issued and allowed the capitalization of internally developed computer software. This allowed researchers to obtain real capitalized R&D costs data for the first time (for the US market). Firms could also opt to report R&D under SFAS No.2 and expense immediately these costs. The most influential researchers in the field, Lev and Sougiannis, who first established the value relevance of R&D costs, in each of their papers address the FASB to consider to amend SFAS No.2, as their findings are in direct contrast with FASB's statement that *"...there is a lack of causal relationship between R&D costs and future revenue..."*.

The study of Aboody and Lev (1998) examined the software industry in the US from 1987 to 1995. Their findings validate the hypothesis that R&D costs are value relevant. More specifically, they examined the value relevance of capitalized R&D using three measures of firm performance, stock returns, stock price and earnings. They found that annual capitalized amount, the value of the software asset and its amortization are related to both capital market variables and subsequent earnings. The major contribution of this paper is not just the validation of value relevance of R&D costs; there was evidence to this direction in prior literature. The most important contribution is that the authors tried to explain the motives behind the choice to capitalize.

2.5 Determinants of R&D expenditures capitalization

Aboody and Lev (1998) were motivated to examine the software industry by SFAS No.86 which allows capitalization and by the petition of SPA (Software Publishers Association) in 1996 to abolish SFAS No.86. The petition is directly quoted as it is in the paper of Aboody and Lev (1998) in order to present the original view of the SPA in the matter. So, in 1996 SPA stated that:

"The rationale underlying the capitalization of software development costs is to recognize the existence of an asset of the corporation. However, an asset should be recognized ... only if ultimate realization of the asset is reasonably assured.... Due to factors such as the ever-increasing volatility in the software marketplace, the compression of product cycles, the heightened level of competition and the divergence of technology platforms, realization of software assets has become increasingly uncertain even at the point of technological feasibility... We do not believe that software development costs are a useful predictive factor of future product sales." [Page 4 of the letter]

"The members of the SPA CFO Committee ... have indicated the substantial majority of their investors, underwriters, and financial analysts believe financial reporting by software companies is improved when all software development costs are charged to expense as incurred. These users of financial statements do not believe the recording of a "soft" asset for the software being developed is particularly relevant and does not aid the user of financial statements. The users of financial statements ... have a high degree of skepticism when it comes to soft assets resulting from the capitalization of software development costs." [Page 5 of the letter]

Aboody and Lev (1998) noted, that in 1985, ADAPSO (Association of Data Processing Service Organizations) was a strong supporter of R&D capitalization. The question which was raised was why there was a shift in the opinion towards capitalization, given that the firms have the flexibility to choose to immediately expense R&D. Analysts also raised objections to capitalization. The empirical results indicated that when firms capitalize, analysts' earnings forecast errors are positively related to the rate of capitalization. As Aboody and Lev noticed, this is interesting, as it is very easy to reverse capitalization by subtracting the annual capitalization charge from the earnings and the capitalized asset from the total assets. After examining the opinions of the stakeholders, Aboody and Lev (1998) divided their sample in ”expensers” and ”capitalizers”. As expensers, they categorized firms that immediately expensed all their R&D costs, while as capitalizers firms that capitalized a portion of the costs. By further examining the sample, they found that some firms capitalized in all years but one and they classified them as capitalizers. They came in the following conclusion regarding the determinants of R&D treatment:

(a) **Size:** proxied by the log of market value of equity. They supported that larger firms spend a big amount of R&D costs on maintenance and upgrade of their already established products. These expenditures are expensed according to SFAS No.86, thus large firms are expected to expense a big portion of their development costs compared to smaller firms.

(b) **R&D intensity:** proxied by the annual development costs to sales ratio. Firms that spend more on development, exploit the economies of scale, and will have, on average, higher success rate in developing novel products, so they are expected to exhibit a higher capitalization rate.

(c) **Profitability:** measured by net income plus R&D amortization, minus annual capitalized R&D divided by sales. Based on analysts' opinion of capitalization, there is the notion that profitable companies will not capitalize in order to avoid negatively affecting analysts' perception of the quality of their earnings.

(d) **Leverage:** measured by long-term debt to BVE less the capitalized R&D asset, which is used as a proxy to loan covenants. Firms that face strict loan restrictions may capitalize R&D in order to increase equity and earnings.

2.5.1 Firm size and R&D intensity: Evidence from the industrial economics and innovation literature

A major research stream on whether economies of scale exist in R&D has emerged in the pharmaceutical industry in the 60s. Comanor (1965) examined the relationship between R&D growth and firm size growth or to put it simple, whether the variation in R&D rate between firms can be explained by their size. The results challenge the common belief that only large firms can drive rapid innovation and technical progress. It is found that larger firms face diseconomies of scale in their R&D. However, it is unclear whether this conclusion extends to all industries. In a later study, Jensen (1987) concludes that the productivity of R&D is not affected by firm size. The increase in R&D expenditures though increases the possibilities of discovering a new drug, which is consistent with the R&D density hypothesis made by Aboody and Lev (1998).

Other researchers found that size is irrelevant and that industry explains half of the variance in R&D intensity (Cohen et al., 1987). Cohen and Klepper (1996) examined a cross-industry sample and found no size-effect in R&D. Their findings support the underlying idea that larger organisations have an advantage in R&D because they can spread the costs of their R&D efforts over a larger amount of output. This enables them to better capitalize on the outcomes of their R&D projects.

DiMasi et al. (1995) motivated by the varying results presented in the relative literature, which are attributed to the aggregative data used in the industrial organization studies, they examined data collected at individual project level for 12 pharmaceutical firms. The smallest firms in the sample had lower clinical development timelines and costs, but significantly higher preclinical expenses, which dominated the overall cost estimates. The findings suggested that economies of scale existed in pharmaceutical R&D, particularly in the preclinical or discovery phase. Shefer and Frenkel (2005) found a negative and significant relationship between firm size and innovation. They concluded that small, young firms tend to spend more in R&D.

R&D intensity is one of the most critical business decisions. Especially in the case of technology firms, managers face challenges such as whether they have invested enough in R&D and whether these investments will bring future economic benefits to the firm (Lin et al., 2006). The hypothesis that R&D intensity of technological firms is positively associated with the financial performance of the firm, proxied by Tobin's Q was tested by Lin et al. (2006). They did not find evidence supporting their hypothesis. They explained that maybe technological firms were not able to obtain a competitive advantage just by increasing their R&D efforts.

Bustinza et al. (2019) examined the relationship between R&D intensity, product–service innovation (servitisation[9]) and performance. Their findings suggest that R&D intensity increases servitisation's positive impact on performance. C.-Y. Lee et al. (2014) made some interesting remarks regarding R&D intensity. The relevant literature considers R&D spending crucial for the survival and growth of the firms. However, in the case of Eastman Kodak, who failed to invest early in digital photography, most of their R&D resources were focused on chemical photography technology. Kodak exhibited high R&D intensity, yet this did not prevent the decline of the company. C.-Y. Lee et al. (2014) concluded that hgih R&D intensity makes the firms to become more exploitative and less explorative, where explorativeness is the degree of the new knowledge used by the firm in order to innovate.

2.5.2 R&D capitalization determinants: Evidence from the accounting literature

Since the seminal paper of Aboody and Lev (1998), numerous studies have emerged examining the R&D capitalization determinants. The majority of these studies were based on the findings of Aboody and Lev. In Table 2.1 we present the relevant literature and the relationship found for the determinants of capitalization. It is noticed that in terms of firm size and leverage, findings are in line with Aboody and Lev (1998) in the majority of the studies. On the other hand, mixed results are found for R&D intensity and profitability. While the effect of firm size and R&D intensity on the decision to capitalize is extensively examined by industrial economics and innovation research,

[9]See Shin et al. (2022) for more details on servitisation, efficiency and performance.

Table 2.1: Summary of studies on firm attributes associated with capitalization

Study	Sample	Size	Inten.	Prof.	Lev.
Aboody and Lev (1998)	US	-	+	-	+
Cazavan-Jeny and Jeanjean (2006)	France	-	-	-	+
D. R. Oswald and Zarowin (2007)	UK	-	-	+	N/A
D. R. Oswald (2008)	UK	-	-	+	+
Markarian et al. (2008)	Italy	-	+	-	-
Cazavan-Jeny et al. (2011)	France	-	-	+	+
Dinh et al. (2015)	Germany	+	+/-	+	+
Eierle and Wencki (2016)	Germany	-	+	-	+
Wang (2016)	China	+	-	+	+
Mazzi et al. (2019)	International	-	+	N/A	+
D. Oswald et al. (2021)	UK	-	-	+	+
Brasch et al. (2022)	UK	-	-	-	+/-

Note: Inten. = R&D intensity; Prof. = profitability; Lev. = leverage. ”+” indicates positive association, ”-” indicates negative association, and ”N/A” indicates not assessed.

There are two distinguishable groups of studies regarding the accounting treatment of R&D costs. The proponents of immediate expensing of R&D support that expensing R&D capitalization requires managerial judgement and it allows for earnings management. On the contrary, proponents of capitalization support that capitalized R&D costs can be used as a signal to the investors about the future performance of the firm (Cazavan-Jeny et al., 2011). Early evidence from the US market, is relied on simulated capitalization data, as capitalization is not permitted. The majority of studies about R&D capitalization determinants focus on countries that allow capitalization under their national GAAP or IFRS. In Table 2.2 we briefly present the capitalization rules for the countries that are examined in these studies.

Table 2.2: Summary of R&D accounting treatment per study

Study	**Country**	**Acc. Stand.**	**Res.**	**Dev.**
K. Ahmed and Falk (2006)	Australia	National	E*	C*
Cazavan-Jeny and Jeanjean (2006)	France	National	E	C*
D. R. Oswald and Zarowin (2007)	UK	National	E	C*
D. R. Oswald (2008)	UK	National	E	C*
Markarian et al. (2008)	Italy	National	E	C*
Cazavan-Jeny et al. (2011)	France	National	E	C*
Dinh et al. (2015)	Germany	IFRS	E	C
Eierle and Wencki (2016)	Germany	National	E	C*
Dinh and Schultze (2022)	Germany	IFRS	E	C
Brasch et al. (2022)	UK	National	E	C*

Note: Acc. Stand. = Accounting standards followed; Res. = research phase; Dev.= development phase; E = expensed as incurred ; E* = expensed as incurred but optionally capitalized; C* = optional capitalization; C = mandatory capitalization.

2.5.2.1 The earnings management theory

Earnings management is the tactic of generating earnings by taking advantage of managerial discretion over accounting policies and operating cash flows (Phillips et al., 2003). Similarly, R. L. Watts and Zimmerman (1990) characterize earnings management as managers' discretion over accounting numbers with or without constraints. Earnings management includes a variety of activities that affect reported accounting earnings or how they are examined. It begins with production and investment decisions, which influence underlying financial results. It continues with the selection of accounting policies and the determination of accrual amounts during the preparation of financial reports, and end with activities that influence how reported earnings are interpreted (Ronen, 2008). The main motives for earnings management include restrictive debt covenants, CEO compensation, and income smoothing (Markarian et al., 2008). According to the earnings management hypothesis, capitalization increases ROA (Return on Assets), improves leverage and smooths earnings (White et al., 2002).

A negative relationship between profitability and capitalization is expected if management capitalized when firm performance is poor (Cazavan-Jeny & Jeanjean, 2006). Furthermore, it is established in the literature that managers seek to smooth the reported earnings (Degeorge et al., 1999; Fudenberg & Tirole, 1995). Findings in the R&D literature support the smoothing hypothesis (Healy et al., 2002; Lev et al., 2005). Finally, according to Aboody and Lev (1998), managers may use capitalization to increase equity and earnings in order to manipulate the reported leverage ratio. Lenders rely on financial statements to set debt covenants (Costello & Wittenberg-Moerman, 2011). When firms are small, high-leveraged and report fewer tangible assets, their loans are more likely to include restrictive covenants (Bradley & Roberts, 2015). Kim et al. (2021) found that earnings management practices could indicate managers' self-serving actions driven by personal ambitions (the managerial opportunism hypothesis[10]).

Cazavan-Jeny et al. (2011), in the French setting, noticed that capitalizers do not necessarily capitalize every year. They run a second determinants test to examine when capitalizers capitalize. Based on the findings of Burgstahler and Dichev (1997), who found that firms avoid reporting losses and earnings decreases, they hypothesized that management could use R&D capitalization in order to meet the zero and last year thresholds. The findings validate their hypothesis. In a similar study (which though did not examine capitalization of R&D per se), Osma and Young (2009) found that firms reduce R&D spending in order to beat earnings targets. In the second part of their study, Cazavan-Jeny et al. (2011) examined the relationship between R&D and three measures of firm performance, sales growth, stock price and earnings. They hypothesized that if capitalisation regulations are correctly implemented, capitalised R&D will more accurately estimate future profitability than expensed R&D. However, the results showed that capitalising R&D expenditures had a neutral or negative impact on future performance. Their findings suggest that, contrary to the accounting standard's intent, managers do not capitalize R&D expenses for projects with a higher chance of success. In a similar study, conducted using a sample from France, Cazavan-Jeny and Jeanjean (2006) found a negative relationship between capitalized R&D and stock returns. They have also conducted a capitalization determinants test, and concluded that managers use R&D capitalization opportunistically.

Dinh et al. (2015) used a German sample of R&D active firms and examined whether

[10]See Chalmers et al. (2002) for more details on the managerial opportunism hypothesis.

managers use R&D capitalization opportunistically. Their study focused on the benchmark beating hypothesis. They found that managers tend to capitalize R&D costs when they want to surpass analysts' earnings forecasts or last year's earnings. Moreover, they found that firms capitalize higher amounts of R&D costs when they are leveraged and have lower growth opportunities. Recent evidence from Germany examined the capitalized R&D under IAS 38 and "as-if" capitalized R&D. Dinh and Schultze (2022) were motivated by two distinct opinions in the literature. While Garanina et al. (2021) examined the benefits of reporting capitalized intangible assets on the balance sheet, Barker et al. (2020) and Penman (2009) suggested that emphasis should be given to the income statement too, as investors can use it to compensate for possibly inadequate balance sheets. According to Dinh and Schultze (2022), "as-if" capitalization removes management's discretion but at the same time it also removes the informativeness to market participants (Riley, 2001). The informativeness of "as-iff" capitalized R&D costs- if it exists- is attributed to the benefits of accrual accounting (Dechow, 1994; Penman & Yehuda, 2009). Dinh and Schultze (2022) findings suggest that capitalized R&D under IAS 38 are associated with forecast errors and that they are as value-relevant as the expensed R&D. They attribute this to the fact that investors believe that capitalization is used for earnings management. The investors make adjustments, meaning they "undo" capitalization- they convert the capitalized R&D to expensed. Finally, "as-if" capitalized costs are value-relevant.

Motivated by the differences that are noted in the literature between private and public firms, Eierle and Wencki (2016) investigated a sample of privately-held German firms. Their findings suggest that R&D capitalization is used for income smoothing when they report low performance or negative earnings. Further, highly leveraged firms tend to capitalize more often. They conclude that despite their differences, private and listed firms do not differ in these specific capitalization determinants. In a similar vein, evidence from UK private firms supports that capitalization of R&D is used to avoid debt covenants restrictions (Brasch et al., 2022).

Evidence from Italy supports the income smoothing hypothesis. More specifically, Markarian et al. (2008) provided evidence in line with studies suggesting that managers capitalize R&D costs when they exhibit lower earnings (ROA). On the contrary, when firms exhibit improved performance, it is more probable that management will expense R&D costs. Their findings indicate that there is a negative relationship between capitalization and ROA.

2.5.2.2 The market signal theory

Signalling theory is useful for describing behaviour when two parties (individuals or organisations) have unequal access to information. In general, one party, known as the sender, selects whether and how to convey (or signal) that information, while the other party, the receiver, must decide how to interpret the signal (Connelly et al., 2011). One of the most influential papers in the field, written by Spence (1973) examines how high quality job applicants distinguish themselves from low quality job applicants using prestigious higher education degrees. Signalling theory focusses on neutralising information asymmetry (Spence, 2002). Akerlof (1970) and Stiglitz (2002) indicated that there is information asymmetry in the markets. Common perception of these authors is that, disregarding the market (labour, car, stock, etc.), sellers have more information about the product than buyers. Dye and Verrecchia (1995) found that when managers have more reporting flexibility, they can signal the market about firm performance. Ball and Brown (1968) investigated whether accounting treatment choices that do not directly impact cash flows are linked to fluctuations in stock prices. They attributed market inefficiency to managerial signalling. Rees et al. (1996) interpreted the abnormal negative accruals during an asset write-off as managerial signal for performance. Accounting choices can serve as a means for more informed insiders to communicate information to less informed parties regarding the timing, size, and risks of future cash flows (Fields et al., 2001).

Hughes and Kao (1991) supported that capitalization is more informative as, despite the required managerial judgement and estimates about future performance, these estimates require verification by an auditor. D. R. Oswald and Zarowin (2007) listed three reasons for which the capitalization of R&D costs may not be informative. First, if the market believes that managers are manipulating earnings, the investors will be sceptical about management's information. Due to the nature of R&D investments and the difficulty to forecast their success, even the auditor's report cannot reduce their concerns. Second, because of the high uncertainty of R&D investments, the information provided by management is based on assumptions and estimates. Thus, the information can be completely misleading about future performance. Finally, expensers could simply choose to disclose the information rather capitalize in order to give a signal to the market.

D. R. Oswald and Zarowin (2007) examined a sample of UK companies, where capitalization was allowed. They measured the informativeness using FERC(Future Earnings Response Coefficient), which is the coefficient on future earnings in a regression of current stock return against current and future earnings. They found that capitalized R&D are more informative than expensed R&D, which is in line with the market signal theory. D. R. Oswald (2008) examined the value relevance of both capitalized and expensed R&D in the UK setting. His study concluded that management will use the most appropriate method (capitalize or expense) in order to convey information to the market.

Evidence from Australia suggests that managers posses superior information and should be allowed to use the R&D costs accounting treatment as a signal (K. Ahmed & Falk, 2006). The authors found that by capitalizing successful R&D and expensing the unsuccessful, information asymmetry would be reduced and the financial statements would be more value relevant. In a similar study, again in Australia, Goodwin and Ahmed (2006), despite the expectation of earnings management when managers have the discretion to capitalize or not, they found that the earnings of capitalizers are more value relevant compared to the expenses, which is in line with the signal theory.

2.6 The bonus plan hypothesis and R&D spending

In the Positive Accounting Theory, R. L. Watts and Zimmerman (1978) made the hypothesis that managers who seek to maximize their bonuses prefer accounting methods that increase the reported income. Waegelein (1988) conducted one of the first empirical studies in the field, by examining the relationship between short-term bonus plans and corporate investment decisions, such as R&D. The short-term bonus plans assess the firm performance for a year. Firm performance is assessed by profitability ratios, such as ROA and ROE. Their sample was formed by US firms, and their findings suggested that short-term bonus plans lead to significant increase in capital expenditures, yet they could not find an association with R&D spending.

Cheng (2004) examined whether compensation committees are able to prevent reductions in R&D expenditures. Two main hypotheses were formed. Changes in R&D spending are more likely to happen when the CEO is near retiring age and when firms report small losses or decline in reported earnings. The findings suggested that those two hypotheses hold. However, there is no significant association with reduced R&D spending. In a similar study, Serfling (2014) found out that older CEOs reduce risk by making less risky investments, and they achieve this by reducing R&D spending. Another study by Duru et al. (2002) suggested that firms do actually protect regular expenses, such as R&D and advertising costs. Furthermore, firms prefer to protect R&D investments over advertising costs. These findings are consistent with earlier research, which indicates that compensation committees protect CEOs from the financial consequences of non-routine occurrences such as restructuring expenditures and unusual losses. Dechow and Sloan (1991) examined the hypothesis that CEO near retiring age will manage discretionary investment expenditures, like R&D, to improve short-term earnings performance. Their findings support this hypothesis. However they find that through CEO stock ownership plans, reductions in R&D expenditures are mitigated.

Cao and Laksmana (2010) found out that firms concerned with income reporting tend to reduce CEO option remuneration when R&D spending rises and reported earnings fall. Furthermore, for firms under similar reporting requirements, the loss in compensation is more evident when they repeatedly fail to fulfil quarterly earnings targets. Overall, findings indicate that the negative impact on CEO compensation leads to short-term, incentive-driven decisions, resulting in the suspension or cancellation of R&D initiatives, even if they have significant long-term potential. Ghosh et al. (2007) found that hat R&D investments and CEO stock options are positively associated at high levels of option holdings.

Dinh et al. (2019) examined a sample of firms in the US. They formulated the following hypothesis. Firms mitigate a decline in earnings by two ways. Either they will increase the capitalized amount of successful R&D spending (those who are allowed to under SFAS No.86) or they will reduce R&D spending. Their findings substantially support the notion that capitalising costs associated with successful software development, as allowed by US accounting standards, minimises the risk of underinvestment due to short-term incentives. Furthermore, firms that capitalise these expenses do not appear to overinvest when they have financial flexibility, unlike organisations that cannot capitalise such costs.

2.7 Value relevance of R&D and the reporting environment

Ali and Hwang (2000) examined accounting data from 16 different countries. Their study explored five country-specific factors that affect value relevance. Among their findings, they found that value relevance is lower for Continental countries compared to the UK and the USA. Similarly, Alford et al. (1993) investigated whether differences in accounting standards and corporate governance affect value relevance. They used the USA as a benchmark, and found that value relevance varies across countries with different national GAAP.

R. Zhao (2002) examined the value relevance of R&D expenses using an international dataset. Their sample was from two countries who allowed conditional capitalization, France and the UK and two countries who required the immediate expensing of R&D, Germany and the USA. Apart from the accounting treatment rules for R&D, the study focused on the different legal environment in these countries, where France and Germany have a code-law system, while the UK and the USA are situated in a common-law environment. Common-law countries typically have dispersed equity ownership, more financial transparency, and higher-quality accounting earnings, whereas code-law countries tend to have more concentrated ownership, less financial transparency, and lower-quality earnings. As a result, the value relevance of R&D reporting in these countries depends on both their financial reporting environments and their R&D accounting standards. In Germany and the USA, where R&D costs are expensed, reporting total R&D costs are value relevant to accounting earnings and book value. Similarly, in France and the UK, where R&D costs are capitalized, the distinction between capitalized and expensed R&D costs further increases the value relevance of R&D.

Another stream of the literature examines the effects of IFRS adoption in 2005. The local GAAP of each country, which were shaped by the local institutions and culture, were replaced by the principle-based IFRS. One of the purposes of IFRS introduction was to enhance financial reporting quality (André et al., 2015). There is a controversy between academics on whether IFRS achieve this goal. For example, academics from the USA believe that IFRS lack rigor and quality (Barth, 2008). Empirical evidence from the European Union countries suggests that a significant number of accounting quality metrics improved when IFRS were adopted in the EU. That is, there is less tendency on managing earnings towards a target, a smaller size of absolute discretionary accruals, and improved accrual quality. However, the findings indicate that firms engage in more income smoothing and recognise big losses in a less timely way in post-IFRS adoption (H. Chen et al., 2010). Similarly, A. S. Ahmed et al. (2013) suggest that indeed, there is evidence towards the direction of increased income smoothing and a decrease in timeliness of loss recognition, on the other hand though, they found that there is no evidence that IFRS adoption lead to increased accounting quality.

One of the first studies that examined the effect of R&D accounting treatment and their value relevance in the pre- and post- IFRS adoption periods was conducted by Dargenidou et al. (2021) in the UK setting. They examined two R&D accounting treatment standards, SSAP 13 and IAS 38. Under SSAP 13, managers had the option to capitalize R&D costs but there were strict rules about this choice (high asset recognition threshold). On the contrary, IAS 38 introduced a lower recognition threshold but greatly reduced managerial discretion on when managers have to capitalize R&D. Stark (2008) supported that the limits that are set by IAS 38 in the managerial discretion limit the managers' ability to send signal to the market about the success of R&D projects. Dargenidou et al. (2021) made the hypothesis that the relationship between current returns and future earnings for the capitalizers is weaker in the post- IFRS period compared to the pre- IFRS period. Their findings suggest that capitalisation under IFRS does not result in current returns that include more future profits information than expensing. Post-IFRS share prices are less information-efficient than pre-IFRS prices due to fewer forward-looking information. They concluded that capitalisation under IFRS may lead to uncertainty, however this is corrected in the future when economic benefits are realised.

In a similar study, Shah et al. (2013) examined the contemporaneous relation between prices or returns and book values, in the pre- and post- IFRS period in the UK. Their findings indicate that capitalized R&D are value relevant throughout the two periods. This does not stand for the expensed R&D. Although capitalized R&D remained value relevant in the post- IFRS period, they noticed a decrease in value relevance in the post- IFRS period. Evidence from South Korea suggests that when firms consistently capitalize R&D costs, those remain value relevant in the post- IFRS period and are a reliable indicator of future economic benefits (Cho & Kim, 2024). Similar conclusions have been made by Tsoligkas and Tsalavoutas (2011) for the UK market. Capitalized R&D remain value relevant to market value in the post- IFRS period, while expensed R&D are negatively value relevant to market value in the post- IFRS period. A recent study by Bhattacharya et al. (2024) examined the efficiency of R&D firms in the pre- and post- IFRS period in Germany. They considered the implementation of IFRS and the shift from full expensing to partial capitalization of R&D expenses as an exogenous shock. Firm efficiency is measured using Data Envelopment Analysis and Stochastic Frontier Analysis.[11] Their findings suggest that German firms became more efficient in the post- IFRS period. Their supplementary analysis suggested the opposite for UK and Australian firms.

[11] See more about DEA in Boussofiane et al. (1991) and for SFA see Koop et al. (1999).

Chapter 3

Methodological framework and data description

3.1 Introduction

In this Chapter the methodological framework is developed. The theoretical models are presented and a detailed description of the data is provided. Furthermore, feature engineering, cross-validation strategies and the evaluation metrics are presented.

3.2 Earnings, stock price, stock returns-R&D relation

All the earnings models in the field are built upon the relationship between earnings and assets, as earnings are generated by assets. Lev and Sougiannis (1996) defined this relationship as:

$$Earn_{it} = f(TangAss_{it}, IntAss_{it}) \tag{3.1}$$

where $Earn_{it}$ are the earnings, $TangAss_{it}$ the tangible assets and $IntAss_{it}$ the intangible assets. In the next formula, they further split the intangible assets in R&D assets and other intangible assets:

$$Earn_{it} = f(TangAss_{it}, RDC_{it}, OIA_{it}) \tag{3.2}$$

where RDC_{it} is the R&D capital (assets) and OIA_{it} are the other intangible assets. Finally, they use operating income as a measure of earnings, as they find that R&D investments are not related to non-operating items. They formulated the following expression:

$$OI_{it} = \alpha_0 + \alpha_1 TangAss_{i,t-1} + \alpha_2 RDC_{i,t-1} + \alpha_3 OIA_{i,t-1} + \varepsilon_{i,t} \tag{3.3}$$

where:

- OI_{it} = annual operating income, before depreciation, advertising and R&D expenses, of firm i in year t, scaled by sales,
- $TangAss_{i,t-1}$ = the value of plant and equipment, inventory, and investment in unconsolidated subsidiaries and goodwill, measured at the beginning-of-year values, scaled by sales,
- $RDC_{i,t-1}$ = R&D capital, measured at the beginning-of-year values, scaled by sales,
- $OIA_{i,t-1}$ = other intangible assets, measured at the beginning-of-year values, scaled by sales[1].

Lev and Sougiannis (1996) followed the suggestion of Kothari and Zimmerman (1995) who suggested that both price and returns models have to be used. They specified the returns-earnings relation as:

$$R_{it} = \alpha_1 + \beta_1 E_{it} + \gamma_1 (AdjE_{it} - E_{it}) + u_{it} \tag{3.4}$$

$$R_{it} = \alpha_2 + \beta_2 E_{it} + \gamma_2 \Delta E_{it} + \delta_2 (AdjE_{it} - E_{it}) + \Omega_2 \Delta (AdjE_{it} - E_{it}) + u_{it} \tag{3.5}$$

$$R_{it} = \alpha_3 + \beta_3 E_{it}^{RD} + \gamma_3 \Delta E_{it}^{RD} + \delta_3 (AdjE_{it} - E_{it}) + \Omega_3 \Delta (AdjE_{it} - E_{it}) + u_{it} \tag{3.6}$$

[1]Other intangible assets were simulated by advertising expenses

where:

- R_{it} = annual stock return from nine months before fiscal t year-end through three months after it,
- $E_{it}, AdjE_{it}$ = reported (GAAP) and adjusted[2] earnings-per-share (before extraordinary items), respectively,
- $AdjE_{it} - E_{it}$ = 'error' or misstatement in reported earnings due to the R&D expensing; this misstatement is equal to $RD_{it} - RA_{it}$, namely the annual R&D outlay minus the R&D amortization, which in turn is equal to the net (amortized) investment in R&D during t,
- $E_{it}^{RD} = E_{it} + RD_{it}$, is reported earnings before the R&D expensing.

Model (3.4) is the basic returns-earnings relation, Model (3.5) includes the first differences of earnings, because differencing yields stationary series and Model (3.6) includes earnings before R&D expensing. All independent variables are divided by beginning of fiscal year share price, $P_{i,t-1}$. Lev and Sougiannis (1996) derived a parsimonious price model:

$$P_{it} = \alpha_4 + \beta_4 E_{it} + \gamma_4 (AdjE_{it} - E_{it}) + u_{it} \tag{3.7}$$

where P_{it} is share price of firm i three months after fiscal year-end. Based on this model they derived the following expression:

$$P_{it} = \alpha_5 + \beta_5 E_{it} + \gamma_5 (AdjE_{it} - E_{it}) + \Omega_5 RDC_{it} + u_{it} \tag{3.8}$$

The studies that have focused in the US setting have the drawback of relying in simulated data to proxy for R&D capital, as R&D capitalization is not allowed in the US GAAP. European studies utilize similar models as Lev and Sougiannis (1996)- any may be more suitable to this research. Most of them have built their models upon

[2] $AdjE_{it} = E_{it} + RD_{it} - RA_{it}$, where RA_{it} is R&D amortization

the Ohlson (1995) valuation model. Cazavan-Jeny and Jeanjean (2006) expressed the relationship between stock price and R&D expenditure as:

$$\begin{aligned} P_{it} = \alpha_0 + \beta_1 ABVPS_{it} + \beta_2 AEPS_{it} + \beta_3 CapRDPS_{it} \\ + \beta_4 ExpRDPS_{it} + \sum YR_{it} + \sum Ind_{it} + \varepsilon_{it} \end{aligned} \tag{3.9}$$

where:

- P_{it}: stock price at the end of fiscal year t for firm i,
- $ABVPS_{it}$: adjusted book value per share, that is, net of capitalized R&D,
- $AEPS_{it}$: adjusted earnings per share, that is, before R&D expense and amortization of capitalized R&D,
- $CapRDPS_{it}$: annual amount of net capitalized R&D per share,
- $ExpRDPS_{it}$: annual amount of expensed R&D per share,
- YR_{it}: time indicator variable,
- Ind_{it}: industry dummy variable.

Following Easton (1999) and Easton and Harris (1991), Cazavan-Jeny and Jeanjean (2006) expressed their returns model as:

$$\begin{aligned} R_{it} = \lambda_0 + \lambda_1 \Delta AEPS_{it} + \lambda_2 \Delta AEPS_{it} + \lambda_3 \Delta CapRDPS_{it} \\ + \lambda_4 \Delta ExpRDPS_{it} + \lambda_5 ABVPS_{it} \\ + \lambda_6 CapRDPS_{it} + \lambda_7 ExpRDPS_{it} + \sum YR_{it} + \sum Ind_{i,t} + \varepsilon_{it} \end{aligned} \tag{3.10}$$

where:

- R_{it}: annual stock return at the end of year t for firm i
- ΔEPS_{it}: change in earnings per share between t and $t-1$
- $\Delta AEPS_{it}$: change in adjusted earnings per share between t and $t-1$

- $\Delta CapRDPS_{it}$: change in capitalized R&D per share

- $\Delta ExpRDPS_{it}$: change in annual amount of expensed R&D per share

All independent variables are scaled by the beginning-of-year market capitalization (apart from the dummies).

Cazavan-Jeny et al. (2011) used a slightly different approach compared to the relevant literature, and apart from using only price and returns models they examined future performance prediction by also using an income model. They used a model similar to L. K. C. Chan et al. (2003), where they modelled future ROA as:

$$\begin{aligned} FUT\_ROA_k &= \beta_0 + \beta_1 RD\_CAP_t + \beta_2 CF\_RD_t + \\ &+ \beta_3 RD\_CAP \times CF\_RD_t + \beta_4 ROA_t + \beta_5 PTB_t + \\ + \beta_6 Size_t + \beta_7 CAPEX_t &+ \beta_8 IMR + \sum Year_k + \sum Industry_k + \varepsilon_t \end{aligned} \tag{3.11}$$

where:

- $FUT\_ROA_k = \frac{\Sigma(\text{ROA}_{t \text{ to } t+k})}{k+1}$,

- RD_CAP_t: indicator variable coded 1 if the firm capitalizes its R&D costs at least once over the examined period, 0 otherwise,

- CF_RD: the cash flow of R&D (irrespective of its accounting treatment) scaled by total assets,

- PTB: Price to Book ratio,

- *Size*: the natural log of total assets, excluding capitalized R&D,

- $CAPEX$: capital expenditures in the year,

- IMR: the Inverse Mills Ratio. See Heckman (1979) for details.

Following L. K. C. Chan et al. (2003), Cazavan-Jeny et al. (2011) included CF_RD as a proxy of R&D intensity. Chan et al. have found that R&D intensity exhibits strong forecasting power. PTB ratio was included as firms with low book-to-market ratio are associated with high growth. *CAPEX* are included as a control, because managers may engage in capital expenditure in order to yield better future growth. To verify their results, Cazavan-Jeny et al. (2011) estimated an alternative income model, where instead of sales growth, they used an income level model, where they used FUT_ROA_k as the dependent variable. FUT_ROA_k is the average future ROA over an one-year or 3-year horizon, measured as: $FUT\_ROA_k = \sum(ROA_{t \text{ to } t+k})/(k+1)$. To further investigate the relationship of R&D and income level, they split the cash flow of R&D into three components, the cash flow expensed by the expensers, the cash flow capitalized by capitalizers and the cash flow expensed by capitalizers. Finally, Cazavan-Jeny et al. (2011) modelled the relationship between stock price and R&D using the valuation model of Ohlson (1995), in a similar way as Aboody and Lev (1998), Cazavan-Jeny and Jeanjean (2006), and R. Zhao (2002). Stock returns are modelled following Easton (1998) as:

$$\begin{aligned} R_{it} = \beta_0 + \beta_1 AE_{it} + \beta_2 \Delta AE_{it} + \beta_3 \Delta RD\_Exp\_Exp_{it} + \beta_4 \Delta RD\_Cap\_Cap_{it} \\ + \beta_5 \Delta RD\_Exp\_Cap_{it} + \beta_6 IMR_i + \sum \text{Year}_{it} + \sum \text{Industry}_{it} + \varepsilon_{it} \end{aligned} \tag{3.12}$$

Where all right-hand side variables are scaled by lagged market value. R_{it} is the annual stock return at the and of year t for firm i and AE_{it} are the earnings before R&D expense and amortization.

3.3 Machine Learning for profitability prediction

There is extensive academic literature on the ability of financial statement information to predict future profitability. Nowadays researchers adopt panel-data methods to predict future profitability (Monahan, 2018). Evidence suggests that traditional regression exhibits poor forecasting performance compared to a random-walk model (Li & Mohanram, 2014). Zarowin (2019) highlighted that machine learning may be a methodological innovation to improve the forecasting accuracy of earnings models.

Jones et al. (2023) suggested that ML (Machine Learning) can be used to earnings forecasting in the following two ways (so far). The first way is to exploit many available features and let the ML algorithm to decide the best predictors from the feature space. In one of the earliest studies that followed this approach, Ou and Penman (1989) used many accounting and market-based ratios to predict changes in next-period earnings with a stepwise logistic regression. Since then though, modern ML algorithms have emerged, such as gradient-boosting and random forests, which allow a vast number of features and provide stable forecasting models (Breiman, 2001; J. H. Friedman, 2001; Hastie et al., 2001). A recent study by X. Chen et al. (2022) used thousands of XBRL[3] items as features and modern ML algorithms to predict earnings. Jones et al. (2023) supported that ML can find hidden patterns in the data that linear methods cannot capture. The second methodological approach that is suggested by Jones et al. (2023) is to use theoretically defined models, training the algorithms in features that are selected by the researchers. In Table 3.1 the most relative studies that use ML algorithms to predict earnings are presented along with the algorithms that have been used.

Table 3.1: Summary of studies, features, targets, and algorithms

Study	**Features**	**Target**	**Algorithm**
Ou and Penman (1989)	Accounting & market-based ratios	ΔEPS‡	LOGR
You and Cao (2021)	Financial statement items, HVZ, SO, LM	EPS	OLS, LASSO, RID, GBR, ANN
X. Chen et al. (2022)	Detailed financial data, DuPont	ΔEPS‡	RF, SGB, LOGR
J. O. S. Hunt et al. (2022)	64 variables from Ou and Penman (1989)	ΔEPS‡	LOGR, RF
Jones et al. (2023)	PZ model, 64 variables from Ou and Penman (1989)	ΔRNOA	TreeNet, OLS
Easton et al. (2024)	FRY, VY, BCG, HVZ, LM	EPS	KNN

Note: LOGR: logistic regression, RID: ridge regression, GBR: Gradient Boosting Regressor, RF: random forest, SGB: stochastic gradient boosting, KNN: k-nearest neighbors. ‡ directional change of earnings, binary classification. HVZ: Hou et al. (2012), SO: So (2013), LM: Li and Mohanram (2014), PZ: Penman and Zhang (2004), FRY: Fairfield et al. (2009), VY: Vorst and Yohn (2018), BCG: Blouin et al. (2010).

[3] eXtensible Business Reporting Language

3.4 Data description and sample formation

The sample is consisted of publicly listed[4] firms from 30 European countries, namely: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Liechtenstein, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Russian Federation, Slovakia, Slovenia, Spain, Sweden, Switzerland and the United Kingdom. We collect all accounting and market data from Worldscope and Datastream International for the period 1988 - 2020. According to the literature, firms classified as banks, insurance, financial services, and oil & gas are excluded from the sample (Cazavan-Jeny et al., 2011; Dargenidou et al., 2021). Financial firms are excluded[5] because they follow different accounting standards; oil & gas firms are given the option to capitalize exploration and evaluation costs under IFRS 6 Exploration for and Evaluation of Mineral Resources. The capitalized amounts may be captured in Datastream as development costs (Dargenidou et al., 2021).

In the next step, firms that do not report either capitalized development costs or research expenses, meaning they do not report any R&D activity are also excluded from the sample. A preliminary analysis has been conducted and in Figure 3.1 the firm-year observations per year are presented. It is noticed that before 2005 there is not a significant amount of available data for capitalized development costs, therefore we limit our analysis for the period from 2005 (the implementation of IAS 38) to 2020. Moreover, firms that do not report in IFRS after 2005 are also excluded. The approach of Anand et al. (2019) is followed, thus only firms that have at least three years of data available are included in the sample.

Although there is some controversy on the identification of outliers and the way they are treated (See: Andrews and Pregibon (2018) and Sullivan et al. (2021)), it is common in the accounting literature to winsorize accounting variables to 1% and 99%

[4]The selected firms are listed on the major stock exchange of each country; both active and inactive firms are included in order to avoid survivorship bias

[5]Firms not classified in any of the industries or industry is a missing value and unquoted equities are also excluded from the sample.

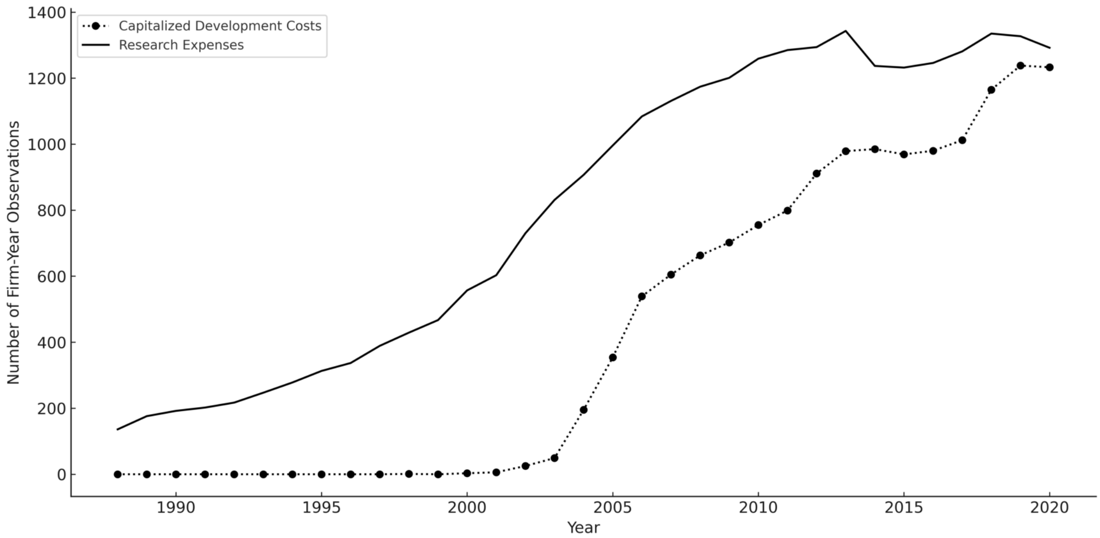

Figure 3.1: Number of firm-year observations for capitalized development costs and research expenses per year

levels, so this approach is followed in this research too. In Table 3.2 the initial sample formation is summarized.

Table 3.2: Initial sample selection process

Criteria	**Firm-years**
European listed firms	232,716
Less: banks, insurance, financial services, oil & gas	(26,334)
	206,382
Less: zero R&D activity	(173,007)
	33,375
Less: firm-year obs. before 2005 and firm-years not reporting in IFRS after 2005	(9,850)
	23,525
Less: firms with less than three years of data	(620)
Total	**22,905**

Artikis et al. (2022) used a sample of 16 European countries in their study: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. Their rationale for choosing this sample is that those countries can be characterized as advanced economies and they share similar legal tradition. Although Artikis et al. (2022) study is on a completely different stream of the accounting and finance literature (asset growth anomaly), their sample selection is similar to R. Zhao (2002) who examined R&D accounting treatment in countries with different legal and cultural characteristics. Artikis et al. (2024) in their study on profitability prediction, conducted a sensitivity analysis by distinguishing between advanced Western countries and the rest of Europe, based on the hypotheses that advanced countries offer higher investor protection and earnings of higher quality. An analysis of the distribution of the firms across countries has been conducted and the results are presented in Table 3.3

Table 3.3: Distribution of firms by country

Country	**Number of firms**	**Firm-year observations**
Austria	33	390
Belgium	58	648
Bulgaria	3	17
Croatia	3	14
Cyprus	3	21
Czech Republic	2	19
Denmark	55	604
Estonia	6	49
Finland	78	960
France	262	2,782
Germany	286	3,259
Greece	61	685
Hungary	7	80
Iceland	1	16
Ireland	19	196
Italy	100	956
Liechtenstein	1	6
Luxembourg	14	125
Malta	3	28
Netherlands	66	656
North Macedonia	1	3
Norway	71	660
Poland	148	1,172
Portugal	17	168
Russian Federation	45	346
Slovakia	1	3
Spain	64	739
Sweden	221	2,055
Switzerland	99	1,091
United Kingdom	504	5,157

Countries with less than 10 firms are dropped from the sample (Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Hungary, Iceland, Liechtenstein, Malta, North Macedonia and Slovakia). Therefore the sample is consisted of 19 European countries.

Sample size is slightly reduced to 22,649 firm-year observations[6].

Cazavan-Jeny et al. (2011) classify a firm as a capitalizer if it reports a non-zero R&D asset for at least one year during the examined period, otherwise it is classified as an expenser.

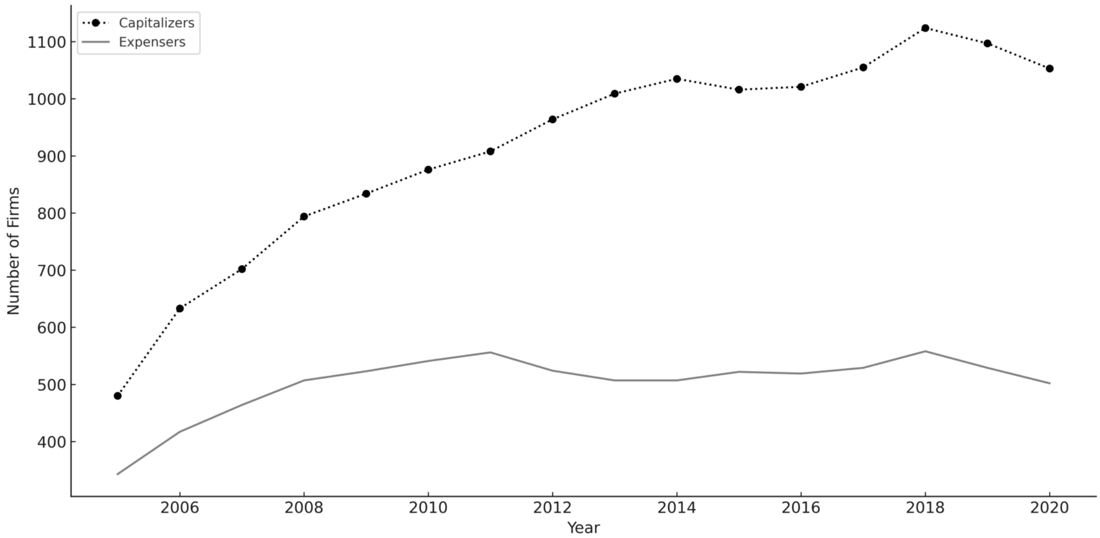

Figure 3.2: Firms classification based on accounting treatment

The effect of IAS 38 is visible in Figure 3.2. We see a steep increase in firms that are classified in the years after 2005. Capitalizers according to the classification that was made do not necessarily capitalize every year. However, in Figure 3.1, we also notice that each year there is an increasing trend in firms that report capitalized development costs in the concurrent year.

To get a clearer picture of the trend, in Figure 3.3 the accounting treatment of R&D for each year is plotted. If a firm capitalizes R&D costs in the concurrent year it is classified as a capitalizer, otherwise as an expenser D. R. Oswald and Zarowin (2007). Until 2011, firms that reported only expensed R&D costs were obviously more compared to those who reported a portion of their R&D costs a capitalized development costs. This behaviour was inverted after 2011. In Table 3.4 the distribution of the firms across the industries is reported.

It is observed that the relative proportions of capitalizers and expensers differs across industries. In the Automobiles & Parts sector nearly 80% of the firms are clas-

[6]After dropping missing values the sample size is reduced to 16,294 firm-year observations

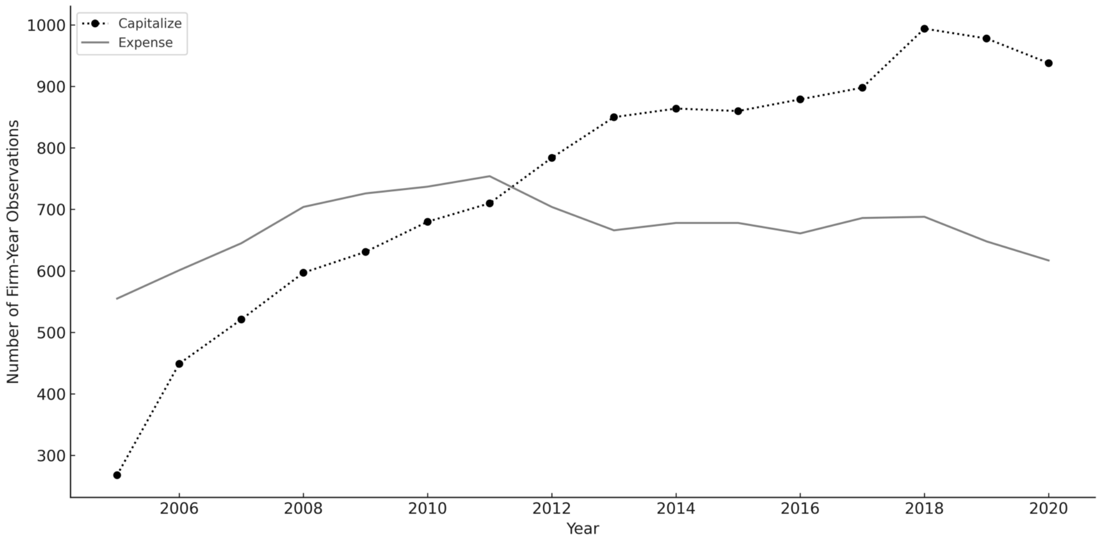

Figure 3.3: R&D accounting treatment per year

Table 3.4: Industry classification

Industry	Total	%	Cap.	Exp.	R&D Inten. (%)
Automobiles & Parts	56	80.36	45	11	0.15
Basic Resources	115	54.76	63	52	0.01
Chemicals	80	51.25	41	39	0.02
Construction & Mats	121	59.49	72	49	0.01
Consumer Prod & Svs	126	62.71	79	47	0.03
Drug & Grocery Stores	36	50.00	18	18	0.01
Energy	100	57.00	57	43	0.02
Food, Bev. and Tobacc	89	35.96	32	57	0.02
Health Care	361	56.79	205	156	0.17
Ind. Goods & Services	441	72.55	320	121	0.07
Media	55	67.27	37	18	0.02
Real Estate	23	43.48	10	13	0.14
Retailers	36	63.89	23	13	0.01
Technology	385	69.86	269	116	0.17
Telecommunications	80	62.50	50	30	0.05
Travel & Leisure	27	70.37	19	8	0.01
Utilities	68	55.88	38	30	0.01

Note: R&D intensity is calculated as the total R&D expenditure (irrespective of its treatment) to sales.

sified as capitalizers while in the Food & Beverage only 36% are capitalizers. Similar observation can be made for the R&D intensity, where Healthcare and Technology exhibit the highest intensity (17%). Industries with a higher proportion of capitalizers do not necessarily exhibit higher R&D intensity. Similarly, in Table 3.5, capitalizers range from 25% in Greece to 86% in Poland and Spain.

Table 3.5: Country classification

Country	Total	%	Cap.	Exp.	R&D Inten. (%)
Austria	33	45.45%	15	18	0.014
Belgium	58	63.79%	37	21	0.03
Denmark	55	63.64%	35	20	0.01
Finland	78	43.59%	34	44	0.05
France	262	63.74%	167	95	0.05
Germany	286	50.35%	144	142	0.08
Greece	61	24.59%	15	46	0.01
Ireland	19	52.63%	10	9	0.02
Italy	100	70.00%	70	30	0.03
Luxembourg	14	57.14%	8	6	0.005
Netherlands	66	72.73%	48	18	0.11
Norway	71	64.79%	46	25	0.05
Poland	148	86.49%	128	20	0.08
Portugal	17	70.59%	12	5	0.002
Russian Federation	45	48.89%	22	23	0.02
Spain	64	85.94%	55	9	0.02
Sweden	221	74.21%	164	57	0.07
Switzerland	99	47.47%	47	52	0.01
United Kingdom	504	63.69%	321	183	0.03

Note: R&D intensity is calculated as the total R&D expenditure (irrespective of its treatment) to sales.

In Figure 3.4 it is noticed that younger firms tend to capitalize compared to older. Firms that are incorporated at least for 30 years, tend to expense their R&D costs. A possible explanation could be that older firms spend more on maintenance costs and upgrades for their already established products.

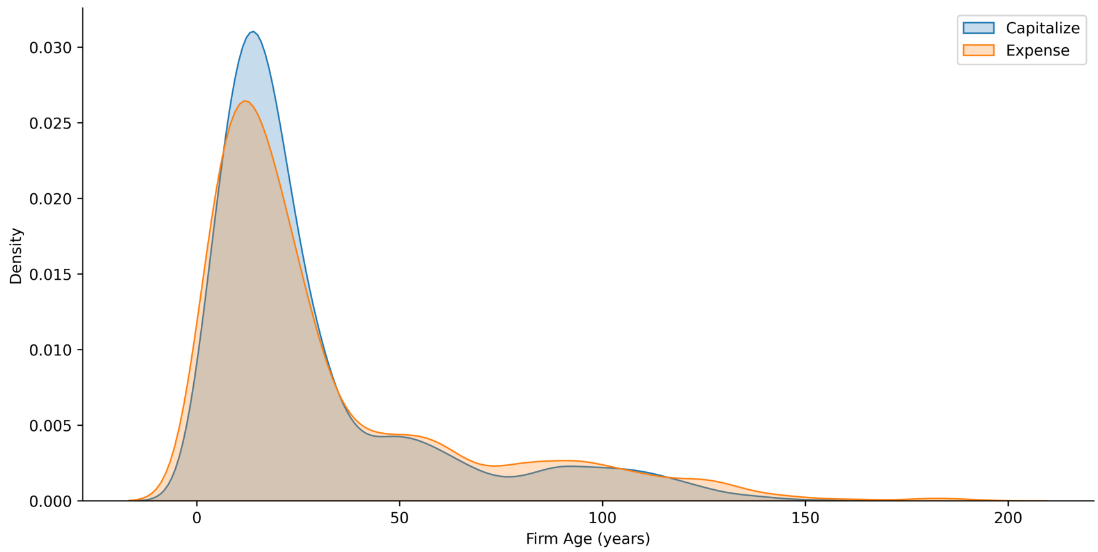

Figure 3.4: Classification based on firm age

Note: Firm age is calculated as the concurrent year minus the year of incorporation

3.5 Methodological approach

In this study, the suggestions of Jones et al. (2023) are followed, thus the directional changes of earnings are forecasted with ML algorithms. Theoretical models that link earnings and R&D are tested. At the same time, the approach of Ou and Penman (1989) and X. Chen et al. (2022) is followed, thus raw accounting items from the financial statements are used to predict earnings. More specifically, the income models of Cazavan-Jeny et al. (2011) is used. Their models were tested in a European setting (where capitalization is allowed) and are suitable for the sample of this study. Moreover, they are one of the few studies that use a sales growth and an income model supplementary to value-relevance models (Lev and Sougiannis (1996) used an operating income model, but they used simulated capitalized R&D costs). The price model of Ohlson (1995) and the returns model of Easton (1998) are used, as modified by Cazavan-Jeny and Jeanjean (2006) and Cazavan-Jeny et al. (2011).

Relevant papers about profitability-earnings forecasting using ML, like X. Chen et al. (2022) and J. O. S. Hunt et al. (2022) predict the directional change of earnings. The justification behind this approach is that predicting the level of earnings or the change of earnings does not yield high prediction performance (Gerakos & Gramacy, 2013; Kothari, 2001; Li & Mohanram, 2014). By predicting the directional change, more accurate forecasts are made, as in this way the variability in earnings changes is reduced (Freeman et al., 1982). Furthermore, the prediction of the directional changes is economically meaningful, as for example portfolios can be constructed based on the direction of earnings' change (X. Chen et al., 2022; Ou, 1990; Ou & Penman, 1989; Wahlen & Wieland, 2011). Anand et al. (2019) back up the idea that making out-of-sample predictions about profitability is an essential part of fundamental research, but even more advanced regression models can't outperform random walks. For the sake of completeness, in this study both the directional changes, the amount of change and the level of earnings are predicted.

X. Chen et al. (2022) justify the use of ML for three reasons. Recent evidence suggests that ML algorithms like random forests and stochastic gradient boosting have been proven quite effective and successful in real-world problems (Liu, 2021; Mullainathan & Spiess, 2017). ML algorithms are designed for prediction, they can find complex relations between predictors and the predicted variable and they can accommodate a large number of predictors compared to traditional regressions. However, the use of complex and advanced ML algorithms cannot always outperform simple linear methods in forecasting directional changes of profitability. Therefore, simple algorithms have to be used as benchmarks and be compared with more complex ones (Belesis et al., 2023).

3.5.1 Direction of profitability changes

Ou and Penman (1989) noticed that earnings increases tend to outnumber decreases. To mitigate for this, they removed the firm-specific drift. They defined the variable $earnings_{it+1} - earnings_{it} - drift_{it+1}$, where the drift was estimated as the mean earnings change over the four years prior to year $t+1$. X. Chen et al. (2022) replicated their approach. They noticed that by removing the drift, the class imbalance problem is mitigated. Class imbalance in ML problems is a known problem that negatively affects certain algorithms' performance (Japkowicz & Stephen, 2002). In addition to that, it is more useful to predict changes without the drift, as some changes are anticipated due to the drift. Following the literature, the drift is removed from the profitability variables and the increase/ decrease is coded after the removal. In Table 3.6 the class balance of each target variable is presented. It is noticed that in the sample, there more decreases in profitability for all the measures. Further, there is not a major change in the class balance after the drift removal.

Table 3.6: Class balance for target variables

Panel A: Without drift		
	Decrease	*Increase*
ΔROA	7,466	8,828
ΔPR	7,016	9,278
ΔRET	7,912	8,382
ΔEPS	7,428	8,868
Panel B: With drift		
	Decrease	*Increase*
ΔROA	8,300	7,994
ΔPR	7,923	8,371
ΔRET	8,397	7,897
ΔEPS	7,063	9,231

Note: Drift removed according to the formula $X_{i,t+1} - X_{it} - drift_{i,t+1}$ of Ou and Penman (1989).

3.6 Descriptive statistics

In Table 3.7 the descriptive statistics for the target variables are reported. According to the descriptive statistics, capitalizers are more profitable (on average) compared to expensers in terms of ROA and OROA. In terms of EPS, expensers appear to exhibit higher earnings, and they have higher price. On average, capitalizers exhibit higher returns, yet the difference between the two groups is not statistically significant. Definitions of variables are provided in Appendix A.

Table 3.7: Descriptive statistics for target variables

	Expensers			Capitalizers			Test	
	N	**Mean**	**Median**	**N**	**Mean**	**Median**	**t-stat**	**p-value**
ROA	5,617	0.055	0.062	10,677	0.099	0.080	-14.598	0.000***
EPS	5,617	3.572	0.561	10,677	2.273	0.512	9.235	0.000***
PR	5,617	54.438	11.508	10,677	27.563	7.416	14.368	0.000***
RET	5,617	0.085	0.021	10,677	0.100	0.012	-1.605	0.109

Note: Variables are winsorized at 1% and 99% levels. Significance levels are denoted by * ($p < 0.1$), ** ($p < 0.05$), and *** ($p < 0.01$).

Table 3.8 reports the R&D characteristics for the full sample and for the sub-samples of expensers and capitalizers. For capitalizers, the mean R&D asset is 6.4% of the total assets. R&D is the 2.37% of the sales for the expensers and 1.779% for the capitalizers. In terms of cash flows of R&D expenditure, they are on average, marginally larger for expensers compared to capitalizers (5.7% versus 5.4% of the total assets). Following Cazavan-Jeny et al. (2011), cash flows from R&D are split in three components, cash flow from R&D expensed by expensers ($CFRDEXP$), cash flow of R&D expensed by capitalizers ($CFRDEXPCAP$) and finally, cash flow of R&D capitalized by capitalizers. Capitalizers can either capitalize or expense their R&D costs, while expensers expense all of their R&D. According to this, capitalizers capitalize 1.1% of their R&D (to total assets) while they expense 4.3%. Taking under consideration that for capitalizers, CFRD is 5.4%, it occurs that they capitalize at about 20% of their R&D and they expense the rest (80%). Expensers on the other hand, based on the condition that they always expense, exhibit equal $CFRD$ and $CFRDEXP$.

In Table 3.9 the descriptive statistics for the economic characteristics of the firms ($CAPEX, PTB, SIZE$) and the variables of the market models are reported. It is noticed that expensers, on average, are larger compared to capitalizers ($SIZE : 13.013 > 12.834$). In terms of Price-to-Book ratio (PTB), despite the marginal difference in their means, this difference is statistically significant. Expensers exhibit on average, nearly double Book Value per Share ($ABVPS : 21.485 > 11.653$). Finally, R&D variables per share descriptive statistics are reported.

Table 3.8: Descriptive statistics for R&D characteristics

	Full sample		Expensers		Capitalizers		Tests	
	N	**Mean**	**N**	**Mean**	**N**	**Mean**	**t**	**Prob.**
		Median		**Median**		**Median**		
RDS	16294	2.370	5617	3.514	10677	1.779	1.689	0.091*
	16294	0.013	5617	0.016	10677	0.012		
RD_ASSET	16294	0.042	5617	0.000	10677	0.064	-16.122	0.000***
	16294	0.000	5617	0.000	10677	0.010		
CFRD	16294	0.057	5617	0.062	10677	0.054	3.183	0.001***
	16294	0.017	5617	0.014	10677	0.019		
CFRDCAPCAP	16294	0.007	5617	0.000	10677	0.011	-6.101	0.000***
	16294	0.000	5617	0.000	10677	0.000		
CFRDEXPCAP	16294	0.028	5617	0.000	10677	0.043	-49.012	0.000 ***
	16294	0.000	5617	0.000	10677	0.010		
CFRDEXP	16294	0.022	5617	0.062	10677	0.000	36.364	0.000 ***
	16294	0.000	5617	0.014	10677	0.000		

Note: Variables are winsorized at 1% and 99% levels. Significance levels are denoted by * ($p < 0.1$), ** ($p < 0.05$), and *** ($p < 0.01$).

Table 3.9: Descriptive statistics for economic characteristics

	Full sample		Expensers		Capitalizers		Tests	
	N	**Mean**	**N**	**Mean**	**N**	**Mean**	**t**	**Prob.**
		Median		**Median**		**Median**		
CAPEXS	16294	0.000	5617	0.000	10677	0.000	-8.331	0.000***
	16294	0.000	5617	0.000	10677	0.000		
ABVPS	16294	15.042	5617	21.485	10677	11.653	12.807	0.000***
	16294	3.623	5617	4.786	10677	3.152		
RDEXPEXPS	16294	0.321	5617	0.931	10677	0.000	33.234	0.000***
	16294	0.000	5617	0.132	10677	0.000		
RDCAPCAPS	16294	0.057	5617	0.000	10677	0.086	-28.355	0.000***
	16294	0.000	5617	0.000	10677	0.000		
RDEXPCAPS	16294	0.323	5617	0.000	10677	0.492	-46.770	0.000***
	16294	0.000	5617	0.000	10677	0.034		
PTB	16294	2.720	5617	2.936	10677	2.606	-5.594	0.000***
	16294	1.830	5617	1.940	10677	1.780		
SIZE	16294	13.013	5617	13.353	10677	12.834	-12.566	0.000***
	16294	12.812	5617	13.387	10677	12.599		

Note: Variables are winsorized at 1% and 99% levels. Significance levels are denoted by * ($p < 0.1$), ** ($p < 0.05$), and *** ($p < 0.01$).

3.7 Categorical variables encoding

Some algorithms require the features to be strictly numerical (SVM) while others, like tree based algorithms, can handle non-numerical features (Coppersmith et al., 1999). A common approach is one-hot-encoding (Myers et al., 2010; O'Grady & Medoff, 1988).

For a categorical variable having *n* unique values, one-hot encoding creates *n* binary columns each representing a category and indicating that the category exists with 1, and otherwise with 0. In this study, the categorical variables are the industry and the country in which each firm belongs. Since those variables do not include many categories, one-hot-encoding is a suitable approach, as the dimensional space will not increase a lot.

Recent evidence suggests that target-based encoding yields better performance for the algorithms. Pargent et al. (2022) examined many encoding techniques and made comparisons using all the traditional supervised learning algorithms. They found that target encoding with regularization performs better in most instances compared to other kind of encoding. Yet, this is a more complex approach. Instead, frequency encoding is used, as proposed by Kosaraju et al. (2023).

3.8 Raw accounting data

Raw accounting data (or detailed financial information, as mentioned by some researchers) have been used by recent studies that use ML in accounting (Bao et al., 2020; Cecchini et al., 2010; X. Chen et al., 2022). Raw accounting data are referred to items from the financial statements as reported. X. Chen et al. (2022) have used thousands of XBRL items, while Bao et al. (2020) and Cecchini et al. (2010) have used a smaller set of 24 to 40 raw data items. Bao et al. (2020) also used a "hybrid" model which combines raw accounting data and financial ratios. Bao et al. (2020) found that using raw accounting items and an ensemble learning model offers better detection performance for accounting fraud compared to theoretically established financial ratios. Similarly, X. Chen et al. (2022) found that detailed financial data outperform conventional models (DuPont).

In this study, there are 31 available raw accounting downloaded from DataStream. These items are reported in Appendix A: Variable definitions in Table A.2. These raw items can be used in three ways according to the literature. The "kitchen sink" approach of X. Chen et al. (2022), where all the items are used to predict the profitability measures without a priori selection. Another approach, suggested by Bao et al. (2020) is to use different sets of raw items, by decomposing the financial ratios of the theoretically established models into the raw items that they were used to construct them. Ou and Penman (1989) used a stepwise regression to select the most appropriate features. This approach though may introduce an issue. Theoretical models have been established following the economic theory. Similarly, raw accounting items which occurred from the decomposition of the ratios used in the theoretical models are ”linked” to the theory (as they occurred from theoretically chosen ratios). If a stepwise approach is used, such as Sequential Feature Selection (SFS), then the algorithm examines all the features, selecting a subset of features based on a metric (e.g. feature importance). Therefore the algorithm has already ”seen” all the available features, and selected those that performed better. Harrell (2015) states that this procedure is unacceptable, as *”[...]it violates every principle of statistical estimation and hypothesis testing”*.

To address this issue of model selection, a more simple approach is selected, the correlation analysis of the raw accounting items. None of the relevant studies has used it before, but it is very simple and straightforward. The first issue that arises is that the target variables are dichotomous, they are coded to take values 0 or 1 (decrease or increase in profitability measure) while the features are continuous. Therefore Pearson correlation is not suitable. On the other hand, the Point-Biserial Correlation[7] (which is a special case of Pearson) is the appropriate method. The correlation tables are reported in Appendix B: Point-Biserial Correlations.

Bao et al. (2020) raised the issue of the difference in scale of raw accounting items. In the theoretical models variables are scaled by total assets, common shares outstanding, sales etc. This is not a case for raw accounting data. Therefore, their approach is followed and the raw items are normalized. The input vector of each firm-year observation which includes the raw data, is standardized in such a way that the output vector's length is one; $x' = \frac{x}{||x||}$, where the divisions are performed element-wise. For example, the vector (2,10) would be normalized as: $||x|| = \sqrt{2^2+10^2} = \sqrt{4+100} = \sqrt{104}$,

[7]See more in Bonett (2020) and Tate (1954) and in Appendix B: Point-Biserial Correlations.

$x' = \left(\frac{2}{\sqrt{104}}, \frac{10}{\sqrt{104}}\right) \approx \left(\frac{2}{10.198}, \frac{10}{10.198}\right) \approx (0.196, 0.980).$

After the calculation of Point-Biserial correlations, for each target variable a group of raw accounting items is derived according to the statistical significance of their relation. Although some variables may be significant but exhibit low correlation with the target, for simplicity and uniformity, all significant variables are included. In Table 3.10 the selected features per target variable are illustrated. Kitchen-sink approach is the use of all available raw accounting data for each target. Interestingly, capitalized R&D have a positive and significant correlation with the market variables and ΔEPS, while the expensed R&D are negatively correlated only with ΔROA. In addition to the Point-Biserial correlations, the Pearson correlations are also calculated for the raw accounting items. In this alternative approach, for each selected set of raw accounting items for each target, high-correlated pairs of features (correlation $> 70\%$) are identified. The feature of the pair with the smallest correlation to the target is removed. Pearson correlations for the features are reported in Appendix B: Point-Biserial Correlations.

Table 3.10: Feature selection according to Point-Biserial correlation

Feature	**Kitchen-sink**	**DROA**	**DEPS**	**DPR**	**DRET**
DEFTA	✓	✓	✓		
TAX	✓	✓	✓		✓
NPS	✓	✓			
TA	✓	✓	✓	✓	
CE	✓				✓
CASHDIVS	✓			✓	✓
NETSAL	✓	✓	✓	✓	
AR	✓	✓	✓	✓	
CAPEX	✓			✓	✓
DEPR	✓	✓	✓	✓	
DEPAM	✓				✓
PPEG	✓	✓		✓	
PPEN	✓	✓		✓	
OINTGA	✓				
INVT	✓	✓		✓	
LTBOR	✓	✓	✓	✓	✓
COGS	✓	✓	✓	✓	
MI	✓	✓			
IBT	✓	✓	✓		✓
ECR	✓	✓	✓		
CA	✓	✓	✓	✓	
CL	✓	✓	✓	✓	
LTL	✓		✓	✓	
TD	✓		✓	✓	
RDCAP	✓		✓	✓	✓
RDEXP	✓	✓			
RDAM	✓		✓	✓	✓
IINC	✓				
STI	✓			✓	
XIT	✓				
OIBDAM	✓	✓	✓		✓
CSHOUT	✓	✓	✓	✓	
IP	✓				
CEQ	✓			✓	
NETINC	✓	✓	✓		✓

3.9 Cross-validation

In ML studies, the sample typically is split into training, validation and test samples. Models are estimated in the training sample, and tested in the validation sample. In this phase, the algorithms are tuned (parameters tuning). Then the best performing model is selected and its real performance is obtained by making predictions in the test sample (X. Chen et al., 2022). A simple approach is to split the sample randomly, e.g. 70% of the data for training-validation and 30% for the test sample, or in a 60%-40% split. Because a 70%-30% split is a rule of thumb (and they do not always work well), cross-validation techniques are used. An k-fold cross validation approach is one of the most useful techniques for parameter tuning. For this technique, a sample is divided into k-folds randomly and the model is trained k-1 times using all folds apart from one fold which is used as an excluded holdout test sample, in which the out-of-sample performance is measured. This process is thoroughly repeated in all possible ways by excluding different folds, so that parameters are finally fitted to optimize average performance of excluded folds. The final out-of-sample performance of the model is the average performance of all test folds. This process is depicted in Figure 3.5, which represents a random 5-fold cross-validation. The model is trained-validated five times, and five out-of-sample prediction scores are obtained.

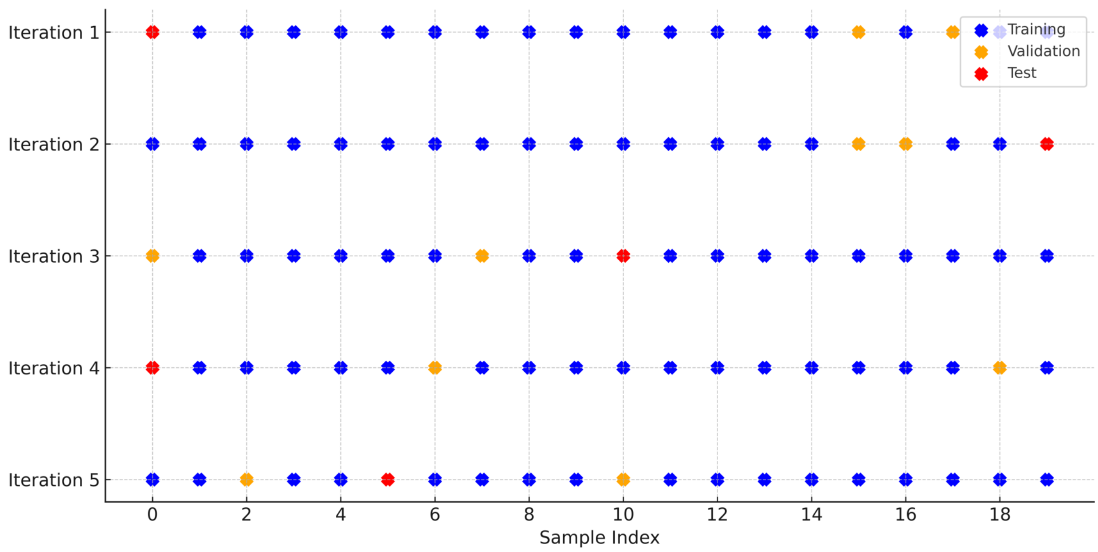

Figure 3.5: Random 5-fold cross-validation

This approach though is not suitable for accounting data, due to their intertemporal nature. Therefore the split must be made chronologically. A single-split approach is depicted in Figure 3.6, which is replicated from Bertomeu et al. (2021). Bertomeu et al. (2021) split their sample from 2001 to 2009 for training, 2010-2011 for validation and 2012-2014 for test.

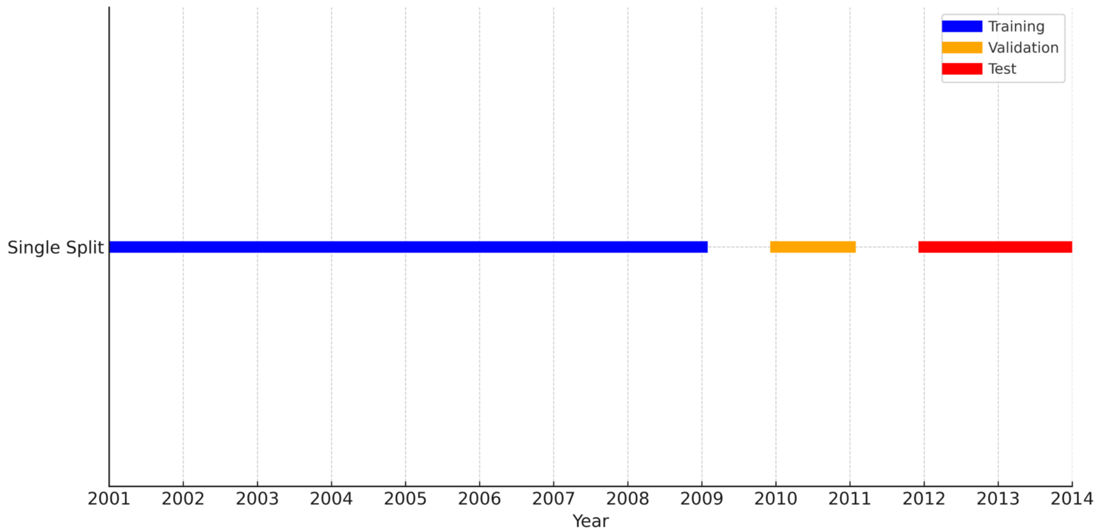

Figure 3.6: Chronological split, replicated from Bertomeu et al. (2021)

Jones et al. (2023) used a k-fold cross-validation approach. Their sample was subdivided into seven training periods and seven test periods. The training period samples are, in increasing order of the end date of the period, 1993-1998, 1993-2001, 1993-2004, 1993-2007, 1993-2010, 1993-2013, and 1993-2016. The corresponding test samples are, in increasing order of the start date of the period, 2000-2002, 2003-2005, 2006-2008, 2009-2011, 2012-2014, 2015-2017, and 2018-2019. The design of the test samples was such to avoid look-ahead. For example, the last year of the first training sample is 1998. In case, there is always a change in profitability a year later that is, in this case it would be in 1999. Therefore, the test sample started in 2000 with the end year of 2002 to avoid the pitfalls of including the actual data one year ahead with the training sample as the latter contains no information from the future. Their approach is illustrated in Figure 3.7.

A similar approach was followed by Bao et al. (2020). Specifically, the training period from 1991 to 2001 was used for the test year 2003, while the period from 1991 to 2002 was used for the test year 2004, and so on. Their approach is presented in Figure 3.8

A slightly different approach was used by X. Chen et al. (2022). In more detail, the employed strategy was a rolling sample splitting, with respective training and validation samples always kept within a fixed span of years, but designed to advance in a forward

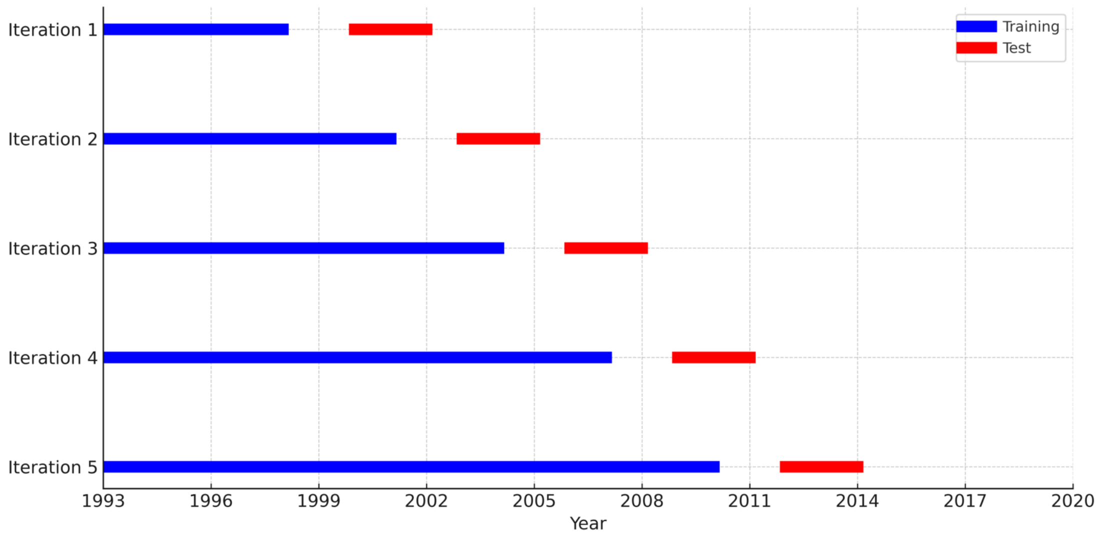

Figure 3.7: Chronological k-fold cross-validation, replicated from Jones et al. (2023)

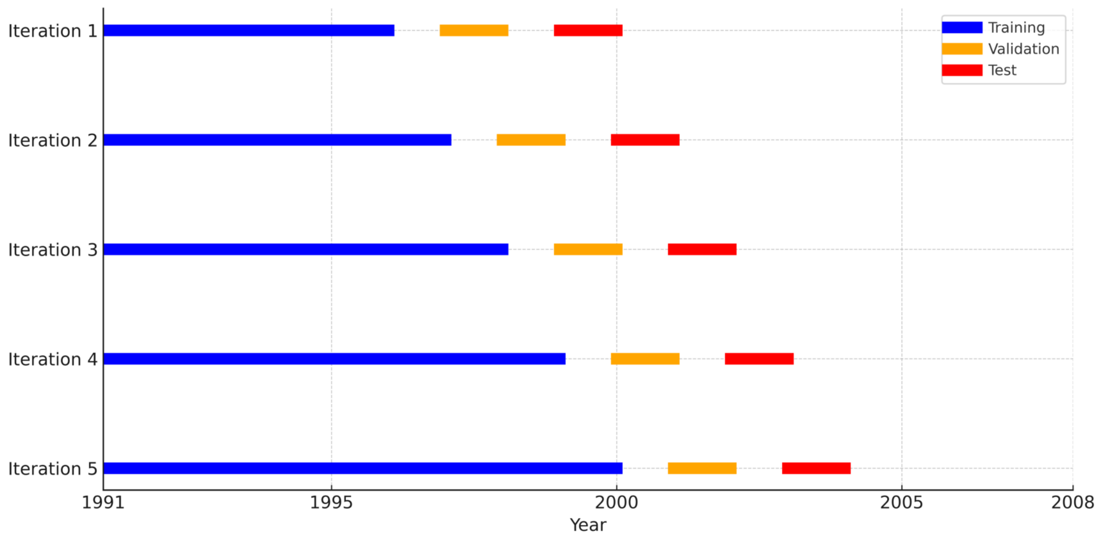

Figure 3.8: Chronological k-fold cross-validation, replicated from Bao et al. (2020)

time order. Each model is trained with a dataset of two years to three years before the targeted year (for example 2012 – 2013 if the target is 2015) with one year preceding the test year, used for validation to adjust the model's parameters (in this example 2014). The second and third preceding years in the training sample are consistently renewed in each iteration, which enables the model to be built using only the most recent information for training. Their approach is illustrated in Figure 3.9.

In this study the approach of X. Chen et al. (2022) is followed for two reasons. Their methodology and data are very similar to this study, and their approach is less computing intensive (the training of the algorithms is performed faster). In the other

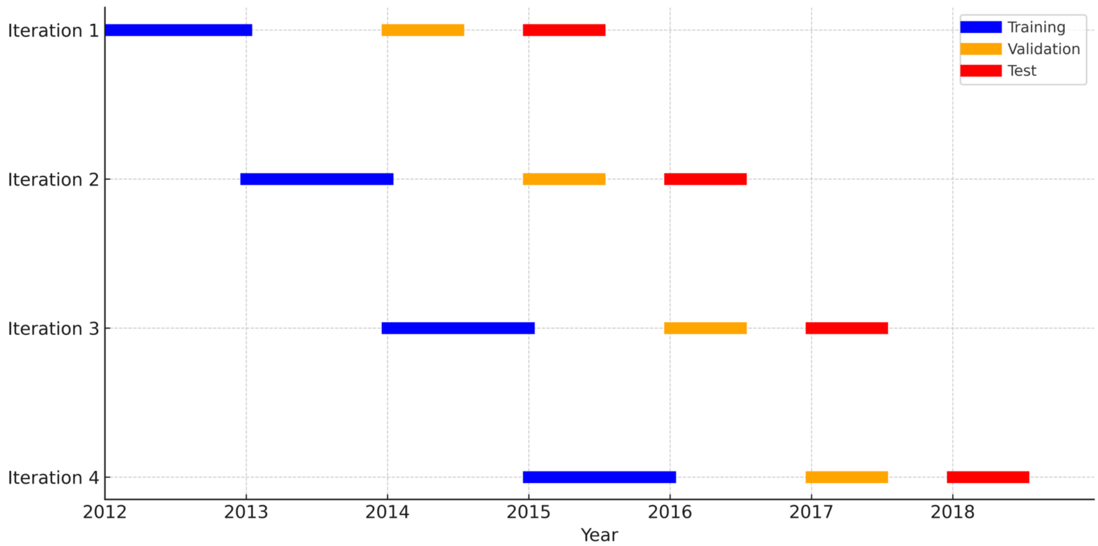

Figure 3.9: Chronological k-fold cross-validation, replicated from X. Chen et al. (2022)

approaches, the training sample is expanding in every iteration, and as it grows it takes more time for the algorithms to train and tune. On the other hand, in X. Chen et al. (2022) approach, the training sample is constant-and not very large. Details on the exact implementation of the cross-validation and how the forecasts are obtained, are available in Appendix C: Python code, in Listing 1.

3.9.1 Parameters

In parameters tuning the approach of X. Chen et al. (2022) is followed. Apart from the logistic regression, which they use as a benchmark, they use random forest and stochastic gradient boosting. Taking under consideration the computational time, instead of using many values for each parameter, they chose to examine values around the default values of each algorithm. Replicating their approach, the parameters grid is presented in Table 3.11, where similar parameters to theirs are reported.

The number of features to be considered by each tree (max features) is randomly selected as $k = \sqrt{p}$ where p is the number of features (Breiman, 2001). We use the same number of trees to grow for both algorithms, following X. Chen et al. (2022). Instead of using the Stochastic Gradient Boosting algorithm, the XGBoost algorithm was used because it was faster to tune.

Table 3.11: Hyperparameters for Random Forest and Stochastic Gradient Boosting

Hyperparameter	Random Forest	XGBoost
Max features	'sqrt'	'sqrt'
# of trees	500, 600, 700, ..., 2,000	500, 600, 700, ..., 2,000
Learning rate	-	0.005, 0.01, 0.05
Tree depth	1, 2, 3, 4	1, 2, 3, 4
Min. # of obs. in a leaf	1, 2, 3, 4	1, 2, 3, 4
Bagging	0.5	0.5

In general it was difficult to find the parameters each study has used, as the authors either do not report the parameters in detail, or they may just report the best parameter value they have found (for SVM, C=20 Bao et al. (2020)). Furthermore, it does not mean that a set of parameters that performed well in a previous study will perform equally well in the sample of this study. Shawi et al. (2025) examined traditional classification algorithms by finding the most important parameters for each classifier. They fitted the algorithms across 200 OpenML datasets. For the SVM they found gamma parameter to be the most important and have an optimal performance in the range of 10^{-4} to 10^{0}. For the adaboost they found max depth to be an important parameter, for the range from 2 to 9. On the contrary, other researchers suggest that leaving the parameters in their default value is non-inferior to tuning (Weerts et al., 2020). Artikis et al. (2024) found that even without tuning, the algorithms perform reasonably well. The parameter tuning is performed using grid-search from Pedregosa et al. (2011) (scikit-learn). During the training phase, the algorithm will test all the combinations of the parameters that are defined in the grid and will obtain the combination of those that performed the best in the validation phase. These parameters will be used to obtain the out-of-sample forecasts in the test set. The parameter grid for each algorithm that has been used is reported in Appendix C: Python code.

3.10 Out-of-sample evaluation metrics

The profitability prediction has been treated as a binary classification problem (increase vs decrease), therefore, the directional change of profitability is evaluated using metrics for classification problems. The most straightforward measure is accuracy. Accuracy is defined as ACC $= \frac{TP+TN}{TP+FN+FP+TN}$, where TP is the number of firm-years that are correctly classified as an increase in profitability (true positive), FN is the number of firm-years that exhibit an increase in profitability but are classified as a decrease in profitability (false negative). FP is the number of firm-years is the number of firm-years that exhibit a decrease in profitability but they are classified as an increase in profitability, and finally (false positive), and finally, TN is the number of firm-years that are correctly classified as a decrease in profitability (true negative). This metric is not suitable for datasets which are imbalanced (see Bao et al. (2020)).

3.10.1 AUC

We follow the studies that examine classification problems and suggest the use of AUC (area under curve), as a more suitable prediction performance metric (Bao et al., 2020; Bertomeu et al., 2021; X. Chen et al., 2022; Larcker & Zakolyukina, 2012). Fawcett (2006) described a ROC (receiving operating characteristics) curve as a two-dimensional depiction of a classification algorithm's performance that combines the true positive rate (*sensitivity*) and the false positive rate ($1 - specificity$) in one plot. Sensitivity is specified as $\frac{TP}{TP+FN}$ and specificity as $1 - \frac{TN}{TN+FP}$. The AUC is a portion of the are of the unit square, and takes values between 0 and 1. A random guess is a straight diagonal line, with an area of 0.5. Fawcett explained that that AUC is interpreted as the probability that a randomly chosen profitability increase firm-year observation will be ranked higher by the classifier than a randomly chosen profitability decrease firm-year observation.

Finally, the out-of-sample score is the average of the evaluation metric that has been used for all the out-of-sample predictions that have been made. This is given by the expression:

$$OOS = \frac{1}{n}\sum_{i=1}^{n} Score_i \tag{3.13}$$

where OOS is the average out-of-sample score, n are the out-of-sample years. This is very straightforward when you want to calculate only the AUC score, but it gets more complicated when the ROC-AUC curve has to be plotted for all test years, as there are more than one ways to average the ROC-AUC curve. In this study the average AUC score should be equal to the AUC score that is calculated in the ROC-AUC curve. More details about the averaging methods for the ROC-AUC curve are provided in Appendix E: Introduction to machine learning.

Chapter 4

Empirical analysis

4.1 Introduction

In this chapter the main empirical findings are presented. Initially, logistic regression is used as the benchmark algorithm. The out-of-sample performance of the more complicated algorithms is compared with the performance of logistic regression. In this way, the question of whether more complicated ML algorithms yield superior prediction performance compared to simple methods is addressed.

Then, it is tested whether raw accounting data can outperform theoretically established models. To do so, the best performing algorithm for each model in the first step is used as a benchmark and it is compared to the best performing algorithm that has been used with the raw accounting items. Furthermore, several sets of raw data items are used.

In the third section of the chapter, it is examined whether capitalized R&D have more predictive power than expensed R&D, which is the main research question of this thesis and an ongoing issue in the relevant literature. To examine this, coefficient magnitudes from the logistic regression, feature importance and SHapley Additive exPlanations (SHAP) are used. In the final section, an alternative analysis is performed, where unadjusted for capitalization variables are used.

4.2 Establishing the benchmark model

In this section, profitability is modeled by using the models which are theoretically specified and have been used in previous studies. All the algorithms are tested, and the average AUC score is used for evaluation. The logistic regression is the benchmark algorithm for each model. In this way, it is attempted to examine the first research question of the thesis, whether more advanced ML algorithms exhibit superior performance compared to traditional econometrics (logistic regression).

4.2.1 Out-of-sample predictions for ROA

In Figure 4.1 the averaged ROC curve and the average AUC scores are presented for ROA Model. The model has been used by Cazavan-Jeny et al. (2011), where the variables of the model have been used as features to predict one-step-ahead ROA directional change. The features and the target are reported below:

- Target: ΔROA_{t+1}
- Features: $RDCAP, CFRD, RDCAP * CFRD, PTB, SIZE, CAPEX,$
  $CNTRY\_freq, LVL3S\_freq$

The R&D costs are decomposed in $CFRDEXP$,$CFRDEXPCAP$,
$CFRDCAPCAP$. The XGB algorithm is the best performing algorithm, and it exhibits 3% better AUC score than the logistic regression, which is the benchmark. Interestingly, SVM performed worse than the logistic regression by 1%.

4.2.2 Out-of-sample predictions for Price

The price model of Cazavan-Jeny and Jeanjean (2006) and Cazavan-Jeny et al. (2011) is used. Target and features are listed below:

- Target: ΔPR_{t+1}
- Features: $ABVPS, EPS, RDEXPEXPS, RDCAPCAPS, RDEXPCAPS,$
  $LVL3S\_freq, CNTRY\_freq$

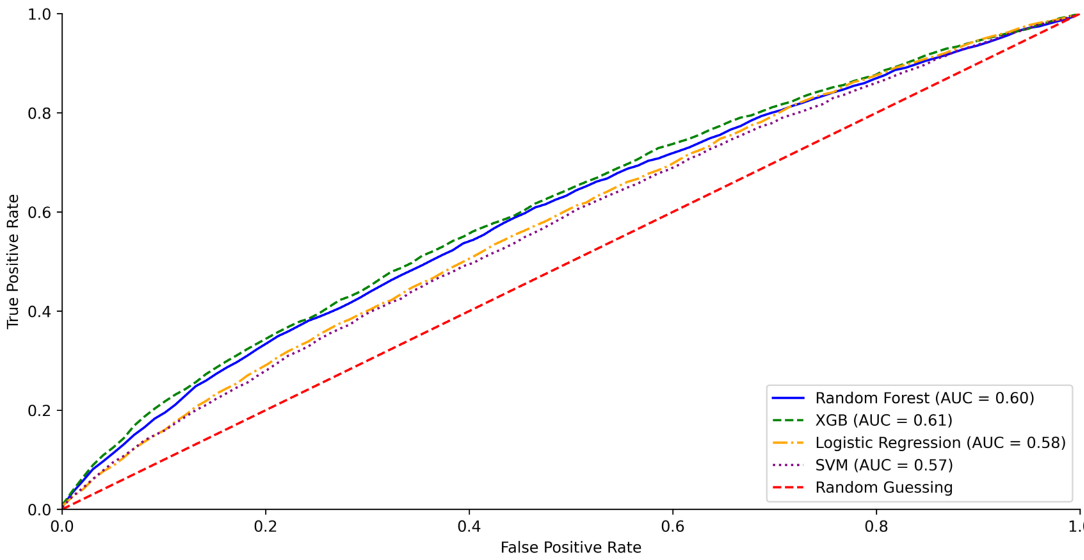

Figure 4.1: ROA Model

All the algorithms exhibited poor performance in this model. Although the random forest and the XGB algorithms performed slightly better compared to the logistic regression (AUC : 51% vs 50%), those results indicate that the algorithms make random guesses. Results are reported in Figure 4.2.

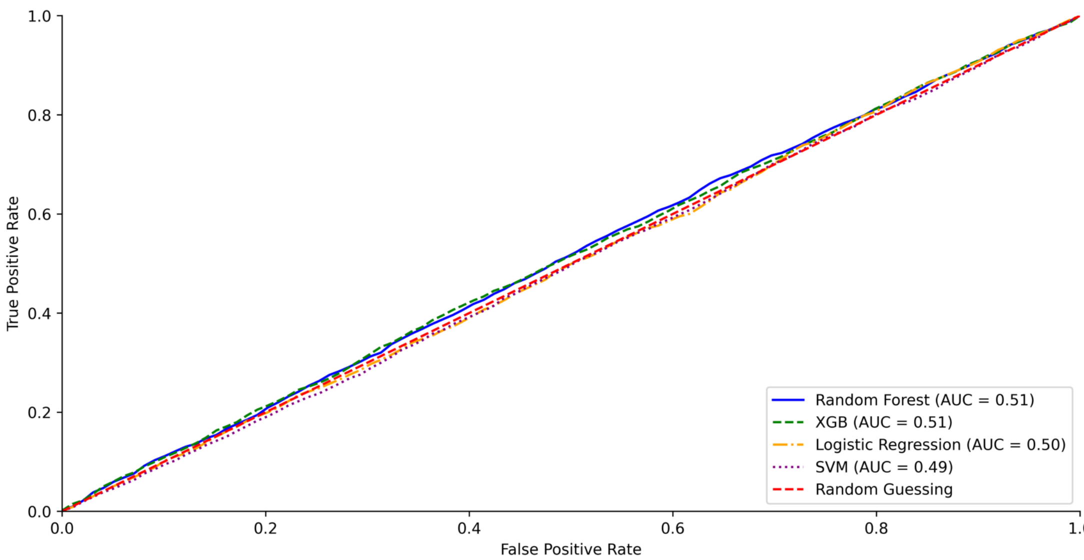

Figure 4.2: Price model

4.2.3 Out-of-sample predictions for Earnings

The features used in the Price model are used to predict one-step-ahead directional changes of earnings, ΔEPS_{t+1}. The results are reported in Figure 4.3.

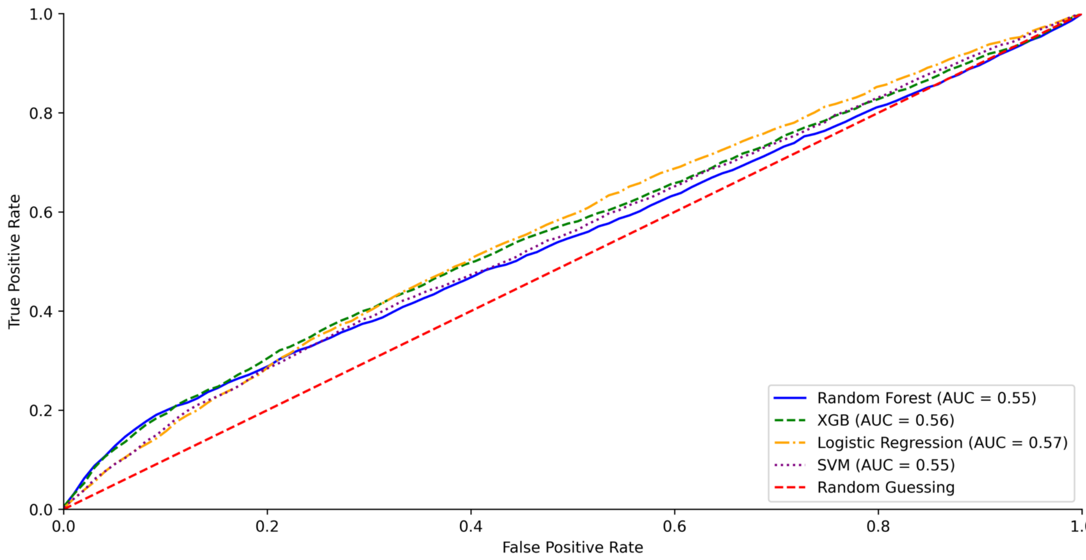

Figure 4.3: EPS model

None of the ML algorithms achieved to outperform the benchmark algorithm. Logistic regression exhibits and out-of-sample AUC score of 57%, while the best performing ML algorithm, XGB, scores 56%.

4.2.4 Out-of-sample predictions for Returns

Returns are modeled according to Cazavan-Jeny and Jeanjean (2006) and Cazavan-Jeny et al. (2011). The target and features are listed below:

- Target: ΔRET_{t+1}
- Features: $AEMV, DAEMV, DRDEXPEXPMV, DRDCAPCAPMV,$
  $DRDEXPCAPMV, LVL3S\_freq, CNTRY\_freq$

The out-of-sample results are presented in Figure 4.4. Random forest and XGB achieved the highest out-of-sample AUC score (57%), 2% above the benchmark model. Once again, SVM did not perform better than the logistic regression.

4.3 Out-of-sample performance of raw accounting data

In this section the out-of-sample performance of raw accounting data is tested. For each model specified by the theory, a set of raw accounting items from the financial

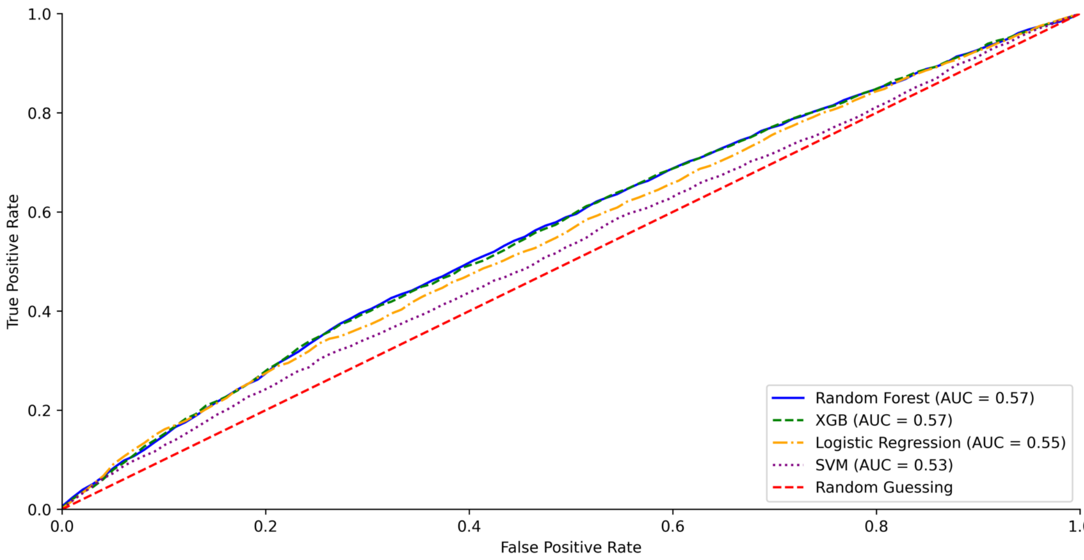

Figure 4.4: Returns model

statements is used as features. As a benchmark, the best performing algorithm of the initial analysis is used. Initially, all the available accounting items are used (kitchen-sink approach). In the second step of the analysis, the raw accounting items that were chosen according to the Point-Biserial correlations are used.

4.3.1 Kitchen-sink approach

In Figure 4.5 the out-of-sample AUC score for the directional changes of ROA is illustrated. All the algorithms apart from the random forest beat the benchmark performance $(AUC = 61\%, XGB)$.

The best performing algorithm is again the XGB ($AUC = 64\%$), while the SVM surpassed the random forest, but still performs 1% worse than the logistic regression. In Figure 4.6 the score for the directional changes of Price is illustrated. A slight improvement has been noticed, as the best performing algorithms, random forest and XGB beat the benchmark. In general, improvement is noticed for all algorithms.

In Figure 4.8, the out-of-sample AUC score for the directional changes of EPS is illustrated. Again, increase in performance is noticed for all the algorithms. The logistic regression is still the better performing algorithm, along with XGB $(AUC = 61\%)$.

On the other hand, regarding the returns, it was not found that a model with raw ac-

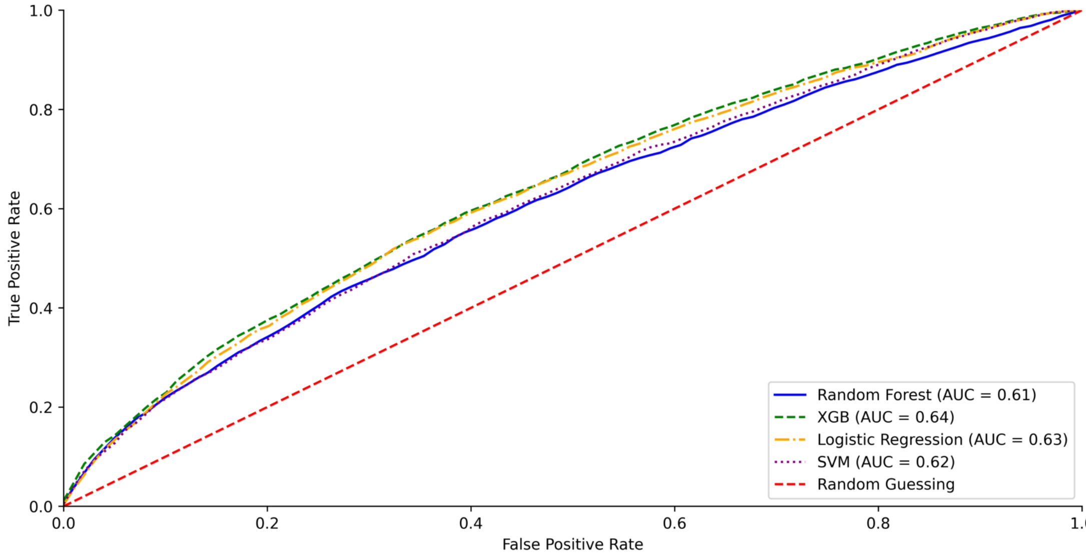

Figure 4.5: ROA-raw accounting items

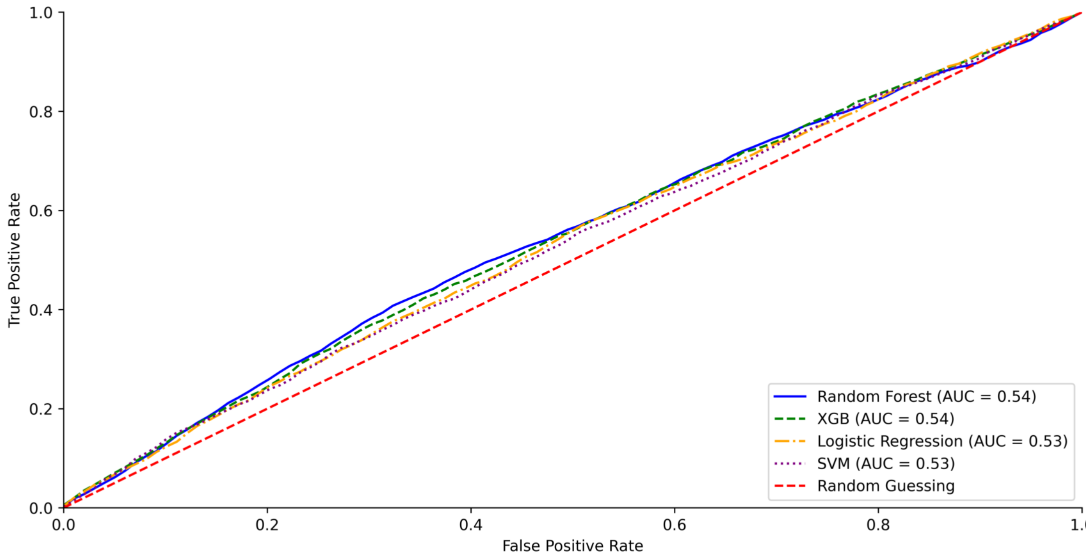

Figure 4.6: Price-raw accounting items

counting items performs better. Although XGB remains the best-performing algorithm ($AUC = 56\%$), it performs 1% less compared to the theoretical model.

4.3.2 Out-of-sample performance of raw accounting data: Point-Biserial approach

In this section the out-of-sample performance of raw accounting items is tested, but for each profitability measure, only the items that were found to have a significant Point-Biserial (PB) correlation with each target are used as features. They are tested both

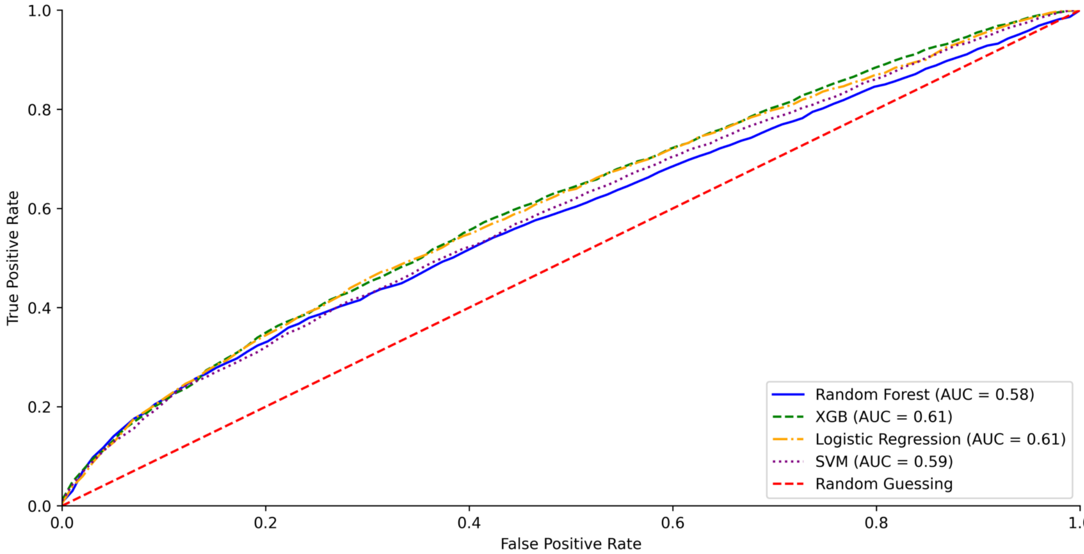

Figure 4.7: EPS-raw accounting items

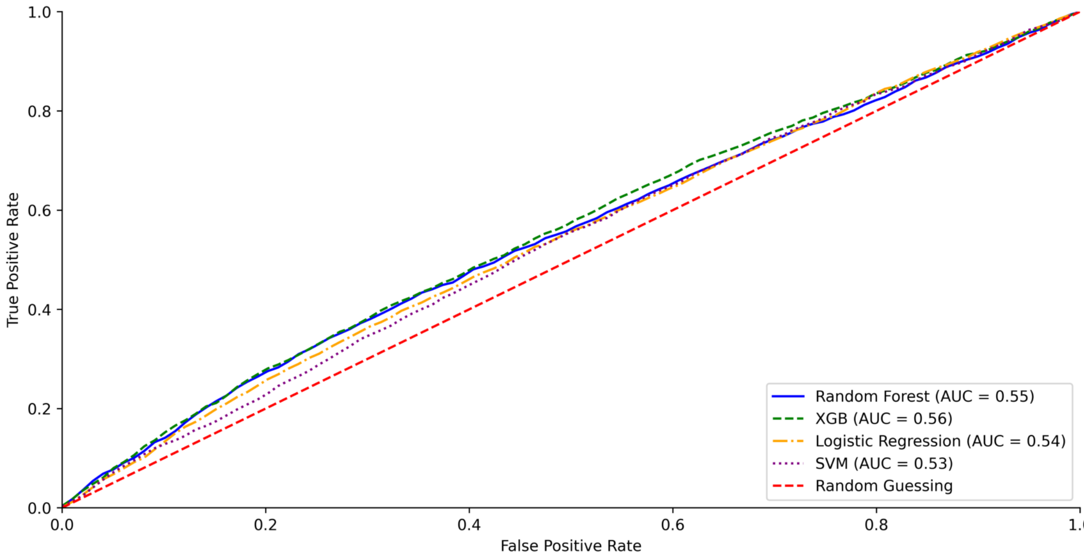

Figure 4.8: Returns-raw accounting items

against the best-performing algorithms that were used in the theoretical models and the kitchen-sink approach.

In the case of ROA, this approach performs better than the theoretical model but it performs marginally worse than the kitchen-sink approach. Similar results are obtained for the Price model. Results are reported in Figures 4.9 and 4.10.

In Figures 4.11 and 4.12 the results for the Returns and EPS models are reported. In the returns model, XGB remains the best algorithm (56%) and there is an 1% increase in the performance of the logistic regression and SVM. The theoretically-specified model

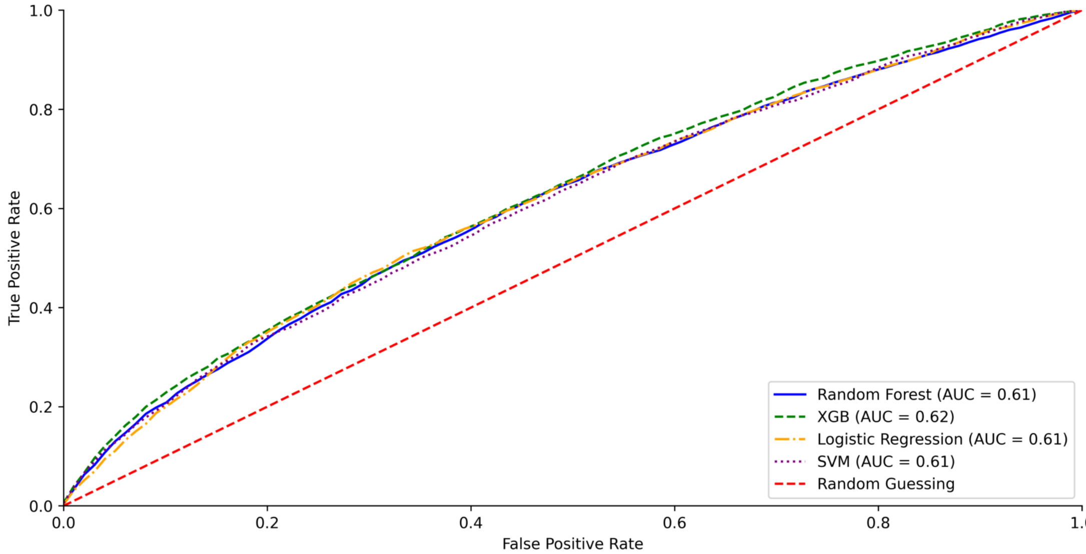

Figure 4.9: ROA- Point-Biserial feature selection

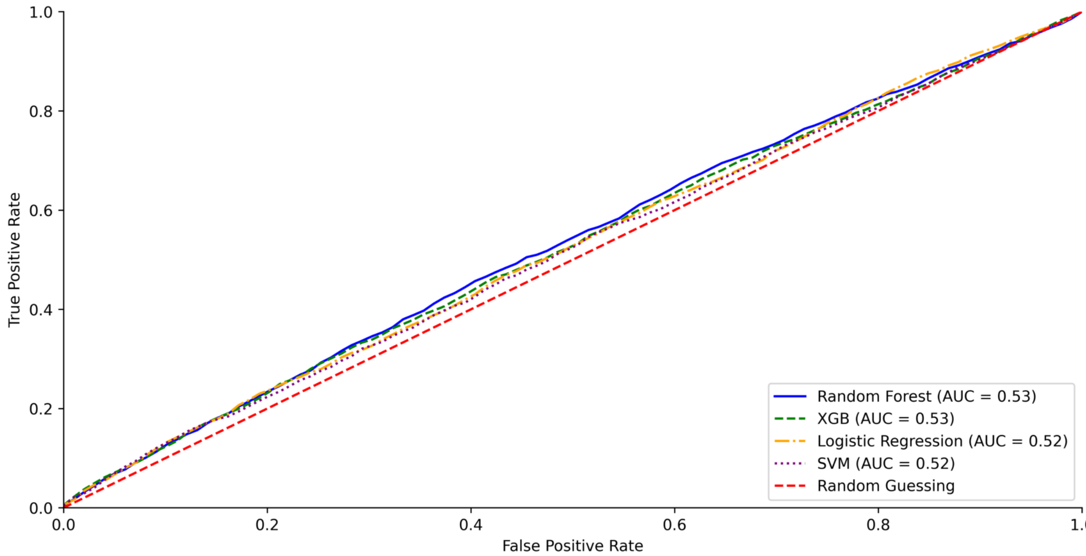

Figure 4.10: Price- Point-Biserial feature selection

remains the best performing model. In the EPS model, XGB performance remains the same, while the logistic algorithm and SVM performance decreased by 1%, yet the performance is still better than the theoretical models.

The findings are in line with the relevant literature (X. Chen et al., 2022), who find that raw accounting items have out-of-sample predictive power. In this study, raw accounting items have better prediction performance compared to the theoretically specified models for all profitability measures apart from the Returns model. Other studies that use raw accounting items, often just test the predictive performance of these items,

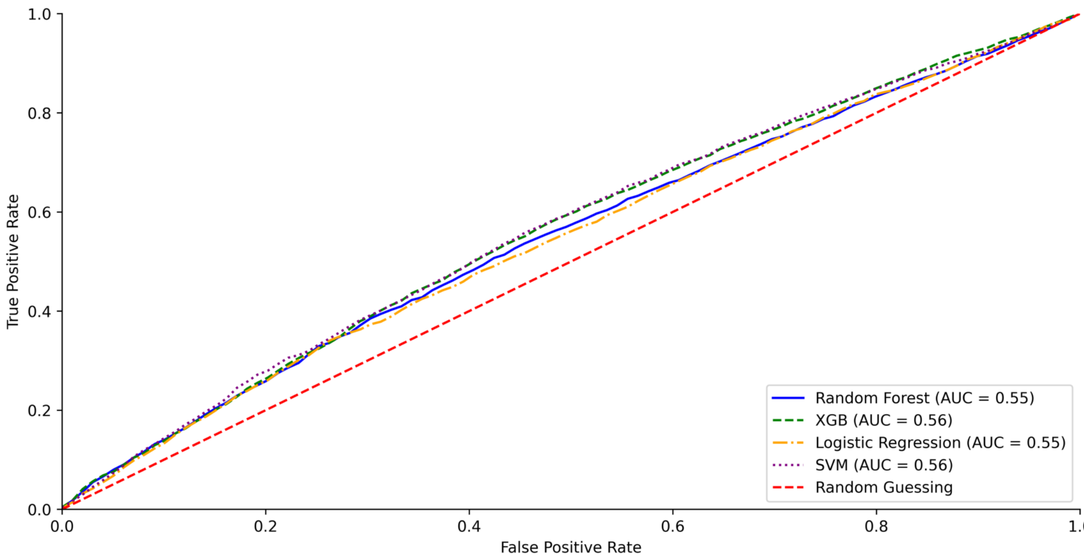

Figure 4.11: Returns- Point-Biserial feature selection

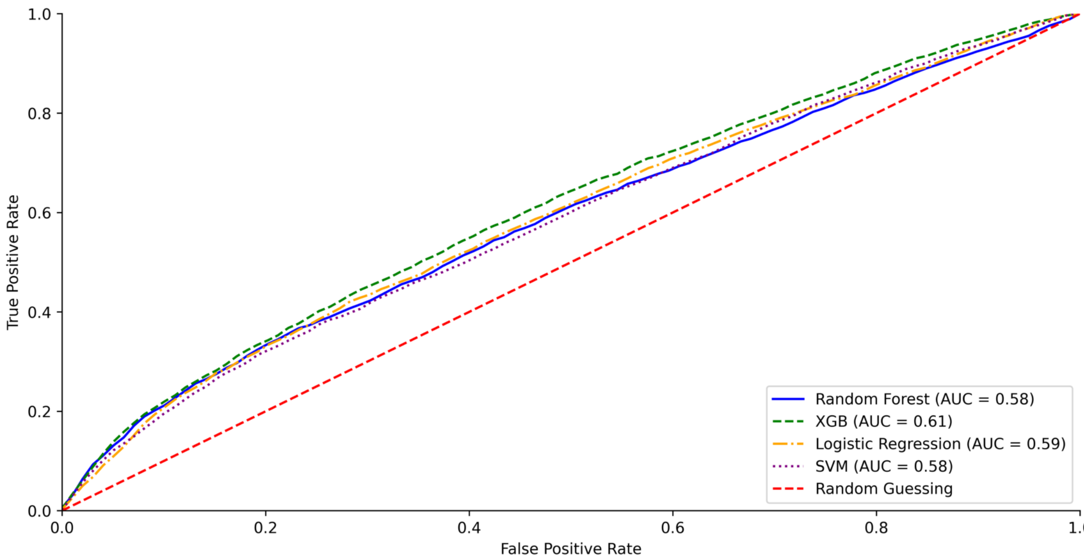

Figure 4.12: EPS- Point-Biserial feature selection

yet they do not discuss or justify in extent why these raw data perform better than the financial ratios.

One of the first studies, if not the first, was conducted by Ball and Brown (1968), who they examined net income against EPS, where they found that the information contained in the income is actually useful. According to Ball and Brown (1968), traditionally, accounting theorists have positioned the effectiveness of accounting practices against the benchmarks of ideal models depicting preferred practices or standards. These models can be as basic as sets of frameworks or as comprehensive as theories and are useful in the evaluation of other practices. However, this approach has a serious limitation: it ignores whether the model actually models the behavior that is encountered in reality. Ample argument has been made as to the inaccuracy of certain aspects of a model however some underlying aspects are always assumed to be included. These models should be employed with caution: although they provide some theoretical knowledge, they may disregard the fine details and subtleties of the actual behaviors which are otherwise important in practice.

Similarly, Lev and Thiagarajan (1993) found that a set of raw financial variables were useful in security valuation. They examined and validated the hypothesis that investors use fundamental variables (instead of ratios) in order to assess earnings growth, and that explains the value relevance of these variables.

Another interest result is the performance of the logistic regression, which is really close to the performance of XGB and random forest. Kirasich et al. (2018) suggested that when there is increased variance in the explanatory and noise variables, logistic regression actually performs better than the random forest. Although random forest is often expected to perform better than logistic regression, this is not true.

4.4 Expensed R&D vs capitalized R&D

It is typical for similar studies in the field to quantify each predictor's importance to the predictive power. This is done by examining feature importance (Bao et al., 2020; Bertomeu et al., 2021; X. Chen et al., 2022; Jones et al., 2023). X. Chen et al. (2022) cautions though that feature importance should not be interpreted as causal inference. Feature importance quantifies how much a variable helps the algorithm to distinguish between the classification outcomes. Jones et al. (2023) pointed out that researchers are more familiar with the statistical significance of the estimated parameters. Variable importance in machine learning informs us how predictive a variable is compared to the other variables in the model. Therefore, the question of whether R&D expenses or capitalized R&D have more predictive power for firm future performance can be answered.

4.4.1 Logistic regression: coefficient magnitude

The logistic regression gives the advantage of not only getting the feature importance but also the coefficient's sign. A positive coefficient is an indication that an increase in the predictor variable increases the possibility of the positive class (increase in profitability), while a negative coefficient indicates the opposite.

For each model that has been estimated, feature importance has been plotted. In Figure 4.13 the coefficients magnitudes for the income model (ROA) are illustrated. Features that concern capitalization exhibit positive coefficients. More specifically, RD_TR, which is the indicator variable of whether the firm is classified as a capitalizer or an expenser has a positive coefficient but small importance. $CFRDCAPCAP$, which is the cash-flow from capitalized R&D (for capitalizers) exhibits an importance of 25%. On the other hand, cash-flow from expensed R&D (expensers), and the expensed R&D of the capitalizers exhibit a negative coefficient. It is significant that the R&D cash-flows of the expensers is the most important feature-among the R&D features.

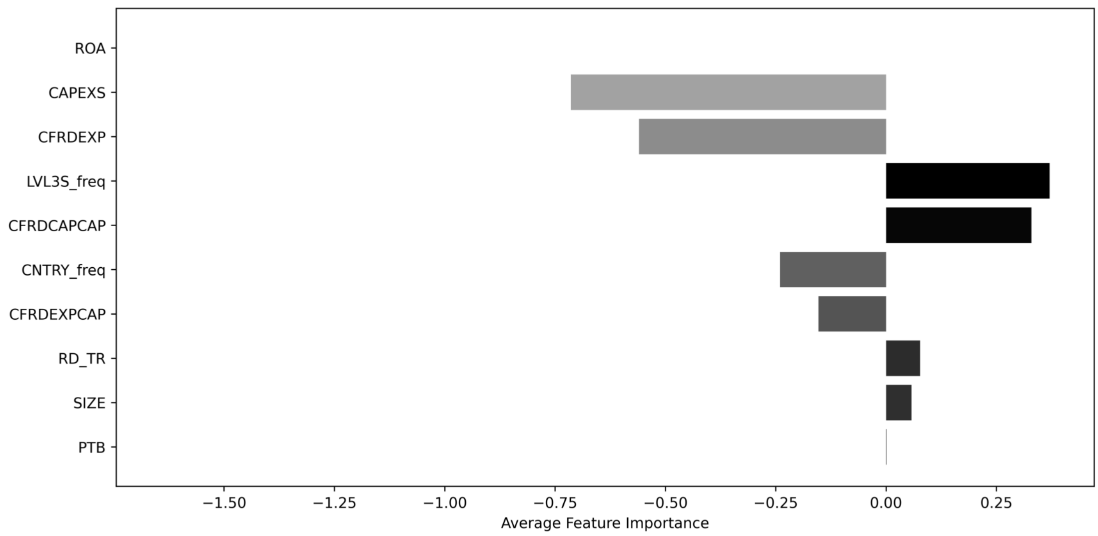

Figure 4.13: ROA-coefficients magnitude

For the Price model, feature importance is illustrated in Figure 4.14. In this model, the most important variable is the capitalized R&D, which also exhibits a positive coefficient. Interestingly, all the expense related R&D features exhibit a positive coefficient, but they have minimal importance.

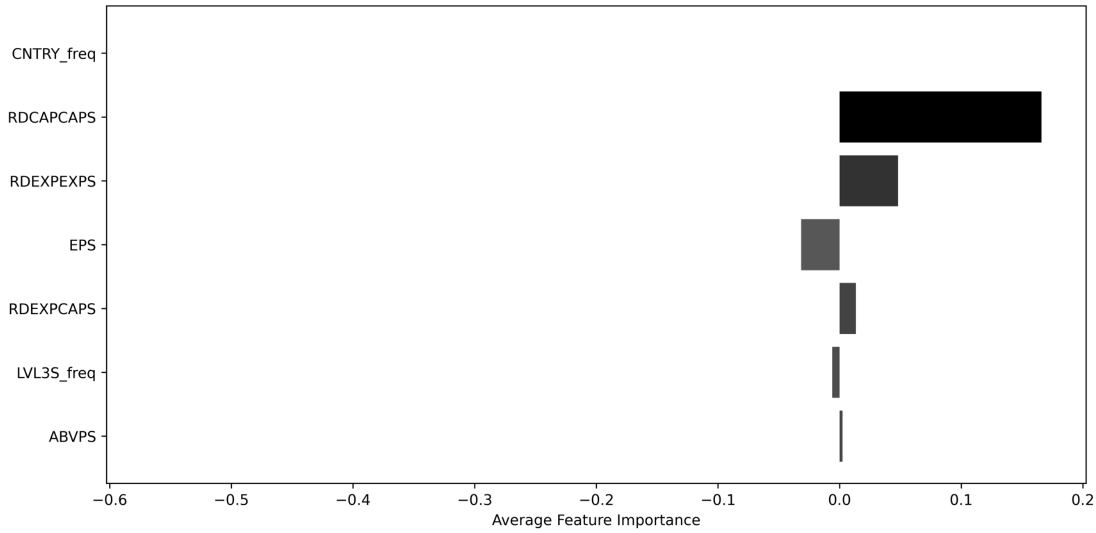

Figure 4.14: Price-coefficients magnitude

For the Returns model, only the expensed R&D exhibit a very strong and positive importance. Results are illustrated in Figure 4.15. Interestingly, all other features exhibit minimal importance compared to the expensed R&D.

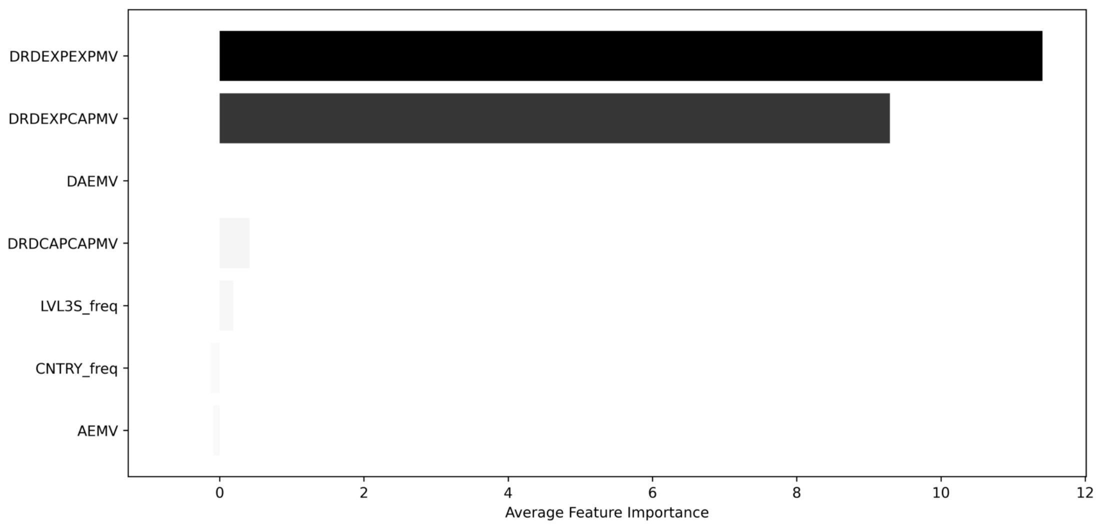

Figure 4.15: Returns-coefficients magnitude

Finally, for the EPS model, again all the R&D variables have a positive importance. However, the most important variable is the change in capitalized R&D for the capitalizers. The R&D expenses of the expensers are more important than the R&D expenses of the capitalizers. Those are illustrated in Figure 4.16.

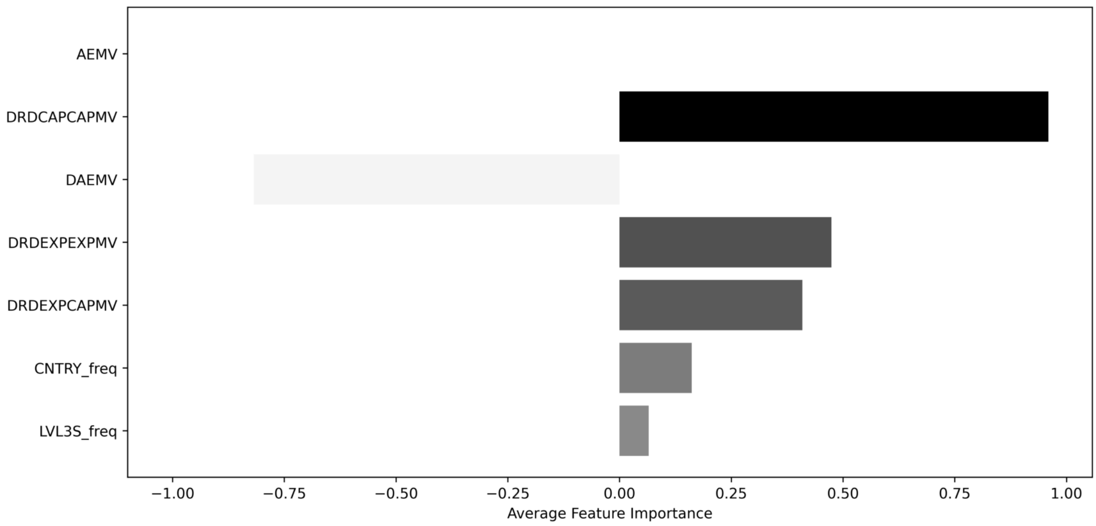

Figure 4.16: EPS-coefficients magnitude

Finally, in Figure 4.17, the coefficients magnitude for the raw accounting items are illustrated. In the ROA model, only R&D expenses were used in the model. They exhibit a negative coefficient and the second higher importance among the features that have been used. In the other models, were only amortized R&D have been used, they exhibit a positive coefficient in all cases apart from the EPS model.

4.4.2 Feature importance: XGB

The analysis is repeated by examining the feature importance of XGB. Although XGB does not show the direction of the importance, in most of the models it has performed, even marginally, better than the logistic regression.

In Figure 4.18 the feature importance for the ROA model is illustrated. Again, the R&D expenses have the greatest importance among the R&D related variables. However, the capitalized R&D have nearly equally importance with the expensed R&D when XGB is used.

In the Price model, feature importance is nearly the same as in the logistic regression. Capitalized R&D and expensed R&D of the capitalizers are more important than the expensed R&D of the expensers. Feature importance for the Price model is illustrated in Figure 4.19

For the Returns model, feature importance is illustrated in Figure 4.20. In this

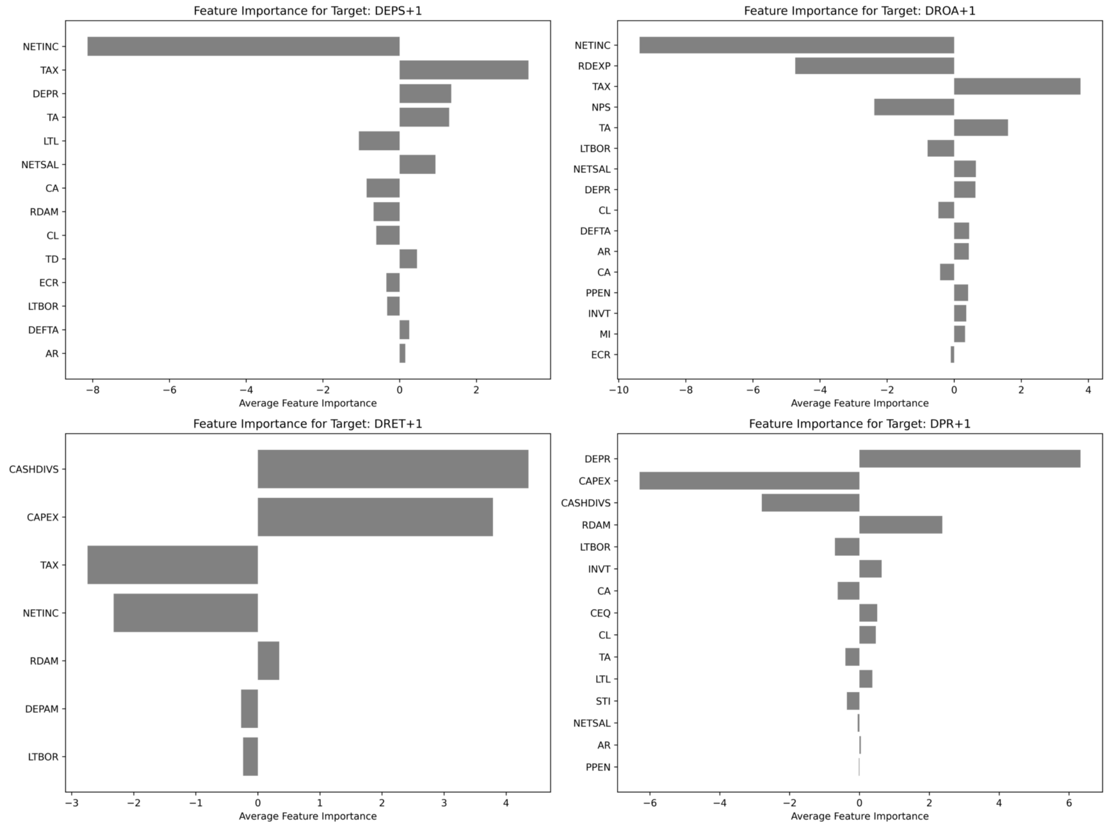

Figure 4.17: Raw accounting items-coefficients magnitude

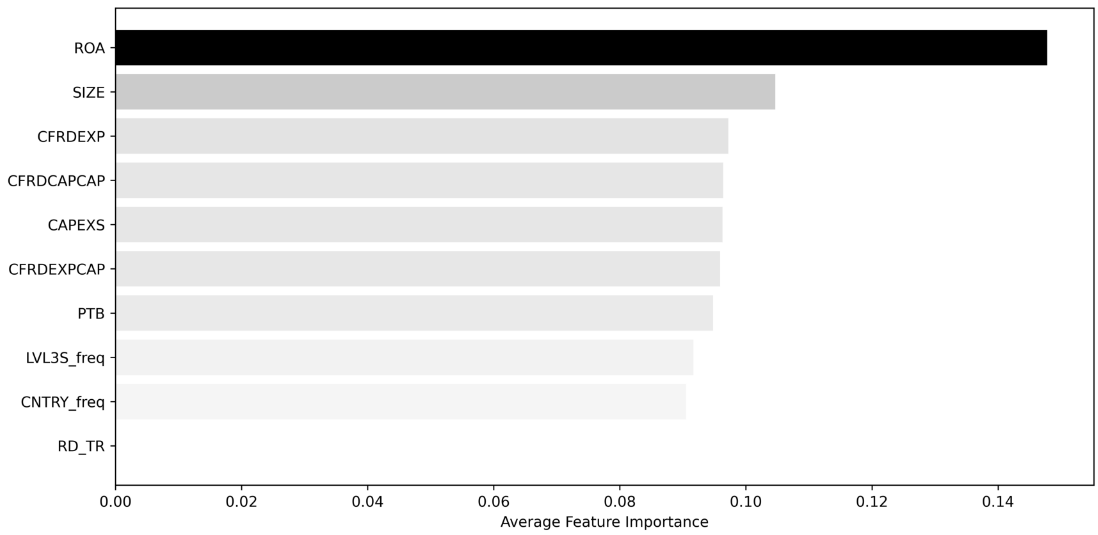

Figure 4.18: ROA-feature importance

model, the expensed R&D of the capitalizers are more important than those expensed by the expensers. On the contrary, when the logistic regression was used, the R&D expenses of the expensers were the most important.

It is interesting that in the case of EPS model, all three R&D variables exhibit nearly identical feature importance. On the contrary, when the logistic regression was used,

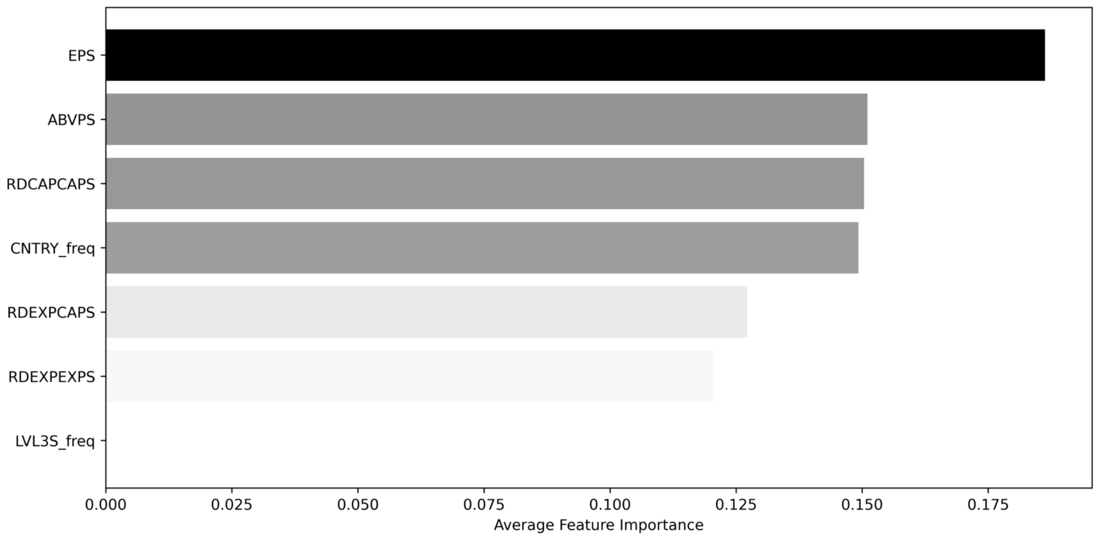

Figure 4.19: Price-feature importance

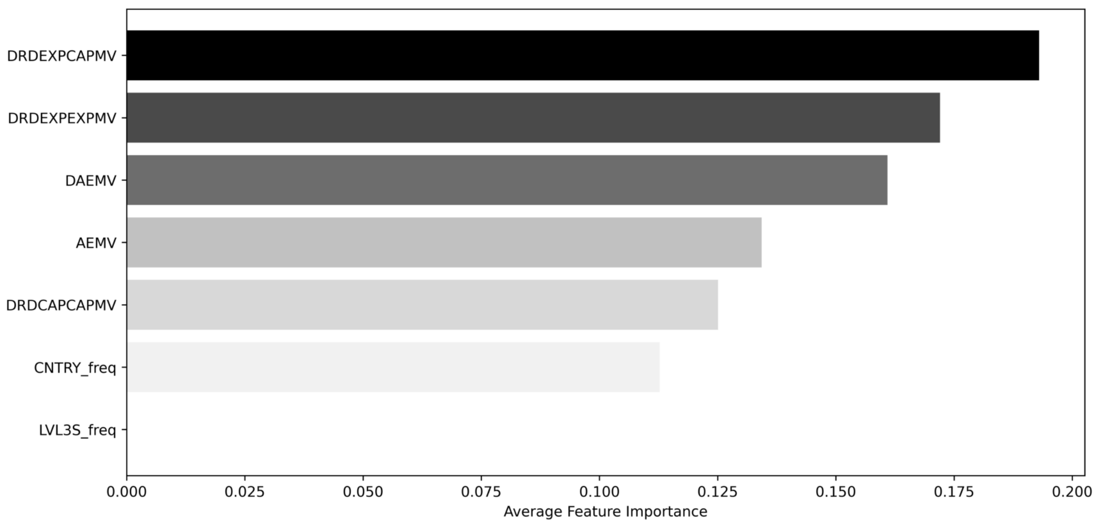

Figure 4.20: Returns-feature importance

capitalized R&D outlays were the most important feature. Feature importance for the EPS model is illustrated in Figure 4.21.

Finally, feature importance for the raw accounting items is illustrated in Figure 4.22. It is noticed that the R&D related items have relatively lower importance compared to the other features, which was not the case when the logistic regression has been used.

Although as explained before, feature importance cannot be used as a measure of causality, features that are important to the prediction and the probability that an increase in the feature's value can increase the probability of positive class (increase in profitability) are very useful aspects for the forecasting task. In the ROA model,

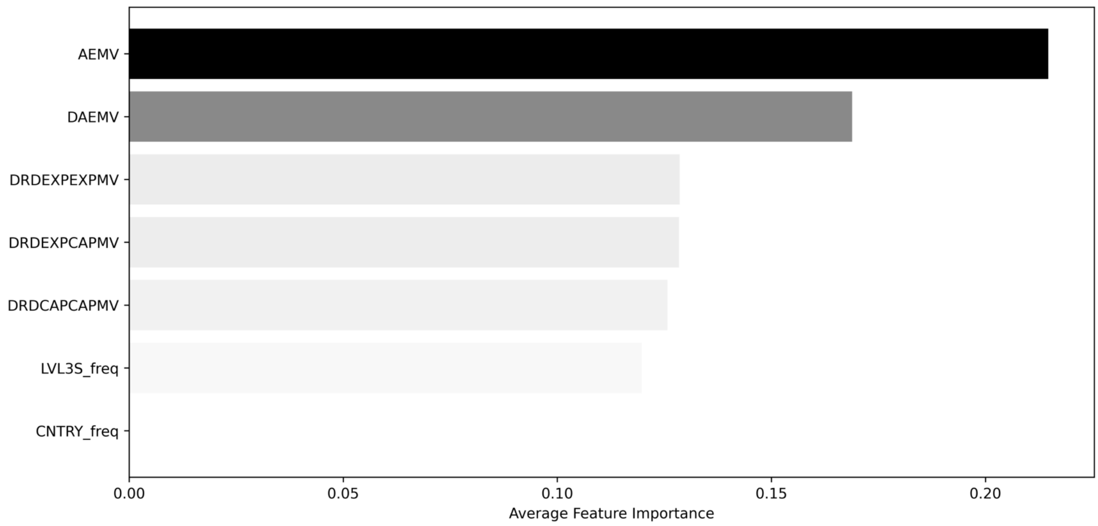

Figure 4.21: EPS-feature importance

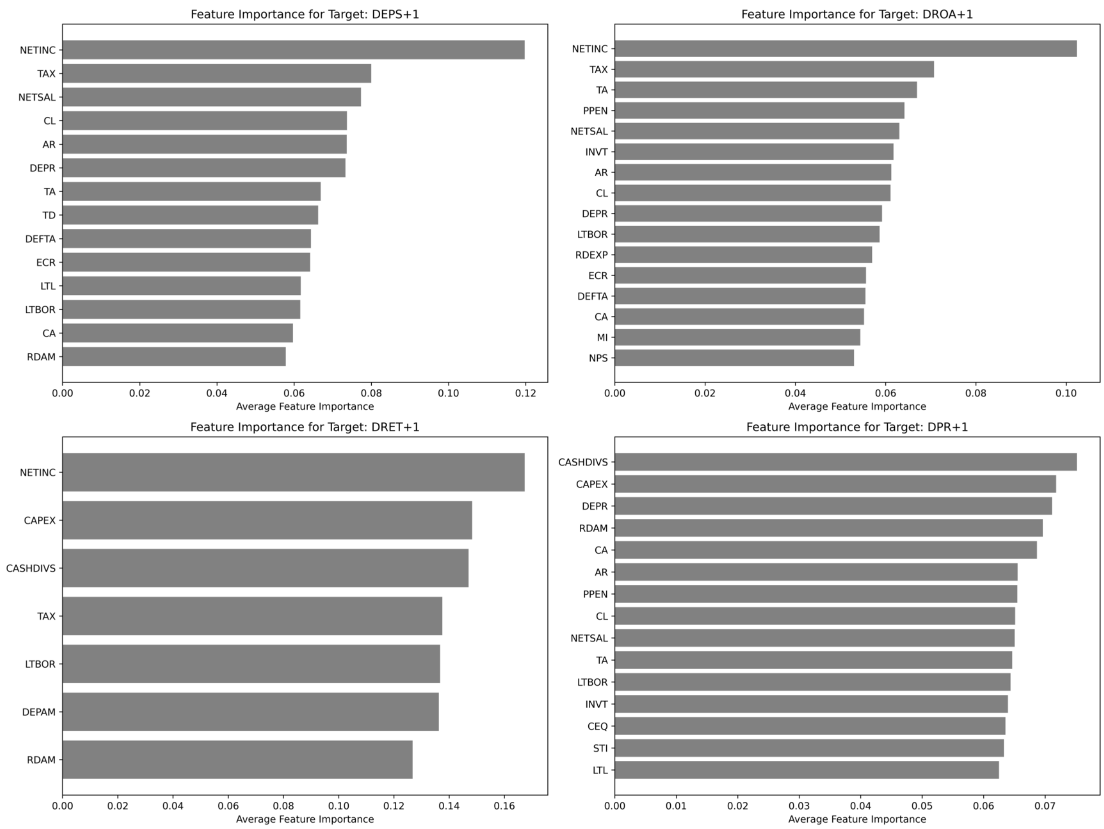

Figure 4.22: Raw accounting items-feature importance

expensed R&D have greater feature importance and a negative coefficient, while the capitalized R&D have a positive coefficient, which is what expected according to the supporters of the R&D capitalization. Capitalized R&D are important in predicting an increase in ROA.

In the Price model, capitalized R&D are very important and exhibit a positive coefficient. As capitalized R&D affect the balance sheet directly, they cause an increase in the assets. This seems to be valued by the investors and there is a positive effect on price. On the contrary, in the Returns model, expensed R&D seem more important compared to capitalized R&D. These expenses represent the company's active investment in innovation and future growth, albeit at the cost of reduced current profitability. However, the positive coefficient reveals that the market may considers that those R&D expenses will lead to successful R&D projects in the future and firms may experience high returns. At the same time, it seems that the capitalized R&D costs are maybe already incorporated in the stock price and that is the reason why they exhibit smaller feature importance.

4.4.3 SHAP- SHapley Additive exPlanations

SHAP values show the contribution of each feature on the prediction of the model. SHAP values explain how a feature contributed to the prediction by comparing to the overall outcome of the prediction. SHAP values were presented by Lundberg and Lee (2017) and are inspired by game theory. Lundberg et al. (2019) suggested that SHAP values are more consistent and reliable compared to feature importance. SHAP values are used as an alternative to feature importance in accounting and finance studies that use machine learning (Bali et al., 2023; Futagami et al., 2021).

In Figure 4.23 the SHAP values for the income model are illustrated. Among all R&D features, the outlays of the capitalizers are the more important, with the expensed R&D of the capitalizers to be the most important one. It is noticed that the lowest values of the feature have almost zero effect in the prediction of profitability increase. The highest values of the feature indicate that they do not consistently lead to either increase or decrease in profitability (they are spread to both negative and positive SHAP values). On the other hand, the highest values of the capitalized R&D lead to the prediction of increase in profitability, while the lowest values have zero effect or lead to the prediction of decrease in profitability. The R&D expenses of the expensers, as they increase (highest values) lead to the prediction of decrease in profitability (the opposite for the lowest values). Their importance though is small compared to the other features.

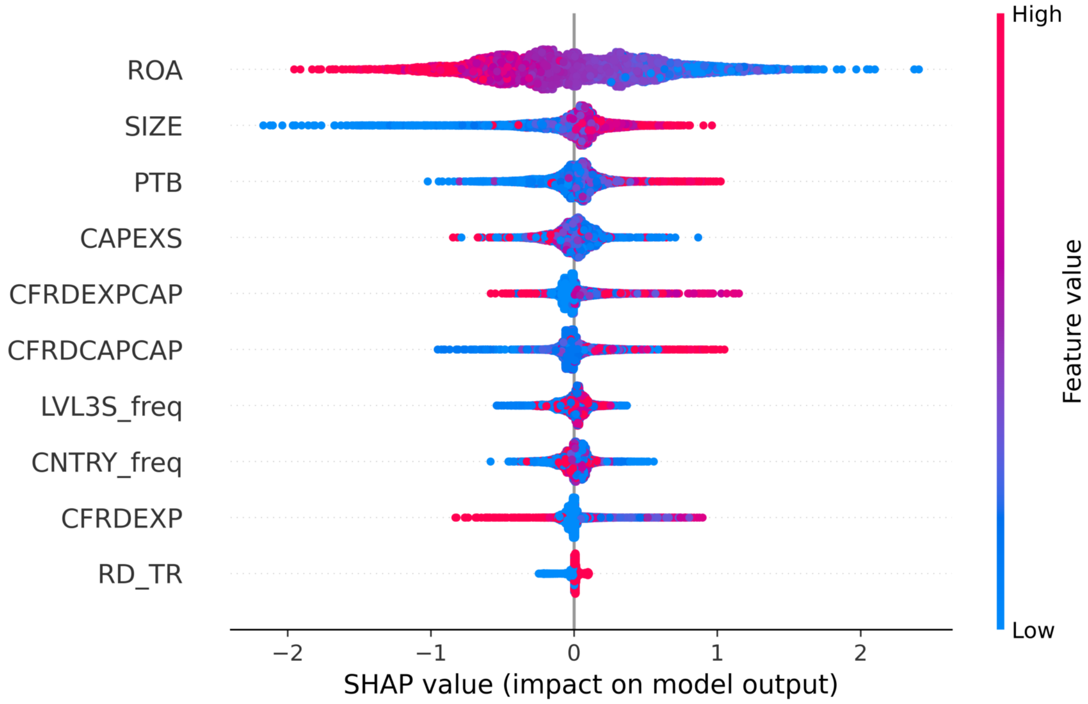

Figure 4.23: ROA-SHAP values

For the Price model, SHAP values are plotted in Figure 4.24. It is noticed that capitalized R&D is the second most important feature in the model. Both low and high values of the capitalized R&D may lead to predict either increase or decrease in profitability. Similar behavior is noticed for the expensed R&D of the capitalizers. On the other hand, expensed R&D of the expensers, when the feature takes high values leads to the prediction of increase of profitability. It has to be noticed though, that both R&D expenses features are less important compared to capitalized R&D.

In the Returns model, expensed R&D for both capitalizers and expensers are the most important R&D features. High values for the expensed R&D of the capitalizers predict an increase in profitability, while lowest values can either predict increase or decrease in profitability. High feature values of capitalized R&D seem to predict decrease in profitability, while low values predict both increase and decrease, but mostly decrease in profitability. SHAP values for the returns model are illustrated in Figure 4.25.

In Figure 4.26, the SHAP values for the EPS model are reported. All three R&D

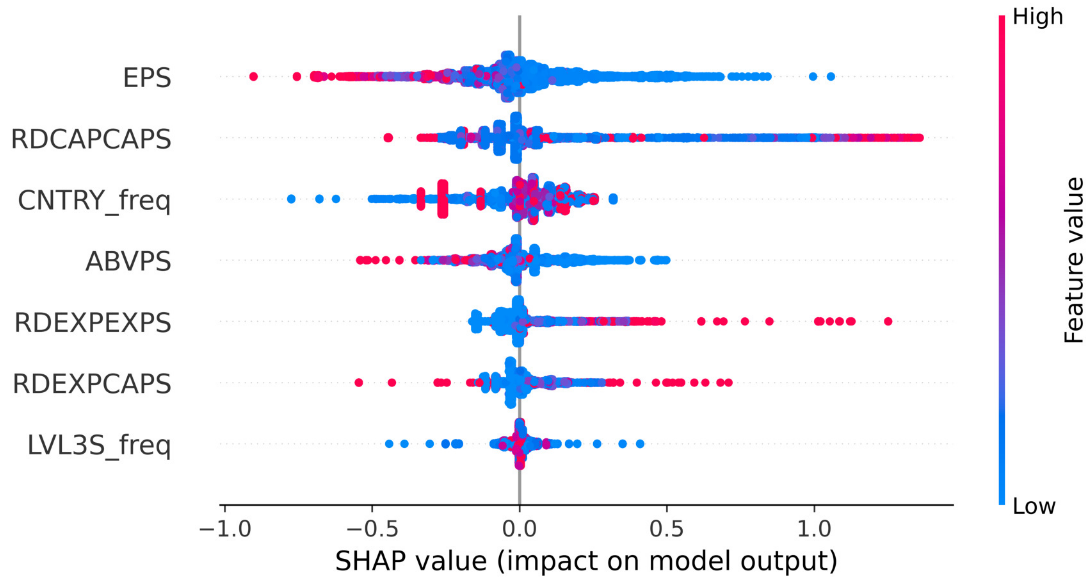

Figure 4.24: Price-SHAP values

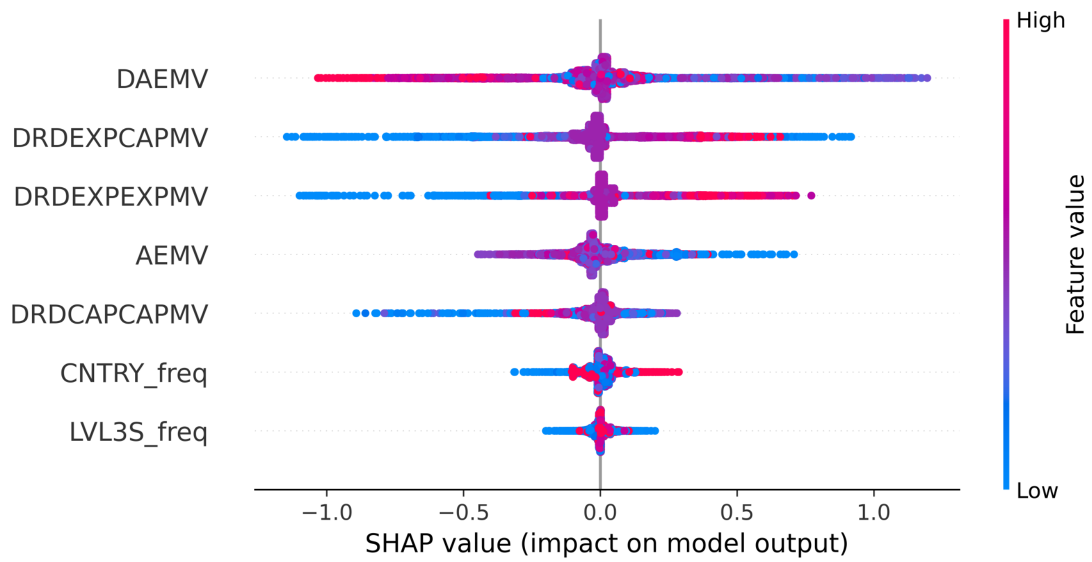

Figure 4.25: Returns-SHAP values

features exhibit the lowest importance among all features. Capitalized R&D is the most important R&D feature. Highest values for the capitalized R&D predict an increase in profitability. The same stands for the rest of the R&D features.

4.5 Supplementary analyses: Unadjusted variables

Aboody and Lev (1998) mentioned that according to empirical evidence, analysts raised objection to capitalization, as they supported that when firms capitalize, earnings fore-

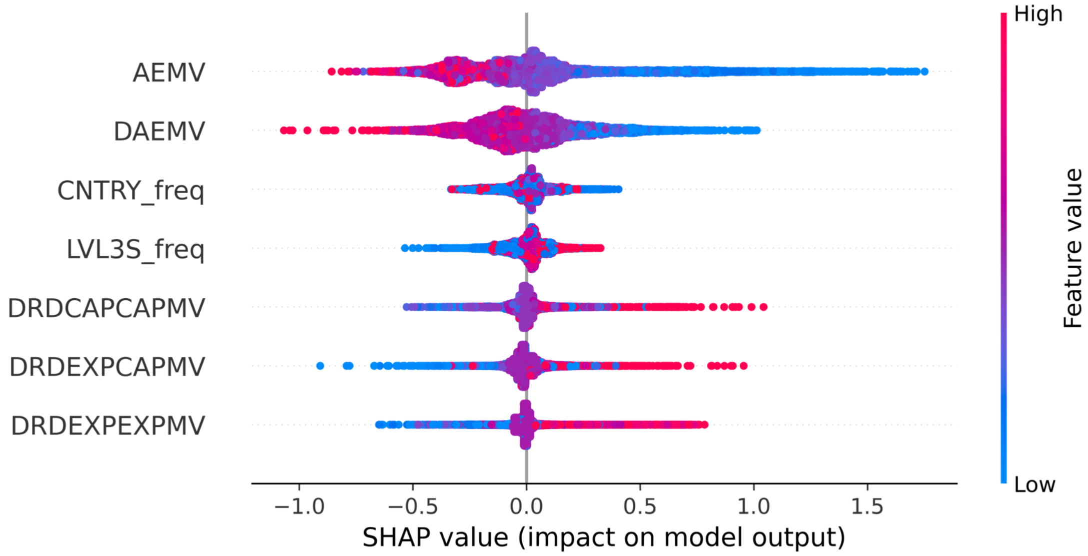

Figure 4.26: EPS-SHAP values

cast errors are positively related to capitalization. Aboody and Lev (1998) objected that it is very easy to reverse capitalization. Since then, studies have adjusted earnings and total assets. In this approach, variables are used unadjusted, without adjustments for the R&D reporting. By comparing the prediction performance of adjusted versus unadjusted models, it can be noticed whether indeed capitalization deteriorates prediction performance. The models are re-estimated using the XGB algorithm, as in general, is the best-performing algorithm.

Out of sample performance is illustrated in Figure 4.27. It cannot be supported that capitalization made prediction performance worse. On the contrary, using unadjusted variables lead to a slight improvement in the forecasts for ROA, Price and EPS. This finding is in line with Aboody and Lev (1998) and in contrast to the opponents of capitalization who support that capitalization increases forecast errors.

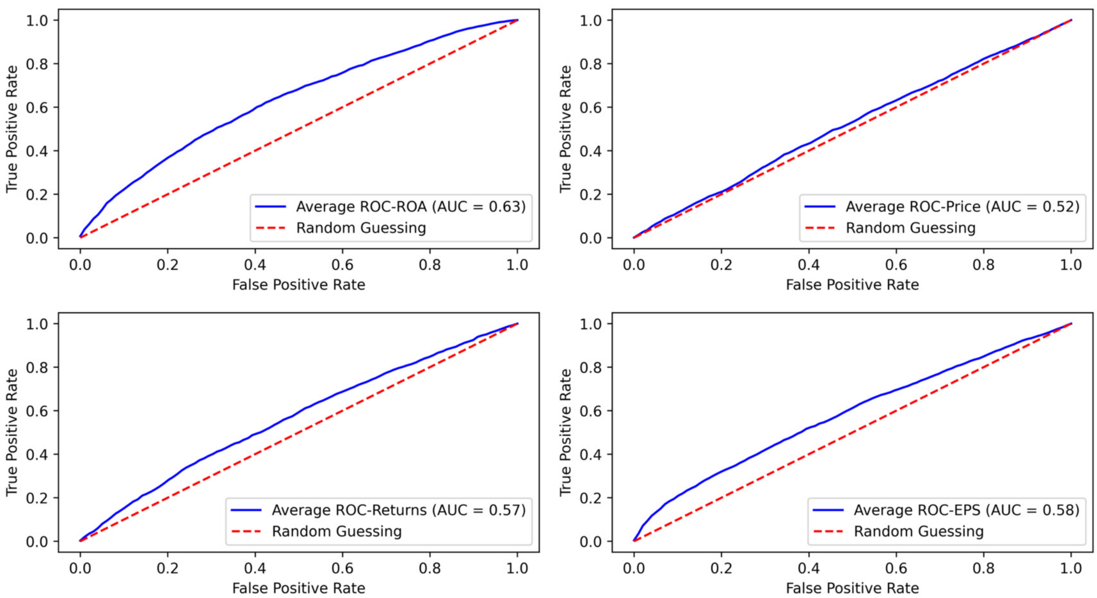

Figure 4.27: Unadjusted models

Chapter 5

Conclusions

The objective of this Ph.D thesis is to shed light on the ongoing debate of the accounting treatment of R&D costs and firm's future financial performance. Relevant literature focuses on explaining if and how R&D accounting treatment affects future performance. On the contrary, in this thesis, out-of-sample predictions of profitability directional changes are obtained; thus, it is examined if indeed expensed R&D costs and capitalized R&D costs have predictive power and which one has the most.

According to the literature, the directional changes of profitability are easier to forecast compared to future profits (Lev & Gu, 2016). Therefore this approach is followed. Further, instead of using only a small set of financial predictors, raw accounting data from the financial statements are used too; similar studies mentioned that theoretically selected financial ratios may not be able to model future earnings adequately (Bao et al., 2020; X. Chen et al., 2022). The small, but growing literature of predicting directional changes of profitability with machine learning was followed. ML algorithms can be used with many predictors, they are designed for forecasting tasks and allow for complex associations between the predictors (features) and the target variable.

In the first step of the analysis, the theoretically specified models were used to obtain out-of-sample predictions of directional changes of profitability. The simplest algorithm, the traditional logistic regression has been used as a benchmark algorithm. Three more complex algorithms, random forest, XGB and SVM have been compared to the logistic regression. Four measures of profitability have been used as targets, ROA, Price, Returns and EPS. Random forest and XGB are in general the best performing algorithms. For the Price model, logistic regression performed better. It has to be noticed though, that logistic regression's performance was very close to the performance of the more complex algorithms. So, can more complex machine learning algorithms enhance prediction performance? Marginally, yes. However, considering the impressive performance of the logistic regression, which is a very simple and fast algorithm to implement, and the most computer-intensive algorithms (they require significantly more time to tune as they have many parameters) that have been used, the trade-off between accuracy and speed has to be taken under consideration. Datasets tend to grow in size, as more and more data from various sources are becoming available, more complex algorithms may need hours or days to be tuned; therefore, the application of these algorithms in large datasets, in practice is impossible (at least for when personal computers are used).

In the second step of the analysis, various sets of raw accounting items have been used as features. The results are in line with previous studies that have used raw accounting items as features Bao et al. (2020) and X. Chen et al. (2022). More specifically, it was found that raw accounting items have predictive power. The best out-of-sample performance was obtained by using the "kitchen-sink" approach, meaning using all the available accounting items. When a selection was made in order to reduce the raw items, the set of features that have been selected performed better than the theoretical models, yet they performed slightly worse than the kitchen-sink approach.

In the third step of the analysis, the main question of this research is answered. Coefficient magnitude from the logistic regression, feature importance and SHAP values from the XGB algorithm are used to examine the predictive power of the R&D variables. As SHAP values provide more insights and are more consistent than feature importance (Lundberg et al., 2019), they are used to interpret the results. For the directional change of ROA, it was found that the most important R&D variable was the expensed R&D of the capitalizers, closely followed by the capitalized R&D. On the other hand, expensed R&D of the expensers and the indicator variable used to distinguish capitalizers from expensers were the least important features. While high values of capitalizers' expensed R&D do not consistently lead to the prediction of either increase or decrease in ROA, the highest values of capitalized R&D lead to the prediction of increase in ROA. This is in line with the supporters of R&D capitalization and the theory that R&D capitalization is used as a signal of improved future performance. It has to be highlighted that only the R&D expenses of the capitalizers are important, and the R&D expenses of the expensers are one of the least important features. This indicates that for a firm which is classified as a capitalizer, its R&D outlays are important in predicting future ROA directional change, while if it is classified as an expenser, its R&D expenses are of little importance.

For the directional change of price, it was found that capitalized R&D is the most important R&D feature. The results indicate that either low or high values of capitalized R&D predict both increase and decrease of the price directional changes. For the returns model, expensed R&D of the capitalizers and expensed R&D of the expensers are the two most important R&D features. High values for the expensed R&D of the capitalizers predict an increase in profitability, while lowest values can either predict increase or decrease in profitability. The same stands for the expensed R&D of the expensers. As for the directional changes of EPS, findings indicate that capitalized R&D are the most important R&D feature. In general, high values of R&D outlays predict increase in EPS.

It is noticed that for accrual-based measures of profitability, ROA and EPS, capitalized R&D is the most important feature and at the same time, high values of capitalized R&D predict profitability increase. For the market-based measures of profitability, price and returns, results are mixed. For the price model, capitalized R&D are the most important, but they predict both increase and decrease in profitability. On the other hand, in the returns model, expensed R&D are the most important and high values of expensed R&D predict both increase and decrease in profitability. Literature suggests that when features exhibit SHAP values ranging in both positive and negative directions, that means that feature's contribution depends on its interaction with other features (Y.-G. Lee et al., 2023).

Finally, according to the literature (Aboody & Lev, 1998), opponents of R&D capitalization support that when firms capitalize, earnings forecasts errors are positively related to capitalization. Therefore, models in the literature use adjusted variables (earnings and total assets). Unadjusted variables were used to examine whether prediction accuracy is lower. Findings suggest that instead of lower accuracy, predictions for the directional changes of ROA, price and EPS was slightly better and for the returns the prediction accuracy was the same. Therefore, analysts indeed can adjust for capitalization, as supported by Aboody and Lev (1998).

This study is unique in that it adopts machine learning (ML) algorithms to forecast the direction of profitability change using R&D models. In this regard, the focus turns away from the conventional paradigm which entails several limitations such as the use of few financial ratios and linear models. In this case, a wide set of raw accounting data is employed together with complex ML algorithms. In this way, it helps to model intricate, non-linear interactions among variables, therefore enhancing overall analysis.

The study provides empirical evidence on the predictive power of expensed versus capitalized R&D costs. Findings suggest that capitalized R&D costs are significant predictors of future profitability, especially for accrual-based measures like Return on Assets (ROA) and Earnings Per Share (EPS). The findings support the stream of the literature which suggests that capitalized R&D costs can signal improved firm future performance.

In the application of ML algorithms in accounting and finance, by comparing simple and complex ML algorithms, this research examines the balance between algorithm's complexity and its predictive power. It is observed that when advanced ML algorithms are used, there is little or no improvement compared to the logistic regression. This contributes to the literature on the practical application of ML in financial prediction tasks.

5.1 Implications

The findings presented in this thesis have multiple implications to the relevant R&D literature. First, the evidence that is presented is obtained by making out-of-sample predictions in an unseen, holdout dataset. Findings suggest that capitalized R&D costs are value relevant to future firm performance in most of the profitability measures that have been used. This is in line with the stream of literature that supports R&D capitalization (Aboody & Lev, 1998; K. Ahmed & Falk, 2006; Lev & Gu, 2016; Lev & Sougiannis, 1999; Lev et al., 2005; Sougiannis, 1994). On the other hand, the empirical results of this thesis oppose the findings of Cazavan-Jeny and Jeanjean (2006), Cazavan-Jeny et al. (2011), Costello and Wittenberg-Moerman (2011), Healy et al. (2002), and Markarian et al. (2008). Although opposing results were found, this does not mean that the explanations for the negative relationship between R&D costs and profitability that were given by the opponents of capitalization, do not stand. This study's main goal is to examine the predicting power of R&D costs rather than explain the relationship with future profitability.

The findings support IAS 38 definition that R&D costs must be capitalized when the intangible asset will generate probable future economic benefits. This is very useful to the accounting standards boards, like IASB. Criticism about the usefulness of the information reported in the financial statements exists as well as negative opinions about R&D capitalization from the stakeholders (see UK Endorsement Board 2022 research project on intangibles). Out-of-sample evidence supportive of R&D capitalization may be important to the ongoing discussion.

In addition to that, implications exist for the investors and financial analysts. The incorporation of raw accounting items with no prior link to the theory, is an easy way to improve their forecasting models. This approach also contributes towards the accounting and finance literature, where the common approach is to use theoretically specified ratios. In this thesis, variable selection mechanisms and feature engineering techniques are presented and employed, thus a methodological framework is developed on how to use raw data for financial modeling.

Last but not least, this study, is one of the first that applies ML algorithms to examine the predictive power of R&D costs regarding future profitability prediction. There is a small but growing number of studies in the field of accounting and finance which use ML algorithms. The application of ML algorithms in accounting and finance and the expansion of Artificial Intelligence and ML prove that these methods will be very relevant to future researchers.

5.2 Limitations and directions for future research

One of the limitations of this thesis is data availability. As datasets increase in size and a growing number of data become available, future researchers may focus on using a bigger list of raw accounting items. Furthermore, as computing power increases, researchers will be able to use even more complex ML algorithms or Neural Networks in the future. Another limitation is the limited number of firms (compared to all listed firms). This is a common issue though in the R&D accounting field.

Further, future researchers should not only examine data from the financial statements but also incorporate macroeconomic data. Finally, another possible research stream would be to examine private firms which report under local GAAP, which may allow for greater managerial discretion; findings in this case may be entirely different.

Bibliography

Aboody, D., & Lev, B. (1998). The value relevance of intangibles: The case of software capitalization. *Journal of Accounting Research*, *36*, 161–191. http://www.jstor.org/stable/2491312

Ahmed, A. S., Neel, M., & Wang, D. (2013). Does Mandatory Adoption of IFRS Improve Accounting Quality? Preliminary Evidence. *Contemporary Accounting Research*, *30*(4), 1344–1372. https://doi.org/10.1111/j.1911-3846.2012.01193.x

Ahmed, K., & Falk, H. (2006). The value relevance of management's research and development reporting choice: Evidence from australia. *Journal of Accounting and Public Policy*, *25*(3), 231–264. https://doi.org/10.1016/j.jaccpubpol.2006.03.002

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500. http://www.jstor.org/stable/1879431

Alford, A., Jones, J., Leftwich, R., & Zmijewski, M. (1993). The relative informativeness of accounting disclosures in different countries. *Journal of Accounting Research*, *31*, 183–223. http://www.jstor.org/stable/2491170

Al-Horani, A., Pope, P. F., & Stark, A. W. (2003). Research and development activity and expected returns in the united kingdom. *Review of Finance*, *7*(1), 27–46. https://doi.org/10.1023/A:1022504029943

Ali, A., & Hwang, L.-S. (2000). Country-specific factors related to financial reporting and the value relevance of accounting data. *Journal of Accounting Research*, *38*(1), 1–21. http://www.jstor.org/stable/2672920

Amir, E., Lev, B., & Sougiannis, T. (2003). Do financial analysts get intangibles? *European Accounting Review*, *12*(4), 635–659.

Anand, V., Brunner, R., Ikegwu, K., & Sougiannis, T. (2019). Predicting profitability using machine learning. *Available at SSRN 3466478*.

André, P., Filip, A., & Paugam, L. (2015). The effect of mandatory ifrs adoption on conditional conservatism in europe. *Journal of Business Finance & Accounting*, *42*(3-4), 482–514. https://doi.org/10.1111/jbfa.12105

Andrews, D. F., & Pregibon, D. (2018). Finding the Outliers that Matter. *Journal of the Royal Statistical Society: Series B (Methodological)*, *40*(1), 85–93. https://doi.org/10.1111/j.2517-6161.1978.tb01652.x

Artikis, P. G., Belesis, N. D., Papanastasopoulos, G. A., & Vasilatos, A. M. (2024). Benchmarking machine learning algorithms to predict profitability directional changes. In A. Emrouznejad, P. D. Zervopoulos, I. Ozturk, D. Jamali, & J. Rice (Eds.), *Business analytics and decision making in practice* (pp. 85–96). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-61589-4_8

Artikis, P. G., Diamantopoulou, L., Papanastasopoulos, G. A., & Sorros, J. N. (2022). Asset growth and stock returns in european equity markets: Implications of investment and accounting distortions. *Journal of Corporate Finance*, *73*, 102193. https://doi.org/10.1016/j.jcorpfin.2022.102193

Auvinen, T. P., Sajasalo, P., Sintonen, T., Takala, T., & Järvenpää, M. (2018). Antenarratives in ongoing strategic change: Using the story index to capture daunting and optimistic futures. In H. Krämer & M. Wenzel (Eds.), *How Organizations Manage the Future: Theoretical Perspectives and Empirical Insights* (pp. 133–151). Springer International Publishing. https://doi.org/10.1007/978-3-319-74506-0_7

Bali, T. G., Beckmeyer, H., Mörke, M., & Weigert, F. (2023). Option Return Predictability with Machine Learning and Big Data. *The Review of Financial Studies*, *36*(9), 3548–3602. https://doi.org/10.1093/rfs/hhad017

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, *6*(2), 159–178. http://www.jstor.org/stable/2490232

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research*, *58*(1), 199–235. https://doi.org/10.1111/1475-679X.12292

Barker, R., Penman, S., Linsmeier, T. J., & Cooper, S. (2020). Moving the conceptual framework forward: Accounting for uncertainty. *Contemporary Accounting Research*, *37*(1), 322–357. https://doi.org/10.1111/1911-3846.12585

Barth, M. E. (2008). Global Financial Reporting: Implications for U.S. Academics. *The Accounting Review*, *83*(5), 1159–1179. https://doi.org/10.2308/accr.2008.83.5.1159

Belesis, N. D., Papanastasopoulos, G. A., & Vasilatos, A. M. (2023). Predicting the Profitability of Directional Changes Using Machine Learning: Evidence from European Countries. *Journal of Risk and Financial Management*, *16*(12), 520. https://doi.org/10.3390/jrfm16120520

Ben-Zion, U. (1977). The investment aspect of non-production expenditures: An empirical test. *Journal of Economics and Business*, *30*.

Berger, P. G., Ofek, E., & Swary, I. (1996). Investor valuation of the abandonment option. *Journal of Financial Economics*, *42*(2), 259–287. https://doi.org/10.1016/0304-405X(96)00877-X

Bertomeu, J., Cheynel, E., Floyd, E., & Pan, W. (2021). Using Machine Learning to Detect Misstatements. *Review of Accounting Studies*, *26*(2), 468–519. https://doi.org/10.1007/s11142-020-09563-8

Bhattacharya, N., Saito, Y., Venkataraman, R., & Yu, J. J. (2024). R&D Reporting Rule and Firm Efficiency. *Journal of Accounting, Auditing & Finance*, *39*(2), 335–363. https://doi.org/10.1177/0148558X211058091

Blouin, J., Core, J. E., & Guay, W. (2010). Have the tax benefits of debt been overestimated? *Journal of Financial Economics*, *98*(2), 195–213. https://doi.org/10.1016/j.jfineco.2010.04.005

Bonett, D. G. (2020). Point-biserial correlation: Interval estimation, hypothesis testing, meta-analysis, and sample size determination. *British Journal of Mathematical and Statistical Psychology*, *73*(S1), 113–144. https://doi.org/10.1111/bmsp.12189

Booz-Allen, S., & Hamilton, S. (1982). *New products management for the 1980s*. Booz, Allen & Hamilton.

Boussofiane, A., Dyson, R., & Thanassoulis, E. (1991). Applied data envelopment analysis. *European Journal of Operational Research*, *52*(1), 1–15. https://doi.org/10.1016/0377-2217(91)90331-O

Bradley, M., & Roberts, M. R. (2015). The structure and pricing of corporate debt covenants. *The Quarterly Journal of Finance*, *05*(02), 1550001. https://doi.org/10.1142/S2010139215500019

Brasch, A., Eierle, B., & Robin, J. (2022). Capitalising or expensing development costs? – mixed methods evidence on the determinants and motives of the accounting policy in the context of uk private companies. *Accounting in Europe*, *19*(3), 363–396. https://doi.org/10.1080/17449480.2022.2049327

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (1999). *Using adaptive bagging to debias regressions* (tech. rep.). Technical Report 547, Statistics Dept. UCB.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & company.

Bublitz, B., & Ettredge, M. (1989). The Information in Discretionary Outlays: Advertising, Research, and Development. *The Accounting Review*, *64*(1), 108–124. Retrieved November 27, 2024, from http://www.jstor.org/stable/248131

Burgstahler, D. C., & Dichev, I. D. (1997). Earnings, adaptation and equity value. *The Accounting Review*, *72*(2), 187–215. http://www.jstor.org/stable/248552

Bustinza, O. F., Gomes, E., Vendrell-Herrero, F., & Baines, T. (2019). Product–service innovation and performance: The role of collaborative partnerships and R&D intensity. *R&D Management*, *49*(1), 33–45. https://doi.org/10.1111/radm.12269

Cao, J., & Laksmana, I. (2010). The effect of capital market pressures on the association between R&D spending and ceo option compensation. *Review of Quantitative Finance and Accounting*, *34*(2), 273–300. https://doi.org/10.1007/s11156-009-0146-7

Cazavan-Jeny, A., & Jeanjean, T. (2006). The negative impact of R&D capitalization: A value relevance approach. *European Accounting Review*, *15*(1), 37–61. https://doi.org/10.1080/09638180500510384

Cazavan-Jeny, A., Jeanjean, T., & Joos, P. (2011). Accounting choice and future performance: The case of R&D accounting in france. *Journal of Accounting and Public Policy*, *30*(2), 145–165. https://doi.org/10.1016/j.jaccpubpol.2010.09.016

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Detecting management fraud in public companies. *Management Science*, *56*(7), 1146–1160. https://doi.org/10.1287/mnsc.1100.1174

Chalmers, J. M. R., Dann, L. Y., & Harford, J. (2002). Managerial opportunism? evidence from directors' and officers' insurance purchases. *The Journal of Finance*, *57*(2), 609–636. https://doi.org/10.1111/1540-6261.00436

Chambers, D., Jennings, R., & Thompson, R. B. (2002). Excess returns to R&D-intensive firms. *Review of Accounting Studies*, *7*(2), 133–158.

Chan, L. K. C., Karceski, J., & Lakonishok, J. (2003). The level and persistence of growth rates. *The Journal of Finance*, *58*(2), 643–684. https://doi.org/10.1111/1540-6261.00540

Chan, L. K. C., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance*, *56*(6), 2431–2456. https://doi.org/10.1111/0022-1082.00411

Chan, S. H., Martin, J. D., & Kensinger, J. W. (1990). Corporate research and development expenditures and share value. *Journal of Financial Economics*, *26*(2), 255–276. https://doi.org/10.1016/0304-405X(90)90005-K

Chauhan, V. K., Dahiya, K., & Sharma, A. (2019). Problem formulations and solvers in linear SVM: a review. *Artificial Intelligence Review*, *52*(2), 803–855. https://doi.org/10.1007/s10462-018-9614-6

Chen, H., Tang, Q., Jiang, Y., & Lin, Z. (2010). The role of international financial reporting standards in accounting quality: Evidence from the european union. *Journal of International Financial Management & Accounting*, *21*(3), 220–278. https://doi.org/10.1111/j.1467-646X.2010.01041.x

Chen, X., Cho, Y. H. (, Dou, Y., & Lev, B. (2022). Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data. *Journal of Accounting Research*, *60*(2), 467–515. https://doi.org/10.1111/1475-679X.12429

Cheng, S. (2004). R&D Expenditures and CEO Compensation. *The Accounting Review*, *79*(2), 305–328. https://doi.org/10.2308/accr.2004.79.2.305

Cho, J., & Kim, H. (2024). Does a consistently capitalized R&D ratio improve information effects of capitalized development expenditures? *International Review of Financial Analysis*, *93*, 103214. https://doi.org/10.1016/j.irfa.2024.103214

Cohen, W. M., & Klepper, S. (1996). Firm size and the nature of innovation within industries: The case of process and product R&D. *The Review of Economics and Statistics*, *78*(2), 232–243. http://www.jstor.org/stable/2109925

Cohen, W. M., Levin, R. C., & Mowery, D. C. (1987). *Firm size and R&D intensity: A re-examination* (Working Paper No. 2205). National Bureau of Economic Research. https://doi.org/10.3386/w2205

Comanor, W. S. (1965). Research and technical change in the pharmaceutical industry. *The Review of Economics and Statistics*, *47*(2), 182–190. http://www.jstor.org/stable/1924065

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*(1), 39–67. https://doi.org/10.1177/0149206310388419

Coppersmith, D., Hong, S. J., & Hosking, J. R. M. (1999). Partitioning Nominal Attributes in Decision Trees. *Data Mining and Knowledge Discovery*, *3*(2), 197–217. https://doi.org/10.1023/A:1009869804967

Corrado, C. A., & Hulten, C. R. (2010). How Do You Measure a ”Technological Revolution”? *American Economic Review*, *100*(2), 99–104. https://doi.org/10.1257/aer.100.2.99

Costello, A. M., & Wittenberg-Moerman, R. (2011). The impact of financial reporting quality on debt contracting: Evidence from internal control weakness reports. *Journal of Accounting Research*, *49*(1), 97–136. https://doi.org/10.1111/j.1475-679X.2010.00388.x

Curth, A., Jeffares, A., & van der Schaar, M. (2024). Why do Random Forests Work? Understanding Tree Ensembles as Self-Regularizing Adaptive Smoothers. https://arxiv.org/abs/2402.01502

Dargenidou, C., Jackson, R. H., Tsalavoutas, I., & Tsoligkas, F. (2021). Capitalisation of R&D and the informativeness of stock prices: Pre- and post-IFRS evidence. *The British Accounting Review*, *53*(4), 100998. https://doi.org/10.1016/j.bar.2021.100998

Darrough, M., & Ye, J. (2007). Valuation of loss firms in a knowledge-based economy. *Review of Accounting Studies*, *12*, 61–93.

Dechow, P. M. (1994). Accounting earnings and cash flows as measures of firm performance: The role of accounting accruals. *Journal of Accounting and Economics*, *18*(1), 3–42. https://doi.org/10.1016/0165-4101(94)90016-7

Dechow, P. M., & Sloan, R. G. (1991). Executive incentives and the horizon problem: An empirical investigation. *Journal of Accounting and Economics*, *14*(1), 51–89. https://doi.org/10.1016/0167-7187(91)90058-S

Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. *The Journal of Business*, *72*(1), 1–33. http://www.jstor.org/stable/10.1086/209601

Dietterich, T. G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning*, *40*(2), 139–157. https://doi.org/10.1023/A:1007607513941

DiMasi, J. A., Grabowski, H. G., & Vernon, J. (1995). R&D costs, innovative output and firm size in the pharmaceutical industry. *International Journal of the Economics of Business*, *2*(2), 201–219. https://doi.org/10.1080/758519309

Dinh, T., Kang, H., & Schultze, W. (2015). Capitalizing research & development: Signaling or earnings management? *European Accounting Review*, *25*(2), 373–401. https://doi.org/10.1080/09638180.2015.1031149

Dinh, T., & Schultze, W. (2022). Accounting for R&D on the income statement? evidence on non-discretionary vs. discretionary R&D capitalization under ifrs in germany. *Journal of International Accounting, Auditing and Taxation*, *46*, 100446. https://doi.org/10.1016/j.intaccaudtax.2022.100446

Dinh, T., Sidhu, B. K., & Yu, C. (2019). Accounting for intangibles: Can capitalization of R&D improve investment efficiency? *Abacus*, *55*(1), 92–127. https://doi.org/10.1111/abac.12149

Doidge, C., Kahle, K. M., Karolyi, G. A., & Stulz, R. M. (2018). Eclipse of the Public Corporation or Eclipse of the Public Markets? *Journal of Applied Corporate Finance*, *30*(1), 8–16. https://doi.org/10.1111/jacf.12272

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Duru, A., Iyengar, R. J., & Thevaranjan, A. (2002). The shielding of ceo compensation from the effects of strategic expenditures. *Contemporary Accounting Research*, *19*(2), 175–193. https://doi.org/10.1506/UM8Q-NVJ6-JKGT-GH5W

Dye, R. A., & Verrecchia, R. E. (1995). Discretion vs. uniformity: Choices among gaap. *The Accounting Review*, *70*(3), 389–415. http://www.jstor.org/stable/248531

Easton, P. D. (1998). Discussion of revalued financial, tangible, and intangible assets: Association with share prices and non-market-based value estimates. *Journal of Accounting Research*, *36*, 235–247. http://www.jstor.org/stable/2491315

Easton, P. D. (1999). Security Returns and the Value Relevance of Accounting Data. *Accounting Horizons*, *13*(4), 399–412. https://doi.org/10.2308/acch.1999.13.4.399

Easton, P. D., & Harris, T. S. (1991). Earnings As an Explanatory Variable for Returns. *Journal of Accounting Research*, *29*(1), 19–36. http://www.jstor.org/stable/2491026

Easton, P. D., Kapons, M. M., Monahan, S. J., Schütt, H. H., & Weisbrod, E. H. (2024). Forecasting Earnings Using k-Nearest Neighbors. *The Accounting Review*, *99*(3), 115–140. https://doi.org/10.2308/TAR-2021-0478

Eberhart, A. C., Maxwell, W. F., & Siddique, A. R. (2004). An examination of long-term abnormal stock returns and operating performance following R&D increases. *The Journal of Finance*, *59*(2), 623–650.

Eierle, B., & Wencki, S. (2016). The determinants of capitalising development costs in private companies: Evidence from germany. *Journal of Business Economics*, *86*(3), 259–300. https://doi.org/10.1007/s11573-015-0778-0

Elliott, G., & Timmermann, A. (2008). Economic Forecasting. *Journal of Economic Literature*, *46*(1), 3–56. https://doi.org/10.1257/jel.46.1.3

Fairfield, P. M., Ramnat, S., & Yohn, T. L. (2009). Do industry-level analyses improve forecasts of financial performance? *Journal of Accounting Research*, *47*(1), 147–178. https://doi.org/10.1111/j.1475-679X.2008.00313.x

Fama, E. F., & French, K. R. (1992a). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (1992b). The economic fundamentals of size and book-to-market equity. *Unpublished working paper. University of Chicago*.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (1995). Size and book-to-market factors in earnings and returns. *The Journal of Finance*, *50*(1), 131–155. https://doi.org/10.1111/j.1540-6261.1995.tb05169.x

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, *51*(1), 55–84. https://doi.org/10.1111/j.1540-6261.1996.tb05202.x

Fawcett, T. (2006). An introduction to ROC analysis [ROC Analysis in Pattern Recognition]. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Feigh, B. (1994). Short-termism on trial: An empirical approach.

Fields, T. D., Lys, T. Z., & Vincent, L. (2001). Empirical research on accounting choice. *Journal of Accounting and Economics*, *31*(1), 255–307. https://doi.org/10.1016/S0165-4101(01)00028-3

Franzen, L., & Radhakrishnan, S. (2009). The value relevance of R&D across profit and loss firms. *Journal of Accounting and Public Policy*, *28*(1), 16–32. https://doi.org/10.1016/j.jaccpubpol.2008.11.006

Franzen, L. A., Rodgers, K. J., & Simin, T. T. (2007). Measuring distress risk: The effect of R&D intensity. *The Journal of Finance*, *62*(6), 2931–2967. https://doi.org/10.1111/j.1540-6261.2007.01297.x

Freeman, R. N., Ohlson, J. A., & Penman, S. H. (1982). Book Rate-of-Return and Prediction of Earnings Changes: An Empirical Investigation. *Journal of Accounting Research*, *20*(2), 639–653. http://www.jstor.org/stable/2490890

Freund, Y. (1995). Boosting a Weak Learning Algorithm by Majority. *Information and Computation*, *121*(2), 256–285. https://doi.org/10.1006/inco.1995.1136

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics*, *28*(2), 337–407. https://doi.org/10.1214/aos/1016218223

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. http://www.jstor.org/stable/2699986

Fudenberg, D., & Tirole, J. (1995). A theory of income and dividend smoothing based on incumbency rents. *Journal of Political Economy*, *103*(1), 75–93. https://doi.org/10.1086/261976

Futagami, K., Fukazawa, Y., Kapoor, N., & Kito, T. (2021). Pairwise acquisition prediction with SHAP value interpretation. *The Journal of Finance and Data Science*, *7*, 22–44. https://doi.org/10.1016/j.jfds.2021.02.001

Garanina, T., Hussinki, H., & Dumay, J. (2021). Accounting for intangibles and intellectual capital: A literature review from 2000 to 2020. *Accounting & Finance*, *61*(4), 5111–5140. https://doi.org/10.1111/acfi.12751

Gerakos, J., & Gramacy, R. (2013). Regression-based earnings forecasts. *Chicago Booth Research Paper No. 12-26*. https://doi.org/10.2139/ssrn.2112137

Ghosal, I., & Hooker, G. (2020). Boosting Random Forests to Reduce Bias; One-Step Boosted Forest and Its Variance Estimate. *Journal of Computational and Graphical Statistics*, *30*(2), 493–502. https://doi.org/10.1080/10618600.2020.1820345

Ghosh, A., Moon, D., & Tandon, K. (2007). Ceo ownership and discretionary investments. *Journal of Business Finance & Accounting*, *34*(5-6), 819–839. https://doi.org/10.1111/j.1468-5957.2007.02011.x

Goodwin, J., & Ahmed, K. (2006). Longitudinal value relevance of earnings and intangible assets: Evidence from australian firms. *Journal of International Accounting, Auditing and Taxation*, *15*(1), 72–91. https://doi.org/10.1016/j.intaccaudtax.2006.01.005

Graham, B., & Dodd, D. (1934). *Security analysis*. McGraw-Hill.

Green, J. P., Stark, A. W., & Thomas, H. M. (1996). UK evidence on the market valuation of research and development expenditures. *Journal of Business Finance & Accounting*, *23*(2), 191–216. https://doi.org/10.1111/j.1468-5957.1996.tb00906.x

Grether, D. M. (1992). Testing bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization*, *17*(1), 31–57. https://doi.org/10.1016/0167-2681(92)90078-P

Gu, F., Lev, B., & Zhu, C. (2023). All losses are not alike: Real versus accounting-driven reported losses. *Review of Accounting Studies*, *28*(3), 1141–1189.

Hall, B. H. (1993). The stock market's valuation of R&D investment during the 1980's. *The American Economic Review*, *83*(2), 259–264. http://www.jstor.org/stable/2117674

Harrell, J., Frank E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, Cham. https://doi.org/10.1007/978-3-319-19425-7

Hastie, T., Friedman, J., & Tibshirani, R. (2001). *The elements of statistical learning: Data mining, inference, and prediction* (1st ed.). Springer New York. https://doi.org/10.1007/978-0-387-21606-5

Healy, P. M., Myers, S. C., & Howe, C. D. (2002). R&D accounting and the tradeoff between relevance and objectivity. *Journal of Accounting Research*, *40*(3), 677–710. https://doi.org/10.1111/1475-679X.00067

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. John Wiley.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161. http://www.jstor.org/stable/1912352

Hirschey, M., & Weygandt, J. J. (1985). Amortization policy for advertising and research and development expenditures. *Journal of Accounting Research*, *23*(1), 326–335. http://www.jstor.org/stable/2490921

Hogan, J., & Adams, N. M. (2023). On averaging ROC curves. *Transactions on Machine Learning Research*. https://openreview.net/forum?id=FByH3qL87G

Hou, K., van Dijk, M. A., & Zhang, Y. (2012). The implied cost of capital: A new approach. *Journal of Accounting and Economics*, *53*(3), 504–526. https://doi.org/10.1016/j.jacceco.2011.12.001

Hughes, J. S., & Kao, J. L. (1991). Economic implications of alternative disclosure rules for research and development costs. *Contemporary Accounting Research*, *8*(1), 152–169. https://doi.org/10.1111/j.1911-3846.1991.tb00840.x

Hunt, E. B., Marin, J., & Stone, P. J. (1966). Experiments in induction.

Hunt, J. O. S., Myers, J. N., & Myers, L. A. (2022). Improving Earnings Predictions and Abnormal Returns with Machine Learning. *Accounting Horizons*, *36*(1), 131–149. https://doi.org/10.2308/HORIZONS-19-125

Inoue, A., & Kilian, L. (2005). In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use? *Econometric Reviews*, *23*(4), 371–402. https://doi.org/10.1081/ETC-200040785

Japkowicz, N., & Stephen, S. (2002). The Class Imbalance Problem: A Systematic Study. *Intelligent Data Analysis*, *6*(5), 429–449. https://doi.org/10.3233/IDA-2002-6504

Jensen, E. J. (1987). Research expenditures and the discovery of new drugs. *The Journal of Industrial Economics*, *36*(1), 83–95. http://www.jstor.org/stable/2098598

Johnson, O. (1967). A consequential approach to accounting for R&D. *Journal of Accounting Research*, *5*(2), 164–172. http://www.jstor.org/stable/2490251

Jones, S., Moser, W. J., & Wieland, M. M. (2023). Machine learning and the prediction of changes in profitability. *Contemporary Accounting Research*, *40*(4), 2643–2672. https://doi.org/10.1111/1911-3846.12888

Joos, P., & Plesko, G. A. (2005). Valuing Loss Firms. *The Accounting Review*, *80*(3), 847–870. https://doi.org/10.2308/accr.2005.80.3.847

Kantarelis, D. (2010). *Theories of the firm* (2nd) [Print ISBN: 0-907776-34-5, Online ISBN: 0-907776-35-3]. Inderscience Enterprises Ltd. https://www.inderscience.com

Kaplan, R. S., & Norton, D. P. (2004). *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*. Harvard Business School Press.

Kim, Y. S., Kim, Y., & Yi, H.-C. (2021). Vice or virtue? the impact of earnings management on bank loan agreements. *International Review of Economics & Finance*, *73*, 303–324. https://doi.org/10.1016/j.iref.2020.12.028

Kirasich, K., Smith, T., & Sadler, B. (2018). Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets. *SMU Data Science Review*, *1*(3), Article 9.

Klein, B. (1983). Contracting costs and residual claims: The separation of ownership and control. *The Journal of Law and Economics*, *26*(2), 367–374. https://doi.org/10.1086/467040

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction Policy Problems. *American Economic Review*, *105*(5), 491–95. https://doi.org/10.1257/aer.p20151023

Koop, G., Osiewalski, J., & Steel, M. F. J. (1999). The Components of Output Growth: A Stochastic Frontier Analysis. *Oxford Bulletin of Economics and Statistics*, *61*(4), 455–487. https://doi.org/10.1111/1468-0084.00139

Kosaraju, N., Sankepally, S. R., & Mallikharjuna Rao, K. (2023). Categorical Data: Need, Encoding, Selection of Encoding Method and Its Emergence in Machine Learning Models—A Practical Review Study on Heart Disease Prediction Dataset Using Pearson Correlation. In M. Saraswat, C. Chowdhury, C. Kumar Mandal, & A. H. Gandomi (Eds.), *Proceedings of international conference on data science and applications* (pp. 369–382). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-6631-6_26

Kothari, S. P., Laguerre, T. E., & Leone, A. J. (2002). Capitalization versus Expensing: Evidence on the Uncertainty of Future Earnings from Capital Expenditures versus R&D Outlays. *Review of Accounting Studies*, *7*(4), 355–382. https://doi.org/10.1023/A:1020764227390

Kothari, S. (2001). Capital markets research in accounting. *Journal of Accounting and Economics*, *31*(1), 105–231. https://doi.org/10.1016/S0165-4101(01)00030-1

Kothari, S., & Zimmerman, J. L. (1995). Price and return models. *Journal of Accounting and Economics*, *20*(2), 155–192. https://www.sciencedirect.com/science/article/pii/0165410195003994

La Porta, R. (1996). Expectations and the cross-section of stock returns. *The Journal of Finance*, *51*(5), 1715–1742. https://doi.org/10.1111/j.1540-6261.1996.tb05223.x

Lafond, R., & Roychowdhury, S. (2008). Managerial ownership and accounting conservatism. *Journal of Accounting Research*, *46*(1), 101–135. https://doi.org/10.1111/j.1475-679X.2008.00268.x

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, *49*(5), 1541–1578.

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting Deceptive Discussions in Conference Calls. *Journal of Accounting Research*, *50*(2), 495–540. https://doi.org/10.1111/j.1475-679X.2012.00450.x

Lee, C.-Y., Wu, H.-L., & Pao, H.-W. (2014). How does R&D intensity influence firm explorativeness? evidence of R&D active firms in four advanced countries. *Technovation*, *34*(10), 582–593. https://doi.org/10.1016/j.technovation.2014.05.003

Lee, Y.-G., Oh, J.-Y., Kim, D., & Kim, G. (2023). SHAP Value-Based Feature Importance Analysis for Short-Term Load Forecasting. *Journal of Electrical Engineering & Technology*, *18*(1), 579–588. https://doi.org/10.1007/s42835-022-01161-9

Lev, B. (2000). *Intangibles: Management, Measurement, and Reporting*. Brookings Institution Press.

Lev, B. (2019). Ending the Accounting-for-Intangibles Status Quo. *European Accounting Review*, *28*(4), 713–736. https://doi.org/10.1080/09638180.2018.1521614

Lev, B., & Gu, F. (2016). *The end of accounting and the path forward for investors and managers*. John Wiley & Sons Inc. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119270041

Lev, B., Sarath, B., & Sougiannis, T. (2005). R&D reporting biases and their consequences. *Contemporary Accounting Research*, *22*(4), 977–1026. https://doi.org/10.1506/7XMH-QQ74-L6GG-CJRX

Lev, B., & Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of R&D. *Journal of Accounting and Economics*, *21*(1), 107–138. https://doi.org/10.1016/0165-4101(95)00410-6

Lev, B., & Sougiannis, T. (1999). Penetrating the book-to-market black box: The R&D effect. *Journal of Business Finance & Accounting*, *26*(3-4), 419–449. https://doi.org/10.1111/1468-5957.00262

Lev, B., & Thiagarajan, S. R. (1993). Fundamental Information Analysis. *Journal of Accounting Research*, *31*(2), 190–215. http://www.jstor.org/stable/2491270

Li, K. K., & Mohanram, P. (2014). Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies*, *19*(3), 1152–1185. https://doi.org/10.1007/s11142-014-9282-y

Liao, S. S. (1979). The matching concept and cost allocation. *Accounting and Business Research*, *9*(35), 228–236. https://doi.org/10.1080/00014788.1979.9729162

Lin, B.-W., Lee, Y., & Hung, S.-C. (2006). R&D intensity and commercialization orientation effects on financial performance [Special Section - The 2005 La Londe Seminar]. *Journal of Business Research*, *59*(6), 679–685. https://doi.org/10.1016/j.jbusres.2006.01.002

Liu, M. (2021). Assessing the Strengths and Weaknesses of Human Information Processing in Lending Decisions: A Machine Learning Approach. https://doi.org/10.2139/ssrn.3483699

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2019). Consistent Individualized Feature Attribution for Tree Ensembles. https://arxiv.org/abs/1802.03888

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Mairesse, J., & Sassenou, M. (1991). R&D productivity: A survey of econometric studies at the firm level. *National Bureau of Economic Research Working Paper Series*, (w3666).

Markarian, G., Pozza, L., & Prencipe, A. (2008). Capitalization of R&D costs and earnings management: Evidence from italian listed companies. *The International Journal of Accounting*, *43*(3), 246–267. https://doi.org/10.1016/j.intacc.2008.06.002

Mazzi, F., Slack, R., Tsalavoutas, I., & Tsoligkas, F. (2019). Country-level corruption and accounting choice: Research & development capitalization under ifrs [Accounting Control, Governance and Anti-Corruption Initiatives in Public Sector Organisations]. *The British Accounting Review*, *51*(5), 100821. https://doi.org/10.1016/j.bar.2019.02.003

Mikko Ranta, M. Y., & Järvenpää, M. (2023). Machine learning in management accounting research: Literature review and pathways for the future. *European Accounting Review*, *32*(3), 607–636. https://doi.org/10.1080/09638180.2022.2137221

Milburn, A. (1971). An empirical study of the relationship of research and development expenditures to subsequent benefits. *Department of Accounting. University of Illinois*.

Mitnick, B. M. (1973). Fiduciary rationality and public policy: The theory of agency and some consequences. *1973 Annual Meeting of the American Political Science Association, New Orleans, LA. In Proceedings of the American Political Science Association*.

Monahan, S. J. (2018). Financial statement analysis and earnings forecasting. *Foundations and Trends® in Accounting*, *12*(2), 105–215. https://doi.org/10.1561/1400000036

Monsen, R. J., & Downs, A. (1965). A theory of large managerial firms. *Journal of Political Economy*, *73*(3), 221–236. http://www.jstor.org/stable/1829037

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Myers, J. L., Well, A. D., & Lorch Jr, R. F. (2010). *Research Design and Statistical Analysis: Third Edition* (3rd). Routledge. https://doi.org/10.4324/9780203726631

Newman, M. (1968). Equating return from R&D expenditures. *Financial Executive*, *36*(4), 26–33.

O'Grady, K. E., & Medoff, D. R. (1988). Categorical Variables in Multiple Regression: Some Cautions. *Multivariate Behavioral Research*, *23*(2), 243–260. https://doi.org/10.1207/s15327906mbr2302_7

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131. http://www.jstor.org/stable/2490395

Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research*, *11*(2), 661–687. https://doi.org/10.1111/j.1911-3846.1995.tb00461.x

Osma, B. G., & Young, S. (2009). R&d expenditure and earnings targets. *European Accounting Review*, *18*(1), 7–32. https://doi.org/10.1080/09638180802016718

Oswald, D., Simpson, A., & Zarowin, P. (2021). Capitalization vs. expensing and the behavior of R&D expenditures. *Review of Accounting Studies*, *27*(4), 1199–1232. https://doi.org/10.1007/s11142-021-09631-7

Oswald, D. R. (2008). The determinants and value relevance of the choice of accounting for research and development expenditures in the united kingdom. *Journal of Business Finance & Accounting*, *35*(1-2), 1–24. https://doi.org/10.1111/j.1468-5957.2007.02060.x

Oswald, D. R., & Zarowin, P. (2007). Capitalization of R&D and the informativeness of stock prices. *European Accounting Review*, *16*(4), 703–726. https://doi.org/10.1080/09638180701706815

Ou, J. A. (1990). The Information Content of Nonearnings Accounting Numbers As Earnings Predictors. *Journal of Accounting Research*, *28*(1), 144–163. http://www.jstor.org/stable/2491220

Ou, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, *11*(4), 295–329. https://doi.org/10.1016/0165-4101(89)90017-7

Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2022). Regularized Target Encoding Outperforms Traditional Methods in Supervised Machine Learning with High Cardinality Features. *Computational Statistics*, *37*(5), 2671–2692. https://doi.org/10.1007/s00180-022-01207-6

Pass, C., & Lowes, B. (1978). Managerial theories of the firm. *Managerial Finance*, *4*(1), 10–25. https://doi.org/10.1108/eb013410

Paton, W. A., Littleton, A., & Association, A. A. (1940). *An Introduction to Corporate Accounting Standards*. American Accounting Association. https://catalog.hathitrust.org/Record/003162605

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peles, Y. (1970). Amortization of advertising expenditures in the financial statements. *Journal of Accounting Research*, *8*(1), 128–137. http://www.jstor.org/stable/2674719

Penman, S. H. (2009). Accounting for intangible assets: There is also an income statement. *Abacus*, *45*(3), 358–371. https://doi.org/10.1111/j.1467-6281.2009.00293.x

Penman, S. H., & Yehuda, N. (2009). The pricing of earnings and cash flows and an affirmation of accrual accounting. *Review of Accounting Studies*, *14*(4), 453–479. https://doi.org/10.1007/s11142-009-9109-4

Penman, S. H., & Zhang, X. (2004). Modeling sustainable earnings and P/E ratios using financial statement information. *New York, NY: Columbia University*.

Penman, S. H., & Zhang, X.-J. (2002). Accounting conservatism, the quality of earnings, and stock returns. *The Accounting Review*, *77*(2), 237–264.

Phillips, J., Pincus, M., & Rego, S. O. (2003). Earnings Management: New Evidence Based on Deferred Tax Expense. *The Accounting Review*, *78*(2), 491–521. https://doi.org/10.2308/accr.2003.78.2.491

Polikar, R. (2012). Ensemble Learning. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (pp. 1–34). Springer New York. https://doi.org/10.1007/978-1-4419-9326-7_1

Prajwala, T. (2015). A comparative study on decision tree and random forest using r tool. *International journal of advanced research in computer and communication engineering*, *4*(1), 196–199.

Provost, F. J., Fawcett, T., Kohavi, R., et al. (1998). The case against accuracy estimation for comparing induction algorithms. *ICML*, *98*, 445–453.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/BF00116251

Rees, L., Gill, S., & Gore, R. (1996). An investigation of asset write-downs and concurrent abnormal accruals. *Journal of Accounting Research*, *34*, 157–169. http://www.jstor.org/stable/2491431

Riley, J. G. (2001). Silver signals: Twenty-five years of screening and signaling. *Journal of Economic Literature*, *39*(2), 432–478. https://doi.org/10.1257/jel.39.2.432

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5), S71–S102. http://www.jstor.org/stable/2937632

Ronen, J. (2008). *Earnings management*. Springer.

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *The American Economic Review*, *63*(2), 134–139. http://www.jstor.org/stable/1817064

Schaffer, C. (1993). Overfitting avoidance as bias. *Machine Learning*, *10*(2), 153–178. https://doi.org/10.1007/BF00993504

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, *5*(2), 197–227. https://doi.org/10.1007/BF00116037

Schapire, R. E. (2003). The Boosting Approach to Machine Learning: An Overview. In D. D. Denison, M. H. Hansen, C. C. Holmes, B. Mallick, & B. Yu (Eds.), *Nonlinear estimation and classification* (pp. 149–171). Springer New York. https://doi.org/10.1007/978-0-387-21579-2_9

Schiff, M. (1966). Accounting tactics and the theory of the firm. *Journal of Accounting Research*, *4*(1), 62–67. https://doi.org/10.2307/2490141

Serfling, M. A. (2014). Ceo age and the riskiness of corporate policies. *Journal of Corporate Finance*, *25*, 251–273. https://doi.org/10.1016/j.jcorpfin.2013.12.013

Shah, S. Z. A., Liang, S., & Akbar, S. (2013). International financial reporting standards and the value relevance of R&D expenditures: Pre and post IFRS analysis. *International Review of Financial Analysis*, *30*, 158–169. https://doi.org/10.1016/j.irfa.2013.08.001

Shawi, R. E., Bahman, M., & Sakr, S. (2025). To tune or not to tune? an approach for recommending important hyperparameters for classification and clustering algorithms. *Future Generation Computer Systems*, *163*, 107524. https://doi.org/10.1016/j.future.2024.107524

Shefer, D., & Frenkel, A. (2005). R&D, firm size and innovation: An empirical analysis. *Technovation*, *25*(1), 25–32. https://doi.org/10.1016/S0166-4972(03)00152-4

Shin, J., Kim, Y. J., Jung, S., & Kim, C. (2022). Product and service innovation: Comparison between performance and efficiency. *Journal of Innovation & Knowledge*, *7*(3), 100191. https://doi.org/10.1016/j.jik.2022.100191

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review*, *61*(4), 66–83. https://doi.org/10.1177/0008125619862257

So, E. C. (2013). A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? *Journal of Financial Economics*, *108*(3), 615–640. https://doi.org/10.1016/j.jfineco.2013.02.002

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, *70*(1), 65–94. http://www.jstor.org/stable/1884513

Sougiannis, T. (1994). The accounting based valuation of corporate R&D. *The Accounting Review*, *69*(1), 44–68. http://www.jstor.org/stable/248260

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374. http://www.jstor.org/stable/1882010

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *The American Economic Review*, *92*(3), 434–459. http://www.jstor.org/stable/3083350

Stark, A. W. (2008). Intangibles and research – an overview with a specific focus on the UK. *Accounting and Business Research*, *38*(3), 275–285. https://doi.org/10.1080/00014788.2008.9663340

Stiglitz, J. E. (2002). Information and the change in the paradigm in economics. *American Economic Review*, *92*(3), 460–501. https://doi.org/10.1257/00028280260136363

Sullivan, J. H., Warkentin, M., & Wallace, L. (2021). So many ways for assessing outliers: What really works and does it matter? *Journal of Business Research*, *132*, 530–543. https://doi.org/10.1016/j.jbusres.2021.03.066

Swets, J. A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. Academic Press.

Tate, R. F. (1954). Correlation between a discrete and a continuous variable. point-biserial correlation. *The Annals of Mathematical Statistics*, *25*(3), 603–607. http://www.jstor.org/stable/2236844

Tsoligkas, F., & Tsalavoutas, I. (2011). Value relevance of R&D in the UK after IFRS mandatory implementation. *Applied Financial Economics*, *21*(13), 957–967. https://doi.org/10.1080/09603107.2011.556588

Van Gestel, T., Baesens, B., & Martens, D. (2010). From linear to non-linear kernel based classifiers for bankruptcy prediction [10th Brazilian Symposium on Neural Networks (SBRN2008)]. *Neurocomputing*, *73*(16), 2955–2970. https://doi.org/10.1016/j.neucom.2010.07.002

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory* (2nd ed.). Springer New York, NY. https://doi.org/10.1007/978-1-4757-3264-1

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Vorst, P., & Yohn, T. L. (2018). Life Cycle Models and Forecasting Growth and Profitability. *The Accounting Review*, *93*(6), 357–381. https://doi.org/10.2308/accr-52091

Waegelein, J. F. (1988). The association between the adoption of short-term bonus plans and corporate expenditures. *Journal of Accounting and Public Policy*, *7*(1), 43–63. https://doi.org/10.1016/0278-4254(88)90004-X

Wahlen, J. M., & Wieland, M. M. (2011). Can Financial Statement Analysis Beat Consensus Analysts' Recommendations? *Review of Accounting Studies*, *16*(1), 89–115. https://doi.org/10.1007/s11142-010-9124-5

Wang, L. (2016). Capitalising or expensing research and development expenditures: A tax perspective explanation. *China Journal of Accounting Studies*, *4*(1), 79–103. https://doi.org/10.1080/21697213.2016.1176416

Watts, R., & Zimmerman, J. (1986). Positive theory of accounting. *Englewood Cliffs, NY: Prentice-Hall.*

Watts, R. L., & Zimmerman, J. L. (1978). Towards a positive theory of the determination of accounting standards. *The Accounting Review*, *53*(1), 112–134. http://www.jstor.org/stable/245729

Watts, R. L., & Zimmerman, J. L. (1979). The demand for and supply of accounting theories: The market for excuses. *The Accounting Review*, *54*(2).

Watts, R. L., & Zimmerman, J. L. (1990). Positive accounting theory: A ten year perspective. *The Accounting Review*, *65*(1), 131–156. http://www.jstor.org/stable/247880

Weerts, H. J. P., Mueller, A. C., & Vanschoren, J. (2020). Importance of Tuning Hyperparameters of Machine Learning Algorithms. https://arxiv.org/abs/2007.07588

White, G. I., Sondhi, A. C., & Fried, D. (2002). *The analysis and use of financial statements*. John Wiley & Sons.

Woolridge, J. R. (1988). Competitive decline and corporate restructuring: Is a myopic stock market to blame? *Journal of Applied Corporate Finance*, *1*(1), 26–36. https://doi.org/10.1111/j.1745-6622.1988.tb00155.x

You, H., & Cao, K. (2021). Financial analysis and machine learning. *Proceedings of the 4th Annual SQA-CFA Society NY Joint Conference Data Science in Finance: Learning from Machine Learning (Virtual).*

Zarowin, P. A. (2019). Book review of Financial statement analysis and earnings forecasting (Foundations and Trends in Accounting) by Steven J. Monahan. *The Accounting Review*, *94*(3), 375–379.

Zhang, Y. (2012). Support Vector Machine Classification Algorithm and Its Application. In C. Liu, L. Wang, & A. Yang (Eds.), *Information computing and applications* (pp. 179–186). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34041-3_27

Zhao, Q., Xu, W., & Ji, Y. (2023). Predicting financial distress of Chinese listed companies using machine learning: To what extent does textual disclosure matter? *International Review of Financial Analysis*, *89*, 102770. https://doi.org/10.1016/j.irfa.2023.102770

Zhao, R. (2002). Relative value relevance of R&D reporting: An international comparison. *Journal of International Financial Management & Accounting*, *13*(2), 153–174. https://doi.org/10.1111/1467-646X.00082

Appendix A: Variable definitions

Table A.1: Definitions and measurements of variables

Variable	Definition	Measurement
RD_CAP	Is an indicator variable coded 1 if the firm capitalizes its R&D costs at least once over the period 1992–2001, 0 otherwise.	1 if conditions are met, 0 otherwise.
CAPITALIZE	Is an indicator variable coded 1 if a capitalizing firm actually capitalizes R&D in year t, 0 otherwise.	Dummy variable coded 1 if the change in R&D costs is positive, 0 otherwise.
TAF	Total assets, free of R&D effects.	Total assets – gross development costs + development cost amortization.
RD_ASSET	R&D asset	Gross development costs - development cost amortization/Avg. TA.
RDS	Expensed R&D divided by sales.	R&D expenses/Avg. sales.
CFRD	Total cash flow of R&D scaled by total assets.	RDS$\times$sales+change in gross development costs/Avg. TA.
CFRDEXP		CFRD for expensers, 0 otherwise.

Table continued on the next page

Variable	Definition	Measurement
CFRDEXPCAP		RDS×sales/Avg. TA for capitalizers, 0 otherwise.
CFRDCAPCAP		Change in gross development costs/Avg. TA for capitalizers, 0 otherwise.
SIZE	The natural logarithm of total assets	ln(TA)
PTB	Price to book value ratio	WC09304
CAPEX	Capital expenditures	Capital expenditures/Avg.TA.
ROA	Return on assets ratio	Net income+net financial expense+R&D amortization+R&D expenses/Avg.TA.
EPS	Earnings per share	Net income+R&D amortization/Avg. common shares outstanding
PR	Period close price	WC05085
RET	Annual stock return	Price in year t- Price in year $t-1$/Price in year $t-1$
ABVPS	Book value per share adjusted for capitalized R&D	Equity capital & reserves/Avg. common shares outstanding
RDEXPEXPS	R&D expenses per share (expensers)	R&D expenses/Avg. common shares outstanding for expensers, 0 otherwise
RDCAPCAPS	Capitalized R&D per share (capitalizers)	Change in gross development costs/Avg. common shares outstanding for capitalizers, 0 otherwise

Table continued on the next page

Variable	Definition	Measurement
RDEXPCAPS	Expensed R&D per share (capitalizers)	R&D expenses/Avg. common shares outstanding for capitalizers, 0 otherwise
AEMV	EPS to market value	EPS/ Avg. Market value
DAEMV		Change in AEMV between t and $t-1$
DRDEXPEXPMV	Change in R&D expenses to Market value (expensers)	Change in R&D expenses/Avg. Market value for expensers, 0 otherwise
DRDCAPCAPMV	Change in capitalized R&D to Market value (capitalizers)	Change in gross development costs/Avg. Market value for capitalizers, 0 otherwise
DRDEXPCAPMV	Change in Expensed R&D to Market value (capitalizers)	Change in R&D expenses/Avg. Market value for capitalizers, 0 otherwise

Table A.2: Raw accounting items

Variable	Description	DataStream Code
CE	Cash & cash equivalents	WC02001
AR	Accounts receivables	WC02051
INVT	Inventories-total	WC02101
STI	Short-term investments	WC02008
CA	Current assets	WC02201
PPEG	Property, plant & equipment- Gross	WC02301
PPEN	Property, plant & equipment- Net	WC02501
TA	Total assets	WC02999
CL	Current liabilities	WC03101
TL	Loans	WC02271
ECR	Equity capital & reserves	WC03501
OIBDAM	Operating income before depreciation & amortization	WC018155
DEFTA	Deferred tax	WC03263
TD	Total debt	WC03255
LTL	Total loan capital	WC03251
NETSAL	Total sales	WC01001
COGS	Cost of goods sold	WC01051
DEPAM	Depreciation, depletion & amortization	WC01151
IINC	Interest income	WC04149
IP	Interest paid	WC04148
IBT	Income before tax, extraordinary items & preferred dividends	WC01401
NETINC	Net income	WC04001

Table continued on the next page

Variable	Description	DataStream Code
TAX	Total tax charge	WC01451
NPS	Net proceeds from sale/issue of common and preferred dividends	WC04251
CAPEX	Capital expenditures	WC04601
CASHDIVS	Cash dividends	WC04551
XIT	Extraordinary items	WC04225
CSHOUT	Common shares outstanding	WC05302
RDEXP	Expensed R&D	WC01201
RDCAP	Gross development costs (capitalized R&D)	WC02505
RDAM	R&D amortization	WC02506
OINTGA	Other intangible assets	WC02649
MI	Minority interests	WC03426
LTBOR	Long-term borrowings	WC04401

Appendix B: Point-Biserial Correlations

The Point-Biserial Correlation coefficient is given by:

$$r_{pb} = \frac{\overline{X_1} - \overline{X_0}}{s_X} \sqrt{\frac{n_1 n_0}{n(n-1)}}$$

Where:

- $\overline{X_1}$ is the mean of the continuous variable for the group where the binary variable is 1.
- $\overline{X_0}$ is the mean of the continuous variable for the group where the binary variable is 0.
- s_X is the standard deviation of the continuous variable.
- n_1 is the number of observations where the binary variable is 1.
- n_0 is the number of observations where the binary variable is 0.
- n is the total number of observations.

Table B.1: Point-Biserial Correlation- ΔROA

Variable	**Correlation**	**P-value**	**Significant** ($p < 0.05$)
DEFTA	0.021	0.011	Yes
TAX	0.063	0.000	Yes
NPS	-0.036	0.000	Yes
TA	0.052	0.000	Yes
CE	-0.014	0.072	No
CASHDIVS	0.001	0.857	No
NETSAL	0.073	0.000	Yes
AR	0.058	0.000	Yes
CAPEX	-0.002	0.817	No
DEPR	-0.034	0.000	Yes
DEPAM	-0.001	0.908	No
PPEG	0.020	0.014	Yes
PPEN	0.031	0.000	Yes
OINTGA	0.012	0.115	No
INVT	0.020	0.010	Yes
LTBOR	-0.052	0.000	Yes
COGS	0.050	0.000	Yes
MI	0.022	0.005	Yes
IBT	0.170	0.000	Yes
ECR	0.040	0.000	Yes
CA	0.030	0.000	Yes
CL	0.032	0.000	Yes
LTL	-0.004	0.600	No
TD	0.002	0.784	No
RDCAP	0.014	0.188	No
RDEXP	-0.022	0.012	Yes
RDAM	0.004	0.719	No
IINC	0.014	0.257	No
STI	0.000	0.997	No
XIT	-0.009	0.278	No
OIBDAM	0.104	0.000	Yes
CSHOUT	-0.064	0.000	Yes
IP	0.034	0.000	Yes
CEQ	0.011	0.149	No
NETINC	0.186	0.000	Yes

Table B.2: Point-Biserial Correlation- ΔPR

Variable	**Correlation**	**P-value**	**Significant** ($p < 0.05$)
DEFTA	-0.015	0.056	No
TAX	-0.005	0.532	No
NPS	0.011	0.173	No
TA	-0.046	0.000	Yes
CE	0.006	0.415	No
CASHDIVS	-0.049	0.000	Yes
NETSAL	-0.037	0.000	Yes
AR	-0.033	0.000	Yes
CAPEX	-0.075	0.000	Yes
DEPR	-0.021	0.009	Yes
DEPAM	-0.010	0.216	No
PPEG	-0.016	0.050	Yes
PPEN	-0.032	0.000	Yes
OINTGA	-0.006	0.468	No
INVT	-0.043	0.000	Yes
LTBOR	-0.031	0.000	Yes
COGS	-0.043	0.000	Yes
MI	-0.014	0.082	No
IBT	0.007	0.348	No
ECR	-0.015	0.053	No
CA	-0.031	0.000	Yes
CL	-0.031	0.000	Yes
LTL	-0.023	0.003	Yes
TD	-0.031	0.000	Yes
RDCAP	0.036	0.001	Yes
RDEXP	-0.005	0.580	No
RDAM	0.045	0.000	Yes
IINC	-0.055	0.000	Yes
STI	-0.029	0.001	Yes
XIT	0.007	0.384	No
OIBDAM	-0.007	0.372	No
CSHOUT	0.060	0.000	Yes
IP	-0.004	0.632	No
CEQ	-0.024	0.002	Yes
NETINC	0.006	0.428	No

Table B.3: Point-Biserial Correlation- ΔRET

Variable	**Correlation**	**P-value**	**Significant** ($p < 0.05$)
DEFTA	-0.011	0.175	No
TAX	-0.028	0.000	Yes
NPS	-0.006	0.431	No
TA	-0.006	0.482	No
CE	0.016	0.036	Yes
CASHDIVS	-0.027	0.001	Yes
NETSAL	-0.014	0.071	No
AR	-0.009	0.227	No
CAPEX	-0.026	0.001	Yes
DEPR	0.015	0.053	No
DEPAM	0.021	0.008	Yes
PPEG	0.003	0.709	No
PPEN	0.002	0.821	No
OINTGA	-0.009	0.246	No
INVT	-0.013	0.094	No
LTBOR	-0.020	0.012	Yes
COGS	-0.014	0.079	No
MI	0.001	0.944	No
IBT	-0.020	0.010	Yes
ECR	-0.004	0.584	No
CA	-0.001	0.947	No
CL	-0.003	0.715	No
LTL	-0.000	0.983	No
TD	0.001	0.876	No
RDCAP	0.028	0.008	Yes
RDEXP	0.011	0.199	No
RDAM	0.032	0.003	Yes
IINC	-0.003	0.799	No
STI	0.005	0.572	No
XIT	0.008	0.331	No
OIBDAM	-0.020	0.010	Yes
CSHOUT	0.011	0.174	No
IP	-0.001	0.939	No
CEQ	0.003	0.680	No
NETINC	-0.016	0.042	Yes

Table B.4: Point-Biserial Correlation- ΔEPS

Variable	**Correlation**	**P-value**	**Significant** ($p < 0.05$)
DEFTA	0.017	0.039	Yes
TAX	0.071	0.000	Yes
NPS	-0.015	0.052	No
TA	0.029	0.000	Yes
CE	-0.010	0.193	No
CASHDIVS	-0.005	0.514	No
NETSAL	0.064	0.000	Yes
AR	0.045	0.000	Yes
CAPEX	-0.012	0.134	No
DEPR	-0.044	0.000	Yes
DEPAM	-0.010	0.221	No
PPEG	0.003	0.675	No
PPEN	0.005	0.506	No
OINTGA	0.012	0.133	No
INVT	0.015	0.063	No
LTBOR	-0.034	0.000	Yes
COGS	0.032	0.000	Yes
MI	0.011	0.153	No
IBT	0.175	0.000	Yes
ECR	0.036	0.000	Yes
CA	0.023	0.004	Yes
CL	0.017	0.034	Yes
LTL	-0.024	0.002	Yes
TD	-0.023	0.004	Yes
RDCAP	0.030	0.005	Yes
RDEXP	-0.002	0.814	No
RDAM	0.021	0.043	Yes
IINC	-0.013	0.322	No
STI	-0.014	0.135	No
XIT	-0.009	0.276	No
OIBDAM	0.105	0.000	Yes
CSHOUT	-0.038	0.000	Yes
IP	0.002	0.807	No
CEQ	-0.006	0.417	No
NETINC	0.188	0.000	Yes

Table B.5: Highly correlated variable pairs (Pearson corr. $> 70\%$)

Variable 1	**Variable 2**	**Correlation**
Panel A: ΔROA		
TA	CSHOUT	-0.751
NETSAL	COGS	0.817
PPEG	PPEN	0.856
IBT	OIBDAM	0.841
IBT	NETINC	0.948
OIBDAM	NETINC	0.785
Panel B: ΔPR		
TA	CSHOUT	-0.751
NETSAL	COGS	0.817
PPEG	PPEN	0.856
LTL	TD	0.879
RDCAP	RDAM	0.924
Panel C: ΔRET		
IBT	OIBDAM	0.841
IBT	NETINC	0.948
RDCAP	RDAM	0.924
OIBDAM	NETINC	0.785
Panel D: ΔEPS		
TA	CSHOUT	-0.751
NETSAL	COGS	0.817
IBT	OIBDAM	0.841
IBT	NETINC	0.948
LTL	TD	0.879
RDCAP	RDAM	0.924
OIBDAM	NETINC	0.785

Note: See Dormann et al. (2013) for the threshold selection of correlation coefficients.

Appendix C: Python code

```
    import numpy as np
    from sklearn.model_selection import GridSearchCV, TimeSeriesSplit
    from sklearn.ensemble import RandomForestClassifier
    from sklearn.metrics import classification_report, confusion_matrix, roc_curve, auc

    test_years = range(2011, 2020)  # Defines the test years (2011 to 2019)
    results = [] # Initially empty, the results of cross-validation will be stored here
    forecasts = [] # Initially empty, the forecasts will be stored here

# The parameters grid is defined here; in this example some parameters for a random forest are used:
param_grid = {
    'n_estimators': [50, 100, 200],
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
    'max_features': ['log2','sqrt'],
}

```

```
# Define training years (three years before the test year, with one year gap between train and test samples)
for test_year in test_years:
      train_years = [test_year - 2, test_year - 3, test_year - 4]
      train_data = df_copy[df_copy['Year'].isin(train_years)]
      # Define test data (holdout set, unseen during cross-validation)
      test_data = df_copy[df_copy['Year'] == test_year]
        # Choose features and target
        X_train = train_data[['feature1','feature2'...,'feature_n']]
        y_train = train_data['target']
        X_test = test_data[['feature1','feature2'...,'feature_n']]
        # Compute the number of firm-years for each year in the training set
       firm_years = train_data.groupby('Year').size()
        # Create a list of fold sizes based on the number of firm-years for each year
        fold_sizes = firm_years.tolist()
        # Create a TimeSeriesSplit object with the number of splits equal to the number of years
        tscv = TimeSeriesSplit(n_splits=len(fold_sizes)-1) #as there are three years in each training fold, there will be two splits.
        # Define the algorithm that will be used:
        model = RandomForestClassifier(random_state=42)
        # Fit the model in the training sample testing all the possible combinations of the parameters
```

```
        grid_search = GridSearchCV(model, param_grid,
           cv=tscv, n_jobs=-1)
        grid_search.fit(X_train, y_train)
        # Store the best combination of the parameters that
           were fitted during cross-validation
        best_model = grid_search.best_estimator_
         # Forecast on the test set
        test_predictions = best_model.predict(X_test)
        # Append predictions to the forecasts list
        forecasts.append((test_year, test_predictions))

# Output the forecasts for all test years with
   classificarion report:
for year, prediction in forecasts:
      # Print classification report
      print("Classification Report:")
      print(classification_report(actual_values,
         prediction))

 #Prints the following in the console:..................
  Out: Test Year: 2011, Forecast: [0 0 1 ... 1 0 1]
  Classification Report:
           precision    recall  f1-score   support

      0       0.50      0.23      0.31       510
      1       0.59      0.83      0.69       682
   accuracy                       0.57      1192

   macro avg      0.54      0.53      0.50      1192
  weighted avg    0.55      0.57      0.53      1192
```

Listing 1: Custom cross-validation and out-of-sample forecasts

```

# parameter grid for Random Forest
param_grid = {
    'n_estimators': list(range(500, 2100, 500)),
    'max_features': ['sqrt'],
    'min_samples_leaf': [1, 2, 3, 4],
    'max_depth': [1, 2, 3, 4],
    'bootstrap': [True],
    'max_samples': [0.5]
}

# parameter grid for XGBoost
param_grid = {
    'n_estimators': list(range(500, 2100, 500)),
    'learning_rate': [0.005, 0.01, 0.05],
    'max_depth': [1, 2, 3, 4, 5],
    'min_child_weight': [1, 2, 3, 4],
    'subsample': [0.5],
    'colsample_bytree': [0.8]
}

# parameter grid for Logistic Regression
param_grid = {
    'C': [0.01, 0.1, 1, 10],
    'solver': ['liblinear'],
    'penalty': ['l1', 'l2']
}

# parameter grid for SVM
```

```
param_grid = {
    'C': [0.1, 1, 10],
    'kernel': ['linear', 'rbf'],
    'gamma': [0.01, 0.1, 1],
}
```

Listing 2: Parameter grid for the algorithms

Appendix D: Averaging ROC curves

Receiver operating characteristic (ROC) curves are a popular method of comparing the adequacy of different classifiers on a two-class problem. There are a variety of situations in which an analyst attempts to combine several ROC curves into one representative curve. Many ways of doing this are available; however, there is a degree of subtlety which is often neglected in deciding which to apply (Hogan & Adams, 2023).

Among all the available options on how to construct averaged ROC curves, the simplest approach is to merge all the scores from all the test folds, mentioned as ”pooling” (Swets & Pickett, 1982). The ROC curve is constructed by merging all the test folds, therefore the TP and FP are calculated in the following way:

$$\tilde{tp} = \frac{\sum_{i=1}^{M} nP_i(t)}{\sum_{i=1}^{M} nP_i} = \frac{nP(t)}{nP}$$

$$\tilde{fp} = \frac{\sum_{i=1}^{M} nN_i(t)}{\sum_{i=1}^{M} nN_i} = \frac{nN(t)}{nN}$$

Provost et al. (1998) proposed the ”vertical averaging”, where FP and TP from each individual’s test fold ROC curve are averaged. The ROC curve is derived by:

$$\bar{R}(fp) = \frac{1}{M}\sum_{i=1}^{M} R_i(fp), \quad 0 \le fp \le 1$$

for which R_i is $tp = R_i(fp)$.

Both approaches have been tested in order to find the one that is closer to the average AUC score. Therefore, four algorithms were fitted, the out-of-sample scores for each year have been obtained, and finally the ROC curve was plotted using both ”pooling” and ”vertical averaging” approaches. In Table D.1 the out-of samples-score (AUC) is reported for several years, along with the average performance across the years, which is the out-of-sample performance for each algorithm.

Table D.1: Out-of-Sample Scores Across All Four Sets (2011-2019)

Year	**RF**	**SGB**	**LR**	**SVM**
2011	0.5977	0.5844	0.5581	0.5846
2012	0.6120	0.6085	0.5578	0.5657
2013	0.6160	0.6244	0.5779	0.5547
2014	0.5926	0.5948	0.5803	0.5632
2015	0.6148	0.6208	0.6389	0.6218
2016	0.5767	0.5712	0.5343	0.5318
2017	0.6709	0.6722	0.6479	0.6258
2018	0.5772	0.5786	0.5434	0.5259
2019	0.6144	0.6101	0.5807	0.5684
Average	0.61	0.61	0.58	0.57

The two ROC curve averaging approaches that were mentioned are plotted in Figure D.1. In the upper plot the ”pooling” approach is depicted, while in the lower plot, the ”vertical averaging” approach. The ”vertical averaging” is the approach that corresponds with the average AUC score calculated in Table D.1.

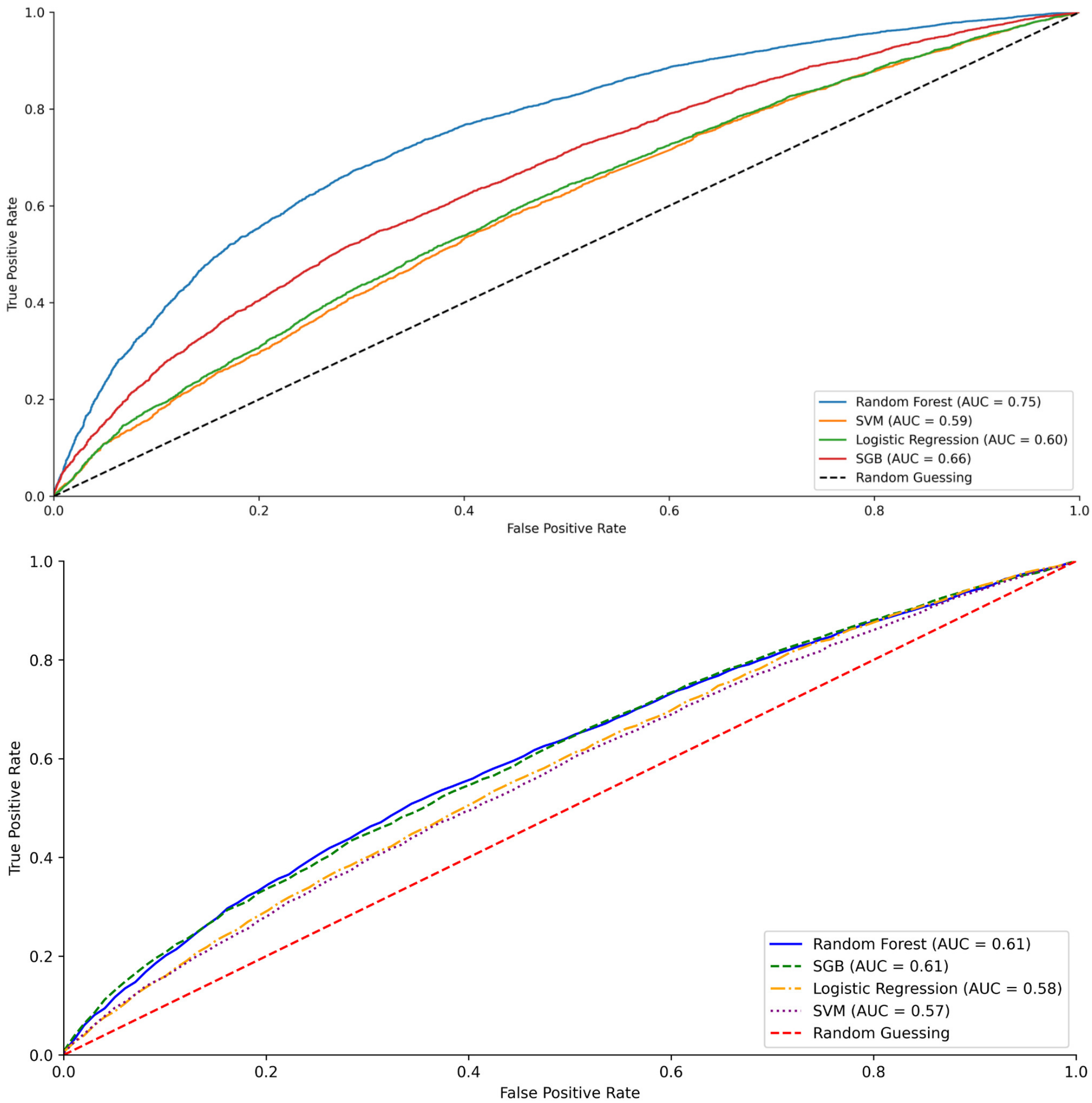

Figure D.1: ROC-AUC curves

Appendix E: Introduction to machine learning

E.1 Introduction to machine learning algorithms

Bertomeu et al. (2021) provided an excellent brief tutorial in which the basic steps of a machine learning exercise are presented. A basic terminology comparison between statistics and machine learning is presented in Table E.1 which is replicated from Bertomeu et al. (2021).

Table E.1: Comparison of terminology in statistics and machine learning (Bertomeu et al., 2021)

Statistics	**Machine learning**
estimator	algorithm, model
observation	example
independent variable, regressor	feature
dependent variable	response, target

The origins of machine learning can be traced back in 1949, in the book of Donald Hebb entitled ”The Organization of Behavior”. Hebb mentioned that the strength of the connection of two neurons is increased when those two neurons are activated simultaneously (Hebb, 1949), which is the principle on which unsupervised learning in neural networks was relied upon. Furthermore, Hebb (1949) mentioned that learning occurs through the association of inputs, a concept that can be parallelized with the idea of finding correlations and patterns in the data in machine learning. Although Hebb’s work and research was in the fields of neurophsysiology and psychology, the terms that he has defined and the relationship between the human neurons have been the basis for machine learning.

E.1.1 Decision tree and random forest

The decision tree is a tool for decision making and it is used both in statistics and machine learning. Hunt's concept learning system framework [1] is considered the patriarch of decision trees. Trees have been used for classification tasks; a classification tree starts with the root of the tree and proceeds down to its leaves (Quinlan, 1986). An illustration of decision tree is given in Figure E.1.

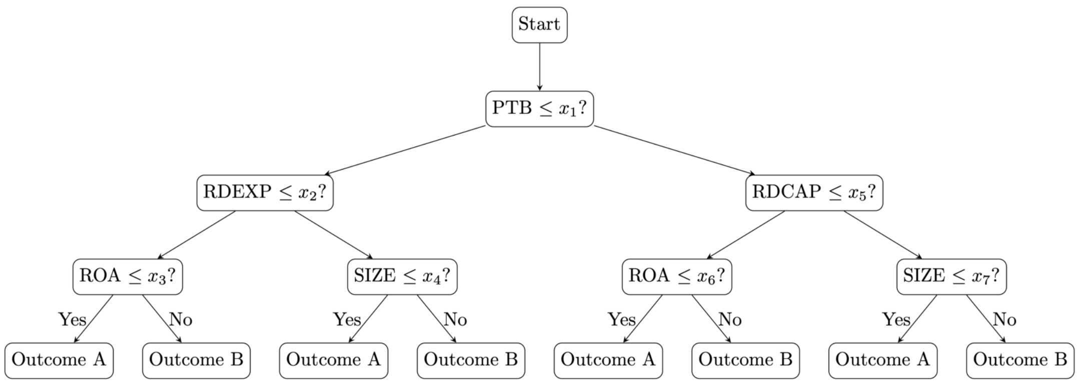

Figure E.1: Example of decision tree

The advantage of decision tree is that is easy to interpret graphically but it can overfit[2] the data very easily (Prajwala, 2015). Several researchers have experimented with constructing ensembles of decision trees (Breiman, 1996, 1999; Dietterich, 2000). Ensemble learning denotes a meta-learning methodology that consolidates predictions from multiple models to attain superior accuracy compared to any individual model. The concept is that a group of varied models can collaboratively enhance predictive accuracy by offsetting one another's deficiencies. This approach is typically classified into two primary types: parallel and sequential ensembles. Parallel approaches train base learners autonomously, whereas sequential methods construct models sequentially,

[1] See E. B. Hunt et al. (1966) for more details
[2] See Schaffer (1993) for more details about overfitting

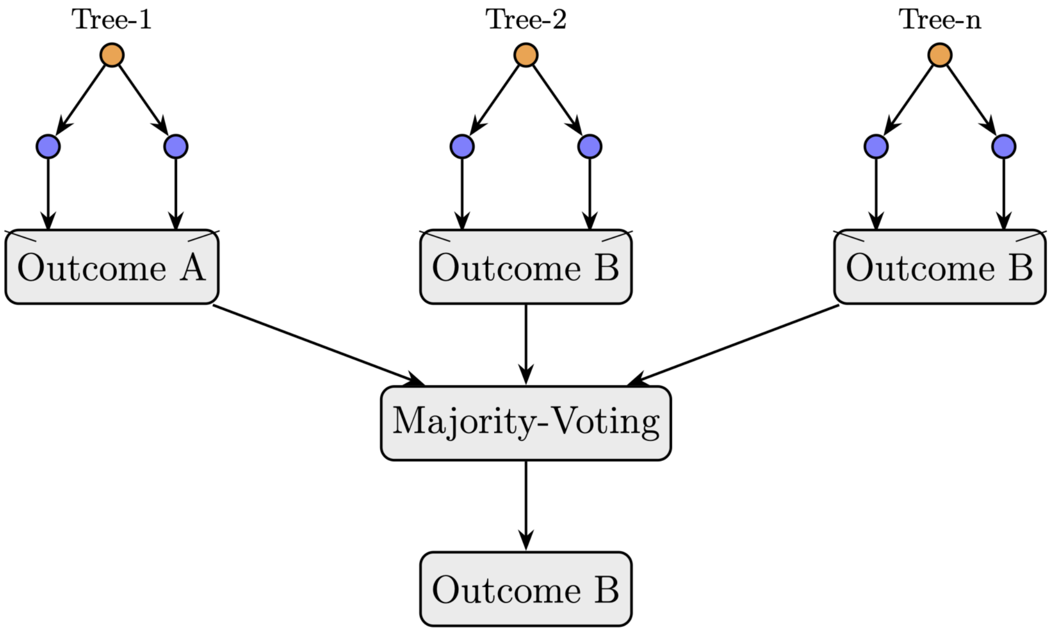

Figure E.2: Example of decision tree

with each subsequent model designed to rectify the errors of its predecessor (Polikar, 2012).

In his seminal paper, Breiman (2001) has introduced the concept of random forests. According to the definition of Breiman (2001), a random forest is "a classifier consisting of a collection of tree-structured classifiers $\{h(x,\theta_k), k = 1,\dots\}$ where the $\{\theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x". An illustration of random forest is given in Figure E.2.

E.1.2 Gradient boosting framework

Boosting is a recently developed framework[3] for the classification methodology (J. H. Friedman, 2001). Boosting was introduced by the computational learning literature

[3] See J. Friedman et al. (2000) and J. H. Friedman (2001)

(Freund, 1995; Freund & Schapire, 1997; Schapire, 1990). Boosting algorithms use ”weak” or ”base” algorithms to generate predictions which are combined by the boosting algorithm into a single prediction, which is expected to be more accurate than the predictions of the base algorithm (Schapire, 2003). The main difference between random forest and a boosting algorithm, like XGBoost, is that random forest builds trees with bagging (parallel) while boosting builds trees sequentially (Curth et al., 2024; Ghosal & Hooker, 2020).

E.1.3 Support vector machine (SVM)

The support vector machine algorithm (SVM) is based on the statistical theory of learning and generalization of Vapnik (2000). SVM is according to Zhang (2012) ”a two dimensional description of the optimal surface evolved from the linearly separable case”. An SVM used for classification ”separates the classes with largest gap (optimal margin) between the border line instances (support vectors)” (Chauhan et al., 2019). SVM can be used with a non-linear kernel for non-linear problems (Van Gestel et al., 2010).