

Can Machine Learning Models "Beat the Market" Around Earnings Announcements? A cross-market comparison between U.S. and Europe.

Filipe Chen

Dissertation

Master in Management

Supervised by

Miguel Augusto Gomes Sousa

José Abílio de Oliveira Matos

2026

Acknowledgements

I would like to thank my supervisors, Professor Miguel Augusto Gomes Sousa and Professor José Abílio de Oliveira Matos, for their guidance throughout this dissertation. Their questions about how the models were specified, how the results were presented, and what the final document needed to meet sent me back to the data and to the manuscript more than once, and the work is considerably stronger for it. The suggestion to assess the economic significance of the predictions, rather than stopping at statistical measures, came from them, and became one of the more interesting parts of the study.

I am also grateful to the Faculdade de Economia da Universidade do Porto for institutional access to LSEG Workspace, without which the earnings-announcement data used in this study could not have been assembled.

To my friends, for the company and the perspective, and for making sure this year was not only about a dissertation.

To my family, for their unconditional support throughout my studies, and for making all of it possible in the first place.

And to my girlfriend, for her support and encouragement throughout, and for lifting me up on the days when this felt beyond me.

Abstract

Corporate earnings announcements (EAs) are major events for companies and investors in capital markets and have historically been associated with abnormal stock price movements. Both the immediate reaction to corporate EAs and subsequent price adjustments, specifically post-earnings-announcement drift, indicate that stock prices moving around EAs may be forecastable. While many prior studies provide significant evidence of abnormal returns (ARs) from corporate EAs, they focus primarily on the United States and rely on linear, parametric models that do not adequately account for potentially nonlinear interactions among accounting fundamentals, valuation measures, and other earnings-related signals. This study examines whether machine learning (ML)-based models can better predict ARs that occur during and after corporate EAs and whether predictive performance differs across institutional environments.

This dissertation combines event-study methodology and supervised predictive models to evaluate the predictive performance of cumulative abnormal returns (CARs) generated for quarterly corporate EAs against a market-model benchmark. The primary event window is the CAR(-1,+1), which represents cumulative ARs occurring from one day before to one day after the announcement date. CAR predictions use firm-level accounting fundamentals and valuation ratios, earnings-related variables, and firm- and market-level characteristics. Predictive performance is compared out-of-sample between the non-linear ML models (Gradient Boosting and Random Forest) and a set of benchmark linear models (Lasso, Elastic Net, and OLS). A cross-market approach contrasts U.S.-listed firms with those listed in Europe, examining whether these relationships differ between firms that report to and are traded in different environments.

No model can forecast CAR(-1,+1) from information available before the event window, in either market. The earnings surprise, revealed at the announcement itself, explains part of it after the fact: Random Forest attains an out-of-sample R^2 of 3.77% against -1.58% for OLS on identical variables, and 3.98% in the United States against 1.47% in Europe.

Key words: Earnings announcements, Cumulative abnormal returns, Machine learning, Return predictability, Event study, Cross-market comparison

Resumo

Os anúncios de resultados constituem eventos de grande relevância para as empresas e investidores nos mercados de capitais, sendo historicamente associados a movimentos anormais nos preços das ações. Tanto a reação imediata como os subsequentes ajustamentos de preço, designadamente a deriva pós-anúncio de resultados, sugerem que estes movimentos poderão ser previsíveis. Embora vários estudos anteriores documentem evidência significativa de retornos anormais, estes focam-se sobretudo nos Estados Unidos e recorrem a modelos lineares e paramétricos que não captam adequadamente as interações potencialmente não-lineares entre fundamentos contabilísticos, métricas de avaliação e outros sinais relacionados com os resultados. Este estudo examina se modelos baseados em machine learning (ML) conseguem prever com maior precisão os retornos anormais em torno de anúncios trimestrais, e se o desempenho preditivo difere entre ambientes institucionais distintos.

Esta dissertação combina metodologia de estudo de eventos e modelos de previsão supervisionados para avaliar o desempenho preditivo de retornos anormais cumulativos (CAR). A janela de evento principal é o CAR(-1,+1), que acumula os retornos anormais do dia anterior ao dia posterior ao anúncio. As previsões utilizam fundamentos contabilísticos, rácios de avaliação, variáveis de resultados e características da empresa e do mercado. O desempenho preditivo é comparado fora da amostra entre os modelos ML não-lineares (Gradient Boosting e Random Forest) e modelos lineares de referência (Lasso, Elastic Net e OLS). É ainda adotada uma abordagem cross-market entre empresas cotadas nos Estados Unidos e na Europa.

Nenhum modelo consegue prever o CAR(-1,+1) a partir da informação disponível antes da janela de evento, em qualquer dos mercados. A surpresa nos resultados, revelada no próprio anúncio, explica parte dela a posteriori: o Random Forest atinge um R^2 fora da amostra de 3,77%, contra -1,58% do OLS com as mesmas variáveis, e 3,98% nos Estados Unidos contra 1,47% na Europa.

Palavras-chave: Anúncios de resultados, Retornos anormais cumulativos, Machine learning, Previsibilidade de retornos, Estudo de eventos, Comparação cross-market

Table of Contents

1. Introduction *1*

2. Literature Review *3*

2.1 Earnings Announcements and Stock Price Reactions **3**

2.1.1 Information Content of Earnings Announcements 3

2.1.2 Abnormal Returns and Event Study Methodology 3

2.1.3 Delayed Market Reactions and Post-Earnings-Announcement Drift 4

2.2 Accounting Fundamentals, Valuation Ratios, and Return Predictability (RP) **5**

2.2.1 Fundamentals-Based Predictability 5

2.2.2 Valuation Ratios, Earnings Surprises, and Forecast Errors 6

2.2.3 Firm Characteristics and Market Conditions 7

2.3 Limitations of Traditional Econometric Approaches **8**

2.3.1 Linear and Parametric Modeling in Earnings Studies 8

2.3.2 Challenges with High-Dimensional and Non-linear Data 9

2.4 Machine Learning in Financial and Accounting Research **10**

2.4.1 Machine Learning for Return Prediction 10

2.4.2 Machine Learning Applications in Accounting and Earnings Prediction 11

2.4.3 Interpretability, Robustness, and Model Discipline 11

2.5 Cross-Market Evidence and International Perspectives **12**

2.5.1 U.S.-Centric Foundations of Earnings-Announcement Research 12

2.5.2 Cross-Country Variation in Earnings Informativeness 13

2.5.3 The Cross-Market Machine Learning Gap 14

2.6 Summary of Literature and Research Gap **14**

3. Methodology *16*

3.1 Methodological Approach **16**

3.2 Research Design **16**

3.2.1 Event Study Framework 16

3.2.2 Predictive Task and Modeling Strategy 17

3.2.3 Cross-Market Design 18

3.2.4 Time Period and Scope 18

3.3 Data Sources and Collection **18**

3.3.1 Earnings Announcement Dates (LSEG Workspace) 18

3.3.2 Daily Stock Prices and Market Index Returns (Yahoo Finance) 19

3.3.3 Fundamental and Valuation Data (Yahoo Finance) 19

3.3.4 Data Collection Procedure 20

3.4 Sample Selection and Data Cleaning **20**

3.5 Feature Engineering **21**

3.5.1 Base Features 21

3.5.2 Full Features 22

3.6 Event Study: Abnormal Return Computation **22**

3.7 Machine Learning Models **23**

3.7.1 Model Specifications 23

3.7.2 Train/Test Split 25

3.7.3 Evaluation Metrics 26

3.8 Economic Significance Assessment **26**
3.8.1 Trading Strategy Design 26
3.8.2 Benchmark Strategies 27
3.8.3 Performance Metrics 27

4. Results and Discussion ***29***

4.1 Descriptive Statistics **29**
4.1.1 Distribution of Cumulative Abnormal Returns 29
4.1.2 Correlation Analysis 30

4.2 Event Study Results **31**

4.3 ML Model Performance: Base Features **32**
4.3.1 Full Sample Results 32
4.3.2 U.S. vs European Market Comparison 33
4.3.3 Feature Importance: Base Features 34

4.4 ML Model Performance: Full Features **35**
4.4.1 Full Sample Results 35
4.4.2 Feature Importance: Full Features 36
4.4.3 Lasso Feature Selection 37
4.4.4 U.S. vs European Market Comparison 37
4.4.5 Base vs Full Feature Comparison 38

4.5 Economic Significance: Trading Strategy Simulation **39**
4.5.1 Full Sample Strategy Performance 39
4.5.2 Cross-Market Strategy Performance 40
4.5.3 Portfolio Growth Simulation 41

4.6 Discussion of Findings **43**
4.6.1 Predictive Power of ML Models 43
4.6.2 ML Models vs. Traditional Benchmarks 44
4.6.3 Cross-Market Differences in Predictability 45
4.6.4 Limitations 46

5. Conclusions ***48***

5.1 Summary of Findings **48**

5.2 Theoretical and Empirical Contributions **48**

5.3 Practical Implications **49**

5.4 Limitations and Directions for Future Research **50**

5.5 Concluding Remarks **50**

References ***51***

Annex ***55***

List of Tables

Table 1 - Summary of Literature and Research Gaps 15
Table 2 - Sample Description 21
Table 3 - Base Features (11 features, 100% coverage) 21
Table 4 - Full Features (4 additional features, ~62% of events have EPS data, ~50% have all four) 22
Table 5 - ML Model Specifications 25
Table 6 - Descriptive Statistics of CAR(-1,+1) 29
Table 7 - Out-of-Sample Performance: Base Features, Full Sample (Test: 2025) 32
Table 8 - Out-of-Sample R^2 by Market: Base Features (Test: 2025) 33
Table 9 - Out-of-Sample Performance: Full Features, Full Sample (Test: 2025) 35
Table 10 - Out-of-Sample R^2 by Market: Full Features (Test: 2025) 38
Table 11 - Trading Strategy Performance Metrics, Full Sample (Test: 2025) 40
Table 12 - Trading Strategy Performance by Market (Test: 2025) 40
Table 13 - Portfolio Growth Simulation (Monthly Compounding, 0.2% Transaction Cost, Test: 2025) 42

List of Figures

Figure 1 - Feature Importance Scores: Random Forest Model (Base Features) 34
Figure 2 - Feature Importance Scores: Random Forest Model (Full Features) 37
Figure 3 - Out-of-Sample R^2 Comparison: Base vs. Full Features Across Models and Markets 39
Figure 4 - Cumulative Portfolio Growth by Strategy: Combined Sample (2025 Test Period) 41
Figure 5 - Cumulative Portfolio Growth by Strategy: U.S. vs. European Subsamples (2025 Test Period) 43

List of Acronyms and Abbreviations

AR – Abnormal Return

CAR – Cumulative Abnormal Return

COVID-19 – Coronavirus Disease 2019

EA – Earnings Announcement

EMH – Efficient Market Hypothesis

EPS – Earnings Per Share

EU – Europe

FE – Forecast Error

FS – Financial Statements

IFRS – International Financial Reporting Standards

LSEG – London Stock Exchange Group

MAE – Mean Absolute Error

ML – Machine Learning

OLS – Ordinary Least Squares

PEAD – Post-Earnings-Announcement Drift

R^2 – Coefficient of Determination

RF – Random Forest

RIC – Reuters Instrument Code

RMSE – Root Mean Squared Error

RP – Return Predictability

RQ – Research Question

U.S. – United States

1. Introduction

Corporate earnings announcements (EAs) are the primary means by which firms disclose their financial performance to investors who hold the company's stock. There has been extensive research on the effects of EAs on stock price movements, and most studies have found significant impacts. Research has also shown that accounting information, such as EAs, becomes reflected in asset prices over time (Ball & Brown, 1968; Kothari, 2001).

Recently, researchers have developed new ways to analyse established financial phenomena using flexible predictive methodologies. Researchers have demonstrated that machine learning (ML) can significantly improve return forecasts in financial markets (Gu et al., 2020) and that ML can be used in accounting-based prediction problems, including those related to earnings (Chen et al., 2022). Additionally, U.S. and European equity markets have become increasingly interdependent while maintaining unique institutional settings, regulatory systems, and investor bases, which provides a reason to investigate whether EA news is evaluated similarly in both markets.

This dissertation examines abnormal stock price movements surrounding EAs. In capital markets research, Abnormal returns (ARs) are generally defined as the difference between the actual return on a particular stock and the return that the stock would have been expected to receive, based on a benchmark model that represents the reaction of the market to a company's specific information (MacKinlay, 1997). The evidence shows that many firms adjust their stock prices at or around the time of their EA (Ball & Brown, 1968). However, the evidence also shows that the adjustment in the stock price is often incomplete, with the stock price continuing to move in the direction of the earnings surprise after the announcement (Bernard & Thomas, 1989, 1990).

Furthermore, subsequent research has tried to explain why ARs associated with earnings-related information were predictable by linking return patterns to both accounting fundamentals and the nature of the EA component (Ou & Penman, 1989; Sloan, 1996). More recently, studies have provided further support for the continued importance of fundamentals-based signals in today's markets (Thakerngkiat et al., 2021). Collectively, the

previous research establishes EA as a prime environment to study return predictability (RP) and the processing of market information.

Although the previous research has provided extensive evidence supporting the predictability of returns around EAs, two gaps remain. First, a significant portion of the early studies in this area utilized U.S.-based data, which created ambiguity regarding whether the EAs-based patterns observed in U.S. markets can be generalized to other markets that have different institutional characteristics and reporting incentives (Sloan, 1996; Ball & Brown, 1968; Bernard & Thomas, 1989; Bushman & Piotroski, 2006). Additional evidence from international reporting reforms suggests that disclosure regimes can alter the information content of EAs (Landsman et al., 2012), implying that return dynamics surrounding EAs may differ between U.S. and European markets. Second, the majority of prior RP research relied on linear and parametric models that may not adequately account for the complexity of high-dimensional data and potential non-linear interactions among predictor variables (Hou et al., 2020; Gu et al., 2020). Although ML offers a flexible approach to RP and has been applied to both U.S. (Gu et al., 2020) and European markets separately (Drobetz & Otto, 2021), no study has directly compared predictive performance across these institutional settings in an earnings-announcement context.

This research addresses both gaps by testing whether ML models can predict ARs surrounding EAs and whether their predictive performance is similar across U.S. and European equity markets. From an empirical perspective, the research design integrates an event-study framework with ML models that use a variety of accounting fundamentals, valuation ratios, and earnings-related variables as predictors and compares predictive performance across markets via out-of-sample evaluations (MacKinlay, 1997).

This dissertation makes three contributions. First, it will provide evidence regarding the use of ML-based prediction of ARs surrounding EAs in an event-driven setting. Second, it will contribute to the cross-market literature by examining whether the predictive relationship between earnings-related variables and subsequent returns will vary between U.S. and European equity markets. Finally, this dissertation will provide additional insight into the relative performance of ML models versus traditional approaches to predicting returns surrounding EAs.

2. Literature Review

2.1 Earnings Announcements and Stock Price Reactions

2.1.1 Information Content of Earnings Announcements

EAs are one of the most important types of periodic disclosures made by corporations to their investors and other stakeholders in the capital markets. In the finance and accounting literature, the announcement of EAs is viewed as value-relevant because it contains new information that could lead to changes in investors' beliefs about firm values (Ball & Brown, 1968). Research has shown that stock prices react to EAs, indicating that EA information is reflected in stock prices (Ball & Brown, 1968).

Ball and Brown (1968) showed that a large portion of the stock price reaction to EAs occurs around the same time that the EAs are announced, which implies that EAs reduce uncertainty surrounding firm performance and causes investors to reassess their opinions of firm value. Research has supported Ball and Brown's views on the importance of EAs in the corporate disclosure process. Kothari (2001) pointed out that EAs are considered the most closely followed disclosures because they provide an overview of a corporation's recent financial performance and provide clues as to what might happen in the future. With regard to market efficiency, researchers expect that stock prices will reflect publicly available information, such as that provided in an EA, although the speed and extent of price adjustment could differ depending upon the firm, the market, and other firm-specific factors (Fama, 1970).

2.1.2 Abnormal Returns and Event Study Methodology

Stock price reactions to EAs have been measured using abnormal returns (ARs). In capital markets research, ARs are defined as the difference between a firm's actual return and its predicted return using an appropriate benchmark asset pricing model, with ARs representing the portion of price change related to firm specific information, rather than market-wide effects (MacKinlay, 1997).

Event study methodology represents a structured method for analysing stock price reactions to EAs. Specifically, event study methodology is used to measure ARs over a well-defined period around a known event date (i.e., an EA). MacKinlay (1997) formally outlined the event study methodology as follows: (a) the estimation window, (b) the expected return model, and (c) the event window during which ARs are calculated. Because the timing of EAs is known in advance and the information disclosed is both discrete and firm-specific, event studies remain the standard empirical method for investigating the relationship between stock price reactions and EAs (Ball & Brown, 1968; MacKinlay, 1997).

2.1.3 Delayed Market Reactions and Post-Earnings-Announcement Drift

Although EAs contain valuable, relevant information, most of the literature indicates that stock prices do not always adjust fully or instantly to this information. A common finding in the literature is post-earnings-announcement drift (PEAD), which refers to the tendency for stock prices to continue moving in the direction of the earnings surprise after the announcement.

Bernard and Thomas (1989, 1990) first documented PEAD and found that firms with positive earnings surprises at the announcement and in the days, weeks, and months after the announcement had positive ARs, whereas firms with negative earnings surprises had negative ARs. Therefore, the slow price reaction to earnings news is contrary to the semi-strong form of the Efficient Market Hypothesis (EMH) (Fama, 1970), which states that all publicly available information is quickly and fully reflected in prices.

Moreover, Fink (2021) conducted an extensive literature review on PEAD and concluded that it is an economically significant phenomenon that persists over time and is observed across many market types and under various institutional arrangements. Furthermore, based on the evidence provided, the results suggest that a considerable portion of the price response to earnings news is delayed beyond the immediate reaction to the earnings release and unfolds gradually over time, generating further predictable ARs. The delayed reaction to earnings news suggests that even though EAs contain useful information, market participants do not always process it efficiently.

Overall, the evidence presented above demonstrates that EAs are valuable, relevant information events. ARs provide a valid framework for quantifying the market's response to EAs, and many systematic patterns of post-announcement returns (e.g., PEAD) persist over time. Therefore, these results support further research to determine why earnings-related RP persists and why prices respond so slowly to earnings news.

2.2 Accounting Fundamentals, Valuation Ratios, and Return Predictability (RP)

Beginning with the evidence that Earnings Announcements (EAs) generate Abnormal Returns (ARs) and delayed price adjustments, an extensive amount of research has examined what type(s) of information determine the returns pattern of EAs. A primary focus of this line of research has been the study of accounting fundamentals, valuation ratios, earnings surprises, and firm- and market-level characteristics that affect how investors process earnings information.

2.2.1 Fundamentals-Based Predictability

A major strand of research has examined whether accounting fundamentals can explain cross-sectional differences in stock price reactions to EAs. Initial studies indicate that financial statement data have predictive power for future stock returns; therefore, investors have not fully or instantaneously incorporated all relevant accounting data into stock prices. Ou & Penman (1989) provided initial evidence of fundamentals-based predictability by demonstrating that combinations of multiple accounting ratios are significantly correlated with subsequent average returns; i.e., accounting fundamentals provide a basis for forecasting future returns.

This basic idea was further refined through examination of the persistence of earnings components. Sloan (1996) found that both the cash flow and accrual components of earnings exhibited different levels of persistence, and investors tended to overvalue the future performance of the accrual component relative to its actual future performance; the resulting misvaluation produced predictable ARs, thereby reinforcing the role of accounting fundamentals in determining return dynamics.

In more recent years, the existence of fundamentals-based predictability has also been supported in modern financial markets. Thakerngkiat et al. (2021) demonstrated that differences in accounting characteristics (e.g., abnormal accruals, earnings smoothness, book-to-market) between pairs of stocks explain cross-asset return predictability, with this predictability varying with funding liquidity and market sentiment. These results are consistent with the ideas of gradual information diffusion and limited arbitrage, which suggest that it takes time for information contained within accounting fundamentals to be incorporated into stock prices.

2.2.2 Valuation Ratios, Earnings Surprises, and Forecast Errors

The literature beyond the accounting fundamentals demonstrates how market expectations affect stock price reactions to EAs. An early body of research demonstrates that the response of returns to EAs is primarily driven by the earnings surprise component, rather than by the level of earnings. In this regard, Bernard & Thomas (1989) demonstrated that positive (negative) earnings surprises are followed by positive (negative) ARs, indicating that investors react to deviations of actual earnings from the expectations embedded in the stock's price.

Earnings surprises have been commonly operationalized through the use of forecast error (FE), which is measured as the difference between the actual earnings of a firm and the expected earnings of a firm, typically based upon analyst estimates or model-based expectations (Livnat & Mendenhall, 2006). FEs represent a direct measure of the amount of unexpected earnings information and have been widely used to explain variation in EA-related returns (Bernard & Thomas, 1989). Typically, larger FEs are associated with larger ARs, underscoring the importance of the magnitude of unexpected earnings information in shaping market reactions.

The literature has also shown that the effect of earnings surprises can vary depending on the firm's valuation at the time of the EA. Alvarado et al. (2021) investigated the degree to which a combination of earnings surprises, revenue surprises, valuation metrics, and reporting characteristics influences post-announcement returns. The authors' results demonstrate that the valuation metrics (i.e., price-to-earnings ratios) condition the market response to earnings-related news. In addition, the authors found that companies with high valuations experience larger stock price changes when their reported performance differs from analysts'

expectations. Complementary findings by Schnaubelt and Seifert (2020) also demonstrated the joint effects of prior valuation and unexpected earnings on average ARs.

2.2.3 Firm Characteristics and Market Conditions

Additionally, the return patterns around earnings announcements (EA) vary systematically for both firms and the broader market. Firm characteristics, including size, number of analysts covering the firm, and quality of the information environment, affect how quickly and efficiently earnings information is reflected in stock prices. As would be expected based on the premise that smaller firms and those with fewer analysts covering them likely face greater information frictions and therefore take longer to disseminate earnings information (Hong et al., 2000), smaller firms and firms with less analyst coverage generally exhibit stronger post-announcement ARs. Similarly, differences in accounting quality and disclosure practices across firms affect investors' ability to understand what is communicated through reported earnings (Dechow & Dichev, 2002).

Finally, as other researchers have previously established, market conditions also influence the return patterns associated with EA. For example, research indicates that ARs vary with market conditions and that stronger reactions typically occur in periods of greater uncertainty or volatility, when investors rely more heavily on firm-specific information. Choi (2019) found that increased market uncertainty had a significant positive relationship with trading volumes around EAs, providing evidence that investors' attention and the friction involved in processing earnings information are important factors affecting how quickly and accurately earnings information is priced.

More recently, the interaction between firm characteristics and the overall market state has been identified as influencing EAs' return patterns. For example, Kim et al. (2024) demonstrated that EAs provide information about future market-wide volatility and, therefore, affect investors' uncertainty during the announcement period.

In conclusion, the empirical research presented in this section demonstrates that earnings-related ARs are influenced by multiple factors, including accounting fundamentals, the firm's valuation, analysts' expectations, and the context in which the announcement occurs. These results highlight the complexities of measuring returns relative to EAs and suggest that

simple, additive models of returns may be inappropriate for capturing the complex relationships and non-linearities present in EA environments, thereby motivating the methodological discussion provided below.

2.3 Limitations of Traditional Econometric Approaches

Empirical research has shown that a variety of factors (including accounting fundamentals, valuation measures, FEs, and firm- and market-level characteristics) influence investors' reactions to EAs. However, as discussed in the previous sections, most of these studies have relied on econometric techniques that assume linearity, which may be inappropriate for modeling such complex interactions. In addition, traditional econometric techniques can be problematic when dealing with large amounts of non-linear or multivariate financial data.

2.3.1 Linear and Parametric Modeling in Earnings Studies

Early empirical studies focused on EAs and ARs used linear and parametric econometric methodologies, such as ordinary least squares (OLS), portfolio sorting, and factor-based asset pricing models. The early studies were important because they showed that earnings-related information is associated with stock prices, which react systematically to EAs and, in turn, produce predictable ARs.

Although the linear models have been empirically successful in demonstrating that earnings-related information is associated with stock prices reacting systematically to EAs and subsequently producing predictable ARs, they make several assumptions about the functional relationship between variables. Specifically, they assume that the relationships between the independent variables are linear and additive. A series of empirical studies also shows that the assumption of additivity may be violated in EA environments. Specifically, several studies have demonstrated that the market's reaction to EAs often depends on interactions among a firm's accounting fundamentals, valuation level, and characteristics (Alvarado et al., 2021; Schnaubelt & Seifert, 2020). Therefore, due to the presence of conditional and/or interactive relationships, the ability of linear models to capture both the non-linear effects and the conditional relationships between variables may be limited (Chen et al., 2022).

Substantial empirical evidence demonstrates the limitations of linear models. Harvey et al. (2016) show that when a large number of candidate predictors is tested within a linear framework, conventional significance thresholds produce a high rate of false positives. Hou et al. (2020) reach a related conclusion from a different direction: replicating 452 published anomalies, they find that roughly 65% lose statistical significance once NYSE breakpoints and value-weighted returns are applied. Taken together, these results indicate that conventional econometric methods struggle to distinguish genuine predictive relationships from statistical artefacts when the number of potential predictor variables is large. In addition, Chen et al. (2022) present evidence suggesting that the predictive power of earnings-related information may be a function of interaction effects that linear models cannot capture. They find that a combination of valuation ratios, financial ratios, and firm characteristics can explain the direction of future earnings changes, but that rigid parametric specifications mask potentially important nonlinear relationships.

2.3.2 Challenges with High-Dimensional and Non-linear Data

The growing volume and granularity of financial data mean that predicting earnings-related returns is increasingly a high-dimensional problem. Contemporary analyses include large numbers of accounting variables, valuation ratios, FEs, and firm-specific attributes, and at these levels of dimensionality, established econometric methods suffer in terms of inference quality and model stability.

The first difficulty is the sheer number of candidate predictors. Harvey et al. (2016) show that testing a large set of candidates within a linear framework raises the likelihood of false positives even at conventional significance thresholds, so that in-sample significance does not reliably translate into out-of-sample performance. Hou et al. (2020) reach the same conclusion empirically: their large-scale replication shows that the majority of previously documented return predictors do not survive a change in portfolio weighting and breakpoint convention, implying that apparent predictive signals in high-dimensional financial data are frequently artefacts of research design rather than stable features of returns.

The second difficulty is functional form. Gu et al. (2020) argue that linear specifications cannot capture the complex relationships in financial data, since the effect of one variable often depends on the levels of others or on prevailing market conditions. Chen et al. (2022)

provide earnings-specific evidence showing that interactions among accounting fundamentals, valuation ratios, and FEs matter for future earnings changes, and that models permitting non-linearity outperform those constrained to linearity. Kaiser (2019) documents similarly poor out-of-sample accuracy when traditional regression methods are applied to rich financial statement data in event-driven settings, and Bryzgalova et al. (2025) show that the problem is not confined to earnings prediction: factor models and predictive regressions are challenged by both high dimensionality and non-linear dependency structures.

Together, these results point to a need for methods able to accommodate non-linear relationships, interaction terms, and high-dimensional predictor spaces.

2.4 Machine Learning in Financial and Accounting Research

2.4.1 Machine Learning for Return Prediction

Recent research has demonstrated the superiority of ML-based models over traditional linear models for predicting cross-sectional stock returns. Gu et al. (2020) established that ML-based models can identify complex non-linear relationships between firm characteristics that traditional parametric models cannot and that return predictability arises primarily from high-dimensional interactions among features rather than from isolated linear relationships. They also emphasised the importance of evaluating such models out-of-sample and of establishing whether any predictive ability they possess is economically relevant.

Flexibility, however, carries its own risks. As Section 2.3.2 discusses, documented return predictors are fragile to reasonable variations in research design (Hou et al., 2020), and the risk of false discovery rises sharply with the number of predictors tested (Harvey et al., 2016); both concerns apply with greater force to models with more capacity to fit noise. Howard (2024) demonstrates this directly, showing that return-prediction performance is highly sensitive to how a model is developed, including the choice of training scope (group-specific versus pooled models) and the definition of the prediction target (absolute versus relative returns). Predictive power attributed to ML may therefore reflect implementation choices rather than any underlying relationship between the explanatory variables and the outcome.

2.4.2 Machine Learning Applications in Accounting and Earnings Prediction

While accounting data have traditionally been analysed using linear econometric methods to improve the prediction and explanation of earnings-related outcomes, there is increasing use of ML to process large volumes of financial statement (FS) data and identify non-linear relationships within it. The primary advantage of using ML in this context is that it enables the processing of large amounts of FS data and the identification of complex, non-linear relationships across accounting data, a key advantage given that earnings outcomes reflect complex interrelationships among multiple aspects of firm characteristics, accruals, and reporting choices.

The primary contribution of Chen et al. (2022) is their application of ML to detailed FS data to predict changes in future earnings. This analysis shows that ML methods result in higher predictive accuracy than linear models based on traditional FS data and identify additional information beyond common accounting ratios. These findings suggest that accounting data exhibit high-dimensional, non-linear structures that are not fully captured by traditional econometric techniques.

Complementary evidence for the use of ML to identify and forecast earnings management comes from Veganzones and Séverin (2025). They used ML techniques to identify and predict earnings management through accounting numbers. The authors found that a model based on a flexible algorithm identifies more patterns of earnings manipulation than traditional regression methods, as it can capture the complex ways a company may report its financials to manipulate earnings. Additionally, Kaiser (2019) illustrates both the potential and practical challenges of applying ML to earnings-announcement prediction tasks.

2.4.3 Interpretability, Robustness, and Model Discipline

A major concern in the literature is overfitting and false positives. Harvey et al. (2016) show that the conventional significance threshold becomes inadequate once a large number of candidate predictors have been tested, and Hou et al. (2020) demonstrate empirically how few documented predictors survive a change in methodological convention. Neither result is specific to machine learning, but both apply with particular force to flexible models, which

have far greater capacity to fit noise and therefore require correspondingly stricter out-of-sample validation.

Recent research also demonstrates considerable differences in how researchers implement various aspects of an ML model. Research by Howard (2024) has shown that a researcher's use of different training methods is likely to yield significantly different predictive results on the same data set. As such, researchers cannot reliably assume that the improvements they have seen in the performance of their models reflect their models' modeling of the underlying relationships in the data, rather than simply the model itself.

In addition to identifying issues with interpretation, researchers have been developing more interpretable ML models. For example, Veganzones and Séverin (2025) contend that researchers should utilize explanation and visualization tools to improve the transparency of ML models employed in forecasting earnings. However, Bryzgalova et al. (2025) show that while traditional factor models and flexible predictive models face challenges in high-dimensional settings, decision-tree-based grouping can recover interpretable, low-dimensional cross-sections with out-of-sample Sharpe ratios that substantially exceed those of standard factor models.

In summary, the literature provides evidence that ML models can overcome some of the limitations of traditional econometric models, such as their inability to include non-linear relationships, high-dimensional feature spaces, and complex interaction effects. However, ML is not a solution to all methodological problems, as the predictive power of ML models depends on the quality of the model development process, including model selection, rigorous validation, and full disclosure. In addition, most existing evidence is based on U.S. data and/or single-market data, raising the question of whether the predictive power of ML models will hold across different institutional environments. This is discussed in the next section.

2.5 Cross-Market Evidence and International Perspectives

2.5.1 U.S.-Centric Foundations of Earnings-Announcement Research

The majority of empirical research regarding EAs, ARs, and RP have been conducted in the context of U.S. equity markets. All the seminal works in this area were limited to U.S. data.

Ball and Brown (1968) and Bernard and Thomas (1989, 1990) both provided evidence that EAs conveyed value-relevant information, reflected in U.S. equity prices both contemporaneously and in subsequent price changes. Subsequently, Sloan (1996) demonstrated that differences in the persistence of cash flows and accruals created predictable ARs within the U.S. institutional framework.

Overall, the body of research provides significant evidence of how EAs affect asset pricing. However, the vast majority of the research is based on an environment of extremely high liquidity, a very large number of analysts covering individual stocks, significant protections afforded to investors, and generally consistent regulatory requirements for publicly traded companies. Therefore, it is unclear to what degree the findings from the majority of the existing research can be generalized beyond the U.S. context.

2.5.2 Cross-Country Variation in Earnings Informativeness

Bushman and Piotroski (2006) provided some of the first evidence on the international implications of the relationship between differences in a country's legal/political environment and the incentives for financial reporting and the timeliness of accounting information used to reflect changes in a firm's economic performance. Therefore, it is suggested that the degree of earnings informativeness is likely to vary significantly across countries and regulatory regimes that govern EA issuance.

Subsequent research examines capital markets' reaction to the implementation of international accounting standards. For example, Landsman et al. (2012) examined market reactions to the annual EAs associated with the mandatory implementation of the International Financial Reporting Standards (IFRS). They found that both abnormal return volatility and trading volume increased following the implementation of IFRS. As such, their findings suggest that improvements in the comparability of reported financial information and in the overall quality of financial reporting disclosures enhance the capital markets' ability to process and respond to earnings information. Additionally, their findings show that the increase in earnings informativeness following IFRS adoption is conditional on the strength of legal enforcement, which operates through three mechanisms: a reduction in reporting lag, an increase in analyst following, and an increase in foreign investment.

2.5.3 The Cross-Market Machine Learning Gap

International research has established a clear relationship between the institutional environment and both the ability of earnings to convey useful information and the way returns develop over time. The majority of empirical research relies on standard econometric techniques and single-market designs. At the same time, as Section 2.4 documents, much of the current evidence supporting the improved predictability of returns via ML models comes from either the U.S. or from pooled samples, which do not directly compare institutional characteristics. While ML evidence exists for European equity markets specifically (Drobetz & Otto, 2021), no study has directly compared ML-based cumalative abnormal returns (CAR) prediction around earnings announcements across U.S. and European markets in a unified framework.

Therefore, it is unclear if the predictive advantages of ML models will be preserved across markets characterized by significantly different regulatory systems, disclosure requirements, and investor bases. The importance of the potential for institutional differences (in particular, U.S. vs. European) to affect both how investors process earnings information and the form of the earnings-related predictive relationships is further emphasised by the extent to which ML models identify common or market-specific patterns of earnings-related signals that remain largely unexamined.

2.6 Summary of Literature and Research Gap

The literature reviewed in this chapter shows that earnings announcements (EAs) generate systematic abnormal returns (AR), that price adjustment is often delayed, and that return predictability (RP) is shaped by accounting fundamentals, valuation measures, forecast errors (FEs), and information environments. Recent research further demonstrates that ML methods can improve predictive performance relative to traditional econometric approaches, particularly in high-dimensional and nonlinear settings.

Table 1 - Summary of Literature and Research Gaps

Table 1 synthesizes the main literature streams reviewed in this chapter, highlighting their core theoretical insights, key empirical findings, and remaining limitations. This synthesis clarifies the central research gap addressed in this dissertation: whether ML models can reliably predict earnings-announcement-related ARs and whether their predictive performance is robust across distinct institutional contexts.

Literature stream	Core theoretical insight	Key findings and limitations
Earnings announcements & ARs	EA convey value-relevant earnings information that should be incorporated into prices	Strong immediate price reactions; incomplete adjustment leads to PEAD; evidence largely U.S.-centric
Accounting fundamentals & RP	Accounting information contains predictive signals not fully impounded by markets	Fundamentals, accruals, and cash-flow persistence predict ARs; mechanisms rely on linear models
Valuation, surprises & expectations	Market reactions depend on deviations from expectations embedded in prices	Earnings surprises and forecast errors interact with valuation; conditional effects often linearized
Firm characteristics & market conditions	Information environments and uncertainty affect information processing	RP varies with size, coverage, volatility, and attention; interaction effects hard to model
Traditional econometric methods	Linear, parametric models offer interpretability and inference	Limited ability to handle non-linearity, interactions, and high dimensionality; fragile inference
Machine learning approaches	Flexible models capture complex, non-linear predictive structures	Improved predictive performance, but results sensitive to implementation; evidence mostly U.S.-based
Cross-market evidence	Institutional context shapes earnings informativeness	Earnings effects vary across countries; ML generalizability across markets remains unexplored

3. Methodology

3.1 Methodological Approach

This chapter outlines the quantitative empirical approach adopted in this dissertation to examine how stock prices react to earnings announcements and to evaluate the predictive accuracy of machine learning models. The study employs an event-driven methodology, which enables precise isolation of abnormal stock price movements around a defined event date, the quarterly earnings announcement.

This study employs two complementary methodologies: event studies to evaluate abnormal stock price movements surrounding earnings announcements and five supervised machine learning (ML) regression models to identify complex relationships between earnings-related information, firm characteristics, and market responses. The five models employed are Ordinary Least Squares (OLS), Lasso, Elastic Net, Random Forest, and Gradient Boosting. This dual approach is adopted because traditional methods are limited by their inability to capture non-linearities and interaction terms.

Therefore, this study compares the predictive associations among earnings announcements, market response, and performance of ML models across different institutional settings of U.S. and European markets. The methodological strategy allows the study to address three primary research questions: (RQ1) Can ML models predict abnormal returns around earnings announcements? (RQ2) Do ML models outperform traditional linear models? (RQ3) Do predictive relationships differ between U.S. and European markets?

3.2 Research Design

3.2.1 Event Study Framework

To determine CARs, a market model is used to estimate normal returns for each company's stock by modelling how it has historically moved relative to the rest of the market before the event window. Once anormal returns (ARs) have been identified, CARs can be calculated by summing all ARs that occurred within the event window.

For the purposes of this study, the main event window for calculating CARs is CAR(-1,+1). Day 0 represents the day when quarterly earnings announcements are made. Day -1 measures the potential effect of "leakage" of information or anticipation related to the EA. Day +1 assesses the market's delayed reaction to the EA. This particular time frame of CAR(-1,+1) is widely employed in the literature on EAs (MacKinlay, 1997) to provide strong evidence of firms' immediate stock price reactions and to minimise the possibility that other relevant information is released at the same time.

It should be noted that CAR(-1,+1) begins one trading day before the earnings release date (Day 0). Any predictor that uses the return of Day -1 would therefore share a component with the target variable, and an investor could not, in any case, act on the Day -1 close before the event window had already started. For this reason, all price-based features are computed using information available at the close of Day -2, the last trading day before the event window, so that no feature overlaps with CAR(-1,+1). The earnings-per-share (EPS) based features, by contrast, become known only when earnings are released on Day 0; they are included in a second, retrospective specification whose purpose is to measure how much of the announcement reaction is explained by the earnings surprise itself, not to produce a tradeable forecast.

3.2.2 Predictive Task and Modeling Strategy

The empirical problem is formulated as a regression problem, in which ML algorithms forecast CARs, taking into account pre-announcement information on EAs, accounting fundamentals, valuation measures, and firm and market characteristics. This formulation allows direct comparisons with traditional linear forecasting methods, as well as flexible modelling of non-linear relationships and interaction terms.

Five different ML algorithms were selected that have been used extensively in prior studies of financial markets. OLS is included as the traditional linear benchmark. Lasso is selected for its regularisation property, performing automatic feature selection by shrinking irrelevant predictors to zero. Elastic Net extends Lasso by combining L1 and L2 penalties to address correlated predictors. Together, OLS, Lasso and Elastic Net form a regularisation family enabling systematic comparison of linear approaches with increasing regularisation discipline. Lastly, Random Forest and Gradient Boosting were chosen as they are widely used non-

linear ensemble methods in the empirical finance literature (Gu et al., 2020), allowing them to capture complex non-linear relationships and to model variable interactions.

3.2.3 Cross-Market Design

The research uses a comparative cross-market approach comparing companies listed on U.S. stock exchanges with those listed on European stock exchanges, represented by the S&P 500 and the STOXX Europe 600 indexes, respectively. This design allows for the determination of whether differences exist in the predictive value of EA-related models and ML algorithms across different institutional and legal environments.

3.2.4 Time Period and Scope

This research is based on data from all four quarterly fiscal periods in the years 2020 through 2025. Since it captures all earnings announcements during these fiscal quarters, it has captured a range of different market environments, which include a major global economic disruption as a result of the COVID-19 pandemic during 2020, the interest rate hike cycle (2022), and the artificial intelligence-driven market expansion (2023–2024), providing a diverse and challenging out-of-sample environment.

3.3 Data Sources and Collection

Secondary financial and accounting data for this research project are drawn from London Stock Exchange Group (LSEG) Workspace and Yahoo Finance via the yfinance package. The specific types of data needed include earnings announcement dates, daily stock prices, market index returns, and firm-level accounting and valuation variables for publicly traded companies in both the United States and European countries. All data extraction, cleaning and modelling was carried out in Python, using Google Colab as the computing environment.

3.3.1 Earnings Announcement Dates (LSEG Workspace)

Quarterly earnings announcement dates were extracted from LSEG Workspace (Datastream) for all firms in the sample using each firm's Reuters Instrument Code (RIC) as the identifier. LSEG Workspace was selected as the primary source for earnings announcements because of its institutional-grade data quality, standardised date adjustment method, and full coverage of both U.S. and European equity markets. The data that was extracted included the adjusted announcement date, the time of release (either Before Market Open or After Market Close),

the fiscal quarter, the fiscal year, and the region. In total, 11,748 U.S. earnings events and 11,301 European earnings events have been extracted.

3.3.2 Daily Stock Prices and Market Index Returns (Yahoo Finance)

Daily adjusted closing stock prices were downloaded using the yfinance package. LSEG Workspace was initially considered the primary source for historical price data; however, bulk extraction of daily prices across the full sample of 1,102 firms proved technically infeasible given the available institutional access constraints. As an alternative, Yahoo Finance was used via the yfinance package to retrieve adjusted daily close prices from January 2020 to December 2025 for all sample firms.

Market index prices were retrieved in the same manner, using the S&P 500 (Yahoo Finance ticker: ^GSPC) as the benchmark for U.S. firms and the STOXX Europe 600 (ticker: ^STOXX) as the benchmark for European firms. The resulting price files contained 1,507 trading days across 503 U.S. stocks and 1,544 trading days across 599 European stocks.

3.3.3 Fundamental and Valuation Data (Yahoo Finance)

Quarterly EPS data (actual reported EPS, analyst consensus estimate, and EPS surprise) were extracted from Yahoo Finance using the yfinance package. RIC codes obtained from LSEG were mapped to the corresponding Yahoo Finance ticker symbols using positional alignment between the two extraction lists. Firm-level characteristics, including market capitalisation, sector classification, and shares outstanding, were also retrieved from Yahoo Finance.

Historical market capitalisation per event was computed as the product of the adjusted closing price on Day -2 and shares outstanding at the time of the event. The historical P/E ratio per event was computed as the price on Day -2 divided by the sum of the four most recently reported quarterly EPS values (trailing twelve-month P/E). EPS data coverage reached approximately 62% at the event level and 58% at the firm level (639 of 1,102 firms), reflecting a structural limitation: quarterly EPS was available for 501 of the 503 U.S. firms but for only 138 of the 599 European firms, as many European firms report semi-annually rather than quarterly.

3.3.4 Data Collection Procedure

The data collection process was executed in the following sequential steps:

Step 1 - RIC Extraction. Reuters Instrument Codes were extracted from LSEG Workspace for all firms in the S&P 500 and STOXX Europe 600, yielding 503 U.S. and 599 European firm identifiers.

Step 2 - Earnings Announcement Dates. The quarterly announcement dates for all firms were obtained from the LSEG workspace, covering the period from January 2020 to December 2025.

Step 3 - Daily Stock Prices. Daily adjusted closing prices were downloaded for all 1,102 firms using the yfinance package. Market index prices for the S&P 500 and STOXX Europe 600 were downloaded in the same step.

Step 4 - EPS and Fundamental Data. Quarterly EPS data, market capitalisation, sector classification, and shares outstanding were retrieved from Yahoo Finance using the yfinance package.

Step 5 - Data Integration. All datasets were merged at the event level using RIC and event date as common keys, producing a single event-level dataset supporting both the event study and ML methodologies.

3.4 Sample Selection and Data Cleaning

The final sample comprises publicly listed U.S. and European firms from the S&P 500 and the STOXX Europe 600, covering the period from January 2020 to December 2025. The following cleaning steps were applied:

Step 1 - Duplicate removal**.** A total of 74 duplicate U.S. earnings events and 4 duplicate European earnings events were removed based on duplicate RIC and event date combinations.

Step 2 - Insufficient estimation-window data. Events with fewer than 30 trading days of historical stock price data available in the estimation window [-120, -10] were removed. This resulted in the removal of 2,668 of the 22,971 events processed, leaving 20,303.

Step 3 - Unmatched and invalid observations. A further 280 events were removed. Of these, 120 could not be matched to complete announcement metadata, as the fiscal quarter and fiscal year identifiers were unavailable; the remaining 160 fell outside the study period 2020 to 2025, comprising truncated year codes and observations dated 2019 or 2026. This reduced the sample to 20,023 events.

After all cleaning steps, the final analytical sample comprises 20,023 firm-event observations: 10,396 U.S. events and 9,627 European events, spanning 1,102 firms across both markets. No industry restrictions were imposed, ensuring broad sectoral coverage. Table 2 summarises the composition of the final sample.

Table 2 - Sample Description

Market	Index	Firms	Events (raw)	Events (final)
United States	S&P 500	503	11,748	10,396
Europe	STOXX Europe 600	599	11,301	9,627
Total	-	1,102	23,049	20,023

3.5 Feature Engineering

Fifteen predictor variables (features) were constructed for use as inputs in the ML models. All price-based features use only information available at the close of Day -2, the last trading day before the CAR(-1,+1) event window, so that no feature overlaps with the target variable; the EPS-based features are known only on Day 0 and are used in the retrospective specification described in Section 3.2.1. Features are organised into two sets: eleven base features, constructed from price and market data, with 100% event coverage, and four additional fundamental features, derived from EPS and valuation data, with approximately 62% event coverage.

3.5.1 Base Features

Table 3 lists the eleven base features, their construction, and their rationale.

Table 3 - Base Features (11 features, 100% coverage)

#	Feature	Formula / Source	Rationale
1	region_enc	EU=0, US=1	Captures cross-market differences in earnings informativeness and market efficiency

2	quarter_enc	Q1=1, Q2=2, Q3=3, Q4=4	Captures seasonal earnings patterns; Q4 full-year results tend to generate larger reactions
3	year	2020–2025 (integer)	Captures macro-economic time trends across the sample period
4	is_amc	1 if after market close, 0 otherwise	After-close announcements shift the price reaction to Day +1
5	pre_return_2d	$(P_{t-2} - P_{t-3}) / P_{t-3}$	Day -2 return; captures unusual price movement on the last trading day before the event window
6	pre_return_5d	$\prod(1+R_t)-1$ over days -6 to -2	Weekly pre-announcement drift; stocks with strong run-ups may have less remaining surprise
7	momentum_30d	$\prod(1+R_t)-1$ over days -31 to -2	Medium-term momentum; captures expectations embedded in prices before announcement
8	volatility_30d	$\sigma(R_t)$ over days -31 to -2	Pre-announcement uncertainty; higher volatility firms generate larger reactions
9	market_return_5d	$\prod(1+R_{m,t})-1$ over days -6 to -2	Controls for general market direction before the announcement
10	market_volatility_30d	$\sigma(R_{m,t})$ over days -31 to -2	Captures macro turbulence; high volatility periods dampen individual earnings reactions
11	log_mktcap_hist	$\log(P_{t-2} \times$ Shares Outstanding)	Log market cap controls for firm size and analyst coverage

3.5.2 Full Features

Table 4 lists the four additional fundamental features.

Table 4 - Full Features (4 additional features, ~62% of events have EPS data, ~50% have all four)

#	Feature	Formula / Source	Rationale
12	eps_surprise_pct	(EPS Actual − EPS Estimate) / |EPS Estimate| × 100	Core fundamental signal; positive surprise typically generates positive CAR
13	eps_actual_lag	EPS Actual from previous quarter	Baseline earnings level; captures historical profitability trend
14	eps_growth	(EPS Actual − EPS Actual Lag) / |EPS Actual Lag|	Quarter-over-quarter earnings trajectory; markets reward accelerating earnings
15	pe_ratio_hist	P_{t-2} / Trailing 12-month EPS	Valuation level at announcement; high-P/E stocks require larger beats to react positively

3.6 Event Study: Abnormal Return Computation

Abnormal returns are computed using the market model as the expected return benchmark. For each firm-event, the market model parameters are estimated over the estimation window of [-120, -10] trading days relative to the announcement date (Day 0), consistent with

MacKinlay (1997). A minimum of 30 valid trading days in the estimation window is required for inclusion.

The market model is estimated using Ordinary Least Squares regression for each firm-event:

$$R_{i,t} = a_i + \beta_i \times R_{m,t} + \varepsilon_{i,t} \quad (3.1)$$

where $R_{i,t}$ is the daily return of firm i on day t, $R_{m,t}$ is the daily return of the market index, α_i is the firm-specific intercept, β_i is the firm's sensitivity to market movements, and $\varepsilon_{i,t}$ is the residual. The S&P 500 is used as the market index for U.S. firms and the STOXX Europe 600 for European firms.

Abnormal returns for each day in the event window [-1, 0, +1] are computed as:

$$AR_{i,t} = R_{i,t} - (a_i + \beta_i \times R_{m,t}) \quad (3.2)$$

The Cumulative Abnormal Return (CAR), which serves as the target variable (Y) in the ML models, is computed by summing the abnormal returns over the three-day event window:

$$CAR_i(-1,+1) = AR_{i,-1} + AR_{i,0} + AR_{i,+1} \quad (3.3)$$

3.7 Machine Learning Models

3.7.1 Model Specifications

Five supervised regression models are implemented using Python's scikit-learn library. All predictor variables are standardised prior to model training using StandardScaler, transforming each feature to a zero mean and unit standard deviation. Standardisation is applied because the L1 and L2 penalties used by Lasso and Elastic Net act on the raw magnitude of the coefficients and are therefore scale-dependent; without it, the penalty falls most heavily on variables measured in small units, irrespective of their predictive content. The scaler is fitted on the training set only and then applied to the test set, so that no test-period information enters the transformation. The transformation is neutral for OLS and for the two tree-based models and is applied uniformly for comparability.

Ordinary Least Squares (OLS) serves as the linear benchmark, estimating the coefficients that minimise the squared difference between observed and predicted CAR:

$$min_\beta \sum_i (CAR_i - X_i\beta)^2 \quad (3.4)$$

where CAR_i is the realised cumulative abnormal return for event i, X_i is the corresponding vector of predictors, β is the vector of estimated coefficients, and the sum runs over all events in the training sample. No penalty is applied to the coefficient vector, making OLS the most interpretable model but also the most susceptible to overfitting in high-dimensional settings.

Lasso extends OLS by adding an L1 penalty to the objective function:

$$min_\beta \sum_i (CAR_i - X_i\beta)^2 + \lambda \sum_j |\beta_j| \quad (3.5)$$

where λ is the penalty parameter, which drives some coefficients exactly to zero and yields a sparse model (Tibshirani, 1996). This is appropriate in the present study, where many of the fifteen features may carry little or no predictive signal.

Elastic Net combines both L1 and L2 penalties:

$$min_\beta \sum_i (CAR_i - X_i\beta)^2 + \lambda[\varrho \sum_j |\beta_j| + (1-\varrho)\sum_j \beta_j^2] \quad (3.6)$$

where ϱ is the mixing parameter that sets the balance between the two penalties, implemented as l1_ratio in scikit-learn. At l1_ratio = 0.5 it is more robust than Lasso when predictors are mutually correlated (Zou & Hastie, 2005), a relevant concern given that pre_return_2d, pre_return_5d and momentum_30d capture overlapping return windows. Both regularised models use alpha = 0.001, the scikit-learn implementation of λ.

Random Forest averages the predictions of decision trees grown on bootstrap samples of the training data, with only a randomly selected subset of features considered at each split. This double randomisation decorrelates the individual trees and substantially reduces prediction variance without meaningfully increasing bias (Breiman, 2001), which suits the noisy, high-dimensional character of financial event data where no single predictor dominates. The parameters are n_estimators=200, max_depth=5 and min_samples_leaf=20, a deliberately conservative depth and leaf size given the low signal-to-noise ratio of the target.

Gradient Boosting instead builds the ensemble sequentially, fitting each new tree to the residuals of the current model:

$$f_m(x) = f_{m-1}(x) + \eta \cdot h_m(x) \quad (3.7)$$

where $f_m(x)$ is the ensemble prediction after m boosting stages, $h_m(x)$ is the regression tree fitted to the residuals of the preceding model $f_{m-1}(x)$, and η is the learning rate, which controls the contribution of each stage (Friedman, 2001). A low learning rate of 0.05 and shallow trees (max_depth = 3) limit that contribution and reduce the risk of overfitting. Gradient Boosting is more sensitive to hyper-parameter choices than Random Forest and carries a higher overfitting risk, but both are retained to allow a complete comparison of linear and non-linear model families across both markets. Table 5 summarises the specification of each model.

Table 5 - ML Model Specifications

Model	sklearn class	Key Parameters
OLS	LinearRegression	Default
Lasso	Lasso	alpha=0.001
Elastic Net	ElasticNet	alpha=0.001, l1_ratio=0.5
Random Forest	RandomForestRegressor	n_estimators=200, max_depth=5, min_samples_leaf=20, random_state=42
Gradient Boosting	GradientBoostingRegressor	n_estimators=200, max_depth=3, learning_rate=0.05, random_state=42

3.7.2 Train/Test Split

Using an 80/20 train/test split based on time will help prevent data leakage and better reflect how the model would perform in real-world applications. The model is trained on data from January 2020 to December 2024, and its performance is then evaluated on data from January 2025 to December 31, 2025. In this way, there is sufficient historical data to train the model while retaining enough out-of-sample data to evaluate predictive accuracy.

3.7.3 Evaluation Metrics

Model performance is assessed on the out-of-sample test set using three complementary regression metrics:

Root Mean Squared Error (RMSE): Penalises large errors more heavily and is expressed in the same units as the CAR.

$$RMSE = \sqrt{[\sum(CAR_i - \hat{CAR}_i)^2 / n]} \quad (3.8)$$

Mean Absolute Error (MAE): The average absolute difference between predicted and actual CAR values.

$$MAE = \sum |CAR_i - \hat{CAR}_i| / n \quad (3.9)$$

Coefficient of Determination (R^2): The proportion of CAR variance explained by the model.

$$R^2 = 1 - \sum(CAR_i - \hat{CAR}_i)^2 / \sum(CAR_i - \overline{CAR})^2 \quad (3.10)$$

All three metrics are computed on the out-of-sample test set only, ensuring that reported performance reflects the model's ability to generalise to new, unseen data.

3.8 Economic Significance Assessment

Statistical metrics such as R^2 and RMSE do not directly answer whether a model has practical economic value for an investor. Following Gu et al. (2020), an economic significance assessment is conducted through a trading strategy simulation.

3.8.1 Trading Strategy Design

The trading strategy uses only the base feature set to generate buy signals, ensuring that only information available at the close of Day -2 is used. The Random Forest model trained on 2020–2024 data generates a predicted CAR for each event in the 2025 test period. A buy signal is generated when the model predicts a positive CAR; no position is taken when the model predicts a negative CAR:

$$Signal_{i,t} = Buy \ \ if \ \ \hat{CAR}_{i,t} > 0; \ \ \ No \ position \ \ if \ \ \hat{CAR}_{i,t} \leq 0 \quad (3.11)$$

The investor purchases the stock at the close of Day -2, the last trading day before the event window, and closes the position at the close of Day +1. A transaction cost of 0.2% per trade is applied to account for brokerage commissions and bid-ask spreads. The net return per trade is:

$$r_{i,t}^{net} = CAR_{i,t} - c, \quad where\ c = 0.002 \quad (3.12)$$

Monthly average net returns are computed by averaging trade returns within each calendar month:

$$\bar{r_m} = (1 / N_m) \times \sum r_{i,t}^{net}, for\ i \in \Omega_m \quad (3.13)$$

where N_m is the number of trades in month m and Ω_m is the set of events in month m for which a buy signal was generated. Portfolio value is then compounded monthly from an initial capital W_0 = €10,000:

$$W_m = W_{m-1} \times (1 + \bar{r_m}), W_0 = €10{,}000 \quad (3.14)$$

3.8.2 Benchmark Strategies

Three benchmark strategies are used for comparison:

Always Buy: The investor purchases every stock before every earnings announcement regardless of any signal, representing an uninformed investor.

Momentum Rule: The investor buys only stocks with a positive five-day pre-announcement return (pre_return_5d > 0), representing a simple rule-based strategy requiring no ML model.

Random Strategy: The investor randomly selects approximately 50% of events to invest in, providing a baseline for random chance.

3.8.3 Performance Metrics

The economic significance of each strategy is evaluated using the following three metrics, each computed per trade on the realised CAR so that all strategies are compared on the same basis. The 0.2% transaction cost is applied in the portfolio growth simulation reported in Section 4.5.3.

Sharpe Ratio: Average return per unit of risk, relative to the risk-free rate r_f ≈ 0.01% per 3-day holding period:

$$Sharpe = (\bar{r} - r_f) / \sigma_r \quad (3.15)$$

where $\bar{r}$ and σ_r are the mean and standard deviation of the CAR across all trades executed by the strategy.

Profit Factor: Ratio of total gains to total losses across all trades:

$$Profit\ Factor = G / |L| \quad (3.16)$$

where G is the sum of all positive CARs and L is the sum of all negative CARs across the test period.

Hit Rate: Percentage of trades where the model correctly predicted the direction of the CAR:

$$Hit\ Rate = n^+ / N \quad (3.17)$$

where n^+ is the number of trades with a positive CAR and N is the total number of trades executed.

The analysis is conducted for the full sample and separately for U.S. and European subsamples to assess cross-market differences in economic value.

4. Results and Discussion

4.1 Descriptive Statistics

4.1.1 Distribution of Cumulative Abnormal Returns

A description of the distribution of the target variable CAR(-1,+1) is provided in Table 6, which provides an overview of the full sample and both subsamples (U.S. and Europe). In total, there were 20,023 observations. The average CAR was +0.20%, suggesting stocks had a slight positive bias in their response to earnings announcements during 2020-2025. However, the standard deviation for the full sample (6.31%) indicates significant cross-sectional variability in individual stock reactions to earnings announcements, confirming that the prediction task is inherently challenging.

Table 6 - Descriptive Statistics of CAR(-1,+1)

Statistic	Full Sample	US Only	EU Only
Events	20,023	10,396	9,627
Mean CAR	+0.20%	+0.17%	+0.22%
Std Dev	6.31%	6.66%	5.91%
Minimum	-51.42%	-51.42%	-42.01%
25th Percentile	-3.29%	-3.46%	-3.11%
Median	+0.15%	+0.13%	+0.19%
75th Percentile	+3.58%	+3.74%	+3.43%
Maximum	+56.31%	+56.05%	+56.31%

The distribution of CARs is centred close to zero, with a small positive mean of +0.20% and a median of +0.15%. There are, however, differences in the volatility of stock reaction to earnings announcements. U.S. stocks exhibited greater variance (a standard deviation of 6.66%) than European stocks (a standard deviation of 5.91%). This suggests that U.S. firms face more volatile analyst expectations and trade in a more liquid market, resulting in faster and larger reactions. Figure A1 in the Annex plots the two distributions side by side. Both are far more sharply peaked and more fat-tailed than a normal distribution with the same mean and standard deviation, with excess kurtosis of 4.4 and 3.2 respectively: most announcements move prices very little, while a small number move them a great deal.

The slightly positive mean CAR (+0.20%) across the sample is indicative of the well-established pattern that firms tend to lower analysts' expectations prior to announcing earnings, resulting in average positive surprises. This is commonly referred to as the "expectations management" effect in the academic literature and has previously been reported by Bartov et al. (2002). Additionally, firms have an incentive to avoid reporting negative earnings surprises as outlined in the same study.

4.1.2 Correlation Analysis

The correlations between the features and CAR are all within ±0.04 of zero. The largest in absolute terms are volatility_30d (r = 0.04), quarter_enc (0.03) and log_mktcap_hist (-0.03), while the pre-announcement return variables, pre_return_2d (-0.02), pre_return_5d (-0.01) and momentum_30d (-0.02), are indistinguishable from zero. No feature has a meaningful linear relationship with the target, consistent with the semi-strong form of the efficient market hypothesis and anticipating the near-zero R^2 values reported in Section 4.3. The full matrix is reported as Figure A2 in the Annex.

Among the features themselves, a few correlations are moderate. pre_return_5d is correlated with pre_return_2d (r = 0.46), with momentum_30d (0.35), and most strongly, with market_return_5d (0.50), while volatility_30d is correlated with market_volatility_30d (0.41); region_enc and is_amc are correlated (0.40) because after-market-close announcements are far more common in the U.S. than in Europe. These correlations reflect overlapping return windows and the common market component in individual stock returns and volatility. None of them is high enough to raise serious multicollinearity concerns in the linear models, but they do mean that the firm-specific and market-wide variables partly convey the same information, which is one reason Elastic Net was retained alongside Lasso.

Two of the variables in the matrix are not continuous. For the binary indicators region_enc and is_amc the coefficient is a point-biserial correlation, which is Pearson's coefficient applied to a dichotomous variable and is interpreted in the same way; quarter_enc is ordinal, so its coefficient assumes equal spacing between quarters and should be read as indicative only. The distributions of the individual features are reported as Figures A3 and A4 in the Annex.

4.2 Event Study Results

The event study was conducted over 22,971 earnings events after removal of duplicate firm-date observations. Of these, 20,303 events (88.4%) had sufficient price data in the estimation window [-120, -10] to compute reliable market model parameters. After removing events outside the 2020-2025 study period, the final sample comprises 20,023 events, of which 51.3% generated a positive CAR and 48.7% a negative one.

The beta estimates from the market model average 0.93 for U.S. firms and 0.99 for European firms, with alpha estimates close to zero in both markets, as expected for a diversified sample of large-cap firms over a long estimation window.

The market-model regressions from which these estimates are drawn are themselves well identified. Re-estimating equation (3.1) for every firm-event, the slope coefficient is significant at the 5% level in 92.6% of cases and at the 1% level in 88.4%, with a median t-statistic of 6.18, and the regressions explain on average 28.2% of the daily variation in firm returns over the estimation window. The two markets are closely comparable: 92.0% of U.S. regressions and 93.3% of European ones are significant at the 5% level, with mean R^2 values of 0.293 and 0.269 respectively. Two implications follow. First, the benchmark against which abnormal returns are measured is estimated with precision, so the CARs analysed below reflect firm-specific price movement rather than estimation error. Second, because the benchmark is equally well estimated in both markets, any cross-market difference in predictability reported in Sections 4.3 and 4.4 cannot be attributed to differences in the quality of the underlying market model.

The finding that 51.3% of events generate positive CARs is broadly consistent with the slightly positive mean CAR documented in Section 4.1.1. However, it also shows an almost even split between stocks generating positive and negative CARs, highlighting how difficult it is to predict direction. This could provide some insight into the hit ratio results reported in Section 4.5.

4.3 ML Model Performance: Base Features

4.3.1 Full Sample Results

Table 7 presents the out-of-sample results for the full sample using the eleven base features. None of the five models achieves a positive R^2: Lasso is the least negative (-0.59%), followed by Random Forest (-0.85%), Elastic Net (-1.01%), OLS (-1.70%) and Gradient Boosting (-2.67%). A negative out-of-sample R^2 means that the model's predictions are further from the realised CARs than the average of the test-period CARs would have been. In other words, price-based information available at the close of Day -2 carries no predictive power for the three-day announcement reaction in the 2025 test year.

Table 7 - Out-of-Sample Performance: Base Features, Full Sample (Test: 2025)

Model	RMSE	MAE	R^2
OLS	0.06928	0.05052	-1.70%
Lasso	0.06889	0.05015	-0.59%
Elastic Net	0.06904	0.05029	-1.01%
Random Forest	0.06898	0.05022	-0.85%
Gradient Boosting	0.06960	0.05086	-2.67%

Part of the negative level is mechanical. The average CAR over the 2020–2024 training period was +0.29%, whereas the average reaction in 2025 was -0.23%, so a model that simply predicted the training-period mean for every event would have obtained an R^2 of -0.59% on the test set. Lasso, whose coefficients are shrunk almost to zero, does essentially that, which is why it is the least negative model. The remaining models deviate from the constant prediction in proportion to how much freedom they have to fit noise: the depth-limited Random Forest deviates a little, OLS, which fits eleven unpenalised coefficients, more, and Gradient Boosting, which keeps fitting residuals through 200 sequential trees, most of all.

In practical terms the five models are equally uninformative. The differences in RMSE, from 0.06889 for Lasso to 0.06960 for Gradient Boosting, are about one hundredth of the 6.31% standard deviation of CAR (Table 6), and the same is true of MAE. This is the outcome the semi-strong form of the Efficient Market Hypothesis (Fama, 1970) predicts: past price behaviour is public information and should not forecast the reaction to news that has not yet been released.

The result also sets a clear benchmark for Section 4.4. Any positive R^2 obtained with the full feature set cannot be attributed to the price-based variables, and must come from the fundamental information that is added to them.

4.3.2 U.S. vs European Market Comparison

Table 8 presents the R^2 decomposition by market. In the U.S. subsample every model is clearly negative, from -1.72% (Random Forest) to -4.30% (OLS). In the European subsample four of the five values are within two tenths of a percentage point of zero, from 0.01% (OLS) to -0.17% (Random Forest), the exception being Gradient Boosting (-4.16%), which again over-fits the training data. The European models are therefore closer to the zero benchmark than the U.S. models, but this is not evidence of European predictability: a value of 0.01% means the model explains nothing, and the gap between the markets has a simple explanation, discussed below.

Table 8 - Out-of-Sample R^2 by Market: Base Features (Test: 2025)

Model	US R^2	EU R^2	US RMSE	EU RMSE
OLS	-4.30%	0.01%	0.07432	0.06402
Lasso	-2.00%	-0.16%	0.07349	0.06408
Elastic Net	-2.83%	-0.05%	0.07379	0.06404
Random Forest	-1.72%	-0.17%	0.07339	0.06408
Gradient Boosting	-2.37%	-4.16%	0.07363	0.06534

The more negative U.S. values reflect the tone of the 2025 announcement season rather than a worse model. U.S. earnings reactions averaged -0.45% in 2025 against +0.31% over the training period, so a model anchored on the training-period mean starts the test year with an R^2 handicap of about 1.1 percentage points before any feature is used; in Europe the average reaction moved far less (from +0.28% to -0.01%) and the corresponding handicap is about 0.2 percentage points. Once this level effect is taken into account, the two markets tell the same story: within each market the five models differ by less than 0.002 in RMSE, a negligible fraction of the dispersion of CAR, and none of them predicts.

The absence of any positive R^2 in either market is a substantive finding in its own right, and it is sensitive to the information-timing convention. As discussed in Section 4.6.4, allowing the return of Day -1 into the feature windows produces positive European R^2 values that are entirely attributable to the overlap between that return and the target. With the windows correctly closed at Day -2, the cross-market comparison with price-based features does not support the hypothesis that European markets are less efficient than U.S. markets in the days

before an announcement; on this evidence, neither market allows the announcement reaction to be forecast from prior price behaviour.

4.3.3 Feature Importance: Base Features

The out-of-sample results above establish that the base features carry no predictive information. The importance rankings show which of them the model relied on nevertheless.

Figure 1 shows the feature importance results from training the Random Forest model on all data points using just the base variables. No single variable dominates. The three most important features, pre_return_2d (approximately 19% of the model's total feature importance), volatility_30d (approximately 18%) and pre_return_5d (approximately 18%), are practically tied, followed by log_mktcap_hist (approximately 14%) and momentum_30d (approximately 10%). The two market-wide variables contribute roughly 6% each, and the four structural features (region, quarter, year and the after-market-close indicator) roughly 8% combined.

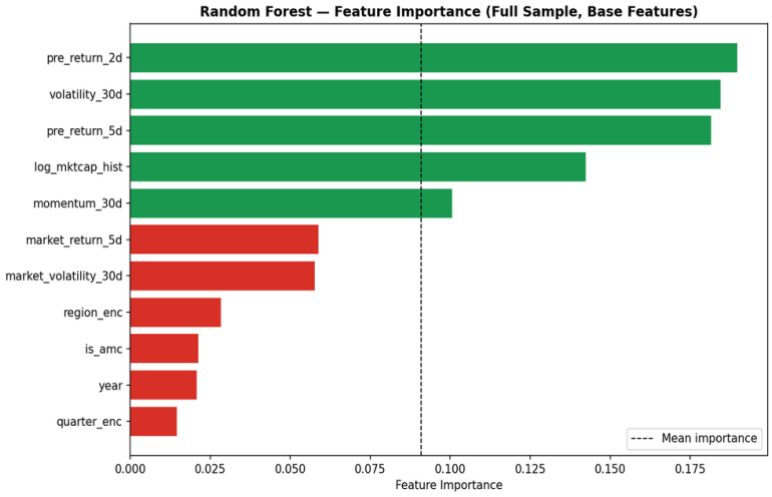


Figure 1 - Feature Importance Scores: Random Forest Model (Base Features)

An important aspect of this flat ranking is that it is what a Random Forest produces when it splits on noise. The algorithm has to place its splits somewhere, and continuous variables with many distinct values, such as returns and volatility, attract them even when they carry no information about the target. As Section 4.3 shows, the model built on these features has no out-of-sample explanatory power, so the ranking in Figure 1 describes what the model used, not what predicts the announcement reaction. The low ranking of the structural

features (region, quarter, year, timing) means that the characteristics of the announcement itself are not what the model relies on either.

4.4 ML Model Performance: Full Features

4.4.1 Full Sample Results

Table 9 presents the out-of-sample results for the full sample using all fifteen features. The inclusion of EPS surprise, earnings growth, lagged EPS, and the historical P/E ratio qualitatively changes the results. Because these variables require a complete set of the four fundamental variables, the full-feature analysis is carried out on a reduced sample of 10,108 events (7,920 in the 2020–2024 training set and 2,188 in the 2025 test set), compared with 20,023 events for the base features; approximately 9,900 events, predominantly European, are therefore excluded. On this sample, the Random Forest model achieves an out-of-sample R^2 of 3.77%, Gradient Boosting 2.20%, and Lasso 0.71%, while Elastic Net (-0.08%) and OLS (-1.58%) remain at or below zero. Relative to the base-feature results, the two tree-based models improve by 4.62 and 4.87 percentage points, respectively; Lasso by 1.30; Elastic Net by 0.93; and OLS by only 0.12.

Table 9 - Out-of-Sample Performance: Full Features, Full Sample (Test: 2025)

Training and test sets are reduced to 10,108 events (7,920 training, 2,188 test) with a complete set of fundamental features. The 'vs Base Features' column shows the change in R^2 from Table 7; because the two feature sets are evaluated on different test samples, these changes should be read as indicative rather than exact.

Model	RMSE	MAE	R^2	vs Base Features
OLS	0.07123	0.05209	-1.58%	+0.12pp
Lasso	0.07042	0.05130	0.71%	+1.30pp
Elastic Net	0.07070	0.05157	-0.08%	+0.93pp
Random Forest	0.06933	0.05113	3.77%	+4.62pp
Gradient Boosting	0.06989	0.05171	2.20%	+4.87pp

The contrast between the model families is the central result of this section. The information in the fundamental features is real, since it lifts the Random Forest from -0.85% to 3.77%, but it is not linear: OLS, which is given exactly the same fifteen variables, essentially gets nothing from them. Two properties of the earnings surprise explain this. First, the relationship between the surprise and the reaction is known to be non-linear, with an S-shaped response in which small surprises produce disproportionately large reactions and very large surprises are discounted (Kothari, 2001), and it interacts with firm characteristics such as volatility and size, which a single linear coefficient cannot represent. Second,

eps_surprise_pct is scaled by the absolute value of the consensus estimate and therefore takes extreme values when the estimate is close to zero (Figure A4); an unpenalised linear fit is pulled towards those observations, whereas a tree simply places them in a terminal region of their own. The regularised linear models sit in between, with Lasso extracting a modest 0.71% by shrinking the noisiest coefficients. The practical lesson is that adding an informative variable only helps a model that is flexible enough to use it.

Gradient Boosting, which was the weakest model with base features, is second with the full set (2.20%) and shows the largest improvement of all models (+4.87 percentage points), although its MAE (0.05171) remains the second highest, indicating that its gains come from the larger reactions rather than from uniformly smaller errors. The sequential boosting procedure, which amplifies noise when there is no signal to fit, becomes competitive once a strong signal such as the earnings surprise is present, but it still does not match the more stable averaging of the Random Forest (RMSE 0.06933 against 0.06989).

4.4.2 Feature Importance: Full Features

The importance rankings show where that improvement comes from. When the EPS-based features are added (Figure 2), the picture changes completely. eps_surprise_pct alone accounts for close to half of total importance (approximately 49%), more than the eleven base features combined, and eps_growth (approximately 7%) is the third-ranked feature. The price-based variables fall between 2% and 8% each (volatility_30d roughly 8%, pre_return_2d 6%, momentum_30d 5%, and pre_return_5d 3%). pe_ratio_hist contributes roughly 5%, and eps_actual_lag roughly 2%. This is the expected result: the sign and size of the earnings surprise are the information the market is reacting to over the event window, and a model given it will use it. Consistent with Ball and Brown (1968) and Bernard and Thomas (1989), the findings confirm that the earnings surprise is the dominant explanatory factor for CARs, with the important qualification that this variable is only known at the announcement, so Figure 2 describes what explains the reaction rather than what could have predicted it.

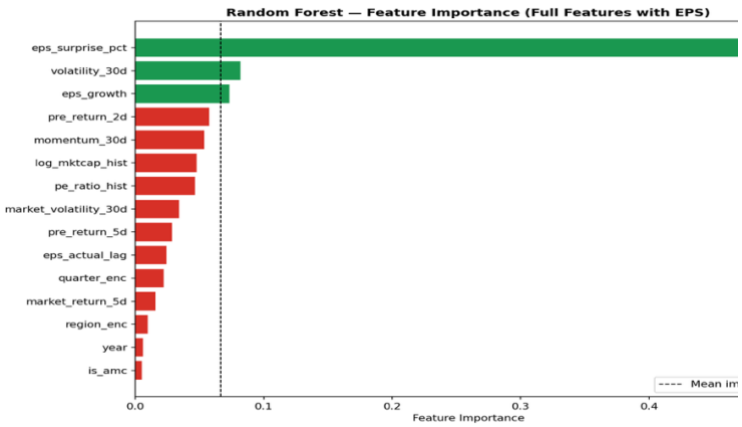


Figure 2 - Feature Importance Scores: Random Forest Model (Full Features)

4.4.3 Lasso Feature Selection

A second reading of the same question comes from the regularised linear models. The Lasso model, by design, shrinks the coefficients of irrelevant predictors to exactly zero. Applied to the full sample with base features, Lasso sets 6 of the 11 coefficients to zero after regularisation, retaining only quarter_enc, year, pre_return_2d, volatility_30d, and log_mktcap_hist, and even these with very small coefficients. With the full feature set, 8 of the 15 coefficients are zeroed, and eps_surprise_pct and eps_growth are the only fundamental variables retained. The overlap between the variables Lasso keeps and the top of the Random Forest rankings in Figures 1 and 2 shows that the two model families are reading the data in the same way, but in the base-feature case, this should not be over-interpreted: as Sections 4.3 and 4.4.1 have shown, the retained price-based variables do not translate into out-of-sample predictive power.

4.4.4 U.S. vs European Market Comparison

Table 10 presents the full-feature results by market. The U.S. subsample drives the full-sample result: Random Forest reaches 3.98% and Gradient Boosting 2.45%, while the linear models are close to zero or negative (Lasso 0.15%, Elastic Net -0.67%, OLS -2.26%). In the European subsample, Random Forest (1.47%) and Lasso (1.24%) are positive and Elastic Net marginally so (0.45%), while OLS is negative (-0.91%). These are the only meaningfully positive R^2 values obtained for European firms in this study (the 0.01% of OLS in Table 8 being indistinguishable from zero), and they appear only when the earnings surprise is

included, which confirms that the fundamental information, not the price history, is what carries the signal in both markets.

Table 10 - Out-of-Sample R^2 by Market: Full Features (Test: 2025)

Model	US R^2	EU R^2
OLS	-2.26%	-0.91%
Lasso	0.15%	1.24%
Elastic Net	-0.67%	0.45%
Random Forest	3.98%	1.47%
Gradient Boosting	2.45%	-53.73%*

* EU Gradient Boosting with full features produced a large negative R^2 due to the very small EU training sample (1,269 events with EPS data), causing model instability. This result is retained for transparency but excluded from substantive conclusions.

The European Gradient Boosting result (R^2 = -53.73%) merits specific discussion. With full features, the European training set contains only 1,269 events, an insufficient sample size for a sequential boosting algorithm with 200 estimators to generalise reliably: the model fits the training data closely and its errors on the 362 test events are almost 25% larger than those of the other models (RMSE 0.07398 against 0.05923–0.05994). The large negative R^2 is a case of model breakdown rather than a finding about European market predictability, and it is documented for methodological transparency, as an illustration of a practical limitation of boosting methods on small subsamples. The other European models are not affected: their RMSE values are the lowest in the study, reflecting the lower volatility of European CARs (Table 6).

4.4.5 Base vs Full Feature Comparison

Figure 3 presents a visual comparison of R^2 across the base and full feature sets for the five models on the full sample; the corresponding market-level figures are those of Tables 8 and 10. The comparison yields three findings. First, the fundamental features are the only source of out-of-sample predictive power in this study: with price-based features alone every bar in Figure 3 lies below zero, and with the full set the two tree-based models move clearly above it. Second, the benefit is concentrated in the non-linear models: Random Forest gains 4.62 percentage points on the full sample and 5.70 in the U.S. subsample, Gradient Boosting 4.87 and 4.82, whereas Lasso gains 1.30 and 2.15, Elastic Net 0.93 and 2.16, and OLS 0.12 and 2.04 percentage points while remaining negative everywhere. Third, the cross-market comparison rests on very uneven samples: Random Forest ranks first in both markets with

full features, but the U.S. values are higher (3.98% against 1.47%), which is consistent both with the far richer EPS coverage of U.S. quarterly reporters and with the small European training sample.

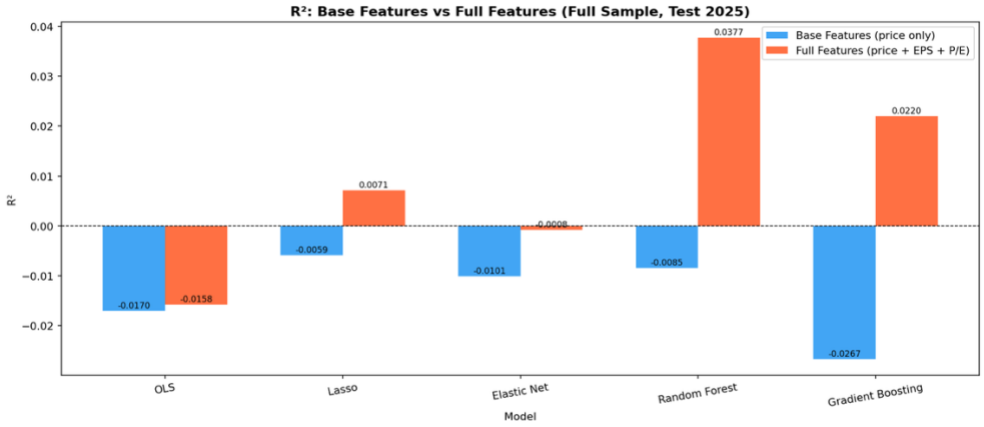


Figure 3 - Out-of-Sample R^2 Comparison: Base vs. Full Features Across Models and Markets

4.5 Economic Significance: Trading Strategy Simulation

Following the approach of Gu et al. (2020), the statistical performance of the models is translated into economic terms via a trading-strategy simulation. The simulation uses the Random Forest model with base features only so that every signal is generated from information available at the close of Day -2, before the event window opens; the full-feature models cannot be used for this purpose because the earnings surprise is known only on Day 0. Given the results of Section 4.3, the simulation should be read as a test of whether a model with no statistical predictive power can nonetheless add economic value through the direction of its predictions, rather than as an attempt to demonstrate a profitable strategy.

4.5.1 Full Sample Strategy Performance

Table 11 presents the performance of the four strategies over the 2025 test period. The model strategy generates a buy signal for 3,575 of the 3,654 test events, that is, for 98% of them: with no informative features, the Random Forest's predictions cluster around the positive training-period mean of CAR, so the rule of buying when the predicted CAR is positive amounts to buying almost everything. The strategy, therefore, earns an average CAR of -0.25% per trade with a hit rate of 49.0%, practically identical to the always-buy benchmark (-0.23% and 49.1%), which confirms that 2025 was a poor year for earnings reactions and that the model provides no filter on it.

Table 11 - Trading Strategy Performance Metrics, Full Sample (Test: 2025)

Strategy	Trades	Avg CAR/Trade	Hit Rate	Sharpe Ratio	Profit Factor
Model (RF)	3,575	-0.25%	49.0%	-0.037	0.91
Momentum Rule	2,150	-0.30%	49.3%	-0.045	0.88
Always Buy	3,654	-0.23%	49.1%	-0.036	0.91
Random	1,840	-0.02%	49.4%	-0.005	0.99

The Momentum Rule, which buys after a positive five-day pre-announcement return, does slightly worse than buying everything (-0.30% per trade, hit rate 49.3%, Profit Factor 0.88), and the Random strategy, which by construction buys half of the events without any information, happens to obtain the least negative average (-0.02%), a reminder of how much sampling noise a single test year contains. All four Sharpe ratios are negative, and all four Profit Factors are below one. There is, in short, no economic edge in any price-based signal once the information available at the close of Day -2 is all that is used.

In plain terms, an investor following the RF model's signals during 2025 would have lost, on average, €25 per €10,000 invested around each earnings announcement, compared with €23 for an investor who invested indiscriminately at every earnings event. The difference, about €2 per €10,000, is well within the noise of a single year.

4.5.2 Cross-Market Strategy Performance

Table 12 displays the performance of the strategies separately for U.S. and European events. In both markets, the model strategy again buys almost every event (1,823 of 1,878 in the U.S. and 1,752 of 1,776 in Europe), and its results are indistinguishable from the always-buy benchmark: -0.47% against -0.45% per trade in the U.S. and -0.01% against -0.01% in Europe, with essentially the same hit rates. The Momentum Rule is the weakest strategy in both markets (-0.60% per trade in the U.S., -0.03% in Europe).

Table 12 - Trading Strategy Performance by Market (Test: 2025)

Strategy	Market	Trades	Avg CAR/Trade	Hit Rate	Sharpe	Profit Factor
Model (RF)	US	1,823	-0.47%	47.7%	-0.066	0.84
Momentum Rule	US	1,038	-0.60%	47.6%	-0.082	0.80
Always Buy	US	1,878	-0.45%	47.7%	-0.063	0.84
Model (RF)	EU	1,752	-0.01%	50.4%	-0.004	0.99
Momentum Rule	EU	1,112	-0.03%	50.8%	-0.006	0.99
Always Buy	EU	1,776	-0.01%	50.5%	-0.003	1.00

The cross-market difference shown in Table 12 is between the markets, not between the strategies. European earnings events in 2025 were, on average, close to neutral (the EU always-buy strategy earns -0.01% per trade, with a hit rate of 50.5% and a Profit Factor of 1.00), whereas U.S. events destroyed value on average (-0.45%, hit rate 47.7%, Profit Factor 0.84). Whatever an investor did with price-based signals, the outcome was set by that market-wide difference in the tone of the 2025 announcement season, and the model neither improved nor worsened it in either market.

4.5.3 Portfolio Growth Simulation

Figure 4 presents the cumulative portfolio growth over the 2025 test period for each strategy, starting from an initial capital of €10,000. Monthly average net returns are computed by averaging trade returns within each calendar month, subtracting a transaction cost of 0.2% per trade, and compounding monthly. This methodology avoids the misleading sequential compounding artefact that would arise from treating all trades as independent sequential investments.

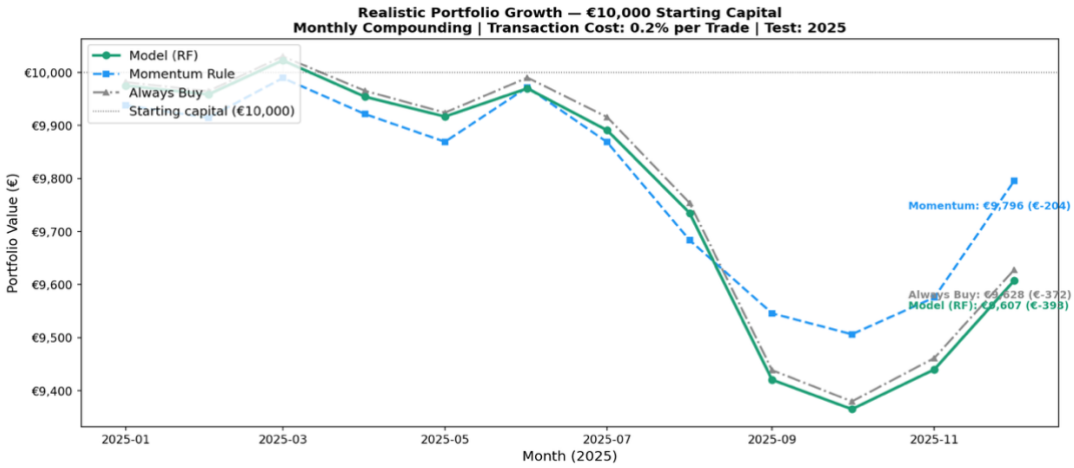


Figure 4 - Cumulative Portfolio Growth by Strategy: Combined Sample (2025 Test Period)

Over the course of the 2025 test period (Table 13), the Random Forest portfolio declined from €10,000 to €9,607 (-3.93%), marginally below the always-buy portfolio at €9,628 (-3.72%): the two lines in Figure 4 are almost superimposed throughout the year, both peaking in March, falling through the third quarter and recovering in the last two months. The Momentum Rule portfolio ends higher, at €9,796 (-2.04%), even though its average return per trade is the lowest of the three (Table 11); the difference arises from the monthly

compounding of the simulation, which gives the same weight to a month with 25 events as to one with 600: in the thin months of June, September, and December the rule's handful of trades happened to do better than the average event (in September, for example, its 15 trades averaged -1.23% against -3.03% for all 45 events), and this is enough to lift the compounded value. In the market-level simulations, the picture is the same: the U.S. model portfolio ends at €9,278 (-7.22%) against €9,315 (-6.85%) for the U.S. always-buy, and the EU model portfolio at €9,815 (-1.85%) against €9,823 (-1.77%) for the EU always-buy.

The absolute portfolio declines across the RF and always-buy strategies reflect the particularly challenging environment for earnings-based investing in 2025, in which an uninformed strategy of buying every earnings event lost 3.72% of capital over the year. The model's inability to reduce that downside is the economic counterpart of the negative R^2 values in Section 4.3: a signal that explains none of the variation in CARs cannot be expected to select the better events.

Table 13 - Portfolio Growth Simulation (Monthly Compounding, 0.2% Transaction Cost, Test: 2025)

Strategy	Starting Capital	Final Value	Net Return	vs Always Buy
Model (RF) - Full	€10,000	€9,607	-3.93%	-0.21pp
Momentum Rule - Full	€10,000	€9,796	-2.04%	+1.68pp
Always Buy - Full	€10,000	€9,628	-3.72%	-
Model (RF) - US	€10,000	€9,278	-7.22%	-0.37pp
Always Buy - US	€10,000	€9,315	-6.85%	-
Model (RF) - EU	€10,000	€9,815	-1.85%	-0.08pp
Always Buy - EU	€10,000	€9,823	-1.77%	-

Figure 5 presents the U.S. and EU portfolio growth separately. In both markets, the model and always-buy lines are almost indistinguishable, which visually reflects a model that buys nearly every event. The U.S. portfolios fall steadily from the start of the year through October, then partially recover, while the European portfolios rise above €10,000 in February and remain there until August, peaking at around €10,250 in June before giving back their gains. The average monthly net return for the RF model on the combined sample is -0.33% per month after transaction costs, compared with -0.31% for always-buy and -0.17% for the Momentum Rule.

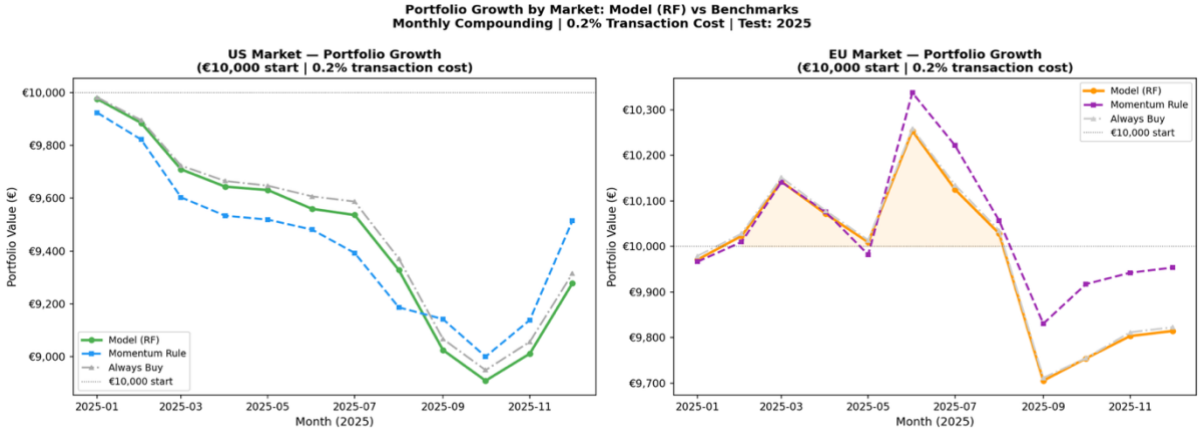


Figure 5 - Cumulative Portfolio Growth by Strategy: U.S. vs. European Subsamples (2025 Test Period)

4.6 Discussion of Findings

4.6.1 Predictive Power of ML Models

The results give a clear and, in one respect, negative answer to RQ1. When the feature set is restricted to information genuinely available before the event window, price-based variables have no out-of-sample predictive power for CAR(-1,+1): every model, in every sample, has an R^2 at or indistinguishable from zero (Tables 7 and 8), the correlations between these variables and CAR are within ±0.04 of zero (Figure A2), and a trading rule based on the model's predictions is indistinguishable from buying every event (Section 4.5). This is the outcome the semi-strong form of the Efficient Market Hypothesis (Fama, 1970) predicts, and it is worth stating plainly because the opposite result is easy to obtain by accident: a feature window that reaches into the event window borrows part of the target, and the apparent predictability that results is an artefact rather than a market inefficiency (Section 4.6.4).

The positive answer concerns the fundamental features. The earnings surprise, together with earnings growth and the P/E ratio, explains 3.77% of the out-of-sample variance of CAR with a Random Forest, 3.98% in the U.S. and 1.47% in Europe, and eps_surprise_pct alone accounts for close to half of that model's feature importance. These magnitudes are low in absolute terms, but out-of-sample R^2 values of a few per cent are the norm in return prediction (Gu et al., 2020, report values below 1% for monthly individual-stock returns), and they are obtained on a target of which the largest part is idiosyncratic noise. The feature importance analysis of Sections 4.3.3 and 4.4.2 and the correlation analysis of Figure A2 point

to the same conclusion: what moves prices over the announcement window is the content of the announcement, consistent with Ball and Brown (1968) and Bernard and Thomas (1989), and not the price behaviour that precedes it.

The answer to RQ1 is no, with one qualification. ML models cannot predict CAR(-1,+1) from information available before the event window. No model in either market achieves an out-of-sample R^2 distinguishable from zero. They can explain a modest share of the reaction once the earnings surprise is observed, but because the surprise is released at the announcement itself, that is a description of how the market processes earnings news rather than a forecast.

4.6.2 ML Models vs. Traditional Benchmarks

Whether ML models outperform OLS depends entirely on whether there is a signal to extract. With base features, the answer is no: the five models are within 2.1 percentage points of one another and all below zero (Table 7), and their ordering, with the most constrained models closest to zero, reflects how much each one over-fits noise rather than how well it predicts. With the full features, the answer is a clear yes: Random Forest reaches 3.77% against -1.58% for OLS on the full sample and 3.98% against -2.26% in the U.S. subsample, with Gradient Boosting second in both cases. The regularised linear models, Lasso (0.71%) and Elastic Net (-0.08%), sit between the two, confirming that regularisation helps a linear model survive noisy predictors but cannot supply the non-linearity that the earnings-surprise relationship requires.

These findings are consistent with Gu et al. (2020), who document that tree-based models outperform linear benchmarks in return prediction, and with the emphasis of Hou et al. (2020) and Howard (2024) on rigorous out-of-sample evaluation: the advantage of the tree-based models appears only where there is genuine information to model and vanishes where there is none, which is what one would expect of real signal extraction rather than of an implementation artefact. The comparison is not one of magnitude, since a three-day, event-specific target contains far less structured variation than the monthly cross-sectional returns studied by Gu et al. (2020), but the pattern, non-linear models extracting what linear ones cannot, is the same.

The trading strategy results add caution rather than nuance. Because the economically usable feature set is the one without predictive power, the statistical advantage of the tree-based

models does not translate into an investable edge in this study: the Random Forest strategy tracks the always-buy benchmark in every sample (Tables 11 to 13), the Momentum Rule loses less overall only because it trades less often in a poor year, not because it predicts better

The answer to RQ2 is yes, conditionally. Where a genuinely informative, non-linearly related predictor is available, the tree-based models outperform the linear benchmarks by a wide margin: Random Forest achieves an out-of-sample R^2 of 3.77% compared with -1.58% for OLS, given identical variables. Where no such predictor is available, the five models are indistinguishable from one another and none has meaningful predictive power. The advantage of flexible models is therefore real, but explanatory rather than tradeable.

4.6.3 Cross-Market Differences in Predictability

With price-based features, the cross-market comparison yields a null result in both markets. The European models are closer to zero than the U.S. ones (Table 8), but a model that explains 0.01% of the variance is no more useful than one that explains -1.72%, and the gap is largely explained by the larger fall in average U.S. reactions between the training period and 2025 (Section 4.3.2) rather than by any European predictability. The trading simulations confirm this: the model strategy is indistinguishable from always-buy in both markets, and the difference in average returns between the EU (-0.01% per trade) and U.S. (-0.45%) strategies reflects the tone of the 2025 announcement season in each market, not the informativeness of the signal.

With the fundamental features, a difference appears: Random Forest explains 3.98% of the variance in U.S. CARs but 1.47% in European CARs, and Gradient Boosting, which works in the U.S. (2.45%), breaks down in Europe. Two explanations coexist and cannot be separated with the present data. The first is institutional. A broad body of cross-country evidence establishes that legal and reporting frameworks shape the properties of earnings information (Bushman & Piotroski, 2006; Landsman et al., 2012), and the semi-annual reporting convention of many European firms means that a quarterly EPS surprise is a less standard and less closely followed signal in Europe than in the U.S., where quarterly guidance, analyst coverage and consensus estimates are denser.

The second explanation is statistical. The European full-feature training set contains 1,269 events against 6,651 in the U.S., and the EPS data that exist for European firms are less complete. With five times fewer observations, a Random Forest has less opportunity to learn

the same non-linear relationship, so part of the gap may simply be a sample-size effect. The direction of the difference is consistent with Drobetz and Otto (2021), who find ML-based predictability in European equities to be present but weaker than in U.S. studies, though its magnitude should not be overinterpreted.

The answer to RQ3 is no, not on this evidence. With information available before the event window, there is nothing to compare, since neither market is predictable. With the earnings surprise included, the U.S. point estimate exceeds the European one (3.98% versus 1.47%), but the European figure is based on 362 test events and cannot distinguish a weaker relationship from no relationship at all. The cross-market question, therefore, remains open, and settling it would require European coverage comparable to that available for U.S. firms.

4.6.4 Limitations

Several limitations of this study should be acknowledged, and one methodological point deserves emphasis because it changed the results. First, the information-timing convention. A preliminary specification of this study computed the price-based features up to the close of Day -1 so that the return of Day -1 entered both the predictors and the target CAR(-1,+1). That specification produced positive out-of-sample R^2 values with base features, ranging from 1.1% to 1.5% for four of the five models in the European subsample, a dominant pre-announcement return feature, and an apparent trading edge over the always-buy benchmark. All of these disappeared when the windows were moved to end at the close of Day -2. The reason is mechanical: the Day -1 return has a correlation of 0.23 with CAR(-1,+1) but of 0.01 with CAR(0,+1), so the apparent signal was the overlap itself, and even a five-day window that merely contained Day -1 inherited enough of it (a correlation of 0.10 with CAR) to generate the pattern. The lesson, which applies beyond this dissertation, is that an event-window target requires every predictor to be measured strictly before the window opens. Second, the EPS coverage asymmetry (62% of events have EPS data, and only about half have all four fundamental features, with an approximate 5:1 U.S.-to-European ratio in the full-feature training set, 6,651 versus 1,269 events) means that full-feature results may reflect U.S.-dominated patterns rather than genuinely cross-market findings. Third, the full-feature models use the earnings surprise, which is known only at the announcement, so their explanatory power is not a forecast, and they cannot be used in the trading simulation. Fourth, the trading strategy simulation adopts a stylised framework that does not account for short-

selling constraints, portfolio concentration limits, or market impact costs beyond the 0.2% transaction cost assumption. Fifth, the study covers a single out-of-sample year (2025), and results may be sensitive to the specific market conditions prevailing during this period. Sixth, the study uses only the CAR(-1,+1) event window; robustness checks with alternative windows (e.g., [0,+1] or [-3,+3]) would strengthen the generalisability of the findings. Finally, the negative R^2 reported by the Gradient Boosting model on the EU full-feature sample (R^2 = -53.73%) is retained for transparency, but should be interpreted as a case of model breakdown under small-sample conditions rather than as a meaningful finding about European market predictability.

5. Conclusions

5.1 Summary of Findings

The central question of this dissertation was whether machine learning models can predict cumulative abnormal returns around earnings announcements and whether predictive relationships differ across U.S. and European equity markets. Chapter 4 answers its three research questions as follows: No, ML models cannot predict CARs from information available before the event window; yes, they outperform the linear benchmarks, but only where an informative and non-linearly related predictor is available, and that advantage is explanatory rather than tradeable; and no, not on this evidence, since the two markets cannot be reliably distinguished. Taken together, the announcement reaction is not forecastable from prior price behaviour in either market, while the content of the announcement explains a modest share of it once released, a result about how markets process earnings news rather than about their predictability.

5.2 Theoretical and Empirical Contributions

This dissertation makes three contributions. First, it extends the ML return-predictability literature (Gu et al., 2020; Chen et al., 2022) into an event-driven setting and shows that the advantage of tree-based methods over linear benchmarks documented in broad cross-sectional studies also holds for a narrow announcement-window return, but only where a genuinely informative, non-linearly related predictor is available. The out-of-sample evaluation framework and the time-based train/test split, with 2020–2024 as the training set and 2025 as the held-out test set, provide a disciplined test that guards against the overfitting concerns raised in the literature.

Second, the cross-market comparison provides new evidence that the institutional environment shapes the informativeness of earnings announcements (Bushman & Piotroski, 2006; Landsman et al., 2012) within an ML framework: the tree-based models, given the same variables, explain more of the reaction of U.S. firms than of European firms, and price-based information explains essentially none of it in either market. The finding that the price-based cross-market difference obtained by a preliminary specification was an artefact of

feature-window overlap is itself a contribution, as it documents a pitfall that is easy to fall into when ML methods are applied to event-window targets.

Third, the trading strategy simulation translates statistical findings into a practically interpretable metric. By comparing RF-based trading performance against three benchmarks, always-buy, Momentum Rule, and Random selection, the dissertation shows that a model with no out-of-sample predictive power produces no economic value either, providing a replicable framework for evaluating the economic significance of ML predictions in event-driven settings and a benchmark against which future, more informative signals can be assessed.

5.3 Practical Implications

For market participants, the results carry a cautionary message: short-term pre-announcement price behaviour, however cheap to monitor, does not help to select the earnings events that will generate positive abnormal returns in either the U.S. or Europe. Where value exists, it lies in the analysis of earnings expectations, since the size and sign of the surprise relative to the consensus explain a measurable share of the reaction, especially in the U.S., and in the ability of non-linear models to represent that relationship. Strategies that seek to profit from earnings announcements, therefore, need a genuine informational advantage about the surprise itself, not a better reading of the price history.

For researchers, the results reinforce the importance of rigorous out-of-sample validation, of multiple benchmarks, and, above all, of a strict information-timing convention when the target is an event-window return. The performance ranking of models in the full-feature setting, with RF outperforming Gradient Boosting, Lasso, Elastic Net and OLS in both the full sample and the U.S. subsample, suggests that the predictive advantage of tree-based methods in this setting reflects genuine signal extraction rather than overfitting to within-sample noise or implementation-specific artefacts of the kind documented by Howard (2024). At the same time, the lack of predictive power of price-based features serves as a reminder that earnings-announcement returns remain difficult to forecast, even with flexible nonlinear methods.

5.4 Limitations and Directions for Future Research

The primary limitations of this study are discussed in Section 4.6.4. Three directions for future research follow directly from them. First, the fundamental features that carry the signal in this study are known only at the announcement; future work could replace the realised surprise with pre-announcement analyst forecast revisions or the dispersion of analyst estimates to test whether any part of the fundamental relationship is forecastable before Day 0. Second, quarterly EPS data cover almost all U.S. events but only about a fifth of European ones, reflecting semi-annual reporting conventions. The European full-feature training sample is therefore 1,269 events against 6,651 in the U.S., which limits the fundamental-based analysis in Europe, alternative measures of earnings expectations would address this. Third, the 2020–2025 sample period includes the COVID-19 pandemic and subsequent market dislocations, and the out-of-sample evaluation rests on a single year, raising questions about the stability of the estimated relationships across different macroeconomic regimes. Extending the analysis to a longer historical sample and to alternative event windows, such as CAR(0,+1), would allow for more formal tests of predictive stability.

5.5 Concluding Remarks

This dissertation set out to determine whether machine learning models can predict abnormal stock price movements around earnings announcements and whether those relationships hold across institutional environments. The evidence is that the reaction cannot be forecast from price information available before the announcement in either market, but that part of it can be explained by the content of the announcement once released, and mainly by models flexible enough to capture the non-linear relationship between the earnings surprise and the reaction. Perhaps the most useful lesson is methodological: in an event-driven setting, apparent predictability is only as credible as the timing of the information used to obtain it.

References

Alvarado, J. I., Clark, L. C., & Gutierrez, J. A. (2021). Stock performance subsequent to combinations in quarterly revenue surprise, earnings surprise, guidance, valuation, and report time. *Journal of Economics and Finance, 45*(1), 95–117. https://doi.org/10.1007/s12197-020-09531-7

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research, 6*(2), 159–178. https://doi.org/10.2307/2490232

Bartov, E., Givoly, D., & Hayn, C. (2002). The rewards to meeting or beating earnings expectations. *Journal of Accounting and Economics, 33*(2), 173–204. https://doi.org/10.1016/S0165-4101(02)00045-9

Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research, 27*, 1–36. https://doi.org/10.2307/2491062

Bernard, V. L., & Thomas, J. K. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics, 13*(4), 305–340. https://doi.org/10.1016/0165-4101(90)90008-R

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

Bryzgalova, S., Pelger, M., & Zhu, J. (2025). Forest through the trees: Building cross-sections of stock returns. *The Journal of Finance, 80*(5), 2447–2506. https://doi.org/10.1111/jofi.13477

Bushman, R. M., & Piotroski, J. D. (2006). Financial reporting incentives for conservative accounting: The influence of legal and political institutions. *Journal of Accounting and Economics, 42*(1–2), 107–148. https://doi.org/10.1016/j.jacceco.2005.10.005

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

Choi, H. M. (2019). Market uncertainty and trading volume around earnings announcements. *Finance Research Letters, 30*, 14–22. https://doi.org/10.1016/j.frl.2019.03.002

Dechow, P. M., & Dichev, I. D. (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review, 77*(s-1), 35–59. https://doi.org/10.2308/accr.2002.77.s-1.35

Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: Evidence from the European stock market. *Journal of Asset Management, 22*(7), 507–538. https://doi.org/10.1057/s41260-021-00237-x

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383–417. https://doi.org/10.2307/2325486

Fink, J. (2021). A review of the post-earnings-announcement drift. *Journal of Behavioral and Experimental Finance, 29*, 100446. https://doi.org/10.1016/j.jbef.2020.100446

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Harvey, C. R., Liu, Y., & Zhu, H. (2016). …and the cross-section of expected returns. *The Review of Financial Studies, 29*(1), 5–68. https://doi.org/10.1093/rfs/hhv059

Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance, 55*(1), 265–295. https://doi.org/10.1111/0022-1082.00206

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies, 33*(5), 2019–2133. https://doi.org/10.1093/rfs/hhy131

Howard, C. (2024). Choices matter when training machine learning models for return prediction. *Financial Analysts Journal, 80*(4), 81–107. https://doi.org/10.1080/0015198X.2024.2388024

Kaiser, D. (2019). *Machine learning based financial statement analysis predicting the market reaction to earnings announcements* (MSc by Research thesis, University of Oxford). https://ora.ox.ac.uk/objects/uuid:a0aa6a5a-cfa4-40c0-a34c-08f779bb4bda

Kim, J., Schonberger, B., Wasley, C. E., & Yang, Y. (2024). Forecasting market volatility: The role of earnings announcements. *The Accounting Review, 99*(4), 251–279. https://doi.org/10.2308/TAR-2021-0351

Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics, 31*(1–3), 105–231. https://doi.org/10.1016/S0165-4101(01)00030-1

Landsman, W. R., Maydew, E. L., & Thornock, J. R. (2012). The information content of annual earnings announcements and mandatory adoption of IFRS. *Journal of Accounting and Economics, 53*(1), 34–54. https://doi.org/10.1016/j.jacceco.2011.04.002

Livnat, J., & Mendenhall, R. R. (2006). Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research, 44*(1), 177–205. https://doi.org/10.1111/j.1475-679X.2006.00196.x

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature, 35*(1), 13–39. https://www.jstor.org/stable/2729691

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

Schnaubelt, M., & Seifert, O. (2020). *Valuation ratios, surprises, uncertainty or sentiment: How does financial machine learning predict returns from earnings announcements?* (FAU Discussion Papers in Economics No. 04/2020). Friedrich-Alexander-Universität Erlangen-Nürnberg.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review, 71*(3), 289–315. https://www.jstor.org/stable/248290

Thakerngkiat, N., Nguyen, H. T., Nguyen, N. H., & Visaltanachoti, N. (2021). Do accounting information and market environment matter for cross-asset predictability? *Accounting & Finance, 61*(3), 4389–4434. https://doi.org/10.1111/acfi.12736

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Veganzones, D., & Séverin, E. (2025). Earnings management visualization and prediction using machine learning methods. *International Journal of Accounting Information Systems, 56*, 100743. https://doi.org/10.1016/j.accinf.2025.100743

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Annex

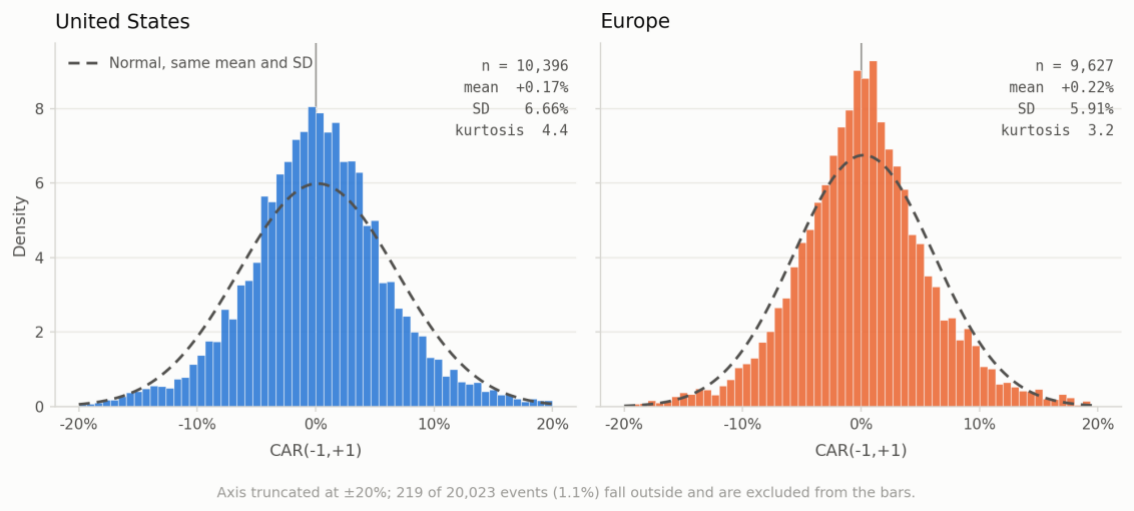


Figure A1 – Distribution of CAR(-1,+1) by market

Density histograms of cumulative abnormal returns over the three-day event window, for 10,396 U.S. and 9,627 European earnings announcements between 2020 and 2025. The dashed line is a normal distribution with the same mean and standard deviation as the observed data. Both distributions are more sharply peaked and more fat-tailed than the normal benchmark, with excess kurtosis of 4.4 and 3.2 respectively. The axis is truncated at ±20%; 219 events (1.1%) fall outside it and are excluded from the bars.

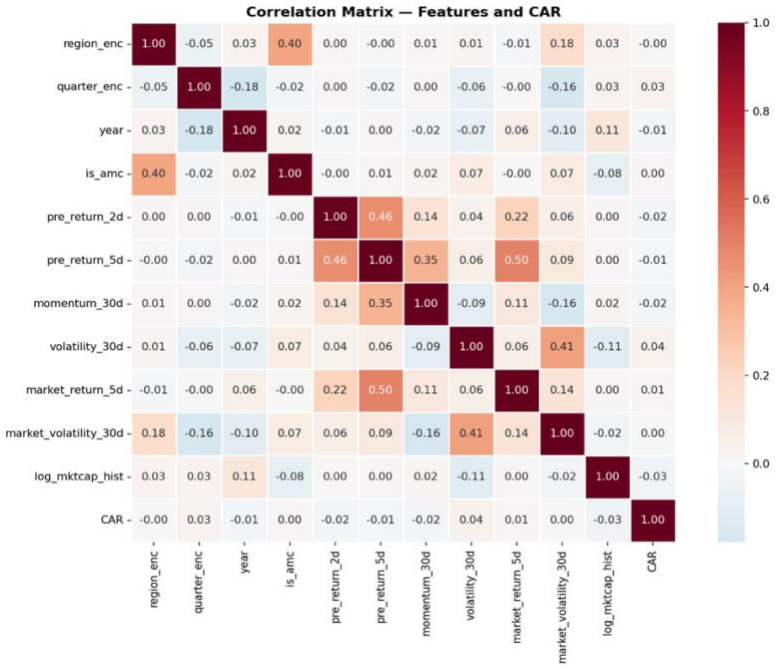


Figure A2 – Correlation matrix: base features and CAR(-1,+1)

Pearson correlation coefficients among the eleven base features and the target variable, computed on the full sample of 20,023 earnings announcements. For the binary indicators region_enc and is_amc the coefficient is a point-biserial correlation; quarter_enc is ordinal, so its coefficient assumes equal spacing between quarters and should be read as indicative only.

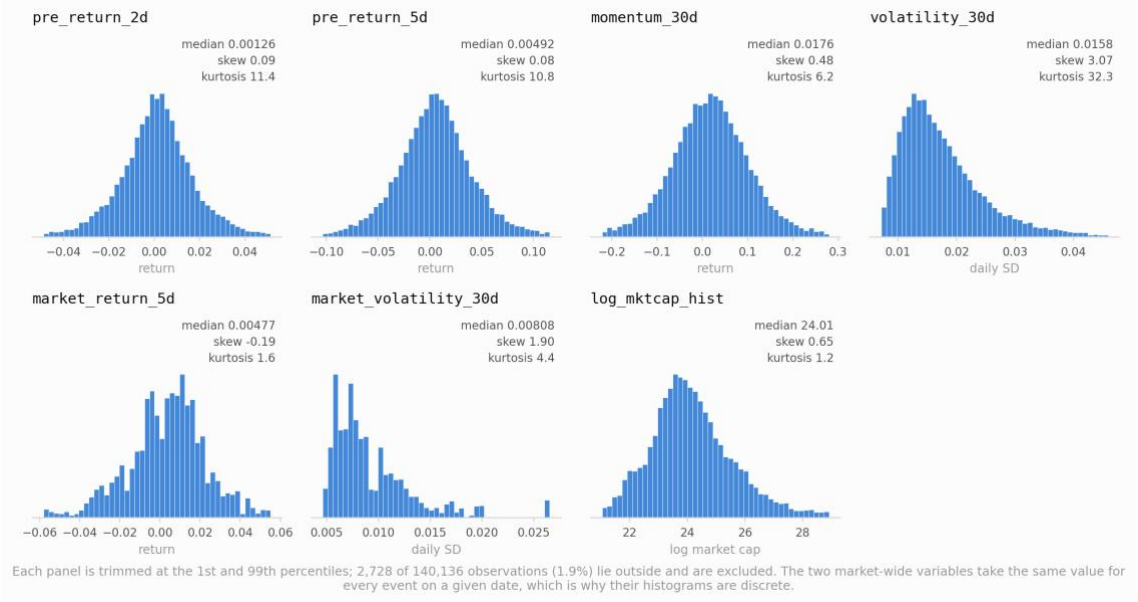


Figure A3 – Distribution of the base features

Histograms of the continuous base features across 20,023 earnings announcements, 2020–2025. Each panel is trimmed at the 1st and 99th percentiles; the median, skewness and excess kurtosis reported in each panel are computed on the untrimmed series. The two market-wide variables take a single value per date across all events, which is why their histograms are discrete.

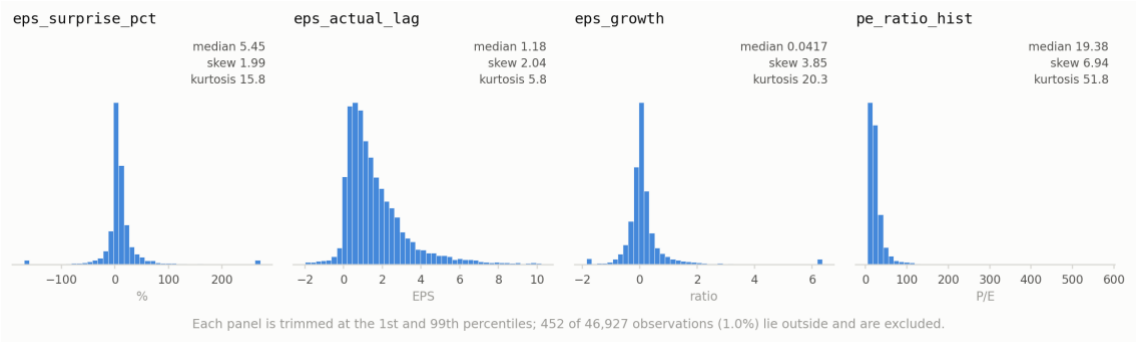


Figure A4 – Distribution of the fundamental features

Histograms of the four fundamental features, each on the subsample for which it is available. Each panel is trimmed at the 1st and 99th percentiles; the median, skewness and excess kurtosis are computed on the untrimmed series. All four are strongly right-skewed with heavy tails, most markedly eps_surprise_pct, which is scaled by the absolute value of the consensus estimate and therefore takes extreme values when that estimate is close to zero.