Master in Management

Major in International Finance

Academic Year 2025–2026

Master Research Paper

Predicting Acquisition Quality in Europe: Machine Learning Evidence on Cross-Sectional Returns

Keywords: Mergers and acquisitions, machine learning, post-acquisition performance, European equity markets, long-run abnormal returns

Luca LIAUBET-PALOMO

Ahmed Bahaeddine LOUATI

Under the Supervision of: Prof. François DERRIEN

President of the Jury: Prof. Takaya SEKINE

1st June 2026

Acknowledgments

We are grateful to Professor François Derrien for his guidance and valuable comments, which substantially improved the analysis.

We thank Professor Takaya Sekine for serving as President of the Jury and for his time and attention in evaluating this work.

We are grateful to HEC Paris and the Master in International Finance program for providing the academic environment, resources, and institutional support that made this research possible.

We acknowledge LSEG for access to the Deals database and Compustat Global for the accounting data used in this study. Fama-French and momentum factors for European markets were obtained from the Kenneth R. French Data Library. Equity return data were obtained from Refinitiv Datastream.

The authors used AI language model assistance during the preparation of this research paper for editing, proofreading, and formatting purposes. All empirical analysis, data collection, model specification, econometric implementation, interpretation of results, and substantive writing reflect the authors' own work. The authors take full responsibility for the content of this paper.

Abstract

This study examines whether machine learning models can predict the cross-sectional variation in long-run post-acquisition returns of European acquirers. Using 6,540 transactions (1993–2024), seven models trained on pre-2018 data are evaluated on 1,210 out-of-sample deals. Despite negative out-of-sample R^2, cross-sectional ranking ability is preserved: Spearman ρ reaches +0.272, the Q5-Q1 spread +35.4 percentage points (t=8.36), and an illustrative Fama-French four-factor alpha of +22.82% per year (t=5.42). SHAP analysis identifies Euribor and acquirer profitability as the most influential predictors, with a distinctive nonlinear interaction between them. All results are exploratory given the short out-of-sample window and limited independent return realizations.

Contents

1 Introduction

Mergers and acquisitions represent one of the most consequential capital allocation decisions a firm can make. In Europe, acquirers collectively deploy several hundred billion euros annually in acquisition activity, yet the empirical record is sobering: acquiring firm shareholders earn, on average, returns that are close to zero or modestly negative in the two years following announcement. The puzzle is not that acquisitions fail on average, it is that outcomes are enormously dispersed. Some acquisitions generate sustained outperformance; others destroy a third or more of market value. If even a fraction of this cross-sectional variation is predictable from information available at announcement, the implications for capital allocation, deal screening, and investment strategy are substantial. This study asks whether machine learning models trained on observable deal, acquirer, and macroeconomic characteristics can identify, before the fact, which acquisitions are likely to outperform their peers.

The prediction problem is inherently difficult. Long-run abnormal returns are noisy, and markets process publicly available information rapidly at announcement. The determinants of acquisition quality have been extensively documented in the theoretical and empirical literature, but their joint dynamics are unlikely to be linear or additive. With 154 candidate predictors spanning deal structure, acquirer fundamentals, and macroeconomic conditions, conventional regression-based approaches face both dimensionality and misspecification challenges. Campbell et al. (2025) demonstrate that machine learning methods can overcome these challenges in the US setting, generating positive out-of-sample quintile spreads where linear benchmarks fall short. No equivalent analysis exists for Europe. European M&A spans heterogeneous governance systems, from common law in the United Kingdom to civil law in France and Germany and Nordic models in Sweden, that have no close American parallel. The monetary environment is equally distinctive: the three-month Euribor entered negative territory in 2015, remained there until 2022, and was followed by the sharpest European rate-hiking cycle in four decades.

This study addresses these questions using 6,540 completed acquisitions by publicly listed European acquirers announced between January 1993 and December 2024. The dependent variable is the 24-month CAPM-adjusted buy-and-hold abnormal return computed relative to the STOXX Europe 600, a methodological refinement that removes the confound introduced by systematic variation in acquirer market sensitivity. Seven supervised learning models are estimated on transactions announced before 2018 and evaluated on the 1,210 deals announced between 2018 and 2024, a test period encompassing the COVID-19 shock, persistent inflation, and an unprecedented European monetary tightening cycle.

Four main results emerge. First, all seven models generate negative out-of-sample R^2 values, with XGBoost producing the least negative at −0.044. This reflects the severity of the macro regime shift between training and test periods rather than an absence of cross-sectional signal: when the common annual component is removed, R^2 turns positive for all tree-based models. Second, Spearman rank correlations are positive for all seven models, with XGBoost reaching +0.272. The models retain meaningful ranking ability even when level forecasts are systematically downward-biased. Third, the Q5-Q1 quintile spread reaches +35.4 percentage points (t=8.36), is positive in all six annual cohorts with a complete outcome window, and strengthens to more than +43 percentage points during the 2022-2023 rate-hiking years. Fourth, a calendar-time long-short strategy generates a Fama-French four-factor alpha of +22.82% per year (t=5.42), with economically negligible factor loadings.

A SHAP decomposition identifies the three-month Euribor as the most influential predictor, followed by acquirer return on assets and pre-announcement return volatility. Their interaction is one of the study's most distinctive findings: accommodative monetary conditions amplify the underperformance of weak acquirers in a way that neither variable predicts in isolation, consistent with the view that negative rates relaxed the financing constraints that would otherwise have prevented marginal acquisitions.

At its core, this study documents that cross-sectional ranking ability in European M&A survives severe macro regime shifts, driven by a nonlinear interaction between financing conditions and acquirer quality. Four more specific contributions follow. To our knowledge, it provides the first large-scale ML-based analysis of long-run post-acquisition performance in Europe. It contributes methodologically through CAPM-adjusted return measurement, SHAP-based economic interpretation, and the documentation of the Euribor × ROA interaction as a novel empirical mechanism. Finally, the persistence of the FF4 alpha is consistent with a slow and partial incorporation of acquisition quality information into European equity prices, and may have implications for the asset pricing literature on corporate events. The remainder of the paper is organized as follows. Section 2 reviews the relevant literature and states the research hypotheses. Section 3 describes the data, variable construction, and empirical methodology. Section 4 presents and interprets the empirical results. Section 5 concludes.

2 Literature Review and Research Hypotheses

2.1 Post-M&A Performance and the Acquirer Return Puzzle

A large body of evidence documents that acquisitions generate, on average, modest or negative long-run returns for acquiring firm shareholders, while targets capture most of the announced synergies through the acquisition premium. Jensen and Ruback (1983) and Andrade, Mitchell and Stafford (2001) establish this pattern across three decades of US data: combined announcement returns are positive, but acquirer abnormal returns remain small and often statistically indistinguishable from zero. Short-term announcement returns may therefore provide an incomplete picture of acquisition quality when integration outcomes unfold gradually over time, a consideration that motivates the 24-month BHAR used in this study.

These averages mask substantial heterogeneity in outcomes. Loughran and Vijh (1997) show that acquirers underperform matched control firms by an average of 15.9% over five years following completion, with the penalty concentrated almost entirely in stock-financed transactions. Rau and Vermaelen (1998) show that underperformance is concentrated among high price-to-book acquirers. Moeller, Schlingemann and Stulz (2005) document that the 1998-2001 merger wave destroyed more than 200 billion dollars of acquirer shareholder value, concentrated in a small number of very large deals. This asymmetry in outcomes suggests that modelling acquisition performance as a homogeneous linear process may obscure economically important tail dynamics. Using a large-scale meta-analysis, King et al. (2021) confirm that no single variable consistently explains post-acquisition outcomes across settings, consistent with the hypothesis that predictive information is distributed across many interacting characteristics.

The European literature confirms similar patterns alongside greater institutional complexity. Goergen and Renneboog (2004), Martynova and Renneboog (2008), and Renneboog and Vansteenkiste (2019) document acquirer underperformance across European markets, with payment method, governance, and deal relatedness emerging as the most consistent predictors. Institutional heterogeneity across governance systems, ownership structures, and financing environments may amplify informational frictions, potentially increasing the persistence of cross-sectional variation in outcomes relative to more integrated capital markets.

The literature therefore identifies a tension: many determinants of acquisition quality are individually well documented, yet post-acquisition performance remains difficult to forecast in practice. Acquisition outcomes likely depend less on isolated variables than on complex interactions between firm characteristics, financing conditions, and macroeconomic regimes, interactions that may be difficult to capture within conventional linear specifications.

2.2 Theoretical Determinants of Acquisition Quality

Three broad mechanisms explain why acquirer quality, financing conditions, and observable deal characteristics jointly determine post-acquisition performance.

Agency and behavioral motives. Jensen's (1986) free cash flow hypothesis argues that managers disposing of cash in excess of productive investment opportunities deploy it in value-reducing ways, including acquisitions. Firms with weak profitability are therefore more likely to pursue acquisitions as a substitute for deteriorating organic performance. Roll (1986) proposes a complementary mechanism: managerial hubris leads acquirers to overestimate synergies and overpay for targets. Malmendier and Tate (2008) document this empirically, showing that overconfident CEOs make more acquisitions and generate more negative market reactions. Pre-announcement return volatility may partially proxy for the uncertainty surrounding acquirers operating from a position of weakness.

Financing conditions and monetary policy. Myers and Majluf (1984) show that managers who consider their equity overvalued prefer stock financing, conveying a negative signal to outside investors. At the aggregate level, Harford (2005) documents that merger waves emerge when industry shocks meet sufficient capital liquidity: cheap financing lowers the hurdle rate and makes marginal deals feasible. Bernanke and Gertler (1989) formalize this mechanism, showing that easy monetary conditions reduce the screening intensity applied to marginal investment projects. Altavilla et al. (2022) document this channel empirically for European firms under negative policy rates. Relative deal size, leverage, valuation multiples, and pre-announcement market dynamics also carry information about post-acquisition integration risk. The interaction between financing conditions and acquirer fundamentals that emerges from this literature is precisely what Section 4.3 documents: accommodative monetary conditions appear to amplify the underperformance of weak acquirers in a way that neither variable predicts in isolation.

The literature identifies these determinants individually, but they combine in ways that are difficult to represent within conventional linear specifications, particularly when candidate predictors are numerous and the consequences of financing conditions may differ sharply once acquirer profitability falls below certain thresholds. Recent advances in machine learning are particularly well suited to this type of high-dimensional, interaction-rich environment.

2.3 Machine Learning and Prediction in Finance and M&A

Mullainathan and Spiess (2017) draw a foundational distinction between causal inference and prediction, arguing that machine learning methods are particularly valuable when the functional form relating predictors to outcomes is unknown and the number of candidate variables is large. Gu, Kelly and Xiu (2020) apply this logic to equity return prediction, showing that nonlinear tree-based methods capture interaction effects that linear factor models miss and generate economically meaningful out-of-sample gains. Giglio, Kelly and Xiu (2022) synthesize this literature, documenting that flexible ML methods consistently outperform linear benchmarks in high-dimensional financial prediction tasks. In such environments, the objective is often less to predict exact outcome levels than to discriminate reliably between relatively strong and relatively weak observations. Bao et al. (2020) extend this logic to corporate events, showing that ML models trained on standard accounting variables substantially outperform logistic regression in detecting financial statement fraud, a setting structurally similar to post-acquisition performance prediction in terms of dimensionality and low signal-to-noise ratio.

The M&A prediction problem exhibits many of these characteristics simultaneously. This study draws on 154 variables spanning deal structure, acquirer fundamentals, and macroeconomic conditions, a dimensionality at which ordinary least squares may struggle to extract stable cross-sectional relationships and standard variable selection approaches may fail to preserve predictors whose informational value emerges primarily through interactions with other variables. The relevant relationships are also plausibly nonlinear: the economic consequences of moving from negative to positive acquirer profitability are unlikely to be proportional across the full distribution of acquirer quality. Tree-based methods are particularly well suited to these features because they capture nonlinearities and interaction effects without requiring them to be specified ex ante. Recent advances in model interpretability further allow these methods to be linked back to economically meaningful mechanisms rather than treated as purely predictive black boxes, a dimension explored directly in Section 4.3.

Campbell et al. (2025) demonstrate the feasibility of this approach for US M&A data. Their study serves as the direct benchmark for this research. Whether similar cross-sectional predictability exists in Europe, where institutional fragmentation, governance heterogeneity, and a distinctive monetary policy cycle introduce additional complexity, remains an open empirical question.

The objective of this study is therefore explicitly predictive rather than causal: the question is not why certain acquisitions outperform, but whether observable announcement-date characteristics contain sufficient information to rank deals by expected quality.

2.4 European Context, Market Frictions, and Research Hypotheses

Faccio and Masulis (2005) show that the choice of payment method in European M&A depends on governance systems, ownership concentration, and legal traditions in ways that have no direct American equivalent. Rossi and Volpin (2004) document that acquisition activity and premia vary substantially with investor protection quality, motivating the inclusion of country-level controls. Erel, Liao and Weisbach (2012) establish that cross-border deal flows depend on institutional proximity and information availability, consistent with the possibility that cross-border acquisitions generate stronger cross-sectional predictability. This diversity motivates the country-specific analysis presented in Section 4.4.

The European monetary policy cycle introduces an additional layer of complexity. The three-month Euribor entered negative territory in 2015 and remained there until 2022, a sustained environment with no close American parallel, creating conditions in which financing constraints were relaxed precisely for the marginally creditworthy acquirers that Jensen (1986) and Bernanke and Gertler (1989) predict will make value-destructive acquisitions. Shleifer and Vishny (1997) and Mitchell, Pulvino and Stafford (2002) establish that informational and operational frictions allow predictable patterns in returns to persist: the analytical infrastructure required to process a large number of interacting variables at scale became economically and operationally accessible only relatively recently, and the 24-month outcome horizon extends well beyond the evaluation cycles of most active managers. A detectable cross-sectional signal in post-acquisition returns is therefore not necessarily inconsistent with a broadly functional market.

Three testable hypotheses follow from these arguments.

H1: Machine learning models trained on observable announcement-date characteristics generate positive and statistically significant Spearman rank correlations on out-of-sample European M&A data. Flexible ML methods may be better suited to extracting information from high-dimensional, interaction-rich data than conventional linear specifications.

H2: The ranking ability of the model translates into economically meaningful differences in subsequent acquisition performance, reflected in statistically significant Q5-Q1 BHAR spreads, particularly when macro-financial conditions amplify the consequences of acquirer quality.

H3: The calendar-time long-short alpha generated by the ranking strategy is not fully absorbed by the Fama-French four-factor model, consistent with the limits-to-arbitrage framework of Shleifer and Vishny (1997).

The following sections describe the data construction, empirical design, and validation framework used to evaluate these hypotheses.

3 Data and Methodology

3.1 Sample Construction and BHAR

This study examines the long-run stock market performance of European acquirers in transactions announced between January 1993 and December 2024. Transaction-level data are drawn from LSEG Deals; accounting variables are collected from Compustat Global; equity returns from Datastream; and European Fama-French and momentum factors from the Kenneth R. French Data Library.

The sample retains completed acquisitions by publicly listed European acquirers with a disclosed deal value of at least USD 1 million. After applying these filters, the LSEG universe contains 6,540 acquisitions spanning 31 years of European M&A activity. Accounting variables and equity return data are merged at the deal level, with observations requiring sufficient pre-announcement return history for beta estimation. This 31-year span across multiple monetary regimes is a meaningful advantage relative to US-focused studies: the European setting combines heterogeneous institutional environments, bank-dominated financing structures in some markets, and a monetary policy cycle with no direct American equivalent.

The dependent variable is the acquiring firm's 24-month CAPM-adjusted buy-and-hold abnormal return (BHAR), measured from the month following the announcement date. For each acquirer, the market beta is estimated over the 36 pre-announcement months using European market returns, and expected monthly returns are computed as:

$$E\left(R_{i,t}\right) = Rf_t + \beta_i(Rm_t - Rf_t) \tag{1}$$

where $Rf_{,t}$ is the monthly European risk-free rate and $Rm_{,t}$ the STOXX Europe 600 return. Monthly abnormal returns are then compounded over the 24-month horizon. The CAPM adjustment prevents highly cyclical acquirers from mechanically appearing to outperform during bull markets regardless of acquisition quality, a confound that simple market-adjusted benchmarks cannot remove. The specification also constitutes a methodological refinement relative to Campbell et al. (2025), whose benchmark relies on market-adjusted returns without controlling for firm-level systematic risk.

All predictors used in the models are constructed exclusively from information observable at or before the announcement date; the 24-month BHAR is the outcome to be predicted, not an input.

The sample requires return data covering the full 24-month post-announcement horizon. Acquirers that delist within 24 months of announcement (due to bankruptcy, subsequent acquisition, or voluntary delisting) are excluded from the estimation sample. To the extent that early-delisting firms are disproportionately concentrated among weaker acquirers, this treatment may introduce a conservative bias in Q1 BHAR estimates. The precise number of excluded observations and their predicted quintile distribution are not reported.

The estimation dataset is partitioned chronologically: transactions announced between 1993 and 2017 form the training set (5,330 observations, approximately 81% of the estimation sample), and those announced between 2018 and 2024 constitute the out-of-sample test period (1,210 observations). The primary year-by-year analysis is limited to the six annual cohorts with a complete 24-month outcome window (2018–2023). The 18 transactions announced in 2024 are retained in the dataset and contribute to the rolling-window analysis, but their incomplete horizon is acknowledged as a limitation. A random partition would allow information from future market environments to contaminate the estimation process, artificially inflating apparent predictive performance. As documented in Section 3.3, average acquirer BHAR shifted from +2.7% during the training period to −10.7% during the test period, a gap of approximately 13 percentage points driven by the COVID-19 shock, inflationary pressures, and the sharpest European rate-hiking cycle in four decades. Models trained on a period with mildly positive average acquisition returns must therefore demonstrate cross-sectional ranking ability in a fundamentally altered macroeconomic environment, a more demanding test of out-of-sample generalization than a stable-regime evaluation would provide.

3.2 Explanatory Variables

The predictive framework draws on 154 variables observable at announcement, grouped into three economically motivated categories. Accounting variables are matched to the latest fiscal year-end available before announcement.

Deal characteristics capture the structure, scale, and strategic nature of the transaction. Payment method is central: following Myers and Majluf (1984), stock-financed acquisitions may signal equity overvaluation, while cash financing reflects managerial conviction in the deal's intrinsic value. Variables include cash and stock payment shares, deal value (log), relative deal size, and indicators for hostile bids, tender offers, competing offers, cross-border transactions, and same-industry deals.

Acquirer characteristics measure financial condition and market positioning. Following Jensen's (1986) free cash flow hypothesis, weak profitability may incentivize acquisitions as a substitute for deteriorating organic performance, while financially stronger acquirers are better positioned to identify and integrate value-creating targets. Variables include return on assets, EBITDA margin, leverage, market-to-book ratio, EV/EBITDA, price-to-earnings, the Altman Z-score, and pre-announcement stock return, volatility, trading volume, and price range.

The one-day announcement CAR is included as a market-consensus input: it captures the aggregate information processing of investors at announcement and is measured over the single trading day of the announcement itself, prior to the start of the 24-month BHAR measurement window. It is therefore observable without look-ahead bias and conditions the model on the market's initial assessment of deal quality.

Macro-financial environment captures conditions prevailing at announcement, a dimension of direct relevance given the sample's span across several European monetary regimes. Variables include the three-month Euribor, the German 10-year Bund yield, investment-grade and high-yield corporate yields, the VIX, EUR/USD, HICP inflation, industrial production, and Brent oil prices. Following Harford (2005), aggregate liquidity conditions partly drive merger waves and affect the discipline with which acquisitions are undertaken. The credit spread is excluded as it is mechanically derived from the investment-grade and high-yield yield series and would introduce near-perfect multicollinearity. Macro-financial variables are measured as annual averages for the announcement year, a trade-off that introduces a potential partial look-ahead bias for deals announced early in the year, particularly in years of sharp intra-year rate variation such as 2022–2023, at the cost of timing precision.

Categorical variables (acquirer country, target country, acquirer industry, target public status, and deal form) are transformed into binary indicators with one reference category omitted per group. After encoding, the full feature matrix comprises 154 variables.

The central premise of the ML framework is that these three dimensions combine non-linearly. Acquisition quality may depend less on any single characteristic than on its joint configuration with others: a financially weak acquirer undertaking a large stock-financed transaction during a period of unusually accommodative financing conditions represents a fundamentally different risk profile than the same deal structure executed by a profitable, conservatively financed firm in a normalized rate environment. The empirical results in Section 4.3 support this hierarchy: deal structure variables contribute less predictive information than acquirer quality and macro-financial conditions, consistent with the theoretical mechanisms outlined above.

3.3 Descriptive Statistics and Regime Shift Evidence

Table 1 reports summary statistics for the main variables. Across the 6,540 transactions in the estimation dataset, the average 24-month CAPM-adjusted BHAR is close to zero at 0.2%, with a median of -1.1% and a standard deviation of 50 percentage points. BHAR ranges from -131.8% to +161.7%, confirming the high noise-to-signal ratio that characterizes long-run post-acquisition returns and the challenge this poses for point prediction.

Deal-level variables display substantial heterogeneity. Cash is the predominant form of consideration: among transactions with a cash component, the average cash share is 86.8% with a median of 97.2%, reflecting the dominance of all-cash and cash-heavy structures in European M&A. Cross-border transactions represent 54.5% of the sample, same-industry deals 51.9%, hostile bids 0.58%, and tender offers 13.9%. Relative deal size has a mean of 13.1 but a median of 0.29, reflecting a distribution dominated by bolt-on acquisitions alongside a small number of very large transformative transactions that introduce extreme right-tail skewness. Private targets account for 74.9% of the sample.

UK acquirers represent the largest geographic share at 50.0% (3,270 transactions), followed by France (8.6%), Sweden (7.9%), Germany (5.2%), the Netherlands (3.6%), Spain (3.6%), and Italy (3.5%), with several additional European markets contributing sizeable subsamples. The breadth of governance environments captured, from common law and shareholder-oriented systems in the UK to civil law with varying stakeholder orientation across Continental Europe and Nordic governance models in Sweden, provides the cross-country variation exploited in the robustness analysis of Section 4.4. Sectoral coverage is equally broad, with Software and IT Services (659 deals), Machinery and Industrial Equipment (551), Real Estate Operations (550), and Professional and Commercial Services (538) forming the largest groups.

Acquiring firm fundamentals confirm the sample's financial heterogeneity. Mean EBITDA margin is 11.9% with a median of 14.1%, while leverage averages 18.2%, suggesting moderate balance sheet risk on average. Mean acquirer ROA is 5.2% with a median of 5.5%, though both variables display substantial dispersion including negative values for weaker acquirers. Pre-announcement return volatility averages 9.7% monthly with a median of 7.9%, though the distribution has a long right tail reflecting a subset of acquirers entering transactions after periods of significant market turbulence. The mean one-day acquirer CAR of +2.25% is above the typical acquirer average, consistent with the sample's 74.9% private-target share, as acquirers of private targets typically generate positive announcement returns (Fuller, Netter and Stegemoller, 2002). On the macro side, the three-month Euribor ranges from -0.58% during the post-2015 negative-rate period to +7.58% during the recent tightening cycle, while high-yield corporate yields span from 4.0% to 21.8%, capturing both the accommodative conditions associated with European merger waves and the credit stress of crisis periods.

Figure 1 plots average BHAR by announcement year, with a vertical line at the 2018 train-test boundary. During 1993-2017, average BHAR is +2.7%. Over 2018-2024, it turns to -10.7%, a shift of approximately 13 percentage points driven by the COVID-19 shock, persistent inflation, and the sharpest European rate-hiking cycle in four decades. Within the test period, the deterioration accelerates after 2020, reaching -30.7% in 2021 as acquirers that had expanded during the low-rate environment faced sharp multiple compression during the tightening cycle.

This shift motivates a distinction that recurs throughout the empirical analysis. Predicting the absolute level of 24-month abnormal returns is difficult when the unconditional return environment changes so sharply between periods. The relevant question is whether models retain the ability to rank acquisitions within the test period, identifying relatively stronger deals in a universe of predominantly negative outcomes. A model that correctly assigns the highest scores to the acquirers that will subsequently outperform their peers delivers economically meaningful information even when all point predictions are downward-biased. The descriptive evidence in Figure 1 therefore contextualizes, but does not dismiss, the negative out-of-sample R^2 values reported in Section 4.1, whose interpretation is examined directly through a year-demeaning exercise that isolates the cross-sectional component of predictive performance.

3.4 Model Specifications and Rationale for Machine Learning

The empirical analysis compares seven supervised learning models. OLS provides the traditional linear benchmark; Ridge, LASSO, and ElasticNet introduce L2, L1, and combined regularization, each addressing the instability that arises when 154 correlated predictors are estimated on 5,330 observations. Ridge retains all variables while shrinking coefficients, isolating the gain from regularization alone. LASSO drives some coefficients to zero, functioning as embedded feature selection. ElasticNet combines both penalties, offering greater stability when predictors share overlapping economic information. Random Forest, XGBoost, and LightGBM form the tree-based family; all three capture non-linearities and interactions that linear models cannot represent without explicit specification. XGBoost and LightGBM both use gradient boosting but differ in tree construction: near-identical results across the two suggest that any non-linear evidence is not algorithm-specific. The baseline XGBoost specification uses a learning rate of 0.005, 200 estimators, a maximum tree depth of 6, and a minimum child weight of 20.

Four structural features of the prediction problem suggest that tree-based methods may outperform their linear counterparts. First, with 154 correlated predictors and a moderate training sample of 5,330 observations, OLS coefficient estimates exhibit elevated variance and are susceptible to overfitting when predictors are correlated; linear shrinkage only partially resolves this, as it cannot capture threshold effects or interactions. Second, the relationship between acquirer characteristics and long-run performance is unlikely to be globally linear: the difference between a firm with negative ROA and one that is marginally profitable is economically far more consequential than an equivalent difference between two already profitable firms. Third, acquisition outcomes are driven by combinations of variables: with 154 explanatory variables, 11,781 pairwise interactions are possible, and higher-order combinations increase combinatorially. Tree-based methods explore this space implicitly through recursive splitting, capturing combinations such as the Euribor × ROA interaction documented in Section 4.3, without requiring manual engineering. Fourth, extreme BHAR observations (below -100%) disproportionately influence OLS through the squared-residuals objective, whereas decision trees assign observations to leaves based on characteristics rather than residual magnitudes and are structurally robust to such outliers.

3.5 Metrics, Validation Framework, and SHAP Interpretation

Within the training sample, hyperparameters are selected using TimeSeriesSplit cross-validation with a one-year gap between folds, preventing deals announced in adjacent periods from contaminating estimation and validation windows and guarding against short-term market conditions leaking across folds.

Five metrics provide complementary views of predictive performance. The out-of-sample R^2:

$$R^2_{OOS} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_{train})^2} \tag{2}$$

measures level prediction accuracy; values below zero indicate underperformance relative to a naive historical-mean forecast and, in the present context, reflect the regime shift rather than an absence of signal. The Spearman rank correlation is the primary metric of this study:

$$\rho = 1 - \frac{6\sum {d_i}^2}{n(n^2 - 1)} \tag{3}$$

where d_i denotes the difference in rank between the predicted score and the realized BHAR for observation i, and n is the number of test-period deals.

Unlike R^2, the Spearman rank evaluates whether the model correctly orders acquisitions from stronger to weaker outcomes independently of the overall return level, capturing the cross-sectional ranking ability that generates economic value for an investor who need not forecast return magnitudes to benefit from correct ordering. The AUC from a binary classification (BHAR positive vs negative) measures discriminatory power that is robust to level shifts and extreme observations.

Quintile portfolio analysis provides direct economic evaluation. Test-period transactions are sorted by predicted score and divided into five equal groups ranging from Q1 (lowest predicted quality) to Q5 (highest). The Q5-Q1 spread, the difference in mean realized BHAR between these two groups, measures whether the model's ranking translates into an economically significant performance differential. A positive and statistically significant spread would suggest that the model's ranking contains information about relative acquisition quality that is not explained by common time variation in returns. The statistical significance of the Q5-Q1 spread is assessed using a two-sample Welch t-test, which does not assume equal variances across quintiles, comparing realized BHAR distributions between the highest- and lowest-predicted quintiles.

Finally, calendar-time long-short alphas test whether the signal survives conventional asset-pricing controls:

$$R_t - Rf_t = \alpha + \beta_M MKT_t + \beta_S SMB_t + \beta_H HML_t + \beta_W WML_t + \varepsilon_t \tag{4}$$

where MKT, SMB, HML, and WML denote the market, size, value, and momentum factors respectively. The regression uses European factors from the Kenneth R. French Data Library (French, 2025) and Newey-West HAC standard errors (Newey and West, 1987) with 24 lags to account for autocorrelation from overlapping 24-month holding periods. The intercept measures abnormal returns unexplained by conventional factor exposures to the market, size, value, and momentum premia, providing a stringent test of whether the signal reflects cross-sectional predictive ability or merely proxies for known systematic risk exposures.

SHAP values, introduced by Lundberg and Lee (2017), interpret tree-based predictions by attributing to each variable its marginal contribution to an individual prediction relative to a baseline expectation. The sum of all SHAP values equals the total prediction gap, and variables whose values most influence individual predictions receive larger absolute contributions. Unlike impurity-based importance measures, SHAP is local in that it explains why a specific acquisition receives its predicted score, and less biased toward high-cardinality variables, making it well suited to detecting interaction effects across deals and macroeconomic environments. Global feature importance is derived by aggregating local contributions across all observations, typically as the mean absolute SHAP value, providing a sample-level ranking of predictor influence that remains grounded in individual-prediction decompositions rather than aggregate correlation measures. SHAP values should be interpreted as a model-interpretation tool: they quantify the structure of the model's predictions but do not constitute evidence of causal importance.

4 Empirical Results and Discussion

The empirical analysis proceeds from statistical predictability to economic magnitude, mechanism identification, and robustness. Sections 4.1 and 4.2 establish the existence and cross-sectional nature of the signal; Section 4.3 examines the economic mechanisms behind it; Section 4.4 tests its robustness and translates it into investment terms.

4.1 Predictive Performance: Point Prediction versus Ranking

This subsection evaluates whether announcement-date variables contain out-of-sample predictive information about long-run acquisition performance. Table 2 reports the predictive performance of the seven models on the primary 24-month CAPM-adjusted BHAR. All models generate negative out-of-sample R^2 values. XGBoost produces the least negative R^2 at -0.044, followed by LightGBM (-0.048) and Random Forest (-0.057). Linear models perform worse:

OLS reaches -0.136 and Ridge -0.173. These values indicate that no model predicts the absolute level of future BHAR more accurately than a naive historical-mean benchmark.

This result is not surprising given the regime shift documented in Section 3.3. Models trained on a period with average acquirer BHAR of +2.7% are evaluated on a period in which it falls to -10.7%, a gap of approximately 13 percentage points driven by macroeconomic shocks that no announcement-date information set could fully anticipate. A negative R^2 therefore reflects the difficulty of forecasting the aggregate downward shift in post-2018 acquisition returns, not an absence of cross-sectional ranking ability.

Spearman rank correlations tell a more nuanced story. They are positive for all seven models, indicating that predicted scores are directionally aligned with realized outcomes even when level forecasts are downward-biased on average. XGBoost produces the highest rank correlation at +0.272, followed by LightGBM (+0.257), Random Forest (+0.255), LASSO (+0.236), and ElasticNet (+0.234). OLS also generates a positive correlation (+0.176), but remains materially below the tree-based models. The key implication is that models retain cross-sectional ranking ability despite the regime shift: they do not predict the precise level of future abnormal returns, but they assign higher scores to acquisitions that subsequently outperform their peers.

The classification results are consistent with this interpretation. When the problem is reformulated as predicting whether BHAR is positive or negative, Random Forest achieves an AUC of 0.591, followed by XGBoost at 0.587 and LightGBM at 0.566. These values exceed the 0.50 benchmark implied by random classification, though the margin is modest. This is consistent with the nature of the task: distinguishing successful from unsuccessful acquisitions over a 24-month horizon is inherently noisy, particularly in a post-2018 sample dominated by negative outcomes.

An additional dimension of the classification evidence concerns asymmetric predictive precision. Among deals the model classifies as likely to underperform, 73% (compared to a base rate of 61% negative outcomes in the full test sample) indeed generate a negative BHAR over the 24-month horizon (Random Forest and XGBoost). This suggests that the signal is particularly informative as a negative screen: the model identifies probable value destroyers with greater reliability than it identifies probable value creators, a pattern that recurs in the quintile analysis of Section 4.2 and that has direct practical relevance for deal screening applications. The asymmetry also helps explain the negative out-of-sample R^2: when most of the signal concentrates in the bottom of the distribution, level prediction for the full sample deteriorates while ranking ability is preserved.

Table 3 isolates the role of the macro regime by repeating the regression on annually centered BHAR, which removes the common annual component shared by all deals announced in a given year. This transformation does not alter the underlying cross-sectional structure of the data. Once this common component is removed, R^2 turns positive for all tree-based models: LightGBM reaches +0.050, XGBoost +0.047, and Random Forest +0.031. Ridge and LASSO also turn positive at lower levels (+0.013 and +0.005 respectively).

Table 4 places these findings in context by comparing key metrics with Campbell et al. (2025), who study M&A performance prediction in the United States using a similar machine learning approach.

The comparison highlights several structural differences between the two studies, the most consequential of which concern the macro environment and the return benchmark. The divergence in R^2 is informative rather than contradictory. Campbell et al. evaluate their models over a broader and more stable out-of-sample window, whereas this study's test period includes the COVID-19 shock, a severe inflation episode, and the sharpest European rate-hiking cycle in four decades. Their positive level-prediction performance reflects a less severe regime discontinuity between training and test periods. On the primary question of whether tree-based models outperform linear benchmarks, both studies reach the same conclusion. The difference lies in the stringency of the empirical test, not in the direction of the evidence.

The combined evidence establishes the first step of the empirical argument. All seven models fail on level prediction as R^2 values are negative across the board but succeed on relative ranking, with positive Spearman correlations, above-random classification performance, and positive annually centered R^2 for all tree-based models. The distinction is not merely statistical: it is the ranking ability, not the level forecast, that determines economic value. A model that consistently assigns higher scores to acquisitions that subsequently outperform their peers remains exploitable regardless of whether its absolute forecasts are biased. This is consistent with H1, which predicted positive and statistically significant Spearman rank correlations on out-of-sample European M&A data. Section 4.2 tests whether this ranking ability translates into economically large performance differentials.

4.2 Cross-Sectional Predictability

The models struggle to predict the absolute level of post-acquisition BHAR, but retain meaningful cross-sectional ranking ability. The analysis proceeds in three steps: transactions are sorted into predicted performance quintiles over the full test period; the exercise is repeated year by year; and rolling three-year windows verify that the results do not depend on a single test-period cut-off.

Pooled quintile analysis. Table 5 reports the pooled quintile results. For each model, test-period transactions are sorted into five equal groups based on predicted scores, from Q1 (lowest predicted quality) to Q5 (highest). The Q5-Q1 spread, defined as the difference in mean realized 24-month BHAR between the highest- and lowest-ranked quintiles, is the most direct economic test of the model's ranking ability: if predicted scores are informative, realized BHAR should increase across predicted quintiles and the spread should be positive.

The magnitude of the spreads is substantial. XGBoost generates a Q5-Q1 spread of +35.4 percentage points, with a t-statistic of 8.36. Acquisitions ranked in the top predicted quintile outperform those in the bottom quintile by more than one third of market value over the 24-month post-announcement horizon. LightGBM and Random Forest produce closely aligned spreads of +33.2 percentage points (t=7.80) and +32.7 percentage points (t=7.35) respectively. The convergence across the three tree-based models, despite relying on different ensemble-learning architectures, suggests that the finding reflects a broader non-linear signal rather than an artefact of any particular algorithm.

Linear and regularized models also produce positive spreads, though at lower magnitudes: ElasticNet +26.8 percentage points (t=6.49), LASSO +25.0 percentage points (t=6.01), OLS +21.4 percentage points (t=5.26), and Ridge +13.7 percentage points (t=3.43). All seven spreads are statistically significant at the 1% level. Two implications follow. First, the underlying cross-sectional signal is not model-specific: even a simple OLS recovers part of the ordering. Second, model flexibility matters systematically: the gap between tree-based models (+33-35 percentage points) and the linear benchmark OLS (+21 percentage points) represents a 50-65% improvement in economic spread, consistent with the hypothesis in Section 3.4 that acquisition outcomes depend on non-linear combinations of variables that linear models cannot fully capture. The gain from adding regularization alone, when comparing Ridge to OLS, is modest relative to the gain from adding non-linearity, which suggests that the tree-based advantage stems primarily from the capacity to model interactions and threshold effects.

The full quintile distribution for XGBoost reveals the structure of the ranking in detail. Mean realized BHAR across quintiles follows a largely monotonic pattern: -33.3% in Q1, -8.6% in Q2, -2.5% in Q3, -11.0% in Q4, and +2.1% in Q5. The slight reversal between Q3 and Q4 reflects residual noise in the middle of the distribution, where predicted scores are closest to the median and therefore least discriminating; within-quintile standard deviations of BHAR range from 0.34 in Q4 to 0.52 in Q1, confirming that individual deal outcomes are highly dispersed around quintile means. The ordering remains broadly monotonic across quintiles for all tree-based models, with the progression from Q1 to Q5 driven primarily by the large gap between the bottom and top predicted groups rather than by a uniform step-wise increase throughout the distribution.

The sign pattern of the XGBoost quintile means is particularly informative. Q5 deals, the top predicted quintile, are the only group to achieve a positive mean BHAR (+2.1%), while all other quintiles remain negative. In a test period characterized by an average BHAR of -10.7%, Q5 acquisitions are the only cohort to deliver positive abnormal returns on average. Conversely, Q1 deals average -33.3%, more than 35 percentage points below Q5. The fact that Q5 deals achieve a positive BHAR under these conditions underscores the economic content of the predicted ranking.

Year-by-year analysis. Table 6 examines whether the result is stable across annual cohorts. The XGBoost Q5-Q1 spread is positive in each of the six test years for which the full 24-month performance window is available. The spread is small and statistically insignificant in 2018 (+0.95 percentage points, t=0.09) and 2019 (+10.0 percentage points, t=0.73), but becomes large and highly significant from 2020 onward: +39.3 percentage points in 2020 (t=3.39), +30.5 percentage points in 2021 (t=4.28), +43.5 percentage points in 2022 (t=5.34), and +43.1 percentage points in 2023 (t=3.47).

The weak results in 2018 and 2019 should not be interpreted as model failure. Both years exhibit positive average BHAR in the test cohort (+5.5% and +14.1% respectively), financing conditions remained accommodative, and the dispersion in realized acquisition outcomes was limited. In such an environment, there is less cross-sectional variation for the model to exploit: when most acquirers perform adequately, the features that distinguish disciplined from fragile acquirers matter less for realized outcomes. By contrast, the COVID shock, the inflation surge, and the subsequent rate-hiking cycle created conditions in which acquisition discipline mattered far more. Weak acquirers, overextended balance sheets, and transactions announced under distorted financing conditions were more severely penalized. The model appears particularly useful in stressed environments, when separating disciplined acquirers from fragile ones becomes most consequential.

A sub-period decomposition reinforces this pattern. During the pre-COVID period of 2018-2019, the pooled Q5-Q1 spread across both years is +5.9 percentage points (n.s., Spearman $\rho = +0.068$), reflecting the limited dispersion discussed above. The COVID shock of 2020 produces a spread of +39.3 percentage points (Spearman $\rho = +0.327$), as the pandemic differentially penalized acquirers according to their pre-deal financial resilience. The recovery and rate-hiking period of 2021-2022 generates a combined spread of +34.8 percentage points (Spearman $\rho = +0.305$), driven by the sharp reassessment of acquisitions made under negative Euribor conditions. The post-hike period of 2023-2024 maintains an elevated spread of +41.0 percentage points (Spearman $\rho = +0.304$), suggesting that the signal persists as higher financing costs continue to differentiate resilient from fragile acquirers.

The 2022 cohort deserves particular attention. With a Q5-Q1 spread of +43.5 percentage points and a t-statistic of 5.34, it represents the strongest annual result in the study. The signal strengthens precisely in the years when the training-period macro regime is least representative of the test environment, consistent with the possibility that the model identifies features whose economic relevance intensifies when capital becomes more expensive.

Rolling window analysis. Table 7 evaluates XGBoost over five overlapping three-year windows. Spearman ρ remains positive in all five windows: +0.192 in 2018-2020, +0.270 in 2019-2021, +0.284 in 2020-2022, +0.276 in 2021-2023, and +0.263 in 2022-2024. The Q5-Q1 spread is positive and statistically significant in every window: +28.2 percentage points ($t=4.09$) in 2018-2020, +31.9 percentage points ($t=4.89$) in 2019-2021, +35.3 percentage points ($t=6.70$) in 2020-2022, +34.0 percentage points ($t=7.09$) in 2021-2023, and +33.9 percentage points ($t=5.42$) in 2022-2024[1]. The narrowness of this range, from +28.2 percentage points to +35.3 percentage points across five overlapping subperiods, indicates that the result is not driven by any single cohort.

The rolling-window evidence also clarifies the relationship between R^2 and ranking performance. R^2 varies substantially across windows, from +0.037 in 2018-2020 to -0.438 in

[1] The 2022-2024 window includes 18 transactions announced in 2024 for which the 24-month BHAR window extends beyond the data collection date. Their inclusion does not materially affect the results given the small sample size.

2021-2023, a range of nearly 50 percentage points. Spearman correlations and quintile spreads remain consistently positive across the same windows, varying by only 9 percentage points in the case of Spearman ρ (+0.192 to +0.284). This decoupling between level-prediction and ranking performance is the empirical counterpart of the theoretical distinction drawn in Section 4.1: the two metrics are not measuring the same thing, and their divergence in this setting reflects the regime discontinuity between training and test rather than a failure of the model to extract cross-sectional information.

The cross-sectional evidence is consistent across pooled, annual, and rolling-window specifications. This evidence is consistent with H2: the cross-sectional ranking ability translates into economically and statistically significant Q5-Q1 BHAR spreads, particularly pronounced when macro-financial conditions amplify the consequences of acquirer quality.

Section 4.3 turns to the question of mechanism. A large quintile spread establishes that the signal exists, but not why. SHAP values and revealed-preference analysis identify which variables drive the separation between predicted high- and low-quality acquisitions, and how these variables interact with the macroeconomic environment.

4.3 Economic Interpretation

The existence of a large and persistent quintile spread raises a more fundamental question: what does the model actually learn? The analysis draws on SHAP values to decompose model predictions at the variable level, on interaction plots to identify joint effects, and on a revealed-preference comparison between the highest- and lowest-ranked acquisitions.

What the model learns. Figure 2 reports the mean absolute SHAP value for the fifteen most influential predictors in the XGBoost model, computed across the 1,210 test-period deals. The ranking reveals a structural finding with direct economic content: within the model, macro-financial conditions and acquirer quality contribute substantially more predictive information than the observable characteristics of the deal itself. Euribor_3m is the most influential variable, with a mean absolute SHAP of 0.031, followed by acquirer return on assets (0.029), pre-announcement return volatility (0.018), pre-announcement price range (0.012), and the stock payment share (0.009). Deal structure variables such as relative deal size, cross-border status, hostile bid indicators, and same-industry flags contribute relatively little incremental predictive information once acquirer quality and macro-financial conditions are accounted for. This does not necessarily imply that these structural characteristics are irrelevant to deal outcomes; it suggests that, conditional on the information captured by acquirer fundamentals and financing conditions, they add limited discriminatory power in this sample.

The prominence of Euribor is consistent with Harford's (2005) merger wave theory: aggregate liquidity conditions shape both the availability of acquisition financing and the discipline with which it is deployed. The prominence of acquirer profitability is consistent with Jensen's (1986) free cash flow hypothesis: managers of firms with weak operating performance have stronger incentives to pursue acquisitions as a strategic response to deteriorating fundamentals, and weaker capacity to integrate and create value from targets. The SHAP beeswarm plot (figure 3) is consistent with these directional effects: high Euribor values are associated with strongly negative SHAP contributions, and negative acquirer ROA values generate the most negative predicted scores in the sample.

Not all dimensions of the signal admit a simple economic interpretation. Some SHAP relationships, particularly those involving industry dummies and pre-announcement volume dynamics, reflect complex interactions that are difficult to reconcile with standard acquisition theory. The main patterns, however, are coherent and theoretically grounded in the mechanisms described above.

The Euribor × ROA interaction. The most distinctive finding of the SHAP analysis is a non-linear interaction between Euribor and acquirer profitability. Figure 4 plots the SHAP contribution of Euribor_3m for each test-period deal, colored by the acquirer's pre-announcement ROA. For acquirers with positive ROA, Euribor has a negative but moderate effect on predicted BHAR: lower rates are associated with mildly better predicted outcomes, and higher rates with mildly worse ones. For acquirers with negative ROA, the relationship changes in magnitude. In the region combining negative Euribor and negative ROA, SHAP contributions for Euribor fall to between -0.08 and -0.12, roughly three to four times more negative than for profitable acquirers under the same conditions.

Negative-rate environments appear to relax the financing constraints that would otherwise discipline unprofitable acquirers. When short-term rates are below zero, the cost of external capital falls to the point where firms generating insufficient internal returns can nonetheless fund large acquisitions. This extends the logic of Jensen's (1986) free cash flow hypothesis from internal cash generation to aggregate monetary conditions: it is not only firms flush with cash that face weakened capital-market discipline, but also firms with access to unusually cheap external financing. Harford (2005) identifies this liquidity channel as a driver of merger waves at the aggregate level; the SHAP interaction suggests it operates within the cross-section as well, amplifying the performance gap between strong and weak acquirers precisely when financing conditions are most accommodative.

Importantly, the evidence does not imply that low interest rates are intrinsically associated with poor acquisition outcomes. The effect is conditional on acquirer quality. Q5 deals, which generate positive mean BHAR, are also disproportionately announced in low-rate environments: their mean Euribor at announcement is -0.33%. What distinguishes them from Q1 deals is not the macro context but the fundamental condition of the acquirer operating within it. Profitable, cash-disciplined firms appear able to deploy cheap financing productively; operationally weaker firms use it to make acquisitions they would not otherwise be able to finance.

This interaction is unlikely to be captured parsimoniously by linear specifications: with 154 explanatory variables, exhaustive specification of pairwise interactions is infeasible at this sample size, and XGBoost identifies it through recursive splitting without requiring ex ante specification.

A capital allocation interpretation is consistent with this pattern. When the cost of external financing falls to near zero, the effective hurdle rate for acquisition decisions declines with it. Projects that would have generated a negative net present value at a 6% cost of capital may appear value-creating when financed at 0%, not because their underlying economics have changed but because the benchmark against which they are evaluated has. For financially sound acquirers with positive operating returns, this shift matters less: their acquisition decisions are primarily constrained by strategic fit and target availability rather than by financing cost. For operationally weak acquirers, however, negative rates can be transformative. Firms that could not have secured financing for a large acquisition under normal conditions gain access to capital markets on terms that make the transaction appear feasible. This may create a selection effect in the acquisition market during low-rate environments: the pool of active acquirers expands to include firms whose investment discipline would not have cleared a conventional hurdle rate, and whose post-acquisition performance reflects that underlying fragility once financing conditions normalize.

The rate-hiking cycle that began in 2022 provides a useful empirical context for evaluating this dynamic. As Euribor rose from negative territory to above 3%, acquirers with strong pre-deal profitability were better positioned to absorb the simultaneous pressure on target valuations and financing costs. The widening of the Q5-Q1 spread from +5.9 percentage points in 2018-2019 to +43.5 percentage points in 2022 is, in this reading, not a sign that the model became more predictive but rather a sign that the economic consequences of the cross-sectional quality differences the model had identified all along became more visible as the rate cycle unwound the distortions of the prior decade.

This interpretation connects to a broader literature on the real effects of monetary policy on corporate investment. Bernanke and Gertler (1989) document that monetary conditions affect not only the cost of capital but the quality of investment undertaken: easy money reduces the screening intensity applied to marginal projects. The SHAP evidence suggests that the European M&A market was not immune to this dynamic: the model's ability to identify ex ante which acquisitions were made under conditions of weakened capital-market discipline, and which were made despite tight discipline, is precisely what generates the cross-sectional ranking signal documented in Section 4.2.

The feature selection evidence from LASSO provides further corroboration consistent with the SHAP importance ranking. Among 154 candidate variables, LASSO retains only 11 at its optimal regularization parameter: Euribor, acquirer ROA, pre-announcement return volatility, cash payment share, full-stock indicator, EUR/USD, EBITDA margin, high-yield spread, industry acquisitiveness, VIX, and a pharmaceutical industry indicator. This sparse solution is strikingly consistent with the SHAP analysis. The three most important predictors are identical across both methods: Euribor, acquirer ROA, and pre-announcement return volatility. The convergence between a purely data-driven selection method (LASSO) and a model-interpretability method (SHAP) applied to a non-linear model strengthens confidence that the most influential predictors are genuinely informative rather than coincidental.

Revealed preferences: what separates Q5 from Q1. Table 8 compares the mean characteristics of the 241 deals assigned to Q5 (top predicted quality) against the 242 assigned to Q1, with t-tests for each difference. The contrast is economically large.

On acquirer quality, the differences are pronounced. Q5 acquirers have a mean ROA of +7.7% against -7.1% for Q1 acquirers, a gap of 14.8 percentage points (t=12.54). The EBITDA margin difference is wider still: +17.4% for Q5 against -21.8% for Q1 (t=7.62). Pre-announcement return volatility averages 6.2% for Q5 acquirers against 19.1% for Q1 acquirers (t=-13.32), the largest t-statistic in the entire table. High pre-announcement volatility is associated with the bottom predicted quintile, consistent with the view that firms already under significant market pressure are more likely to pursue acquisitions from a position of operational weakness.

On financing structure, the pattern aligns with Myers and Majluf (1984). Q5 deals have a mean cash payment share of 95.7% against 74.2% for Q1 deals (t=11.07). Cash financing is generally interpreted as signalling managerial confidence that the deal price is fair and that the acquirer's equity is not materially overvalued. Stock financing, which is more prevalent in Q1 deals, conveys the opposite: managers who pay in equity may implicitly signal that they consider their shares a relatively cheap currency at prevailing prices.

On macro conditions, Q5 deals are announced when Euribor averages -0.33%, compared with +0.18% for Q1 deals (t=-5.58). As noted above, this does not mean that low rates drive poor performance; the directional difference in Euribor between Q5 and Q1 reflects the fact that both groups of deals are drawn from the same low-rate period, but that Q1 acquirers combine that environment with weak fundamentals while Q5 acquirers do not.

By contrast, several traditional deal descriptors exhibit limited differentiation between Q5 and Q1. Relative deal size, cross-border status, hostile bid indicators, same-industry classification, and the presence of competing offers display economically small or statistically insignificant differences across quintiles. This reinforces the broader SHAP evidence: the model derives most of its discriminatory power from acquirer quality and financing conditions rather than from observable transaction structure. Whatever information these structural variables contain appears to be more readily incorporated into market prices at announcement.

The model appears to do more than recover statistical regularities: the predicted separation between quintiles is consistent with coherent differences in acquirer quality, financing discipline, and macro-financial context. The result is consistent with a decades-old theoretical literature, and its detection appears to depend on the ability of tree-based models to capture non-linear interactions across the feature space.

4.4 Robustness and Investment Implications

The ranking signal established in the preceding analysis is large and persistent. This subsection examines whether those results survive a series of robustness checks, whether they hold across different institutional and geographic settings, and whether they translate into economically meaningful risk-adjusted returns when formed into an investment strategy.

Subsample robustness. Table 9 reports XGBoost Spearman rank correlations across ten distinct subsamples of the test set. The signal is positive in all ten.

Cross-border deals generate the highest rank correlation at +0.321 (n=678), above the full-sample estimate of +0.272. Cross-border acquisitions typically involve greater information asymmetry between acquirer and target, between deal participants and outside investors, and between domestic and foreign regulatory environments. Where information asymmetry is larger, observable proxies for acquirer quality and financing discipline may contain more incremental signal relative to market prices, making the model's ranking task comparatively easier. At the other end of the distribution, the 2018-2019 sub-period produces the weakest result at +0.068 (n=402). As discussed in Section 4.2, this period is characterized by positive average BHAR and limited cross-sectional dispersion, reducing the variation the model has available to exploit.

The remaining subsamples produce Spearman correlations within a relatively narrow range. Domestic deals (n=532) produce a Spearman ρ of +0.210, lower than cross-border but positive and economically non-trivial. Public targets (n=180, ρ=+0.308) are more predictable than private targets (n=1,030, ρ=+0.271), consistent with public acquirer-target pairs generating richer observable pre-deal information. Cross-industry deals (n=611, ρ=+0.304) are more predictable than same-industry deals (n=599, ρ=+0.241), suggesting that the acquirer-quality signal is particularly informative when the strategic rationale of the deal is less transparent to outside observers. Small deals below the median deal value (n=605, ρ=+0.297) are somewhat more predictable than large deals (n=605, ρ=+0.250), consistent with large acquisitions receiving more intensive analyst coverage at announcement. The post-2020 sub-period (n=808, ρ=+0.267) reinforces the decoupling between level-prediction and ranking ability documented in Section 4.1 despite an R^2 of -0.246.

Country-level heterogeneity. Table 10 reports results for the eight acquirer nations with at least 40 test-period deals. Spearman ρ is positive in seven of the eight. Sweden generates the strongest signal at +0.396, with a Q5-Q1 spread of +51.0 percentage points (t=5.33), followed by France (+0.358, Q5-Q1=+32.6 percentage points, t=3.22) and the United Kingdom (+0.224, Q5-Q1=+31.0 percentage points, t=4.70). Spain is marginally significant at the 10% level (ρ=+0.325, t=1.94). The weaker result in Germany may reflect the greater prevalence of bank-centered financing relationships, which tend to reduce the information content of publicly disclosed financial metrics relative to private bilateral relationships between acquirers and their

Hausbank. These patterns are suggestive but should not be over-interpreted given the subsample sizes involved.

Italy stands apart from the broader pattern, with a Spearman ρ of -0.034 over 54 test-period deals — underpowered and not statistically distinguishable from zero. Switzerland (ρ=+0.288, n=42) and Denmark (ρ=+0.227, n=71) are directionally consistent with the majority pattern but too small to support strong conclusions.

Alternative BHAR benchmark. The primary results use a CAPM-adjusted BHAR. Re-evaluating the XGBoost model against a simple market-adjusted BHAR, computed relative to the STOXX Europe 600 without beta adjustment, produces a Spearman rank correlation of +0.277 against +0.272 under the CAPM specification. The difference of 0.005 is economically negligible. The ranking signal is therefore insensitive to the choice of abnormal return benchmark, which allays the concern that it might reflect systematic differences in market beta across acquirers rather than cross-sectional variation in acquisition quality.

Investment implications. The long-short strategy buys the equal-weighted Q5 portfolio and sells short the equal-weighted Q1 portfolio at each announcement, holding both legs for 24 months. Deals in Q2, Q3, and Q4 are excluded. The strategy takes no directional view on aggregate acquisition returns; its performance depends entirely on the model's ability to rank deals within the cross-section. The 24-month CAPM-adjusted BHAR of each deal serves as the return measure for each position, requiring no additional market adjustment.

The long-short exercise should be interpreted as an economic validation of the ranking signal rather than as a directly implementable trading strategy. Implementation frictions, discussed in Section 5, would materially affect net returns.

Over the 2018-2024 test period, the strategy generates a pooled 24-month long-short BHAR of +35.4%, positive in all six available annual cohorts. The long leg (Q5) averages +2.1% over 24 months and the short leg (Q1) averages -33.3%. Most of the spread is therefore driven by the severe underperformance of bottom-quintile acquisitions rather than by unusually strong positive returns among top-quintile deals. This asymmetry is economically informative: the model's primary value may lie in identifying acquisitions likely to destroy value rather than in identifying outright winners, a distinction with practical relevance for deal screening applications where avoiding the worst transactions is often more actionable than identifying the best ones. The event-time Sharpe ratio, net of the mean Euribor rate over the test period, is

0.43, modest by institutional standards, reflecting the concentration and volatility inherent in event-driven strategies with 24-month holding periods.

The calendar-time regression provides a more rigorous assessment. Monthly long-short portfolio returns are regressed on the European market factor, then the Fama-French three factors, and finally the Carhart four-factor model, using Kenneth R. French Data Library factors and Newey-West HAC standard errors (Newey and West, 1987) with 24 lags. Table 11 reports the results of the calendar-time factor regressions.

The estimated alpha is statistically significant across all three specifications, and should be interpreted as exploratory evidence in a short and regime-specific sample. Adding SMB, HML, and WML absorbs only 2.7% of the CAPM alpha. The estimated factor loadings are economically small, suggesting limited exposure to conventional market, size, value, and momentum factors. The estimated FF4 alpha therefore does not appear to be fully explained by conventional systematic risk exposures. This is consistent with H3: the calendar-time long-short alpha is not fully absorbed by the Fama-French four-factor model, consistent with the limits-to-arbitrage framework of Shleifer and Vishny (1997).

Viewed in aggregate, the investment evidence provides a concrete answer to the practical question underlying this study. As a stylized illustration of the ranking signal's economic magnitude, a systematic strategy of buying predicted winners and selling predicted losers, formed exclusively from information observable at announcement, would have generated a Fama-French four-factor alpha of +22.82% per year (t=5.42) over a six-year out-of-sample period that included a global pandemic, persistent inflation, and the sharpest European rate-hiking cycle in four decades. The signal remains positive across ten subsample tests, eight country sub-markets, and an alternative return benchmark. It is not fully explained by standard factor exposures within the sample period considered here. Whether this alpha reflects a persistent and partial incorporation of acquisition quality information into prices, a risk premium not captured by standard four-factor models, or characteristics specific to the post-2018 regime remains an open question that only a longer out-of-sample period can resolve.

5 Conclusion

This study examines whether machine learning models trained on announcement-date variables can predict the long-run cross-sectional performance of European acquirers. Four main findings emerge.

All seven models generate negative out-of-sample R^2 on the primary 24-month CAPM-adjusted BHAR, reflecting the macroeconomic regime shift between training and test periods. When the common annual component is removed through year-demeaning, R^2 turns positive for all tree-based models, suggesting that cross-sectional predictive information was present throughout but obscured by aggregate shocks.

Cross-sectional ranking ability is the central empirical finding. Spearman rank correlations are positive for all seven models and the Q5-Q1 spread reaches +35.4 percentage points, positive in every annual cohort and stable across rolling subperiods. Notably, the spread strengthens during periods of macro stress, suggesting that the model identifies features of acquirer quality whose consequences become most visible when capital-market discipline tightens. These results hold across ten subsample tests and eight European acquirer markets.

The long-short strategy generates a Fama-French four-factor alpha of +22.82% per year (t=5.42), with factor loadings that are economically negligible. The asymmetry between legs, with Q5 averaging +2.1% against -33.3% for Q1, suggests that the model's primary value lies in identifying probable value destroyers rather than exceptional value creators.

The results carry concrete implications for practitioners. For corporate boards and deal committees, the findings suggest that acquirer operating fundamentals and prevailing financing conditions, rather than deal structure, are the primary determinants of long-run acquisition quality. Disciplined deal screening should therefore prioritize acquirer profitability and balance sheet strength, particularly when financing conditions are accommodative. For institutional investors, the model provides a negative screen: deals ranked in the bottom predicted quintile generate an average BHAR of -33.3%, and 73% produce negative abnormal returns, making avoidance of Q1 acquisitions a more reliable strategy than selection of Q5. For risk committees and lenders, the Euribor × ROA interaction implies heightened vigilance during low-rate environments, when capital-market discipline weakens for marginal acquirers.

This study makes four contributions to the existing literature. The first two are empirical and methodological. The analysis provides the first large-scale ML-based study of long-run post-acquisition performance in Europe, covering 6,540 transactions over 31 years across multiple monetary regimes, institutional environments, and governance systems. Methodologically, the use of CAPM-adjusted BHAR removes the confound introduced by systematic variation in acquirer betas, SHAP values connect model predictions to economic mechanisms, and the

Euribor × ROA interaction documented in Section 4.3 constitutes a novel empirical finding and relates the evidence to the literature on merger waves and monetary transmission.

The remaining two contributions are conceptual. The study also documents that cross-sectional ranking ability can survive structural breaks between estimation and deployment environments, a finding relevant to any ML application where the prediction context differs from the training period. The persistence of a +22.82% annualized FF4 alpha over six out-of-sample years, not fully explained by standard factor exposures, contributes to the asset pricing literature on corporate events and is consistent with a slow incorporation of acquisition quality information into European equity prices.

Four limitations bear on the interpretation and implementability of the results.

Regime dependence is the most fundamental concern. The test period coincides with an unusually severe macroeconomic environment, and the persistence of the documented signal outside the post-2018 context remains an open empirical question. Whether the Euribor × ROA interaction and the associated ranking ability would survive in a period of stable, moderate interest rates is not established by the evidence presented here. More broadly, financial relationships are potentially non-stationary: feature importance rankings documented here reflect the 1993–2024 estimation period and may not persist as market structures and monetary regimes evolve.

The machine learning approach itself carries inherent methodological limitations. Tree-based models are sensitive to the choice of train-test boundary and may capture patterns specific to the training regime. A further concern is the potential for data mining and multiple testing: with 154 candidate predictors, seven model specifications, and multiple evaluation windows, the risk of sample-specific relationships cannot be entirely ruled out. The consistency of results across models, subsamples, and rolling windows mitigates this concern, but does not eliminate it. Finally, the predictive framework does not support causal interpretation: the documented association between Euribor, acquirer ROA, and long-run BHAR is consistent with the theoretical mechanisms described in Section 2, but cannot be attributed causally without an identification strategy that the current design does not provide.

Implementation frictions limit practical exploitability. Under a stylized assumption of 200 basis points annually, likely optimistic for a concentrated long-short strategy with mid-cap European shorts, the FF4 alpha would decline to approximately +20.82%; at 400 basis points, to +18.82%. The long-short alpha may also partially reflect a quality or distress premium not captured by the four-factor model, consistent with the characteristics of Q1 acquirers documented in Campbell, Hilscher and Szilagyi (2008) and Asness, Frazzini and Pedersen (2019). A broader factor specification augmented with a quality factor such as QMJ could provide a more complete decomposition of the documented alpha and represents a natural extension of the present analysis. More broadly, the calendar-time regression is estimated on 72 monthly observations with 24-month overlapping holding periods, implying a limited number of independent return realizations. This reduces the effective precision of the factor regressions and reinforces the interpretation of the FF4 alpha as exploratory evidence requiring validation over a longer out-of-sample period.

Finally, the analysis is restricted to publicly listed European acquirers with disclosed deal values above USD 1 million. Deals in which the acquirer is delisted before the 24-month window closes are excluded, which likely understates underperformance in the lower quintiles and makes the documented spread a conservative estimate. Private transactions, very small deals, and non-European markets are also excluded. Whether the documented ranking ability extends to these segments remains an open question. In a live deployment, annual retraining of model specifications would additionally be required to adapt to structural changes in the relationship between announcement-date variables and long-run outcomes.

Four directions follow naturally from the results. The most immediate is an operating performance validation: given that acquirer ROA is the second most important predictor, it would be worth testing whether Q5 firms also outperform Q1 firms on post-integration accounting metrics. A second direction is the incorporation of textual signals from acquisition announcements, whose language likely contains incremental predictive information beyond the quantitative variables used here. Third, replacing annual macro averages with announcement-month values is a necessary robustness check before drawing strong conclusions about Euribor's predictive role: annual averages introduce a potential partial look-ahead bias for deals announced early in years of sharp intra-year rate variation, such as 2022 and 2023, and this limitation is most consequential precisely for the predictor that drives the central interaction documented in Section 4.3. Finally, replicating the analysis on Asia-Pacific or emerging market data would establish whether the documented patterns are general or specific to European acquisition markets.

References

Altavilla, C., Burlon, L., Giannetti, M., Holton, S., 2022. Is There a Zero Lower Bound? The Effects of Negative Policy Rates on Banks and Firms. *Journal of Financial Economics* 144, 885–907.

Andrade, G., Mitchell, M., Stafford, E., 2001. New Evidence and Perspectives on Mergers. *Journal of Economic Perspectives* 15, 103–120.

Asness, C.S., Frazzini, A., Pedersen, L.H., 2019. Quality Minus Junk. *Review of Accounting Studies* 24, 34–112.

Bao, Y., Ke, B., Li, B., Yu, Y.J., Zhang, J., 2020. Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research* 58, 199–235.

Bernanke, B., Gertler, M., 1989. Agency Costs, Net Worth, and Business Fluctuations. *American Economic Review* 79, 14–31.

Campbell, J.L., Elfrink, E.T., Irons, C., Moon, R., 2025. What is the Deal?: Predicting M&A Outcomes with Machine Learning. SSRN Working Paper.

Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In Search of Distress Risk. *Journal of Finance* 63, 2899–2939.

Erel, I., Liao, R.C., Weisbach, M.S., 2012. Determinants of Cross-Border Mergers and Acquisitions. *Journal of Finance* 67, 1045–1082.

Faccio, M., Masulis, R.W., 2005. The Choice of Payment Method in European Mergers and Acquisitions. *Journal of Finance* 60, 1345–1388.

French, K.R., 2025. Kenneth R. French Data Library. Available at: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Fuller, K., Netter, J., and Stegemoller, M., 2002. What do returns to acquiring firms tell us? Evidence from firms that make many acquisitions. *Journal of Finance*, 57(4), 1763–1793.

Giglio, S., Kelly, B., Xiu, D., 2022. Factor Models, Machine Learning, and Asset Pricing. *Annual Review of Financial Economics* 14, 337–368.

Goergen, M., Renneboog, L., 2004. Shareholder Wealth Effects of European Domestic and Cross-Border Takeover Bids. *European Financial Management* 10, 9–45.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical Asset Pricing via Machine Learning. *Review of Financial Studies* 33, 2223–2273.

Harford, J., 2005. What Drives Merger Waves? *Journal of Financial Economics* 77, 529–560.

Jensen, M.C., Ruback, R.S., 1983. The Market for Corporate Control: The Scientific Evidence. *Journal of Financial Economics* 11, 5–50.

Jensen, M.C., 1986. Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *American Economic Review* 76, 323–329.

King, D.R., Wang, G., Samimi, M., Cortés, A.F., 2021. A Meta-Analytic Integration of Acquisition Performance Prediction. *Journal of Management Studies* 58, 1198–1236.

Loughran, T., Vijh, A.M., 1997. Do Long-Term Shareholders Benefit from Corporate Acquisitions? *Journal of Finance* 52, 1765–1790.

Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems* 30, 4765–4774.

Malmendier, U., Tate, G., 2008. Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction. *Journal of Financial Economics* 89, 20–43.

Martynova, M., Renneboog, L., 2008. A Century of Corporate Takeovers: What Have We Learned and Where Do We Stand? *Journal of Banking and Finance* 32, 2148–2177.

Mitchell, M., Pulvino, T., Stafford, E., 2002. Limited Arbitrage in Equity Markets. *Journal of Finance* 57, 551–584.

Moeller, S.B., Schlingemann, F.P., Stulz, R.M., 2005. Wealth Destruction on a Massive Scale? A Study of Acquiring-Firm Returns in the Recent Merger Wave. *Journal of Finance* 60, 757–782.

Mullainathan, S., Spiess, J., 2017. Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives* 31, 87–106.

Myers, S.C., Majluf, N.S., 1984. Corporate Financing and Investment Decisions when Firms Have Information that Investors Do Not Have. *Journal of Financial Economics* 13, 187–221.

Newey, W.K., West, K.D., 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55, 703–708.

Rau, P.R., Vermaelen, T., 1998. Glamour, Value and the Post-Acquisition Performance of Acquiring Firms. *Journal of Financial Economics* 49, 223–253.

Renneboog, L., Vansteenkiste, C., 2019. Failure and Success in Mergers and Acquisitions. *Journal of Corporate Finance* 58, 650–699.

Roll, R., 1986. The Hubris Hypothesis of Corporate Takeovers. *Journal of Business* 59, 197–216.

Rossi, S., Volpin, P.F., 2004. Cross-Country Determinants of Mergers and Acquisitions. *Journal of Financial Economics* 74, 277–304.

Shleifer, A., Vishny, R.W., 1997. The Limits of Arbitrage. *Journal of Finance* 52, 35–55.

Tables

Table 1: Summary Statistics

Variable	Mean	Median	Std Dev	Min	Max
Panel A: Outcome variable					
BHAR 24m (%)	0.2	−1.1	50.0	−131.8	161.7
Panel B: Deal characteristics					
Deal value (EUR million)	521	40			
Cash payment share (%)	86.8	97.2			
Stock payment share (%)	57.1				
Relative deal size	13.1	0.29			
Cross-border (%)	54.5				
Same-industry (%)	51.9				
Hostile bid (%)	0.58				
Tender offer (%)	13.9				
Private target (%)	74.9				
Panel C: Acquirer characteristics					
ROA (%)	5.2	5.5			
EBITDA margin (%)	11.9	14.1			
Leverage (%)	18.2				
Altman Z-score	2.00	1.98			
Acquirer market cap (EUR million)	10,771	221			
Pre-announcement return volatility (%)	9.7	7.9			
Pre-announcement price range (%)	21.7				
Acquirer CAR 1-day (%)	2.25	0.67			
Panel D: Macro-financial conditions					
Euribor 3m (%)	2.26	2.34		−0.58	7.58
Bund 10y (%)	3.08				
Investment-grade yield (%)	5.4				
High-yield yield (%)	8.5			4.0	21.8
VIX	19.6	19.0			

Note: The sample covers 6,540 completed European M&A transactions announced between 1993 and 2024. The training period covers 1993–2017 (5,330 observations) and the test period 2018–2024 (1,210 observations). BHAR denotes the 24-month CAPM-adjusted buy-and-hold abnormal return. Cash share computed among transactions with a cash component; stock payment share computed across all transactions. Relative deal size is deal value divided by acquirer market capitalization. All continuous variables are winsorized at the 1st and 99th percentiles.

Table 2: Out-of-Sample Predictive Performance

Model	RMSE	MAE	R^2OOS	Spearman ρ	AUC
Tree-based models					
XGBoost	0.4726	0.3714	−0.0440	+0.272***	0.587
LightGBM	0.4735	0.3724	−0.0483	+0.257***	0.566
Random Forest	0.4754	0.3754	−0.0565	+0.255***	0.591
Regularized linear models					
LASSO	0.4857	0.3856	−0.1028	+0.236***	—
ElasticNet	0.4860	0.3861	−0.1042	+0.234***	—
Linear benchmark					
OLS	0.4930	0.3821	−0.1363	+0.176***	—
Ridge	0.5009	0.3998	−0.1728	+0.132***	—

Note: Out-of-sample performance over the 2018–2024 test period (n=1,210). R^2OOS is defined relative to the training-period historical mean. Spearman ρ measures rank correlation between predicted scores and realized BHAR. AUC measures binary classification accuracy (BHAR positive vs. negative) and is reported for the three tree-based classifiers only. Among deals classified as likely underperformers by XGBoost and Random Forest, 73% subsequently generate a negative BHAR. *** $p<0.01$.

Table 3: Predictive Performance on Annually Centered BHAR

Model	R^2OOS (Primary)	Spearman ρ (Primary)	R^2OOS (Year-demeaned)	Spearman ρ (Year-demeaned)
LightGBM	−0.0483	+0.257	+0.0499	+0.243
XGBoost	−0.0440	+0.272	+0.0467	+0.222
Random Forest	−0.0565	+0.255	+0.0309	+0.234
Ridge	−0.1728	+0.132	+0.0134	+0.122
LASSO	−0.1028	+0.236	+0.0048	+0.189
ElasticNet	−0.1042	+0.234	+0.0029	+0.186
OLS	−0.1363	+0.176	−0.0582	+0.114

Note: Year-demeaning removes the common annual component shared by all deals announced in a given year, isolating cross-sectional predictive information from aggregate time-series variation. The sign reversal in R^2OOS for all tree-based models confirms that the negative primary R^2OOS reflects the macroeconomic regime discontinuity rather than an absence of cross-sectional predictive information.

Table 4: Comparison with Campbell et al. (2025)

	Campbell et al. (2025)	This study
Geography	United States	Europe
Estimation sample	5,992 deals	6,540 deals
Test period	Broad post-1990 sample	2018–2024
Macro environment	Relatively stable	COVID, inflation, rate-hiking cycle
Number of predictors	70 variables	154 variables
Return benchmark	Market-adjusted	CAPM-adjusted
Best out-of-sample R^2	+0.12 (Random Forest)	−0.044 (XGBoost)
Cross-sectional metric	Quintile portfolio sorts	Spearman ρ = +0.272
Primary economic test	Quintile spread	Quintile spread and FF4 alpha
Factor analysis	Not reported	Fama-French four-factor
SHAP interpretation	Not reported	Full SHAP analysis
Country heterogeneity	Not applicable	8 country sub-markets

Note: The divergence in R^2OOS reflects the greater severity of the regime discontinuity in the European test period rather than a methodological difference. Both studies find that tree-based models outperform linear benchmarks on cross-sectional ranking metrics.

Table 5: Pooled Quintile Analysis

Panel A: Q5–Q1 Spreads by Model

Model	Q5 BHAR	Q1 BHAR	Q5–Q1 Spread	t-stat
XGBoost	+2.1%	−33.3%	+35.4pp	8.36***
LightGBM	+1.4%	−31.8%	+33.2pp	7.80***
Random Forest	+1.4%	−31.3%	+32.7pp	7.35***
ElasticNet	−2.5%	−29.3%	+26.8pp	6.49***
LASSO	−3.4%	−28.4%	+25.0pp	6.01***
OLS	−4.9%	−26.3%	+21.4pp	5.26***
Ridge	−9.0%	−22.7%	+13.7pp	3.43***

Panel B: XGBoost Full Quintile Distribution

Quintile	N deals	Mean BHAR	Std Dev BHAR
Q1 (lowest)	242	−33.3%	0.518
Q2	242	−8.6%	0.510
Q3	242	−2.5%	0.438
Q4	243	−11.0%	0.338
Q5 (highest)	241	+2.1%	0.406
Full test sample	**1,210**	**−10.7%**	

Note: Test period: 2018–2024 (n=1,210). Panel A reports mean realized BHAR for Q5 and Q1 and the Q5–Q1 spread for each of the seven models. Panel B reports the full quintile distribution for the best-performing model (XGBoost). Only Q5 achieves a positive mean BHAR in a test period with an average return of −10.7%. The slight reversal between Q3 and Q4 reflects residual noise in the middle of the distribution. *** p<0.01.

Table 6: Year-by-Year Quintile Analysis (XGBoost)

Panel A: Annual results

Year	N	Q1 BHAR	Q5 BHAR	Q5–Q1 Spread	t-stat	Sig
2018	215	+9.7%	+10.6%	+0.95pp	0.09	
2019	187	+9.9%	+19.9%	+10.0pp	0.73	
2020	187	−29.6%	+9.7%	+39.3pp	3.39	***
2021	256	−51.6%	−21.1%	+30.5pp	4.28	***
2022	221	−52.6%	−9.1%	+43.5pp	5.34	***
2023	126	−47.8%	−4.8%	+43.1pp	3.47	***

Panel B: Sub-period decomposition

Sub-period	N	Q5–Q1 Spread	Sig	Spearman ρ
2018–2019	402	+5.9pp	n.s.	+0.068
2020	187	+39.3pp	***	+0.327
2021–2022	477	+34.8pp	***	+0.305
2023–2024	144	+41.0pp	***	+0.304

Note: Panel A reports the XGBoost Q5–Q1 BHAR spread for each annual cohort with a complete 24-month outcome window. The 2024 cohort (n=18) is excluded as the BHAR window is incomplete. Panel B reports pooled spreads and Spearman rank correlations for four macro sub-periods. *** p<0.01, n.s. = not significant.

Table 7: Rolling Window Analysis (XGBoost)

Window	N	R^2OOS	Spearman ρ	Q5–Q1 Spread	t-stat	Sig
2018–2020	589	+0.037	+0.192	+28.2pp	4.09	***
2019–2021	630	−0.028	+0.270	+31.9pp	4.89	***
2020–2022	664	−0.268	+0.284	+35.3pp	6.70	***
2021–2023	603	−0.438	+0.276	+34.0pp	7.09	***
2022–2024	365	−0.244	+0.263	+33.9pp	5.42	***

Note: Five overlapping three-year windows over the 2018–2024 test period. Spearman ρ is positive in all five windows; Q5–Q1 is positive and statistically significant in all five. The narrow range of Q5–Q1 spreads (+28.2pp to +35.3pp) indicates robustness to the choice of window. *** $p<0.01$.

Table 8: Revealed Preferences: Q5 vs Q1 Deal Characteristics

Variable	Q1 Mean	Q5 Mean	Difference	t-stat	Sig
Acquirer quality					
ROA (%)	−7.09	+7.66	+14.75pp	12.54	***
EBITDA margin (%)	−21.84	+17.37	+39.20pp	7.62	***
Pre-announcement return volatility (%)	19.05	6.24	−12.81pp	−13.32	***
Intangible ratio	0.365	0.252	−0.112	−5.51	***
Log acquirer market cap	5.31	5.96	+0.65	4.40	***
Financing structure					
Cash payment share (%)	74.19	95.74	+21.55pp	11.07	***
Leverage (%)	18.03	21.88	+3.85pp	3.18	***
Log deal value	3.74	4.01	+0.27	2.38	**
Macro conditions at announcement					
Euribor 3m (%)	+0.178	−0.331	−0.509pp	−5.58	***
VIX	21.58	19.14	−2.44	−4.86	***
HICP inflation index	110.01	106.74	−3.27	−6.20	***
Industrial production index	99.48	101.40	+1.92	7.40	***
Deal structure (limited differentiation)					
Stock payment share (%)	53.31	56.82	+3.52	2.08	**
Relative deal size	2.23	0.26	−1.97	−2.01	**
Pre-announcement return (%)	4.34	1.26	−3.08	−4.95	***

Note: Comparison of mean deal and acquirer characteristics for the 242 Q1 deals (bottom predicted quality) and 241 Q5 deals (top predicted quality) as ranked by XGBoost over the 2018–2024 test period. t-statistics test the null of equal means. *** p<0.01, ** p<0.05, * p<0.10.

Table 9: Subsample Robustness (XGBoost)

Subsample	N	R^2OOS	Spearman ρ	Sig
Deal type				
Cross-border deals	678	−0.050	+0.321	***
Domestic deals	532	−0.037	+0.210	***
Target status				
Public targets	180	−0.033	+0.308	***
Private targets	1,030	−0.046	+0.271	***
Industry relatedness				
Cross-industry deals	611	−0.075	+0.304	***
Same-industry deals	599	−0.021	+0.241	***
Deal size				
Large deals (above median)	605	−0.072	+0.250	***
Small deals (below median)	605	−0.022	+0.297	***
Time period				
Post-2020 sub-period	808	−0.246	+0.267	***
2018–2019 sub-period	402	−0.005	+0.068	n.s.

Note: XGBoost Spearman rank correlations across ten subsamples of the 2018–2024 test period. The signal is positive in all ten subsamples. The weakest result (2018–2019) coincides with positive average BHAR and limited cross-sectional dispersion. *** p<0.01, n.s. = not significant.

Table 10: Country-Level Results (XGBoost)

Country	N	Spearman ρ	Q5–Q1 Spread	t-stat	Sig
Sweden	202	+0.396	+51.0pp	5.33	***
France	114	+0.358	+32.6pp	3.22	***
Spain	40	+0.325	+23.5pp	1.94	*
Switzerland	42	+0.288	+14.9pp	1.15	n.s.
Denmark	71	+0.227	+22.9pp	1.16	n.s.
United Kingdom	482	+0.224	+31.0pp	4.70	***
Germany	65	+0.114	+12.6pp	0.71	n.s.
Italy	54	−0.034	+27.1pp	1.11	n.s.

Note: Countries with at least 40 test-period deals, sorted by Spearman ρ. Spearman ρ is positive in 7 of 8 countries. Italy is the only country with a negative point estimate, not statistically distinguishable from zero. *** $p<0.01$, * $p<0.10$, n.s. = not significant.

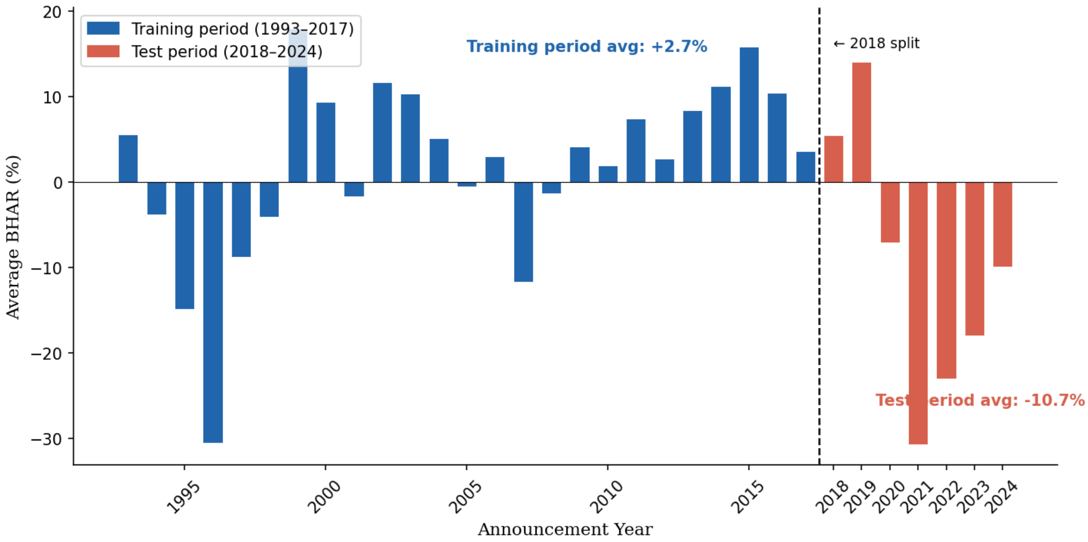
Table 11: Calendar-Time Factor Regressions

	CAPM	FF3	FF4
Alpha			
Alpha (monthly)	+1.771%	+1.755%	+1.728%
Alpha (annualized)	**+23.45%**	**+23.22%**	**+22.82%**
t-statistic	**4.99*** **	**5.52*** **	**5.42*** **
Factor loadings			
Market (MKT) β	+0.009	+0.007	+0.020
Size (SMB) β	—	−0.048	−0.052
Value (HML) β	—	+0.049	+0.062
Momentum (WML) β	—	—	+0.029
Fit statistics			
R-squared	0.003	0.075	0.085
N observations	72	72	72
Alpha decay			
Decay vs CAPM alpha	—	−0.23pp	−0.63pp
Total decay (CAPM → FF4)			**−2.7%**

Note: Monthly long-short portfolio returns regressed on European Fama-French factors. Sample: 72 months, 2018-07 to 2024-06. The long-short strategy buys the equal-weighted Q5 portfolio and sells short the equal-weighted Q1 portfolio, holding both legs for 24 months. Factors from the Kenneth R. French Data Library. Newey-West HAC standard errors (Newey and West, 1987) with 24 lags. Mean active Q5 deals per month: 80; mean active Q1 deals per month: 74. Event-time Sharpe ratio (net of mean Euribor 0.11%): 0.43. *** p<0.01.

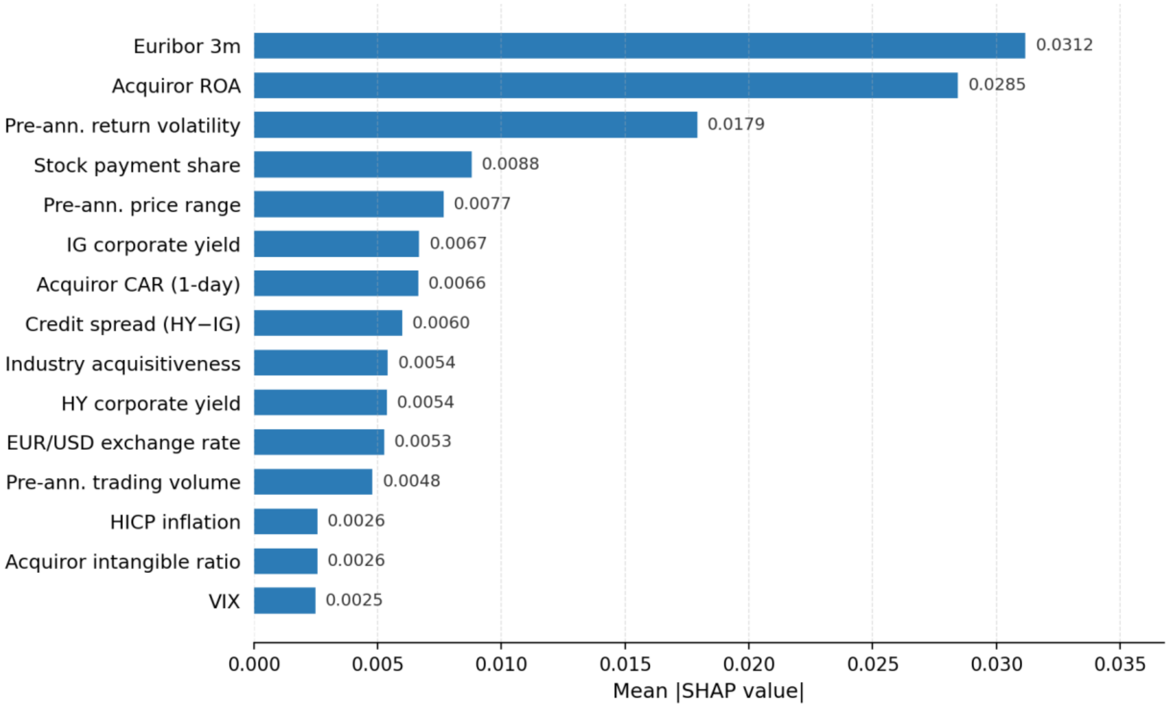
Figures

Figure 1: Average BHAR by Announcement Year



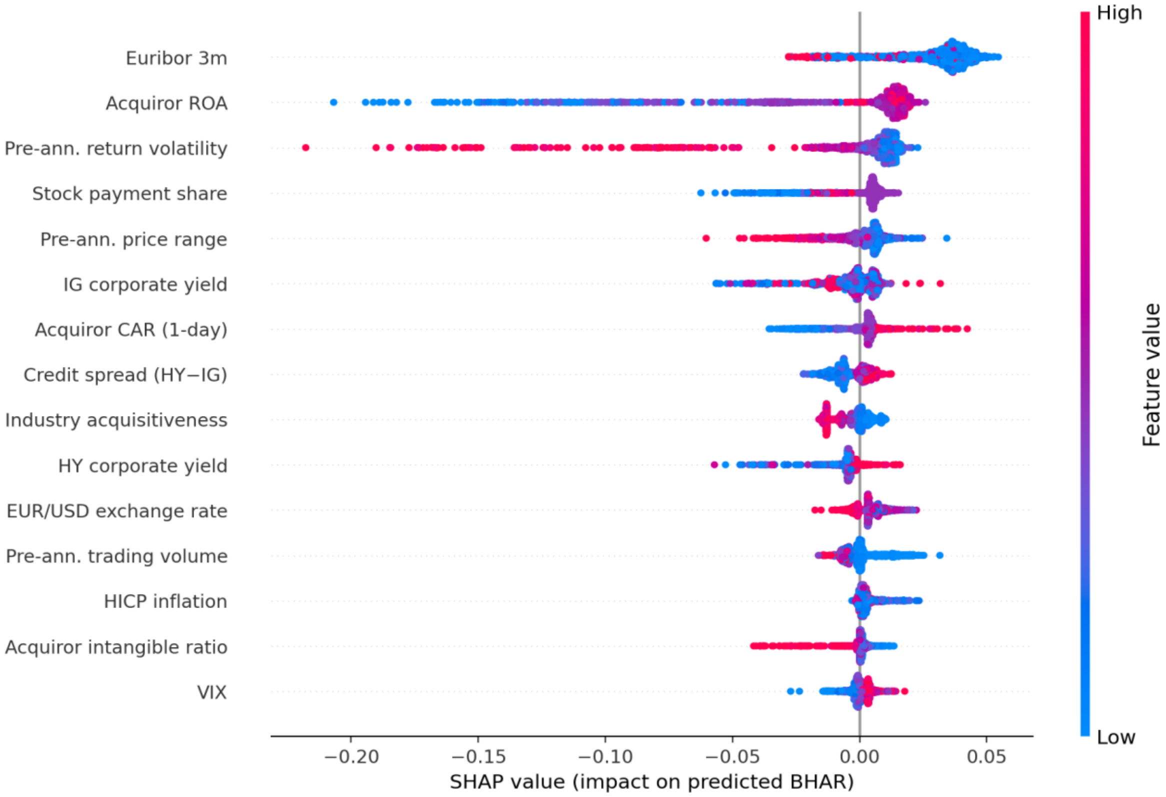
Note: Equal-weighted mean 24-month CAPM-adjusted BHAR by announcement year for the full sample of 6,540 European M&A transactions (1993–2024). Blue bars denote the training period (1993–2017, average BHAR +2.7%); red bars denote the out-of-sample test period (2018–2024, average BHAR −10.7%). The dashed vertical line marks the 2018 train-test split. The 13-percentage-point decline in average BHAR between periods reflects the COVID-19 shock, persistent inflation, and the sharpest European rate-hiking cycle in four decades.

Figure 2: Most Influential Predictors — Mean Absolute SHAP Values (XGBoost, Test Set)



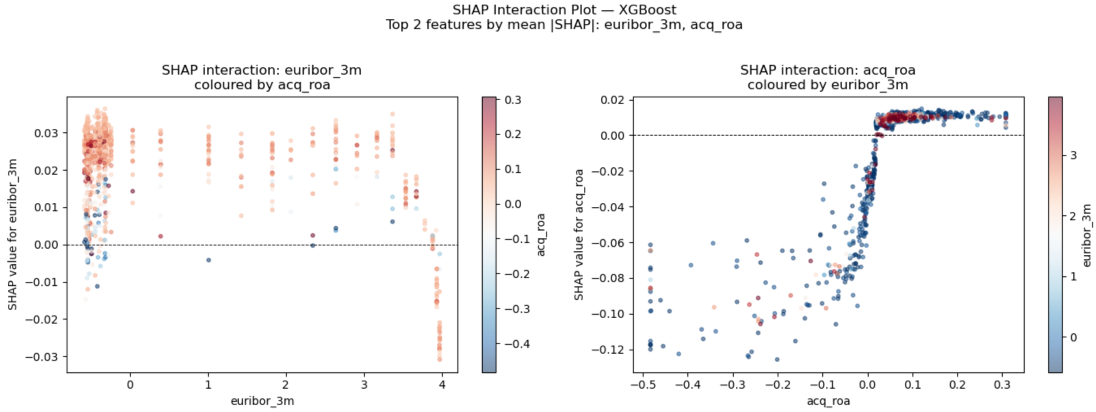
Note: Mean absolute SHAP values for the 15 most influential predictors in the XGBoost model, computed across the 1,210 test-period deals (2018–2024). Each bar represents the average magnitude of a variable's contribution to individual predictions relative to the baseline. Larger values indicate greater predictive influence regardless of direction. Euribor 3m and Acquirer ROA are the two most influential predictors, consistent with the monetary transmission and free cash flow mechanisms described in Section 2. Features are ordered from most to least influential.

Figure 3: SHAP Beeswarm Plot — XGBoost Feature Contributions (Test Set, Top 15 Features)



Note: SHAP beeswarm plot for the 15 most influential predictors in the XGBoost model, computed across the 1,210 test-period deals. Each dot represents one deal; horizontal position indicates the SHAP value (contribution to the predicted BHAR relative to the baseline), and colour indicates the feature value (red = high, blue = low). Positive SHAP values increase predicted BHAR; negative values decrease it. High Euribor values (red dots, top row) are associated with strongly negative contributions. Negative acquirer ROA values (blue dots, second row) generate the most negative predicted scores in the sample.

Figure 4: SHAP Interaction: Euribor × Acquirer ROA



Note: SHAP interaction plots for the two most influential predictors of the XGBoost model, computed across the 1,210 test-period deals. Left panel: SHAP contribution of Euribor_3m for each deal, colored by acquirer ROA (acq_roa). Points colored blue indicate negative ROA acquirers; red indicates positive ROA acquirers. In the region of negative Euribor and negative ROA, SHAP contributions for Euribor fall to between −0.08 and −0.12, roughly three to four times more negative than for profitable acquirers under the same conditions. Right panel: SHAP contribution of acq_roa for each deal, colored by Euribor_3m level. The interaction is consistent with the view that accommodative monetary conditions amplify the negative predicted outcomes for acquirers with weak operating fundamentals.