Machine Learning M&A Strategies[*]

Dragon Yongjun Tang[†] Fangming Xu[‡] Xiaohan Xue[§]

January 30, 2026

Abstract

We study whether machine learning can predict takeover premiums using only information available prior to merger announcements. Combining accounting fundamentals for acquirers and targets, deal and advisor characteristics, and macro–financial indicators, we train and evaluate linear, penalized linear, tree–based, and neural network models on U.S. public–to–public transactions from 1979–2021. Models sort deals into economically distinct strata, the realized premium increases monotonically across predicted percentiles and the top–minus–bottom spread is several times larger for machine learning models than for OLS. Results are robust across premium definitions. The paper provides a transparent, predictive framework for screening and benchmarking M&A premiums while linking model performance to economically interpretable mechanisms.

Keywords: Mergers and Acquisitions, Machine Learning, Explainable Artificial Intelligence
JEL Classification: C10, C45, G30, G34, G24

[*]We appreciate the helpful comments from Sean Cao, Daejin Kim, Hanwen Sun, Sarah Wang, and the conference participants at the 2025 Global AI Finance Research Conferenc.

[†]University of Hong Kong. Email: yjtang@hku.hk
[‡]University of Bristol. Email: fangming.xu@bristol.ac.uk
[§]University of Bath. Email: xx623@bath.ac.uk

1. Introduction

Mergers and acquisitions (M&A) are among the largest and most consequential corporate investments. A central object in these transactions is the *takeover premium*, the price per target share paid by the acquirer relative to the target's pre-deal market value, which embeds expectations of synergies, bargaining power, financing conditions, and market sentiment. Classic evidence shows that targets earn large announcement-window gains while acquirer short-window effects are near zero on average and long-run acquirer performance is often weak, underscoring the economic importance and heterogeneity of premiums (e.g., Asquith et al., 1983; Bradley et al., 1988; Andrade et al., 2001; Agrawal et al., 1992; Loughran and Vijh, 1997; Moeller et al., 2004; Betton et al., 2009). Measuring premiums is itself nontrivial: pre-bid price *runups* mean that benchmarks must be chosen with care (Schwert, 1996, 2000), and recent work argues that fixed look-backs can materially understate premiums in settings with early information leakage (Eaton et al., 2021).

Predicting M&A outcomes, however, has proven to be exceptionally challenging. Decades of empirical work using traditional linear models have reached little consensus on whether acquisitions benefit acquirers or their shareholders. In fact, many studies find that target shareholders realize substantial gains while acquirer shareholders often see zero or even negative abnormal returns on average (Shleifer and Vishny, 2003). This difficulty in prediction arises in part because M&A events are relatively infrequent and highly heterogeneous, limiting the effectiveness of conventional econometric techniques. At the same time, advances in artificial intelligence, and machine learning (ML) in particular, offer new opportunities to discern complex patterns in financial data. Modern machine learning algorithms can flexibly model nonlinear relationships and high-dimensional interactions, potentially uncovering signal in M&A outcomes that linear models miss. Financial institutions have begun to exploit these techniques in practice; for example, Bank of America reported that its machine learning-based investor targeting system achieved over 80% accuracy in identifying likely buyers for equity offerings, far surpassing traditional network-based approaches. These developments underscore the potential of machine learning to identify nuanced patterns that elude more restrictive frameworks.

In this paper, we ask whether machine learning methods can improve the prediction of M&A takeover premiums. This question is not obvious ex ante. Machine learning algorithms tend to excel with large, high-frequency datasets (as in credit card transactions or high-frequency trading), whereas acquisitions are sporadic, idiosyncratic events with relatively limited samples. Nonetheless, machine learning offers two key advantages in this context: it can incorporate a broad array of features, spanning firm fundamentals, deal characteristics, market conditions, and advisor attributes, and it can capture nonlinear relationships and interactions among these features without imposing a rigid functional form (Varian, 2014; Mullainathan and Spiess, 2017). We leverage these properties to build predictive models of takeover premiums using information available prior to deal announcement. In this study, we consider the commonly used algorithms in the supervised machine learning literature: Lasso, Elastic Net, Random Forest, Gradient Boosting Trees (XGBoost), and Neural Networks.[1] To address concerns about interpretability, the criticism that machine learning is a "black box", we also employ tools from the emerging literature on *explainable AI* (*XAI*). In particular, we use Shapley value-based methods (*SHAP*) to quantify the contribution of each predictor to the model's forecasts, thereby shedding light on the economic drivers of predicted premiums.

Our study makes several contributions. First, we provide one of the first systematic applications of machine learning to predicting takeover premiums, and we document that flexible nonlinear models can significantly outperform traditional linear benchmarks. In out-of-sample tests, the best-performing machine learning algorithms (such as tree-based ensembles and neural networks) achieve materially higher predictive accuracy than OLS regressions or regularized linear models. For instance, our preferred machine learning model improves the out-of-sample R^2 from approximately zero (or even slightly negative) under OLS to around 5–6%, and it reduces the root mean squared prediction error by roughly

[1] Further details on the model specifications and discussion are provided in Appendix A.

25–30%. While these figures may appear modest, they represent a notable improvement given the noisy nature of deal-level outcomes and the longstanding difficulty of forecasting M&A success. The result is important because it demonstrates that modern data-driven methods can extract meaningful predictive signals in an area where conventional models have largely failed to do so.

Second, we introduce an empirical framework that combines an unusually rich set of pre-deal variables from diverse sources. Our predictor set integrates firm-level financials of both acquirers and targets, detailed deal characteristics (e.g. hostility, payment method, industrial relatedness), advisor and intermediary information, short-term market performance measures, and macroeconomic indicators. This comprehensive approach extends prior studies that often relied on a narrower range of features (such as a few firm size or valuation ratios, or basic deal attributes). By broadening the information set, we allow the ML algorithms to learn from a more complete snapshot of the deal environment, which in turn helps improve predictive performance. In essence, our design mimics how a seasoned practitioner might assess a deal, by simultaneously weighing firm fundamentals, deal context, and market conditions, but we do so in a systematic, data-driven manner that can accommodate far more complexity than the human mind or linear models could handle.

Third, we confront the "black box" critique of machine learning by deploying state-of-the-art interpretability techniques to open up the model's inner workings. Using SHAP values, we evaluate the relative importance of each predictor and how it influences the predicted premium. This analysis yields economically intuitive insights: for example, we find that characteristics of the target firm, such as its pre-announcement stock price, valuation multiples, leverage, and cash flows, are among the most influential factors in the premium prediction. Credit market conditions also emerge as important: wider credit spreads (indicating tighter financing conditions) are associated with lower predicted premiums. By contrast, acquirer-specific factors (like the acquirer's size or recent stock volatility) and certain deal features (like the payment mix) play a more secondary role once target fundamentals and macro conditions are accounted for. These findings not only validate that the ML model is picking up sensible economic relationships, but they also contribute to the M&A literature by highlighting which dimensions of deal information are most salient for pricing. In short, our use of *XAI* bridges the gap between predictive performance and interpretability, allowing us to translate the complex patterns captured by the ML model into familiar financial factors.

Finally, our results carry practical implications for deal-making. Improved predictions of takeover premiums can inform deal screening and valuation exercises, for instance, helping acquirers and their advisors gauge whether a proposed price is in line with what would be expected given the deal's characteristics and market environment. Better premium forecasts can also aid in risk management by flagging deals that are likely to be overpriced (or underpriced) relative to fundamentals, thus allowing investors or regulators to scrutinize such transactions more closely. By showing that machine learning methods can add value in these ways, our study illustrates the promise of applying advanced analytics in corporate finance decisions. It suggests that investors and policymakers can benefit from incorporating data-driven forecasts alongside traditional valuation techniques when evaluating mergers and acquisitions.

Taken together, our findings demonstrate that machine learning can not only improve the prediction of M&A deal premiums but also enhance our understanding of the forces driving those premiums. By flexibly synthesizing diverse data and uncovering nonlinear patterns, ML algorithms uncover signal in pre-deal information that was previously obscured. And by applying explainability tools, we relate those signals back to economic fundamentals, thus ensuring the insights are transparent and actionable. In doing so, the paper contributes to the literature on corporate investments and M&A, and it provides evidence that modern data-science approaches can complement classic financial analysis in tackling complex, strategic business decisions.

The remainder of the paper is organized as follows. Section 2 reviews the background and prior literature on takeover premiums and introduces the machine learning methodologies we employ, along with performance metrics for evaluation. Section 3 describes the data sources, sample construction, and variable definitions. Section 4 presents the empirical results of our prediction exercise, comparing the out-of-sample performance of various ML models. Section 5 delves into model interpretability and the economic drivers of premiums. Section 6 concludes with a summary of the findings and a discussion of their implications for research and practice.

2. Background and Prior Literature

Takeover premiums are central to both academic inquiry and deal practice, yet their measurement requires care. The empirical convention benchmarks the offer price against the target's *unaffected* pre–announcement stock price observed a fixed number of trading days before the first public bid—most commonly 20, 42, or 63 trading days. This practice is motivated by evidence that target prices begin a pre-bid *runup* well ahead of the announcement, so that an earlier anchor better approximates the target's standalone value (e.g., Schwert, 1996, 2000). Building on this logic, our baseline premium is measured relative to the price 63 trading days prior to the announcement, with a 105–day alternative used as robustness to capture longer pre-announcement processes. Measuring at fixed pre-announcement windows is intended to recover a standalone valuation for the target, at least on average; see, e.g., Fich et al. (2011), Heitzman (2011), Boone and Mulherin (2011), and Masulis and Simsir (2018). Recent work, however, cautions that fixed windows can understate premiums when information leakage or anticipation occurs earlier, with underestimation as large as eight percentage points in some samples (Eaton et al., 2021). These insights motivate our use of alternative look–back horizons in robustness analysis.

Recent work revisits this fixed-window approach and documents that premiums can be materially *understated* when information leakage or deal processes start earlier than the conventional look-back dates; the bias can reach several percentage points in samples with lengthy negotiations or target-initiated deals. Accordingly, we report results under both 63– and 105–day anchors to ensure that inferences are not an artifact of window choice (see Eaton et al., 2021). This design choice also complements our broader empirical strategy—which conditions on process heterogeneity (e.g., hostility, initiator, payment method) and permits nonlinear interactions in the prediction stage.

A large literature characterizes who gains from M&A and by how much. Targets typically earn sizable positive announcement-window abnormal returns consistent with acquirers paying substantial premiums, whereas acquirer short-window effects are, on average, close to zero; over the long run, acquiring firms (or the combined entity) often underperform, giving rise to the post-merger underperformance puzzle (e.g., Rau and Vermaelen, 1998; Moeller et al., 2004). These patterns underscore the economic significance of the premium itself: premiums in public deals routinely reach levels that require ambitious synergy realization and cost savings to avoid value destruction.

Information frictions help explain cross-sectional variation in premiums but yield nuanced predictions. On the target side, greater opacity is associated with *higher* bid premiums—consistent with acquirers compensating for valuation uncertainty or bargaining against better-informed insiders. Conversely, when *bidders* are better informed, the winner's-curse logic predicts *lower* prices paid, and bidder toeholds generally strengthen bargaining power and depress both initial and final terms (e.g., Cheng et al., 2016; Dionne et al., 2015; Bris, 202). These forces interact with process features (e.g., hostility, auction entry) and payment method to produce substantial heterogeneity in realized premiums.

Intermediaries can mitigate information problems and shape outcomes. Classic certification theories suggest that top-tier investment banks should improve price discovery via information production and network reach; early evidence on bidder returns was mixed once deal complexity was controlled for, but more recent studies show that advisor reputation and *network centrality* matter, particularly in public deals where reputational exposure and analytical demands are greatest (e.g., Chemmanur and Fulghieri, 1994; Servaes and Zenner, 1996; Rau, 2000; Golubov et al., 2012; Chaudhry et al., 2022). These findings motivate our inclusion of advisor-tier and process variables as predictors and our use of machine learning tools that can flexibly accommodate nonlinear interactions among firm-, deal-, and intermediary-level characteristics.

Taken together, prior research implies (i) premium measurement must account for pre-bid runups and process length; (ii) premiums are economically large yet heterogeneous across deals; and (iii) information asymmetry, bidder toeholds, and advisory intermediation jointly shape bargaining outcomes. Our empirical framework incorporates these insights by employing multiple pre–announcement benchmarks for premiums and by conditioning on rich firm, deal, and advisor characteristics, with flexible algorithms used to learn the complex, potentially nonlinear mapping from pre-announcement information to realized takeover premiums.

3. Data

3.1. Measuring M&A Deal Premium

An essential step in designing our prediction algorithm is to specify the outcome variable that captures the price impact of the acquisition. We focus on the takeover premium paid to target shareholders as an economically meaningful, market-based measure of deal pricing. Our main outcome variable is the premium computed relative to the target's pre-announcement stock price, consistent with common practice in the M&A literature. Specifically, our baseline measure (Model 1) is defined as

$$\text{Premium}_{63} = \frac{\text{OfferPrice}_i - P_i(-63)}{P_i(-63)}, \tag{1}$$

where OfferPrice_i is the per-share price paid for target i (from SDC), and $P_i(-63)$ is the target's stock price 63 trading days prior to the public announcement date. To address the concern that information leakage or anticipatory price runups may begin earlier than 63 trading days before announcement, we construct an alternative specification (Model 2) following Eaton et al. (2021):

$$\text{Premium}_{105} = \frac{\text{OfferPrice}_i - P_i(-105)}{P_i(-105)}, \tag{2}$$

which anchors the benchmark to the stock price 105 trading days prior to announcement. As in Section 3, these price-based measures are computed only for publicly traded targets for which pre-announcement prices are available.

A potential concern with price-based premiums is that they are unavailable for private targets and can be affected by pre-announcement volatility, rumor-driven runups, or market-wide movements. To mitigate these issues and to expand coverage, we also employ a multiples-based premium that does not rely on public listing status (Model 3), following the comparable transactions method of Officer (2007). For each deal, we form a matched set of transactions within the same two-digit SIC industry, occurring in the three-year window surrounding the focal announcement ($t-1$, t, $t+1$), and with deal value within ±20% of the focal deal. Let Mult_i denote the focal deal's value-to-sales multiple and $\overline{\text{Mult}}_{\mathcal{C}(i)}$ the average multiple of its matched set $\mathcal{C}(i)$. The multiples-based premium is

$$\text{Premium}_i^{\text{CT}} = \frac{\text{Mult}_i - \overline{\text{Mult}}_{\mathcal{C}(i)}}{\overline{\text{Mult}}_{\mathcal{C}(i)}}. \tag{3}$$

This construction allows us to benchmark the focal deal's pricing against contemporaneous, industry- and size-comparable transactions, thereby reducing sensitivity to idiosyncratic pre-announcement price dynamics and broadening the sample to include private targets.

While takeover premiums are typically positive and right-skewed, they can be small or even negative in distressed or unique bargaining environments. We therefore treat premiums as a continuous outcome (which can take negative values) in our supervised learning framework. Our main results use Premium_{63} as the outcome, with robustness to Premium_{105} and the comparable-transactions measure in (3). Across specifications, the algorithm's task is to map pre-announcement information—accounting fundamentals, deal characteristics, and macro-financial conditions—into an out-of-sample prediction of the premium without imposing a specific functional form. This design directly parallels our motivation for machine learning: theory offers limited guidance on the exact interactions among covariates that determine bargaining outcomes, whereas flexible algorithms can accommodate nonlinearities and high-dimensional predictor sets in a disciplined, data-driven manner.

3.2. Sample Selection

Following prior literature, we use *SDC* as our source of M&A transactions. We focus on U.S. mergers from 1979 to December 2021 that are either acquisitions (A), mergers (M), acquisition of assets (AA), acquisitions of material interest (AM), or acquisitions of remaining interest (AR). We only include deals that are between public acquirers and targets, the amount acquired must be greater than 50%, and the total ownership amount after acquisition must also be greater than 50% (i.e., in the deal of interest the acquirer must have purchased more than a controlling interest in the target). Finally, we remove all deals that do not have a transaction value and small deals that have a transaction value less than $1 million.

We measure the takeover premium as the difference between the price paid per share for the target firm (as obtained from *SDC*) and the target firm's stock price 63 trading days prior to the M&A announcement date (Premium 1). To mitigate concerns that the transaction process or the target stock price runup may start earlier than 63 trading days prior to the official announcement date of an acquisition (as recorded by *SDC*), in Premium 2, we also measure the takeover premium using the stock price of the target firm 105 trading days preceding the M&A announcement date, as suggested by Gokkaya et al. (2023) and Eaton et al. (2021).

All predictors are measured using information available at or before the announcement date. Firm characteristics (from *Compustat* unless noted) include size (market equity from *CRSP*), asset and financing structure (current assets, total assets, short- and long-term debt, book debt, book equity), profitability and operating performance (net income, EBIT/EBITDA, OIADP, cash flow and cash flow-to-equity, ROA), investment and cost structure (CAPX, SG&A and its ratio, R&D and its ratio, depreciation), liquidity and working capital (cash/CHE, working capital and its ratio), leverage measures (book leverage, total liabilities-to-assets), valuation (book-to-market, Tobin's q), payout policies (dividends, total dividends, dividend payer indicator, repurchases, combined payout scaled by assets), retained earnings, Altman Z-score, and industry tech status (high-tech indicator). Market-based pre-announcement dynamics (from *CRSP*) comprise stock price runup and volatility (Σ) computed as value-weighted index–adjusted BHAR level and its standard deviation over the $[-205,-6]$ window; "large bidder" is an indicator for acquirer market equity above the sample median. Deal characteristics (from *SDC*) cover hostility, diversifying status (different 2-digit SIC), payment method (all-cash, all-stock, includes-stock) and financing shares (% cash/% stock). Advisor characteristics include a top-tier indicator defined by cumulative advised deal value (top 5 over 1979–2021). Macro-financial controls (from *FRED*)—industrial production, CPI, WTI spot oil, 3-month T-bill, 10-year Treasury yield, and credit spreads (AAA–10Y, BAA–10Y, BAA–AAA)—are lagged one month to ensure availability at the decision date. Continuous variables are winsorized at the 1st/99th percentiles, and ratios are constructed as specified in Appendix A.

3.3. Summary Statistics

Table 1 reports the number of unique deals per year, average takeover premiums (Premium 1 and Premium 2), and the number of unique acquirers and targets. The sample includes 7,183 transactions between 1979 and 2021, with an average Premium 1 of 0.478 and Premium 2 of 0.499. Deal activity is highly cyclical, peaking during the late 1990s merger wave when annual transactions exceeded 450, before declining sharply after the financial crisis. Premiums also vary substantially over time, ranging from above 0.70 in some boom years (e.g., 1982, 1999, 2003) to below 0.30 in downturn periods (e.g., 2005–2008). The distribution of transactions is broad, with over 150 unique acquirers and 160 unique targets on average per year, indicating wide cross-sectional coverage of firms. These patterns highlight the joint dynamics of M&A activity and acquisition pricing across market cycles.

[Insert Table 1 here]

Figure 1 plots the time series of M&A activity and average takeover premiums. The grey bars show the annual number of completed deals, while the blue and red lines display the average values of Premium 1 and Premium 2, respectively. Deal volume exhibits clear cyclicality, with pronounced peaks during the late 1990s merger wave and a sharp decline following the 2008 financial crisis. Average premiums fluctuate substantially across years, ranging from below 0.30 in downturn periods (e.g., mid-2000s and the global financial crisis) to above 0.70 in boom years such as 1999 and 2003. Overall, the figure highlights the joint dynamics of deal frequency and acquisition pricing over time.

[Insert Figure 1 here]

Tables 2 and 3 illustrate that the frequency of both low- and high-premium deals varies systematically with acquirer, target, deal, and advisor characteristics. For example, in Table 2, low premiums (bottom decile) occur more frequently when acquirers are high-tech firms (34.2%) compared with non–high-tech acquirers (26.9%). Similarly, acquirer dividend payers are less likely to be associated with low premiums (22.7%) than non-payers (32.4%). On the target side, dividend-paying targets are linked to substantially fewer low-premium outcomes (21.6%) compared with non-dividend payers (31.3%). At the deal level, transactions involving large bidders and those advised by top-tier financial advisors also exhibit significantly different propensities for falling into the low-premium group.

Table 3 complements these findings by documenting how the likelihood of being in the high-premium group (top decile) differs across firm and deal characteristics. For instance, target dividend payers are substantially less likely to experience high premiums (14.4%) than non-payers (19.2%). Deals with hostile bids are also far less likely to be in the top premium category (12.9%) compared with friendly transactions (18.6%). Finally, the role of advisors appears relevant, as deals involving top-tier advisors are associated with fewer high-premium outcomes (16.2%) than those without (19.4%).

Although some variables appear to be correlated with premium outcomes, theory offers limited guidance as to which specific firm, deal, or advisor characteristics should be expected to drive low versus high premiums, or the functional form of these relationships. For example, it is unclear whether a high-tech, dividend-paying acquirer targeting a large, dividend-paying firm in a hostile bid advised by a top-tier bank should be expected to yield relatively higher or lower premiums. The complexity increases further when multiple covariates interact in nonlinear ways. This motivates the use of machine learning algorithms, which are explicitly designed to uncover patterns in data without imposing strong parametric assumptions, thereby offering a rigorous, data-driven approach to predicting M&A premiums.

Moreover, the patterns in Tables 2 and 3 are consistent with the possibility that several key determinants affect premiums in a distinctly nonlinear manner, with thresholds, diminishing returns, and interaction effects that are difficult to capture in linear specifications. For instance, valuation and size proxies (e.g., equity price, shares outstanding, market equity) can mechanically compress or amplify percentage premiums depending on the denominator, implying that the marginal effect of a one-unit change in these variables need not be constant across the distribution. Similarly, uncertainty and information frictions—proxied by volatility measures (e.g., *Sigma*) and high-tech indicators—may matter primarily in the tails: moderate increases in uncertainty may have limited pricing impact in routine deals, whereas very high uncertainty can trigger discrete changes in bargaining outcomes, due diligence intensity, or the likelihood of competitive bidding. Deal-process characteristics can also operate through "regime switches": hostility, for example, may only materially influence pricing when coupled with financing constraints, weak target fundamentals, or the presence of competing bidders, so that its effect is conditional rather than additive. These considerations reinforce that the mapping from observable deal characteristics to realized premiums is plausibly nonlinear and state-dependent, strengthening the case for flexible algorithms that can learn threshold behavior and interactions directly from the data.

[Insert Tables 2 and 3 here]

4. Research Design and Results

4.1. Model Specification

We employ machine learning algorithms that predict the premiums of the M&A deals. The algorithms use a set of observable firm-level characteristics of acquirers and targets, deal characteristics, macroeconomic indicators, and financial advisor characteristics (see Appendix A) that is available prior to the deal announcement. The algorithms are commonly used in the supervised machine learning literature: Lasso, Elastic Net, Random Forest, Gradient Boosting Trees (XGBoost), and Neural Networks. We first train each algorithm on the 1979–2016 portion of our sample, containing 6764 M&A deals, of which 3586 are unique public acquirers and 6109 public targets. Training involves having the algorithm determine which combinations of variables best predict future performance.[2] We evaluate the models' out-of-sample predictions on the held out 2017–2021 portion of our sample containing 419 M&A deals, of which 360 are unique public acquirers and 402 are unique public targets. We compare these out-of-sample predictions to those from an OLS model. All comparisons are based on predictions for the 2017–2021 subsample of M&A deals, which does not overlap with the 1979–2016 subsample on which the algorithms are trained.

The optimal way to choose the size of the training and test sets depends on the signal-to-noise ratio in the data ad the training sample size. Therefore, establishing a general rule on how much training data is sufficient is challenging. For very large data sets, a 90%/10%split

[2] The algorithms rely on a regularizer that balances in-sample fit versus out-of-sample overfitting.

can be done, although 70%/30% or 80%/20% splits are typically used in practice.[3] We use an 80%/20% split, but our results do not depend on the way in which we split the data into training and test periods.

4.2. Prediction Results

To evaluate how well each model predicts M&A premiums, we examine both the magnitude of forecast errors (RMSE) and the proportion of variance explained out of sample (R^2_{oos}). Table 4 reports these statistics for six predictive models across three premium definitions: Premium 1, Premium 2, and Premium 3. A consistent pattern emerges: machine-learning models substantially outperform traditional linear approaches, and the performance gap widens as the premium measure incorporates more economic information (moving from Premium 1 to Premium 3).

The OLS benchmark performs the weakest across all premium definitions. For Premium 1, OLS yields an RMSE of 0.985 and a large negative R^2_{oos} of –0.332, indicating that a simple linear regression performs worse than a naïve mean forecast. Similar results appear for Premium 2, where RMSE rises slightly to 1.020 and R^2_{oos} remains negative at –0.250. Even for Premium 3, the measure most closely aligned with valuation fundamentals, OLS achieves only a marginally negative R^2_{oos} of -0.051 while still producing the highest RMSE among all models. These findings reinforce that linear relationships alone are insufficient for capturing the complex structure underlying M&A premiums.

Regularized linear models offer only modest improvement. Both Lasso and Elastic Net reduce RMSE relative to OLS across all three premium measures and produce small, positive out-of-sample R^2_{oos} values (ranging from -0.028 to -0.017). These gains suggest that some predictive structure exists but is limited in linear form; once overfitting is controlled through regularization, the remaining linear signal is weak.

In contrast, nonlinear models yield substantially more accurate forecasts. Random Forest

[3] See Hastie et al. (2009) for a discussion of methodological issues involved in choosing training and testing sets.

improves performance noticeably, especially for Premium 3, where RMSE drops to 0.552 and R^2_{oos} increases to 0.076. This suggests that allowing for complex interactions between firm, deal, and market characteristics enhances predictive accuracy.

Tree boosting and neural networks perform best across all three premium definitions. For Premium 1, XGBoost and the neural network achieve RMSE values of 0.558 and 0.589, with R^2_{oos} around 0.15, representing a 30% reduction in forecast error relative to OLS. Performance improves further for Premium 2, where the neural network achieves the lowest RMSE (0.596) and a higher R^2_{oos} (0.141). The largest gains appear for Premium 3: XGBoost attains an RMSE of 0.328 and R^2_{oos} of 0.165, while the neural network yields an RMSE of 0.445 with the highest R^2_{oos} of 0.176. These improvements underscore that nonlinear, flexible models are able to extract meaningful predictive structure from the data even in the presence of substantial noise.

Overall, the results demonstrate that machine-learning methods capture important nonlinearities and interactions that traditional linear models fail to exploit. Although the absolute out-of-sample R^2 values remain modest, reflecting the inherent unpredictability of M&A outcomes, the improvements delivered by tree-based and neural models are economically significant. These findings support the use of advanced machine-learning approaches when forecasting M&A premiums and highlight the value of incorporating flexible functional forms into empirical merger research.

[Insert Table 4 here]

We, additionally, examine pooled premiums to see if, in our prediction time-period sample of 2017–2021, machine learning models identify differences in the deal premiums by quintile. The premiums by quintile grouped by prediction scores from our machine learning models are presented in Figure 1. The figure documents that the average deal premium is an increasing function of the predicted premium for all the machine learning algorithms, but not for the OLS model.

[Insert Figure 2 here]

Figure 2 reports average predicted M&A premiums by quintile for different prediction models. In each panel, realized premiums are grouped into quintiles by prediction score, with quintile 0 representing the lowest predicted premiums and quintile 4 the highest.

Panel (a) shows results for the linear OLS benchmark. While OLS identifies variation across quintiles, the relationship is relatively weak and does not exhibit a clear monotonic pattern. By contrast, the machine learning models in Panels (b)–(f) exhibit more systematic variation. For instance, the lasso (Panel b) and elastic net (Panel c) produce steadily increasing premiums across quintiles, consistent with their ability to impose regularization and select informative predictors. Random forest (Panel d), XGBoost (Panel e), and neural networks (Panel f) also reveal clear monotonic increases in realized premiums with prediction scores. Among these, XGBoost and neural networks exhibit the steepest gradients, with quintile 4 premiums substantially higher than quintile 0, underscoring their ability to capture nonlinearities and interactions among firm, deal, and macroeconomic variables.

Overall, Figure 2 highlights the performance differences between linear and machine learning approaches. Whereas OLS provides limited discrimination between low- and high-premium deals, the machine learning models consistently generate sharper cross-sectional spreads in realized premiums. This suggests that nonlinear algorithms provide superior predictive power by uncovering complex structures in the determinants of M&A deal premiums.

Table 5 evaluates each model's ability to rank deals by premium level by comparing realized premiums across predicted-percentile buckets in the extreme tails. The lower tail illustrates the clearest divergence between linear and nonlinear approaches. When examining deals predicted to fall in the bottom 1% of the premium distribution, the nonlinear models correctly assign substantially negative realized premiums: Random Forest, XGBoost, and the Neural Network yield average realized premiums of –0.163, –0.173, and –0.158, respectively. In contrast, OLS assigns these same deals a positive realized premium of 0.371 (Premium 1), demonstrating poor calibration in identifying genuinely low-premium transactions. Similar patterns arise at the 5% and 10% thresholds, where the machine-learning models keep realized premiums near zero or negative, whereas the linear models continue to produce materially positive values. This suggests that linear specifications struggle to separate low-premium outcomes, while nonlinear models successfully detect characteristics associated with deep discounts in deal pricing.

Turning to the upper tail, realized premiums increase across all models, but the steepness of this increase differs markedly. For the highest predicted-premium deals (the 100% bucket), the Neural Network produces realized premiums of 1.438 for Premium 1, substantially higher than those assigned by the OLS model (0.772 in the raw table, but far less differentiated on a spread basis). XGBoost and Random Forest similarly generate high realized premiums, indicating that these models more accurately recognize the features associated with aggressive bidder valuations. The economic magnitude of the separation between predicted low- and high-premium deals is striking. For Premium 1, the Neural Network produces a top–bottom spread of roughly 1.60, compared with about 0.40 for OLS, representing more than a fourfold improvement in cross-sectional discrimination. These patterns hold, and in some cases strengthen, for Premium 2 and Premium 3. Machine-learning models consistently show sharper monotonic increases in realized premiums from the lower to the upper percentiles, particularly for Premium 3, where valuation fundamentals and market conditions play a larger role.

Overall, the results demonstrate that nonlinear machine-learning algorithms not only improve point predictions (as shown in Table 4) but also excel at ranking and sorting deals by their expected premium. Their ability to generate wider and more monotonic spreads across predicted-percentile buckets suggests that they capture nonlinear relationships and interaction effects that govern takeover pricing. These enhanced ranking properties reinforce the conclusion that flexible, data-driven models provide economically meaningful improvements over traditional linear methods in forecasting M&A premiums.

[Insert Table 5 here]

We compare our predicted premiums to the observed realized premiums for the 1437 observations in our test set. Figure 3 plots a binned scatterplot depicting the relationship between algorithmic predictions by the *XGBoost* model and the observed outcome among all new companies in our representative sample. Each point represents the average realized premium for the M&A deals grouped in bins according to their predicted premiums. Figure 3 illustrates the selected algorithm's ability to predict the distribution of M&A premiums.

[Insert Figure 3 here]

5. Further Analysis

5.1. Feature Importance

To shed light on the signals behind our predictions, we examine SHAP values. Figure 4 (*beeswarm* plot) shows local effects: each point represents a deal, with horizontal position indicating whether the feature raises or lowers the predicted premium, and colour indicating whether the underlying feature value is high (red) or low (blue). This plot shows how the model uses each variable. In contrast, Figure 5 (*mean-absolute SHAP bar* chart) reports a global ranking of feature importance, showing which variables contribute most to predictive accuracy overall.

Figure 4 indicates that the direction of feature effects is economically intuitive. Throughout the whole sample, the most influential signals cluster in two blocks. First are target fundamentals: the target's pre-announcement price (`t_prcc_f`), cash flow to equity (`t_cashflow_to_equity`), operating cost structure (`t_sga_ratio`), valuation (`t_tobins_q`), leverage (`t_leverage`), and shares outstanding (`t_csho`) sit at the top of the ranking. The beeswarm indicates economically sensible directions: higher `t_prcc_f` (a larger denominator in $(\text{Offer} - P_{-k})/P_{-k}$) and stronger target financials tilt SHAP values left, implying lower predicted premiums; weaker balance sheets or higher SG&A tend to push predictions right. Second are macro credit conditions: the Baa–Aaa spread (`baa_aaa_spread`), the Baa–10Y spread (`baa10ym`), and the oil price proxy (`wtisplc`) carry meaningful weight, with tighter credit (wider spreads) generally associated with compressed premiums. Acquirer variables, risk (`a_sigma`), valuation (`a_book_to_market`), price level and recent run-up, contribute non-trivially but rank below target fundamentals and credit spreads, and deal mechanics (cash share `pct_cash`, all-equity indicator) add comparatively small incremental power once fundamentals and macro factors are in the model.

[Insert Figure 4 here]

Figure 5 shows that two groups of variables dominate the global importance ranking. The SHAP diagnostics also clarify why machine learning improves on linear benchmarks. Importance reflects both (i) *direct* nonlinear relations (e.g., diminishing effects of target cash flows at high levels) and (ii) *interactions* (e.g., the effect of acquirer risk being stronger when payment is stock–heavy or when credit spreads are wide). In year-by-year tallies (not shown), target fundamentals (`t_prcc_f`, `t_cashflow_to_equity`) and credit spreads (`baa_aaa_spread`, `baa10ym`) recur most frequently among the top predictors, while financing mix variables rise in importance during easy-credit years. In general, the patterns indicate that the predicted premiums are jointly shaped by the quality of the target and the credit environment, with acquirer traits and the payment structure refining, rather than overturning, the fundamentally-driven ranking. This motivates the next step, where we probe partial/interaction effects using SHAP dependence and conditional plots.

[Insert Figure 5 here]

5.2. Robustness Check

5.2.1. Random Split

As a robustness exercise, we re-estimate the XGBoost model using a standard 80/20 random train–test split, instead of the temporally ordered split used in the main analysis. While random splits are common in prediction tasks, they implicitly allow future observations to inform the training set, violating the natural information structure faced by economic agents. Nonetheless, comparing the two approaches helps assess whether the model's performance depends heavily on time ordering.

The random-split results show only modest changes in the estimated performance metrics. Figure 6 displays the calibration curve based on 20 equal-sized bins of predicted premiums. The relationship between predicted and realized premiums remains clearly increasing, indicating that the model still ranks deals reasonably well. However, the slope is slightly shallower and the dispersion around the fitted line somewhat larger compared to the time-series split. This implies that, although the model retains its directional accuracy, predictive precision is modestly worse under random sampling.

The time-series split respects the temporal evolution of corporate fundamentals, credit conditions, and deal pricing, ensuring that the model is evaluated on genuinely unseen "future" observations. Random sampling weakens this structure by mixing earlier and later deals, artificially improving similarity between train and test sets and thereby reducing the model's ability to generalize to truly out-of-sample time periods.

[Insert Figure 6 here]

Despite minor changes in predictive accuracy, the SHAP importance rankings remain remarkably stable. Figure 7 presents the top ten features under the random split. The leading predictors, such as the target's pre-announcement price (`t_prcc_f`), acquirer stock price run-up (`a_stock_price_runup`), cash-share indicators (`payment_all_cash`), target share count (`t_csho`), and valuation proxies, continue to dominate the model's decision structure. Their relative magnitudes shift slightly, but the identity and ordering of the most influential features remain broadly consistent with those obtained under the time-series split.

This stability indicates that the economic signals driving premium variation are robust to the choice of sampling scheme. Target fundamentals and acquirer characteristics continue to account for the bulk of predictive power, and deal-structure variables retain secondary but meaningful contributions. The key predictors identified in the time-series setting are genuine, not artefacts of temporal ordering. The SHAP rankings change little, confirming that the model consistently relies on the same core fundamentals and macro-deal attributes. The time-series split provides a more conservative and economically appropriate evaluation of forecasting performance.

[Insert Figure 7 here]

5.2.2. Premium Measured by Fundamental Variables

To address potential concerns that Premiums 1 and 2 rely on stock-price information that may be missing or noisy for delisted targets, we construct an alternative measure, Premium 3, that does not depend on public listing status. Following the comparable-transactions approach of Officer (2007), we form for each focal deal a matched set of transactions drawn from the same two-digit SIC industry, occurring in a three-year window around the announcement year ($t-1$, t, $t+1$), and with deal values within $\pm 20\%$ of the focal deal. Premium 3 is then defined as the relative deviation of the focal deal value from the median valuation multiple of its matched set. This approach expands sample coverage and mitigates the influence of short-term market volatility embedded in stock-price–based premiums.

In our baseline specification, the valuation multiple is computed using sales. As a further robustness test, and consistent with the suggestions in Officer (2007), we re-estimate Premium 3 using book equity as the key scaling variable. The two resulting premium definitions allow us to examine whether alternative measures of fundamental value materially affect the predictive performance of the machine-learning model.

Across both versions of Premium 3, the XGBoost predictions exhibit a systematic downward bias relative to realized premiums. The binned realized–predicted scatterplots show that nearly all points lie below the 45-degree line, in stark contrast with Premiums 1 and 2 where the model captures the overall level more accurately. This pattern suggests that Premium 3 behaves differently from market-based premiums: it appears more tightly linked to firms' fundamentals (sales or book equity) and less driven by forward-looking price dynamics.

Because XGBoost is trained on a rich set of accounting, market, and macro predictors—many of which proxy for market perceptions and expectations—the model performs best when the premium itself contains information derived from stock prices. By construction, Premium 3 is more backward-looking and less sensitive to such signals, which likely contributes to the observed underestimation.

More broadly, persistent underestimation of Premium 3 is consistent with a combination of *shrinkage toward the center* and *mismatch between the target and the predictors*. First, supervised learners optimized under squared-error-type objectives (including gradient boosting with regularization) tend to deliver conservative forecasts when the outcome is noisy and heavy-tailed: in the presence of limited signal, the optimal prediction is pulled toward the conditional mean, compressing the upper tail and producing systematic underprediction of unusually high premiums. This effect is amplified by explicit regularization (tree depth penalties, learning-rate shrinkage, subsampling), which is designed to reduce variance but also attenuates extreme fitted values. Second, Premium 3 is constructed from a benchmark multiple that itself contains sampling variation (finite matched sets, time-varying industry pricing, and deal-value filters), introducing measurement error that is largely orthogonal to many pre-announcement predictors. When part of the realized premium reflects idiosyncratic dispersion in the matched-set multiple (rather than firm/deal attributes), the model cannot reliably learn it and will rationally discount it, again biasing predictions toward lower magnitudes. Finally, because Premium 3 removes the forward-looking component embedded in market prices, it mechanically strips out information channels (expectations, anticipation, and sentiment) that several market-based predictors are designed to proxy; the remaining predictable component is more "accounting anchored," leaving a larger irreducible residual that the model treats as noise. Together, these forces imply that even well-tuned ML models will tend to underpredict Premium 3, particularly for deals with exceptionally high fundamentals-based deviations.

[Insert Figures 8 and 9 here]

The SHAP analyses reinforce this interpretation. In both the sales-based and book-equity–based definitions, the key scaling variable used to construct Premium 3 emerges as the dominant predictor. When sales is the basis for the comparable-transactions multiple, `t_sale` overwhelmingly drives the model's predictions. When book equity is used, `t_book_equity` becomes the single most influential feature. This substitution occurs mechanically: because the premium is essentially a scaled deviation from fundamental value, the same fundamental variable directly governs both its magnitude and the structure of its cross-sectional variation. Other predictors, such as leverage, Tobin's q, and various acquirer-level characteristics—remain relevant but rank well below the primary scaling variable.

[Insert Figures 10 and 11 here]

Overall, these results confirm that Premium 3 behaves differently from market-based premiums and that its predictability is fundamentally rooted in the accounting quantity used to construct the comparable-transactions measure. The robustness checks using sales and book equity produce consistent qualitative conclusions, supporting the internal validity of the main findings while clarifying the economic mechanisms behind the weaker predictive accuracy for Premium 3.

5.3. Role of Financial Advisors: Counterfactual Replacement Evidence

Table 6 provides counterfactual evidence that top-tier financial advisors are associated with higher model-predicted M&A premiums. Across all three specifications—target-side advisors (Panel A), acquirer-side advisors (Panel B), and transactions in which both sides engage top-tier advisors (Panel C)—replacing the relevant top-tier advisor indicator(s) with a non–top-tier assignment leads to a systematic decline in the predicted premium. Importantly, this pattern holds throughout the distribution of deals: the predicted premium decreases in every decile, indicating that the advisor effect is not concentrated in a particular segment (e.g., only low- or high-premium transactions), but is instead pervasive across deal types and baseline premium levels.

The magnitude of the decline varies by advisory role. In Panel A, replacing target-side top-tier advisors yields modest but consistently negative changes in predicted premium, with Δ Pred. Prem. ranging from approximately -0.001 to -0.005 across deciles. Panel B shows a similarly uniform decline when replacing acquirer-side top-tier advisors, with somewhat smaller point estimates in most deciles (roughly -0.001 to -0.003). The strongest counterfactual impact appears in Panel C: when both sides' top-tier advisor indicators are replaced, the predicted premium drops materially in every decile, with declines on the order of -0.007 to -0.011. The fact that the "both advisors" case produces the largest reductions is consistent with complementarity, whereby top-tier advisors on each side jointly contribute to deal outcomes that the model associates with higher premiums.

Overall, the replacement exercise supports the interpretation that top-tier financial advisors play an economically meaningful role in shaping deal terms linked to higher premiums. The uniformity of the negative replacement effect across deciles strengthens this conclusion by demonstrating that the estimated advisor contribution is robust throughout the premium distribution rather than being driven by a small subset of extreme transactions.

[Insert Table 6 here]

Table 7 reports the counterfactual advisor-replacement exercise using the alternative premium definition in Model 2, where the benchmark price is measured 105 trading days prior to the announcement date, Premium_{105}, as defined in equation (2). This specification is designed to mitigate concerns that information leakage or anticipatory trading may begin earlier than the 63-trading-day window and therefore contaminate the pre-announcement benchmark price. The results remain qualitatively unchanged under this stricter anchoring: in every decile, replacing the top-tier advisor indicator with zero reduces the model-predicted premium. The monotone sign pattern across the full distribution indicates that the advisor effect is not driven by a narrow set of transactions with unusually large pre-announcement runups; instead, it appears as a broad and stable component of the predictive structure even when the premium is measured from a more distant pre-announcement reference point. Consistent with the construction of price-based premiums, the analysis is necessarily restricted to publicly traded targets with available pre-announcement prices, but within this sample the directional effect is uniform.

The magnitude of the counterfactual decline varies systematically across panels, providing additional structure to the advisor interpretation. When replacing target-side top-tier advisors (Panel A), the reduction in predicted premium is economically small but consistently negative across all deciles (approximately -0.0004 to -0.0010), suggesting a modest incremental contribution of target advisors after conditioning on the deal and firm characteristics captured by the model. The acquirer-advisor replacement (Panel B) produces larger declines, particularly toward the upper tail (e.g., deciles 9–10), implying that top-tier acquirer representation is more strongly associated with outcomes the model links to higher premiums in relatively high-premium transactions. The largest effects arise when both sides' top-tier advisor indicators are set to zero (Panel C), with declines on the order of -0.022 to -0.034 across deciles, consistent with a complementary role of high-quality intermediation on both sides of the transaction. Taken together, the Premium_{105} results reinforce the conclusion that top-tier financial advisors are associated with higher predicted premiums in a way that is robust to using a benchmark price sufficiently far ahead of the announcement to reduce concerns about early leakage and anticipatory runups.

[Insert Table 7 here]

6. Conclusions

This paper examines whether machine learning methods that combine accounting fundamentals, deal characteristics, advisor attributes, and macro-financial indicators can predict M&A takeover premiums using only information available prior to announcements. We construct several measures of premiums—anchored at 63 and 105 trading days—and evaluate a spectrum of models (penalized linear, tree ensembles, and a neural network) in a strictly out-of-sample design. The analysis shows that nonlinear algorithms materially outperform linear benchmarks, both in overall accuracy (lower RMSE, positive R^2_{oos}) and in their ability to sort deals into economically distinct premium strata: realized premiums rise monotonically with model-predicted percentiles and the top–bottom spreads are several times larger for ML models than for OLS. These findings indicate that meaningful predictive structure exists in pre-announcement data and that capturing nonlinearities and interactions is essential for extracting it.

Explainable–AI diagnostics help open the "black box". SHAP results highlight the primacy of target fundamentals (e.g., pre–announcement price, cash flows, valuation, leverage) and macro credit conditions (e.g., Baa-Aaa and Baa-10Y spreads), with acquirer traits and financing mix refining, rather than overturning, the fundamentals-driven ranking. The directional effects are economically coherent (for example, higher pre-announcement target prices mechanically compress percentage premiums), and they are stable across years, underscoring that both firm-level quality and the credit environment jointly shape bargaining outcomes reflected in takeover prices. Together with robustness to alternative premium definitions, these results provide a consistent picture of where predictive signal resides and why nonlinear learners capture it better than linear specifications.

The study has practical and scholarly implications. For practitioners, the models offer a transparent way to screen deals, assess premium risk conditional on market conditions, and benchmark negotiation outcomes; they complement traditional valuation by quantifying how observable features map to expected premiums. For researchers, the evidence (i) underscores the sensitivity of premium measurement to pre-bid runups, (ii) documents economically meaningful but modest predictive power typical of noisy corporate-finance settings, and (iii) shows how interpretable ML can bridge predictive performance and economic insight in settings where theory is agnostic about the correct functional form.

References

Abis, S., 2020. Man vs. machine: Quantitative and discretionary equity management. *Machine: Quantitative and Discretionary Equity Management (October 23, 2020)* .

Agrawal, A., Jaffe, J. F., Mandelker, G. N., 1992. The post-merger performance of acquiring firms: a re-examination of an anomaly. *The Journal of Finance* 47, 1605–1621.

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

Asquith, P., Bruner, R. F., Mullins Jr, D. W., 1983. The gains to bidding firms from merger. *Journal of Financial Economics* 11, 121–139.

Betton, S., Eckbo, B. E., Thorburn, K. S., 2009. Merger negotiations and the toehold puzzle. *Journal of Financial Economics* 91, 158–178.

Boone, A. L., Mulherin, J. H., 2011. Do private equity consortiums facilitate collusion in takeover bidding? *Journal of Corporate Finance* 17, 1475–1495.

Bradley, M., Desai, A., Kim, E. H., 1988. Synergistic gains from corporate acquisitions and their division between the stockholders of target and acquiring firms. *Journal of Financial Economics* 21, 3–40.

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

Bris, A., 202. Toeholds, takeover premium, and the probability of being acquired. *Journal of Corporate Finance* 8, 227–253.

Bubb, R., Catan, E. M., 2022. The party structure of mutual funds. *The Review of Financial Studies* 35, 2839–2878.

Chaudhry, A. N., Kontonikas, A., Vagenas-Nanos, E., 2022. Social networks and the informational role of financial advisory firms centrality in mergers and acquisitions. *British Journal of Management* 33, 958–979.

Chemmanur, T. J., Fulghieri, P., 1994. Investment bank reputation, information production, and financial intermediation. *The Journal of Finance* 49, 57–79.

Cheng, P., Li, L., Tong, W. H., 2016. Target information asymmetry and acquisition price.

Journal of Business Finance & Accounting 43, 976–1016.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* 2, 303–314.

Dionne, G., La Haye, M., Bergerès, A.-S., 2015. Does asymmetric information affect the premium in mergers and acquisitions? *Canadian Journal of Economics/Revue canadienne d'économique* 48, 819–852.

Easley, D., Michayluk, D., O'Hara, M., Putniņš, T. J., 2021. The active world of passive investing. *Review of Finance* 25, 1433–1471.

Eaton, G. W., Liu, T., Officer, M. S., 2021. Rethinking measures of mergers & acquisitions deal premiums. *Journal of Financial and Quantitative Analysis* 56, 1097–1126.

Fich, E. M., Cai, J., Tran, A. L., 2011. Stock option grants to target ceos during private merger negotiations. *Journal of Financial Economics* 101, 413–430.

Freund, Y., 1995. Boosting a weak learning algorithm by majority. *Information and Computation* 121, 256–285.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics* 28, 337–407.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics* pp. 1189–1232.

Gokkaya, S., Liu, X., Stulz, R. M., 2023. Do firms with specialized m&a staff make better acquisitions? *Journal of Financial Economics* 147, 75–105.

Golubov, A., Petmezas, D., Travlos, N. G., 2012. When it pays to pay your investment banker: New evidence on the role of financial advisors in m&as. *The Journal of Finance* 67, 271–311.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. *The Review of Financial Studies* 33, 2223–2273.

Heitzman, S., 2011. Equity grants to target ceos during deal negotiations. *Journal of Finan-*

cial Economics 102, 251–271.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359–366.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., Mullainathan, S., 2018. Human decisions and machine predictions. *The Quarterly Journal of Economics* 133, 237–293.

Li, K., Mai, F., Shen, R., Yan, X., 2021. Measuring corporate culture using machine learning. *The Review of Financial Studies* 34, 3265–3315.

Loughran, T., Vijh, A. M., 1997. Do long-term shareholders benefit from corporate acquisitions? *The Journal of Finance* 52, 1765–1790.

Masulis, R. W., Simsir, S. A., 2018. Deal initiation in mergers and acquisitions. *Journal of Financial and Quantitative Analysis* 53, 2389–2430.

Moeller, S. B., Schlingemann, F. P., Stulz, R. M., 2004. Firm size and the gains from acquisitions. *Journal of Financial Economics* 73, 201–228.

Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives* 31, 87–106.

Officer, M. S., 2007. The price of corporate liquidity: Acquisition discounts for unlisted targets. *Journal of Financial Economics* 83, 571–598.

Rau, P. R., 2000. Investment bank market share, contingent fee payments, and the performance of acquiring firms. *Journal of Financial Economics* 56, 293–324.

Rau, P. R., Vermaelen, T., 1998. Glamour, value and the post-acquisition performance of acquiring firms. *Journal of Financial Economics* 49, 223–253.

Rossi, A. G., 2018. Predicting stock market returns with machine learning. *Georgetown University* .

Schapire, R. E., 1990. The strength of weak learnability. *Machine learning* 5, 197–227.

Schwert, G. W., 1996. Markup pricing in mergers and acquisitions. *Journal of Financial Economics* 41, 153–192.

Schwert, G. W., 2000. Hostility in takeovers: in the eyes of the beholder? *The Journal of*

Finance 55, 2599–2640.

Servaes, H., Zenner, M., 1996. The role of investment banks in acquisitions. *The Review of Financial Studies* 9, 787–815.

Shleifer, A., Vishny, R. W., 2003. Stock market driven acquisitions. *Journal of financial Economics* 70, 295–311.

Varian, H. R., 2014. Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28, 3–28.

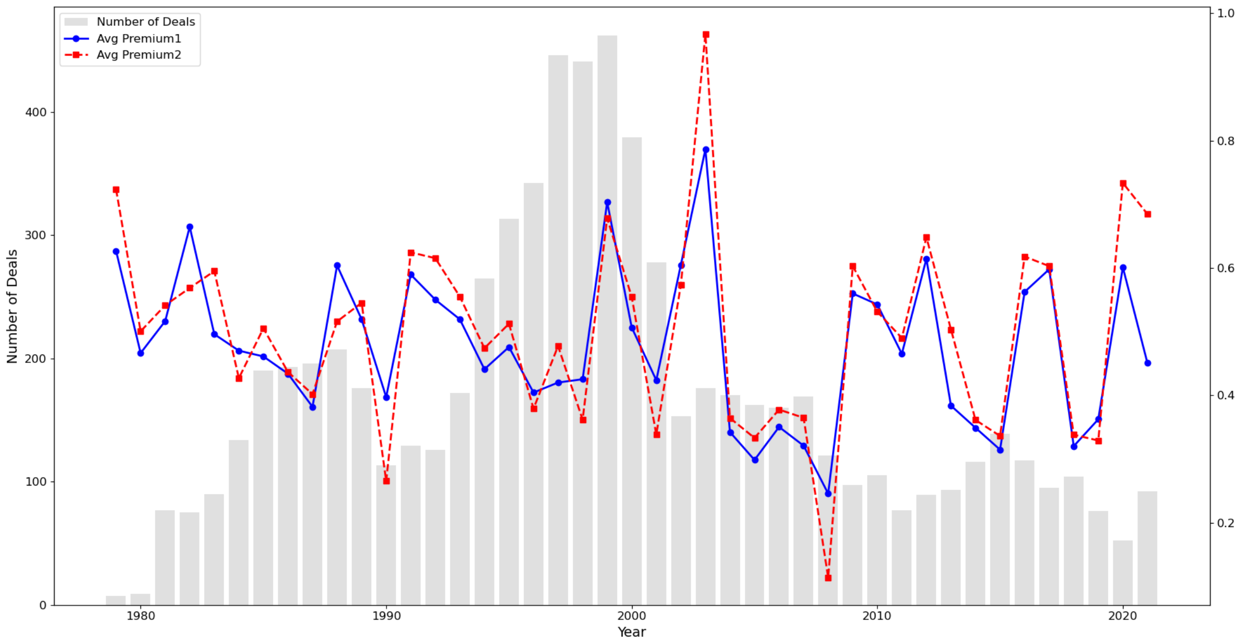
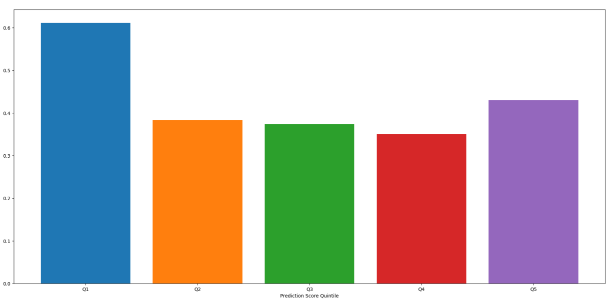
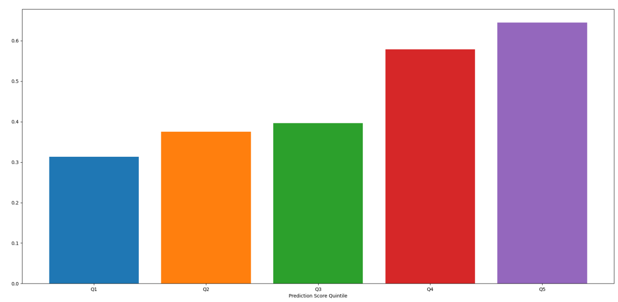


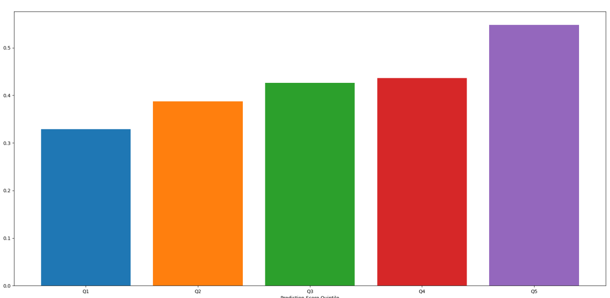
Fig. 1. This figure illustrates the annual number of completed M&A deals and the corresponding average deal premiums. The grey bars (left axis) represent the number of unique transactions per year. The blue solid line and red dashed line (right axis) plot the time-series evolution of the average deal premiums, measured respectively by Premium1 and Premium2. Premiums are computed as the acquisition price relative to the target's pre-announcement market value. The figure highlights the joint dynamics between market activity (deal volume) and the level of takeover premiums.



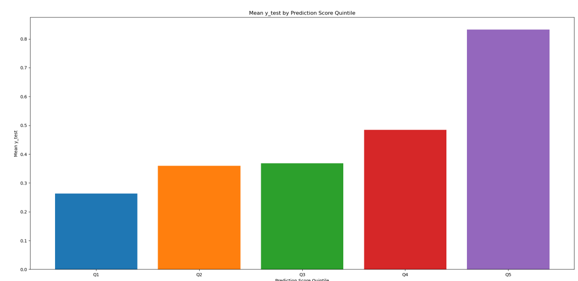
(a) OLS



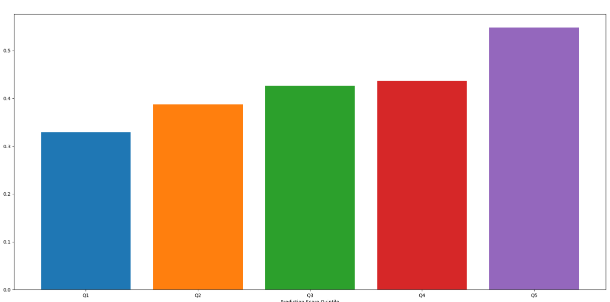
(b) Lasso



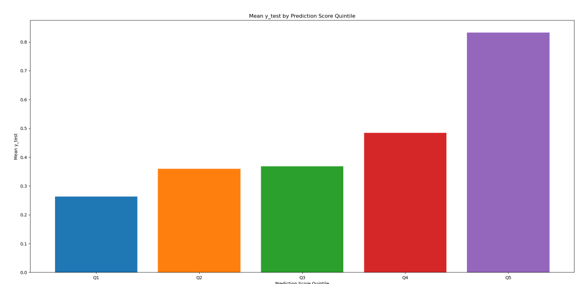
(c) Elastic Net



(d) Random Forest



(e) XGBoost



(f) Neural Network

Fig. 2. This figure illustrates average premiums of M&A deals, in Panels (a) – (f), respectively. We measure the takeover premium as the difference between the price paid per share for the target firm and the target firm's stock price 63 trading days prior to the M&A announcement date. The realized premiums are grouped into quintiles by year by prediction score. The first panel illustrates the predicted M&A premiums using linear OLS predictions, the rest panels utilize linear and non-linear machine learning models.

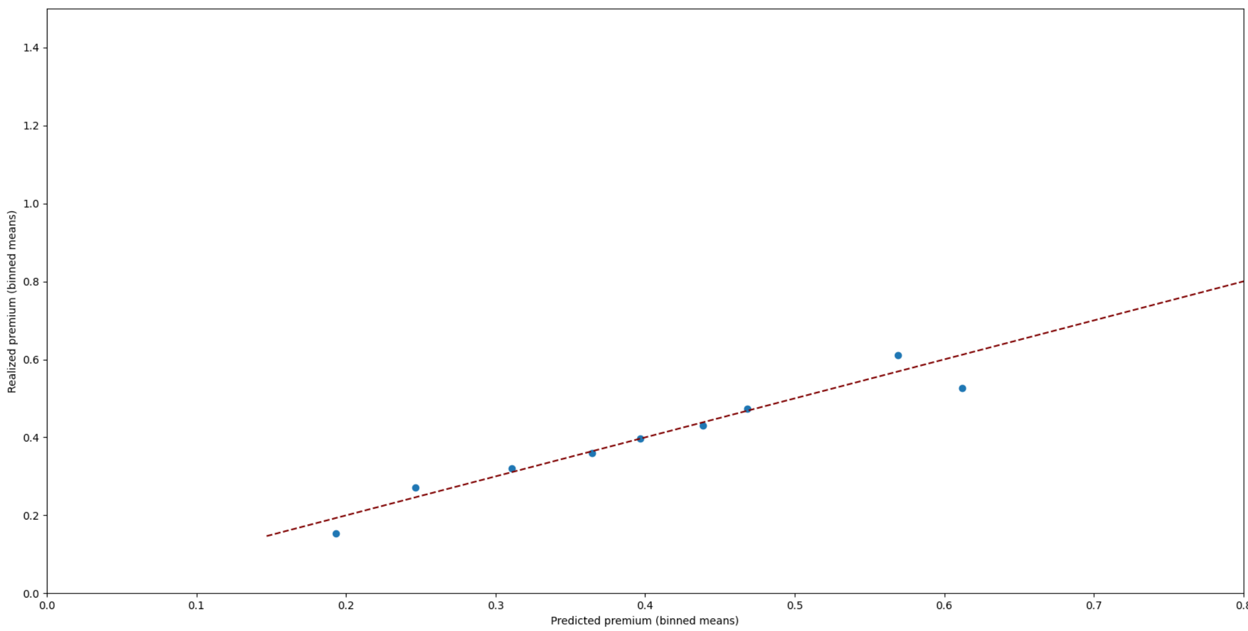


Fig. 3. This figure plots average realized M&A premiums (y-axis) against 10 bins of predicted M&A premiums (x-axis) for the 1437 deals in our test set from 2017 to 2021. Both measures are in ratios. The predictions come from a model trained using 5-fold time-varying cross validation on 5749 deals from 1979 to 2016, restricting the sample to deals with realized premiums.

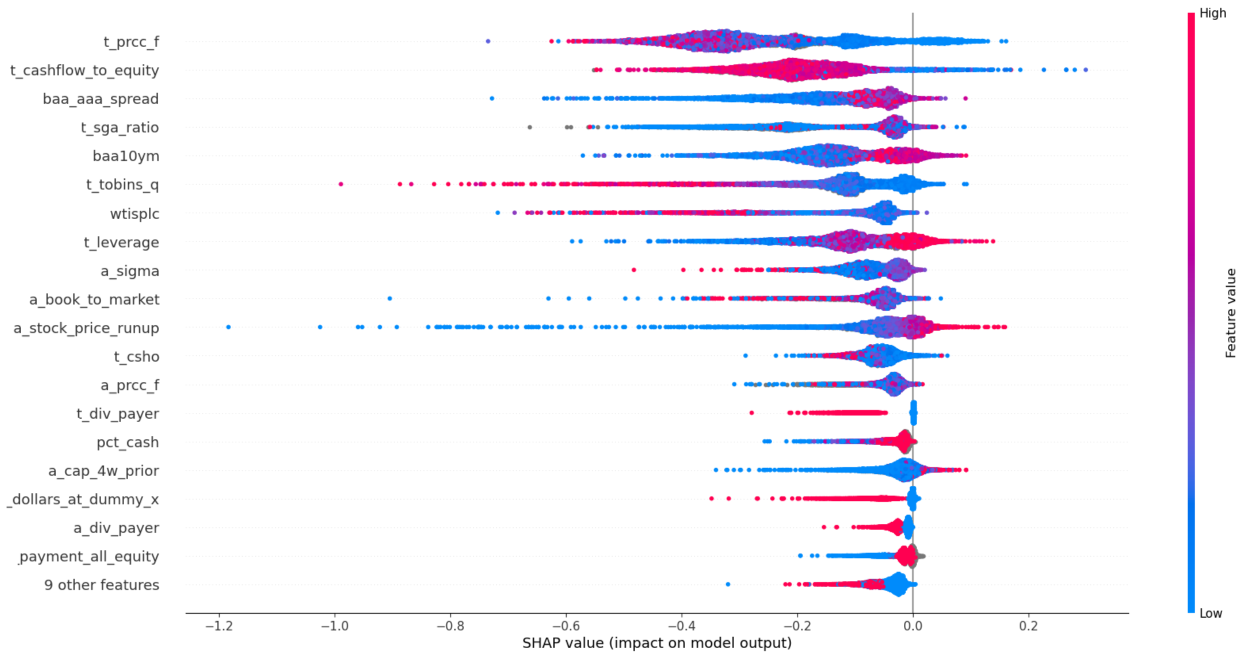


Fig. 4. This figure shows the importance of each feature used in the neural network algorithm. Each dot represents a deal in the test sample. The horizontal axis reports the SHAP value (the *local* contribution of a feature to the model's prediction relative to the model baseline): dots to the right (left) increase (decrease) the predicted premium. Colors encode the feature level in the observation (red = high value, blue = low value). Features are ordered from top to bottom by their overall importance, measured by the mean absolute SHAP value across observations.

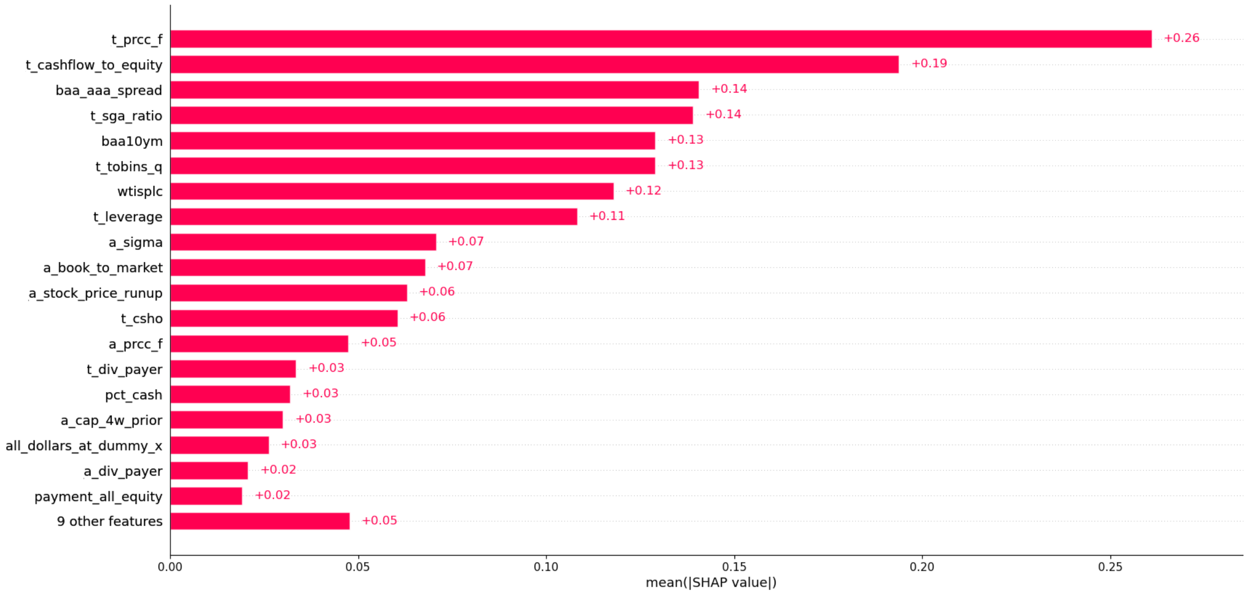


Fig. 5. The bars report mean absolute SHAP values (mean |SHAP|) for the same set of predictors, which rank features by their average *magnitude* of influence on predicted premiums. This metric is sign-agnostic: a large bar indicates a strong effect, positive or negative, on the prediction.

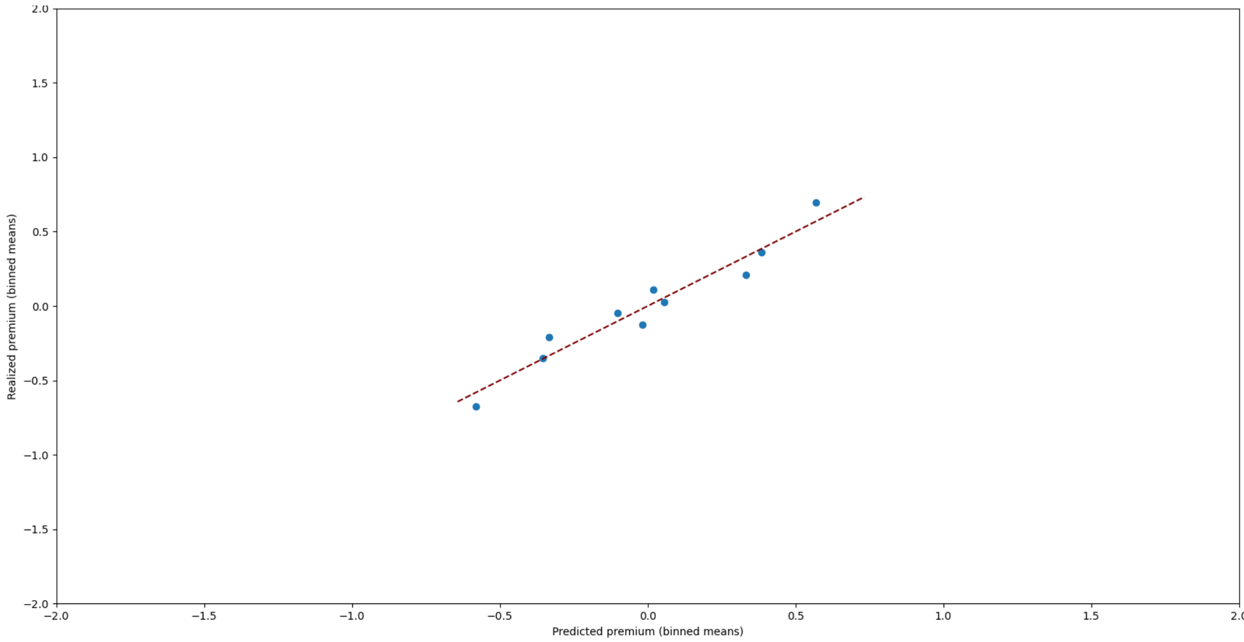


Fig. 6. This figure plots the average realized premium against the average predicted premium across 20 equal-count bins. The dashed line represents perfect calibration. The model preserves a positive relationship between predicted and realized premiums, though with slightly lower calibration accuracy compared to the time-series split.

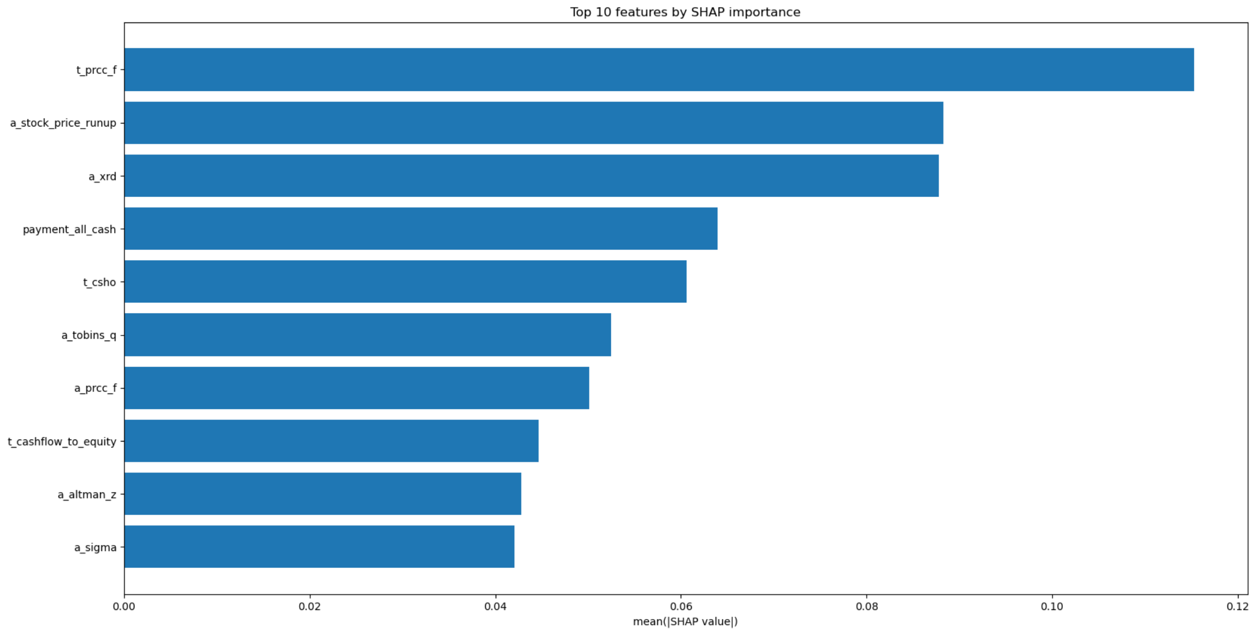


Fig. 7. This figure displays the top ten predictors ranked by mean absolute SHAP values for the XGBoost model using a random train–test split. The most influential features—such as `t_prcc_f`, `a_stock_price_runup`, `payment_all_cash`, and `t_csho`—are broadly consistent with those identified under the time-series split, indicating stable feature importance across sampling schemes.

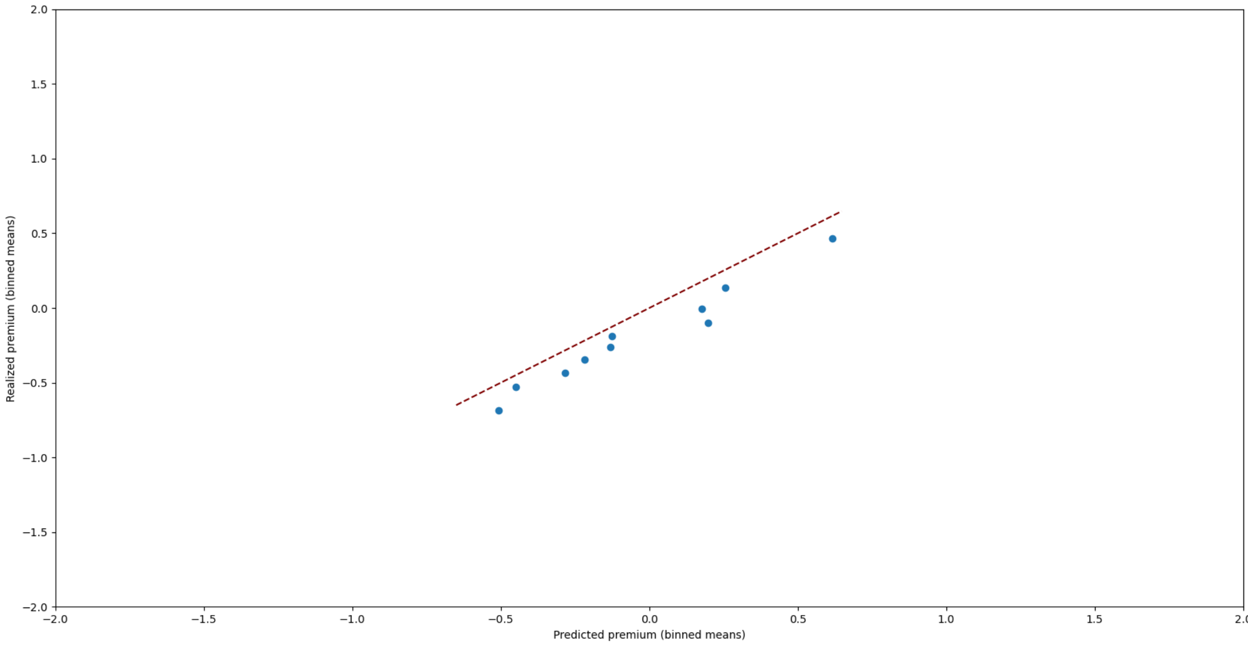


Fig. 8. This figure plots average realized premiums against average predicted premiums across bins of predicted values. Nearly all observations lie below the 45-degree line, indicating systematic underestimation by the XGBoost model. The downward bias highlights the difficulty of predicting a fundamentals-only premium that omits forward-looking information embedded in market prices.

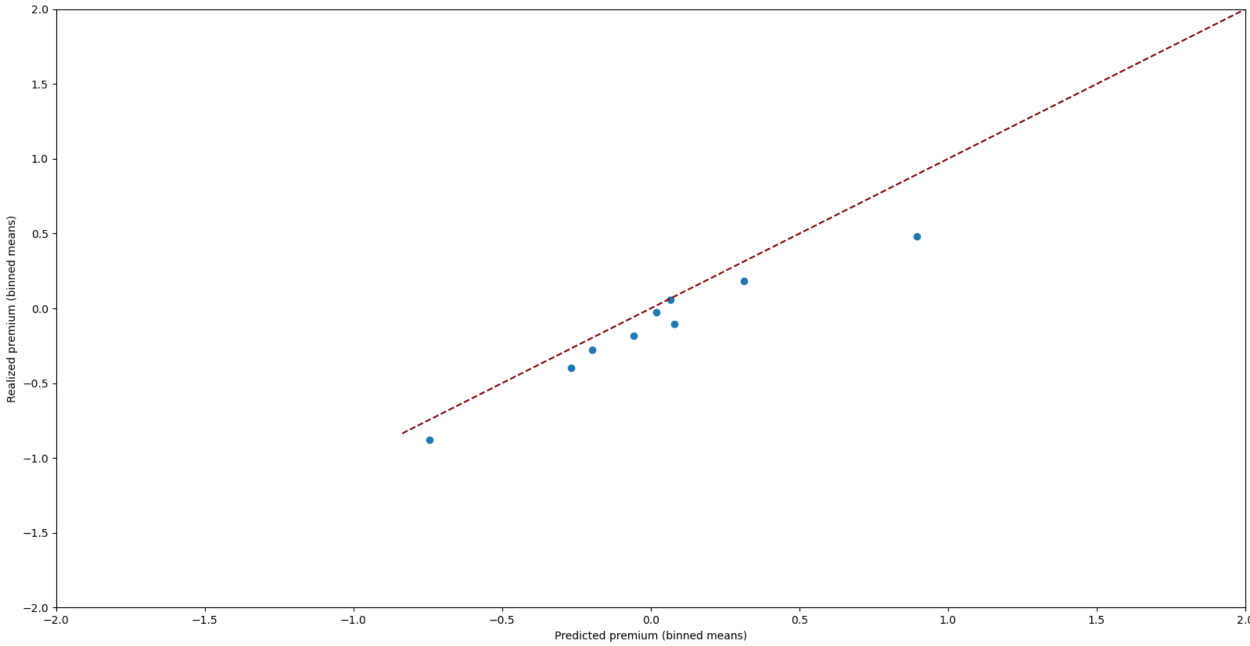


Fig. 9. The realized–predicted relationship is similar when Premium 3 is scaled by book equity: the model persistently underpredicts the level of realized premiums. The bias again reflects the reduced informational content of a purely accounting-based premium measure compared with market-based premiums, which contain richer signals for machine-learning models.

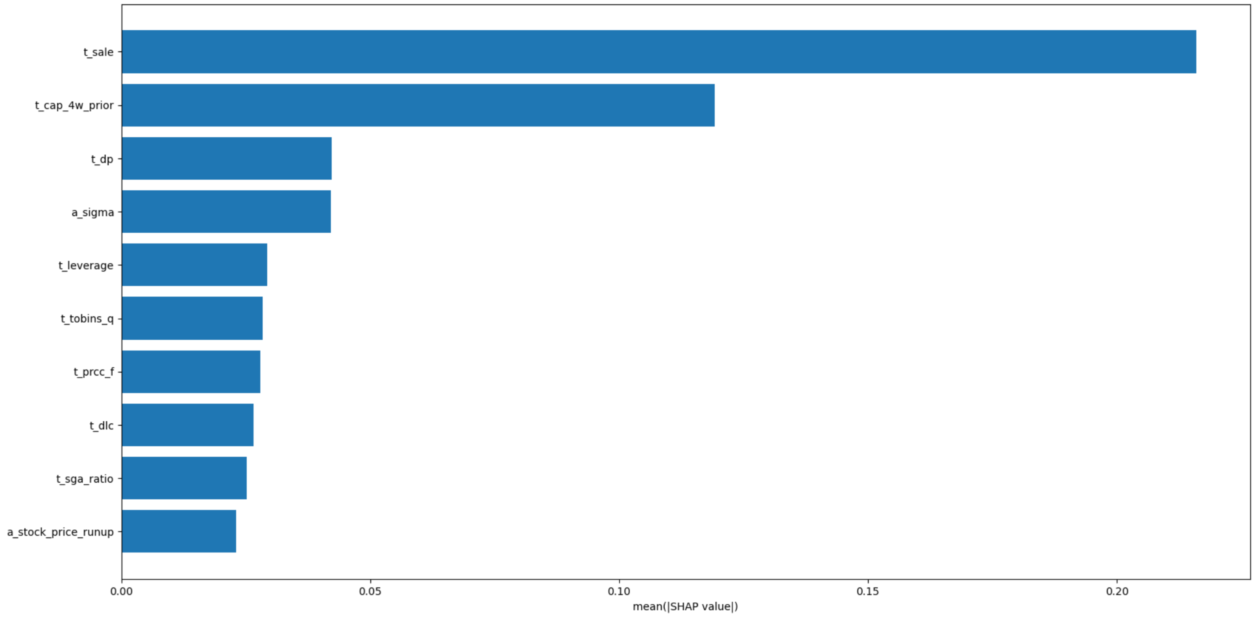


Fig. 10. The figure displays the top predictors in the XGBoost model when Premium 3 is computed using sales. The scaling variable `t_sale` dominates the importance ranking, reflecting its mechanical role in the comparable-transactions multiple.

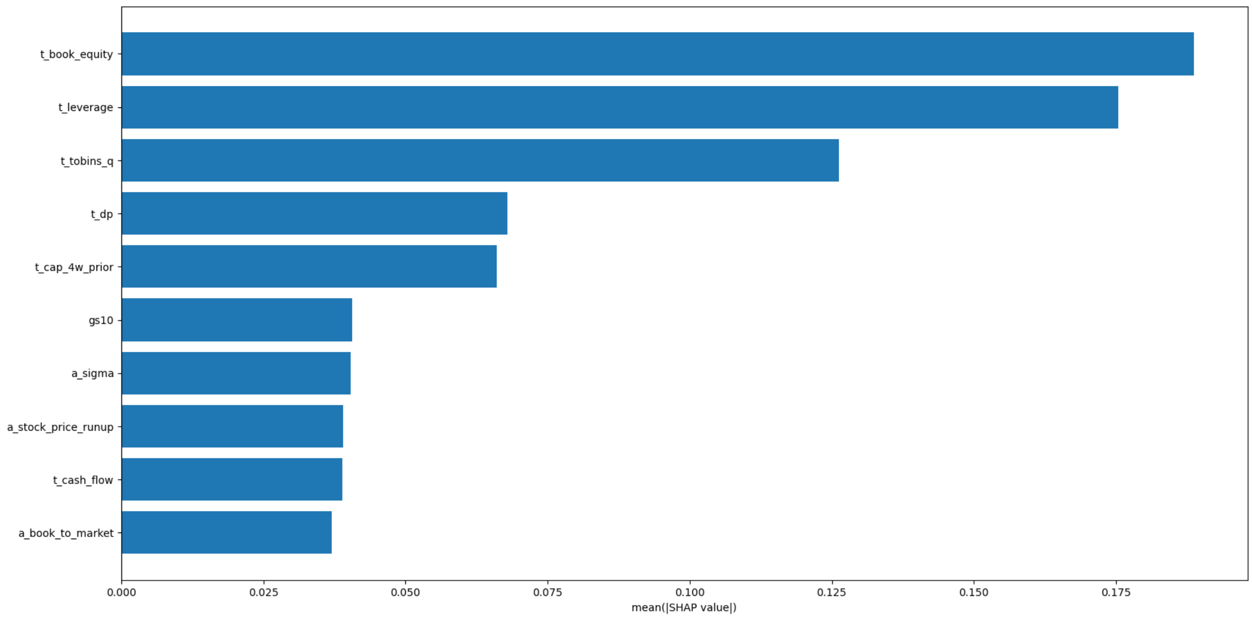


Fig. 11. This figure displays the top predictors in the XGBoost model when Premium 3 is computed using book equity. The scaling variable `t_book_equity` dominates the importance ranking, reflecting its mechanical role in the comparable-transactions multiple.

Table 1: Summary Statistics for the M&A Deal Premiums

This table presents the mean for the premiums of the M&A deals over time, as well as the number of unique public acquirers and targets. Premium 1 is measured as the difference between the price paid per share for the target firm and the target firm's stock price 63 trading days prior to the M&A announcement date. Premium 2 is calculated as the difference between the price paid per share for the target firm and the target firm's stock price 105 trading days prior to the M&A announcement date. The data come from *SDC* and *CRSP*.

	N	Mean premium 1	Mean premium 2	# of unique acquirers	# of unique targets
1979	7	0.627	0.723	7	7
1980	9	0.466	0.501	9	9
1981	77	0.517	0.542	67	66
1982	75	0.665	0.569	70	69
1983	90	0.496	0.595	84	82
1984	134	0.470	0.427	115	120
1985	190	0.461	0.505	165	175
1986	193	0.434	0.437	174	182
1987	196	0.382	0.402	180	179
1988	207	0.604	0.516	187	191
1989	176	0.519	0.545	162	164
1990	113	0.397	0.265	106	110
1991	129	0.590	0.625	116	125
1992	126	0.550	0.616	107	122
1993	172	0.519	0.555	156	160
1994	265	0.441	0.474	227	248
1995	313	0.476	0.512	274	301
1996	342	0.404	0.379	292	328
1997	446	0.420	0.478	374	427
1998	441	0.425	0.362	390	431
1999	462	0.704	0.678	380	450
2000	379	0.507	0.554	340	367
2001	278	0.424	0.339	254	273
2002	153	0.605	0.573	136	151
2003	176	0.787	0.967	160	174
2004	170	0.342	0.364	163	168
2005	162	0.299	0.333	149	156

(continued on next page)

Table 1 (continued from previous page)

2006	160	0.350	0.378	146	153
2007	169	0.321	0.365	160	167
2008	121	0.246	0.114	116	117
2009	97	0.560	0.603	94	94
2010	105	0.543	0.531	99	103
2011	77	0.465	0.490	75	76
2012	89	0.615	0.648	86	89
2013	93	0.384	0.503	88	93
2014	116	0.349	0.361	109	112
2015	139	0.315	0.336	136	137
2016	117	0.562	0.618	106	117
2017	95	0.598	0.603	93	91
2018	104	0.320	0.338	101	101
2019	76	0.362	0.329	76	75
2020	52	0.601	0.734	50	52
2021	92	0.451	0.685	88	89
	7183	0.478	0.499	150.395	160.488

Table 2: Average fraction of lower premiums. This table presents the average fraction of lower premiums for various characteristics of acquirers, targets, deals and financial advisors. A deal is considered to have a low premium if it is in the bottom 10% of the sample.

	Full Sample	Yes	No	Difference p-value
Acquirer level				
High tech	-0.289	-0.342	-0.269	0.000
Dividend payer	-0.289	-0.227	-0.324	0.000
Stock price runup>median	-0.289	-0.262	-0.309	0.005
Sigma>median	-0.289	-0.342	-0.229	0.000
R&D expenditure>median	-0.289	-0.282	-0.314	0.217
Size>median	-0.289	-0.252	-0.318	0.000
Target level				
High tech	-0.289	-0.337	-0.272	0.001
Dividend payer	-0.289	-0.216	-0.313	0.000
Equity price>median	-0.289	-0.249	-0.325	0.000
Shares outstanding>median	-0.289	-0.276	-0.301	0.151
Tobin's q>median	-0.289	-0.308	-0.278	0.107
Cash flow>median	-0.289	-0.256	-0.331	0.000
Common equity>median	-0.289	-0.246	-0.329	0.000
Deal level				
Diversifying	-0.289	-0.285	-0.295	0.519
Large bidder	-0.289	-0.252	-0.312	0.000
All equity	-0.289	-0.294	-0.269	0.216
All cash	-0.289	-0.291	-0.243	0.101
Hostile	-0.289	-0.238	-0.290	0.303
Financial advisor level				
Top tier	-0.289	-0.242	-0.314	0.000

Table 3: Average fraction of higher premiums. This table presents the average fraction of higher premiums for various characteristics of acquirers, targets, deals and financial advisors. A deal is considered to have a low premium if it is in the top 10% of the sample.

	Full Sample	Yes	No	Difference p-value
Acquirer level				
High tech	1.837	2.019	1.764	0.122
Dividend payer	1.837	1.650	1.987	0.007
Stock price runup>median	1.837	1.821	1.831	0.941
Sigma>median	1.837	1.992	1.658	0.016
R&D expenditure>median	1.837	1.924	1.813	0.567
Size>median	1.837	1.849	1.804	0.744
Target level				
High tech	1.837	1.875	1.820	0.717
Dividend payer	1.837	1.439	1.917	0.000
Equity price>median	1.837	1.531	2.176	0.000
Shares outstanding>median	1.837	2.008	1.709	0.041
Tobin's q>median	1.837	1.919	1.880	0.814
Cash flow>median	1.837	1.551	2.207	0.000
Common equity>median	1.837	1.622	2.104	0.001
Deal level				
Diversifying	1.837	1.860	1.794	0.628
Large bidder	1.837	1.817	1.857	0.767
All equity	1.837	1.826	1.792	0.857
All cash	1.837	1.750	1.947	0.384
Hostile	1.837	1.291	1.861	0.000
Financial advisor level				
Top tier	1.837	1.624	1.938	0.006

Table 4: Out-of-sample Performance by Model. This table reports the out-of-sample performance of six predictive models—OLS, Lasso, Elastic Net, Random Forest, XGBoost, and Neural Network—across three premium definitions: *Premium 1*, *Premium 2*, and *Premium 3*. For each premium measure, we present the root mean squared error (RMSE) and the out-of-sample coefficient of determination (R^2_{oos}). Higher R^2_{oos} and lower RMSE indicate better predictive performance.

	Premium 1		Premium 2		Premium 3	
Model	**RMSE**	$\boldsymbol{R^2_{\text{oos}}}$	**RMSE**	$\boldsymbol{R^2_{\text{oos}}}$	**RMSE**	$\boldsymbol{R^2_{\text{oos}}}$
OLS	0.985	-0.332	1.020	-0.250	1.908	-0.051
Lasso	0.892	-0.028	0.903	-0.023	0.979	0.033
Elastic Net	0.856	-0.017	0.915	-0.025	0.912	0.038
Random Forest	0.688	0.054	0.781	0.052	0.552	0.076
XGBoost	0.558	0.151	0.571	0.143	0.328	0.165
Neural Network	0.589	0.137	0.596	0.141	0.445	0.176

Table 5: Realized premiums by predicted percentile and model. This table reports the average realized M&A premiums in the test sample across predicted-premium percentiles for six forecasting models: OLS, Lasso, Elastic Net, Random Forest, XGBoost, and Neural Network. For each model, deals are sorted into percentile groups based on their predicted premium (e.g., bottom 1%, bottom 5%, bottom 10%, top 90%, top 95%, and top 100%). Panels A–C present results separately for Premium 1, Premium 2, and Premium 3. Higher realized values in higher predicted percentiles indicate stronger predictive ordering ability.

	Predicted Percentile of Premiums	OLS	Lasso	Elastic Net	Random Forest	XGBoost	Neural Network
Panel A: Premium 1							
Deals predicted to have low premiums	1%	0.371	-0.085	-0.082	-0.163	-0.173	-0.158
	$< 5\%$	0.689	-0.033	0.093	0.058	0.063	0.052
	10%	0.523	0.096	0.225	0.121	0.128	0.113
Deals predicted to have high premiums	90%	0.656	0.729	1.134	1.216	1.241	1.258
	$> 95\%$	0.680	0.701	1.301	1.339	1.358	1.419
	100%	0.772	0.797	1.371	1.325	1.373	1.438
Panel B: Premium 2							
Deals predicted to have low premiums	1%	0.286	-0.014	-0.115	-0.104	-0.094	-0.082
	$< 5\%$	0.551	0.078	0.107	0.092	0.052	0.078
	10%	0.539	0.136	0.215	0.175	0.183	0.131
Deals predicted to have high premiums	90%	0.662	0.922	1.923	1.716	1.705	1.632
	$> 95\%$	0.691	1.035	1.981	1.830	1.762	1.719
	100%	0.730	1.018	2.034	1.957	1.770	1.825
Panel C: Premium 3							
Deals predicted to have low premiums	1%	-1.328	-2.117	-2.082	-2.910	-3.524	-3.262
	$< 5\%$	-0.851	-1.987	-2.093	-2.804	-1.530	-1.108
	10%	0.413	-1.860	-1.225	-1.776	-1.487	-0.744
Deals predicted to have high premiums	90%	0.565	0.952	2.134	2.228	2.457	2.268
	$> 95\%$	0.472	1.046	2.301	2.320	2.484	2.426
	100%	0.579	1.377	2.371	2.487	3.608	4.969

Table 6: Predicted Premium Changes from Advisor Replacement by Decile. The table reports average model-predicted M&A premium by decile of the baseline predicted premium. "Avg. Pred. Prem." is the mean predicted premium under the observed advisor assignment. "Avg. Pred. Prem. (Repl.)" is the mean predicted premium after replacing the relevant financial advisor(s) as indicated in each panel. Δ Pred. Prem. equals "(Repl.)" minus the baseline value; negative values indicate a decline in predicted premium after replacement. N denotes the number of observations in each decile.

Decile	N	Avg. Pred. Prem.	Avg. Pred. Prem. (Repl.)	Δ Pred. Prem.
Panel A: Target advisors				
1	46	0.2922	0.2874	-0.0048
2	55	0.2656	0.2625	-0.0031
3	60	0.2809	0.2768	-0.0041
4	51	0.3156	0.3139	-0.0017
5	52	0.3272	0.3241	-0.0031
6	64	0.3432	0.3422	-0.0011
7	55	0.3478	0.3441	-0.0037
8	48	0.3614	0.3589	-0.0025
9	48	0.3966	0.3938	-0.0028
10	55	0.4168	0.4123	-0.0045
Panel B: Acquirer advisors				
1	53	0.2951	0.2920	-0.0031
2	53	0.2710	0.2692	-0.0019
3	70	0.3014	0.2998	-0.0017
4	62	0.3135	0.3124	-0.0011
5	65	0.3090	0.3074	-0.0016
6	67	0.3558	0.3550	-0.0009
7	65	0.3542	0.3531	-0.0011
8	67	0.3687	0.3676	-0.0010
9	59	0.3849	0.3824	-0.0025
10	56	0.4180	0.4149	-0.0031
Panel C: Both advisors				
1	70	0.2941	0.2834	-0.0107
2	84	0.2689	0.2594	-0.0094
3	92	0.2981	0.2886	-0.0095
4	80	0.3112	0.3025	-0.0087
5	86	0.3230	0.3145	-0.0085
6	89	0.3495	0.3430	-0.0065
7	91	0.3560	0.3478	-0.0082
8	83	0.3694	0.3616	-0.0079
9	80	0.3929	0.3838	-0.0091
10	78	0.4167	0.4063	-0.0104

Table 7: Predicted Premium (Premium2) Changes from Advisor Replacement by Decile. The table reports average model-predicted M&A premium by decile of the baseline predicted premium. "Avg. Pred. Prem." is the mean predicted premium under the observed advisor assignment. "Avg. Pred. Prem. (Repl.)" is the mean predicted premium after replacing the relevant financial advisor(s) as indicated in each panel. Δ Pred. Prem. equals "(Repl.)" minus the baseline value; negative values indicate a decline in predicted premium after replacement. N denotes the number of observations in each decile.

Decile	N	Avg. Pred. Prem.	Avg. Pred. Prem. (Repl.)	Δ Pred. Prem.
Panel A: Target advisors				
1	52	0.2662	0.2657	-0.0005
2	60	0.2645	0.2639	-0.0005
3	56	0.3078	0.3073	-0.0005
4	51	0.3107	0.3100	-0.0007
5	48	0.3341	0.3335	-0.0006
6	55	0.3292	0.3287	-0.0006
7	51	0.3663	0.3657	-0.0007
8	53	0.3608	0.3603	-0.0004
9	51	0.3687	0.3680	-0.0007
10	55	0.3932	0.3922	-0.0010
Panel B: Acquirer advisors				
1	52	0.2532	0.2487	-0.0045
2	62	0.2730	0.2710	-0.0020
3	60	0.2754	0.2719	-0.0035
4	51	0.3164	0.3127	-0.0037
5	64	0.3232	0.3203	-0.0029
6	73	0.3314	0.3303	-0.0011
7	63	0.3600	0.3573	-0.0028
8	66	0.3714	0.3685	-0.0028
9	58	0.3689	0.3627	-0.0062
10	63	0.4098	0.4003	-0.0095
Panel C: Both advisors				
1	74	0.2635	0.2409	-0.0225
2	91	0.2678	0.2459	-0.0219
3	86	0.2899	0.2681	-0.0219
4	72	0.3204	0.2968	-0.0236
5	78	0.3279	0.3049	-0.0230
6	95	0.3314	0.3086	-0.0228
7	83	0.3629	0.3381	-0.0248
8	83	0.3622	0.3375	-0.0247
9	79	0.3633	0.3381	-0.0252
10	88	0.4027	0.3683	-0.0344

Appendix A. Machine Learning

Machine learning provides a set of computational tools designed to detect hidden patterns in data and generate predictions. Compared with traditional econometric approaches, machine learning is well suited for large-scale and high-dimensional datasets, requires fewer parametric assumptions, and flexibly captures nonlinearities and complex interactions. These advantages have contributed to its rapid growth in both academic research and industry applications over the past decade. Beyond numerical prediction, machine learning has become a cornerstone of text analysis and natural language processing, with large language models (e.g., ChatGPT, Bard) exemplifying recent advances.

The central strength of machine learning lies in its ability to combine predictive accuracy with adaptability. It encompasses a wide spectrum of methods, ranging from regularized linear regression to tree-based ensembles and deep learning architectures. Each class of methods differs in how it balances interpretability, complexity, and flexibility in capturing nonlinear associations. Several papers in the recent economics and finance literature have used machine learning techniques. In a seminal paper, Kleinberg et al. (2018) study judges' bail decisions and show that machine predictions could significantly reduce crime. Machine learning is quickly being adopted as a new methodology in the asset pricing literature (e.g., Rossi (2018), Gu et al. (2020), and Abis (2020) among others) and microstructure (e.g., Easley et al. (2021)). Corporate finance applications are developing (see, e.g., Li et al., 2021; Bubb and Catan, 2022; Easley et al., 2021).

In this study, we apply supervised machine learning algorithms to predict the premiums of M&A deals. Our predictors include observable firm-level characteristics of public acquirers and targets, deal features, and macroeconomic indicators available prior to the deal announcement (see Appendix A). Specifically, we implement and compare five widely used algorithms: *Lasso*, *Elastic Net*, *Random Forest*, *Gradient Boosting Trees (XGBoost)*, and *Neural Networks*. These methods allow us to evaluate how linear, ensemble, and deep learning approaches perform in modeling the determinants of M&A premiums.

A.1. Penalized Linear Models

Traditional linear regression becomes unreliable when the number of predictors is large relative to the sample size. In such settings, models tend to overfit noise rather than identify meaningful signals, a well-known issue in return prediction where the signal-to-noise ratio is low. Penalized linear models address this challenge by imposing regularization penalties on coefficient estimates, thereby discouraging overly complex specifications.

The elastic net penalty, defined as

$$\phi(\theta; \lambda, \rho) = \lambda(1-\rho)\sum_{j=1}^{P}|\theta_j| + \frac{1}{2}\lambda\rho\sum_{j=1}^{P}\theta_j^2, \tag{4}$$

balances between the *lasso* ($\rho = 0$) and *ridge* ($\rho = 1$) penalties. *Lasso* sets some coefficients exactly to zero, enabling variable selection, whereas ridge shrinks coefficients toward zero without eliminating them. *Elastic net* combines both effects, improving performance when predictors are highly correlated.

Key hyperparameters are the penalty strength λ and mixing parameter ρ, optimized using a validation sample. We implement estimation via the accelerated proximal gradient algorithm under both least squares and Huber loss functions. In our application, penalized regression provides a benchmark model that balances interpretability with robustness to overfitting.

A.2. Random Forest

Random forests are ensemble learners that average predictions from many decision trees. Each tree is trained on a bootstrap sample of the data, and splits are chosen recursively to minimize prediction error. While individual trees tend to overfit, averaging across trees reduces variance and improves predictive stability. It is a variation on a more general procedure known as bootstrap aggregation, or "bagging" (Breiman, 2001).

A distinctive feature of *random forests* is the random subset of predictors considered at each split, which reduces correlation across trees. This "feature dropout" mechanism prevents dominant variables from driving all tree splits, thereby enhancing ensemble diversity. The main hyperparameters are the number of trees (B) and the maximum tree depth (L), which are tuned via validation.

In the context of M&A premiums, *random forests* capture nonlinear effects and higher-order interactions among firm, deal, and macroeconomic characteristics, while their ensemble structure provides robustness to noise in individual predictors.

A.3. Gradient Boosting Trees (XGBoost)

Boosting methods sequentially build trees such that each new tree focuses on correcting the errors of the previous ensemble. This iterative procedure allows boosting to reduce bias and capture complex, nonlinear relationships that bagging alone cannot address. Boosting is originally described in Schapire (1990) and Freund (1995) for classification problems to improve the performance of a set of weak learners. Friedman et al. (2000) and Friedman (2001) extend boosting to contexts beyond classification, eventually leading to the gradient boosted regression tree.

XGBoost (Extreme Gradient Boosting) is a widely adopted implementation of gradient boosting that introduces innovations such as regularization to prevent overfitting, shrinkage (learning rate) to control the impact of each tree, and subsampling of rows and predictors to improve generalization. Computational optimizations and parallelization make *XGBoost* particularly efficient for large, high-dimensional datasets.

Key hyperparameters include the number of boosting iterations, maximum tree depth, learning rate, and subsampling ratios, selected using validation-sample performance. For M&A premiums, *XGBoost*'s ability to capture nonlinearities and interaction effects across heterogeneous predictors makes it a powerful tool, while its built-in regularization controls model complexity.

A.4. Neural Networks

Artificial neural networks represent the most flexible class of models in our analysis. They consist of an input layer of predictors, one or more hidden layers that apply nonlinear transformations, and an output layer that aggregates the results (Hornik et al., 1989; Cybenko, 1989). Each hidden layer contains neurons, with weighted connections ("synapses") transmitting signals between layers. Stacking multiple layers enables the network to approximate highly complex functional relationships.

The key hyperparameters include the number of hidden layers, the number of neurons per layer, the activation functions, and the optimization algorithm used for training. These are selected via validation to balance predictive accuracy against computational cost and overfitting risk.

Neural networks are "universal approximators" capable of modeling virtually any smooth predictive relationship. In our application, they allow us to capture intricate nonlinearities in M&A premium determination, though at the cost of reduced interpretability compared with tree-based or penalized linear methods.

A.5. Performance Evaluation Metrics

Following prior work in return and premium prediction (e.g., Gu, Kelly, and Xiu 2020; Feng, Giglio, and Xiu 2022), we evaluate model performance using the out-of-sample R^2 and the root mean squared error (RMSE). These two complementary metrics capture both relative explanatory power and absolute predictive accuracy.

The out-of-sample R^2 compares the predictive accuracy of the model against a naive benchmark of zero premium, following the convention in the asset pricing literature where mean-based benchmarks are unreliable due to high noise in individual returns (see Welch and Goyal 2008). Specifically, for the testing sample T,

$$R^2_{\text{oos}} = 1 - \frac{\sum_{i \in T}(y_i - \hat{y}_i)^2}{\sum_{i \in T} y_i^2}, \tag{5}$$

where y_i is the observed M&A premium and $\hat{y}_i$ is the model prediction. Unlike the conventional in-sample R^2, this metric evaluates only on data that were never used for model estimation or hyperparameter tuning, thereby providing a more stringent assessment of predictive power. A positive R^2_{oos} indicates that the model improves upon the naive forecast of zero, while negative values imply underperformance.

The RMSE measures the absolute magnitude of prediction errors in the original units of the dependent variable. It is defined as

$$\text{RMSE} = \sqrt{\frac{1}{|T|} \sum_{i \in T} (y_i - \hat{y}_i)^2}, \tag{6}$$

where $|T|$ is the number of test observations. RMSE is particularly informative for gauging the economic magnitude of prediction errors, since it expresses forecast errors directly in terms of M&A premium percentage points. Lower RMSE values indicate better predictive accuracy.

By jointly reporting R^2_{oos} and RMSE, we provide a comprehensive evaluation of model performance. The former captures relative improvements over a naive benchmark, while the latter quantifies the absolute scale of forecast errors. Together, they allow us to assess not only whether the models extract signal from firm, deal, and macroeconomic characteristics, but also whether such predictions are economically meaningful in magnitude.

The four classes of machine learning models considered, penalized linear regression, random forests, gradient-boosting trees, and neural networks, represent complementary approaches to prediction. Penalized regression provides interpretability and variable selection in high-dimensional settings. Random forests capture nonlinearities and interactions while remaining robust to noise. Gradient boosting offers enhanced predictive accuracy through sequential refinement and regularization. Neural networks provide the most flexible architecture, capable of approximating complex functional relationships at the cost of reduced transparency. Together, these models allow us to assess the relative strengths of linear, ensemble, and deep learning methods in predicting M&A deal premiums.

Appendix B. Data Definitions

Variables	Description
Firm characteristics	
Current Assets	Firm's total current assets in the fiscal year preceding the acquisition announcement date. Information is from *Compustat*. [**act**]
Total Assets	Firm's total assets in the fiscal year preceding the acquisition announcement date. Information is from *Compustat*. [**at**]
Long-term Debt	Long-term debt obligations in the fiscal year preceding the acquisition announcement date. Information is from *Compustat*. [**dltt**]
Short-term Debt	Short-term debt, including current portion of long-term debt and other borrowings in the fiscal year preceding the acquisition announcement date. Information is from *Compustat*. [**dlc**]
Book Debt	Book value of total debt as defined by Fama and French (2000), calculated as the sum of long-term debt, short-term debt, and preferred stock at liquidating value in the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**dltt + dlc + pstkl**]
Book Equity	Book value of equity as defined by Fama and French (2000), calculated as common equity plus deferred taxes and investment tax credit, minus preferred stock. Measured in the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**ceq + txdb - pstk**]
Net Income	Firm's net earnings after all expenses, including taxes and interest, during the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**ni**]
Sales	Total net sales or revenues generated by the firm during the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**sale**]

(continued on next page)

Table 8 (continued from previous page)

Depreciation	Depreciation and amortization expenses recorded in the fiscal year preceding the acquisition announcement. Represents non-cash charges for asset wear and obsolescence. Information is from *Compustat*. [**dp**]
Dividends	Total amount of dividends paid to common (ordinary) shareholders during the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**dvc**]
Total Dividends	Total amount of dividends paid to both common (ordinary) shareholders and preferred shareholders during the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**dvt**]
Dividend Payer	Indicator variable equal to one if the firm paid dividends to common shareholders during the fiscal year preceding the acquisition announcement (i.e., **dvc** > 0), and zero otherwise. Information is from *Compustat*.
Div Ratio	Ratio of dividends paid to common shareholders to the firm's EBITDA in the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**dvc / ebitda**]
Stock Repurchase	Cash spent by the firm to repurchase its own common and/or preferred shares during the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**prstkc**]
Div Stock Repurchase	Combined shareholder payout, measured as the sum of dividends and stock repurchases scaled by total assets. Information is from *Compustat*. [**(dvc + prstkc) / at**]
CAPX	Capital expenditures by the firm during the fiscal year prior to the acquisition announcement. Information is from *Compustat*. [**capx**]
SG&A	Selling, General, and Administrative expenses incurred by the firm during the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**xsga**]

(continued on next page)

SG&A Ratio	Selling, General, and Administrative (SG&A) expenses scaled by total assets. Information is from *Compustat*. [**xsga / at**]
Extraordinary Items	Extraordinary income or loss reported by the firm during the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**xi**]
R&D	Research and Development (R&D) expenses incurred by the firm during the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**xrd**]
R&D Ratio	Research and Development (R&D) expenses scaled by total assets the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**xrd / at**]
Cash	Cash held by the firm at the end of the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**ch**]
CHE	Cash and short-term investments (cash equivalents) held by the firm. Information is from *Compustat*. [**che**]
EBIT	Earnings Before Interest and Taxes. Information is from *Compustat*. [**ebit**]
EBITDA	Earnings Before Interest, Taxes, Depreciation, and Amortization. Information is from *Compustat*. [**ebitda**]
Cash Flow	Approximation of operating cash flow calculated as EBITDA minus taxes and interest expense, scaled by total assets. Information is from *Compustat*. [**(ebitda - txt - xint) / at**]
Working Cap	Working capital, calculated as current assets minus current liabilities in the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**wcap**]
Working Cap Ratio	Working capital scaled by total assets for the fiscal year preceding the acquisition announcement. Information is from *Compustat*. [**wcap / at**]

(continued on next page)

Table 8 (continued from previous page)

Size	Market value of firm's equity four weeks prior to the acquisition announcement date obtained from SDC. Information market value of equity is obtained from *CRSP*. [**prcc_f * csho**]
Stock Price Runup	*CRSP* value-weighted index adjusted buy-and-hold abnormal return (BHAR) of the firm's stock over the [-205, -6] event window relative to the acquisition announcement date obtained from SDC. Stock price data is from *CRSP*.
Sigma	Standard deviation of the firm's CRSP value-weighted index adjusted buy-and-hold abnormal return (BHAR) over the [-205, -6] event window relative to the acquisition announcement date obtained from SDC. Stock price data is from *CRSP*.
Book Leverage	Total debt (current liabilities plus long-term debt) scaled by book value of total assets in the fiscal year preceding the acquisition announcement date obtained from SDC. Information is from *Compustat*. [**(dlc + dltt) / at**]
Tobin's Q	Market value of the firm's assets divided by book value of its assets in the fiscal year preceding the acquisition announcement date obtained from SDC. The market value of assets is calculated as the sum of the book value of assets and market value of common stock minus the book value of common stock minus deferred taxes in the balance sheet. The data are from *CRSP* and *Compustat*. [**(at + prcc_f * csho - ceq - txdb) / at**]
ROA	Firm's net income divided by the book value of its total assets for the fiscal year preceding the acquisition announcement date obtained from SDC. Information is from *Compustat*. [**ni / at**]

(continued on next page)

Table 8 (continued from previous page)

Book-to-Market	Firm's book value of equity (in the fiscal year before the acquisition announcement) divided by the market value of equity four weeks preceding the acquisition announcement. The data are from *CRSP* and *Compustat*. **[ceq / (prcc_4w * csho)]**
Cash Flows-to-Equity	Income before extraordinary items plus depreciation minus dividends scaled by the book value of assets in the fiscal year before the acquisition announcement date obtained from SDC. Information is from *Compustat*. **[(ib + dp - dvc) / at]**
Leverage	Ratio of total liabilities to total assets for a firm, calculated using balance sheet data in the fiscal year preceding the acquisition announcement. Information is from *Compustat*. **[lt / at]**
Retained Earnings	Cumulative profits retained by the firm after dividend payments in the fiscal year preceding the acquisition announcement. Information is from *Compustat*. **[re]**
OIADP	Operating income after depreciation for the fiscal year preceding the acquisition announcement. Information is from *Compustat*. **[oiadp]**
Altman Z-Score	$Z = \left(\frac{1.2\cdot(\mathrm{act}-\mathrm{lct})-\mathrm{che}}{\mathrm{at}}\right) + \left(\frac{1.4\cdot\mathrm{re}}{\mathrm{at}}\right) + \left(\frac{3.3\cdot\mathrm{oiadp}}{\mathrm{at}}\right) + \left(\frac{0.6\cdot(\mathrm{csho}\cdot\mathrm{prcc\_f})}{\mathrm{lt}}\right) + \left(\frac{\mathrm{sale}}{\mathrm{at}}\right).$ Information on the market value of equity is from *CRSP* and *Compustat*.
High Tech	Indicator variable is one if the acquirer operates in a high-tech industry as defined in Appendix 4 of Loughran and Ritter (2004), zero otherwise. Information is from *Compustat*.
Large Bidders	Indicator variable is one if the market value of an acquirer's equity is above the sample median of this measure, zero otherwise. Information on the market value of equity is from *CRSP*.
Deal characteristics	

(continued on next page)

Table 8 (continued from previous page)

Hostile	Indicator variable is one for hostile acquisitions, zero for unsolicited acquisitions. Information is from *SDC*.
Payment-All Cash	Indicator variable is one if the acquisition is paid for with all cash, zero otherwise. Information is from *SDC*.
% Cash Financing	Percentage of the deal value paid in cash. Information is from *SDC*.
Payment-All Stock	Indicator variable is one if the acquisition is paid for with all stock, zero otherwise. Information is from *SDC*.
Payment-Includes Stock	Indicator variable is one if the acquisition is paid for with some equity, zero otherwise. Information is from *SDC*.
% Stock Financing	Percentage of the deal value paid in stock. Information is from *SDC*.
Diversifying	Indicator variable is one if the acquirer and target do not belong to the same two-digit SIC code, zero otherwise. Information is from *SDC* and *Compustat*.
Financial advisor characteristics	
Top tier Advisor	Indicator variable is one if the acquirer retained a top-tier investment bank for an acquisition or a target, zero otherwise. To define top-tier banks, we calculate the total value of deals advised by each investment bank over 1979 and 2021 and then define an investment bank as top-tier if it ranks in the top 5 based on this measure. Information is from *SDC*.
Macro variables	
Industrial Production	Monthly Industrial Production Index, lagged one month. Information is from *St. Louis Fed.* [***indpro***]
Consumer Price Index	Monthly Consumer Price Index, lagged one month. Information is from *St. Louis Fed.* [***cpiaucsl***]
Spot Oil Price (WTI)	Monthly West Texas Intermediate crude oil spot price, lagged one month. Information is from *St. Louis Fed.* [***wtisplc***]

(continued on next page)

Table 8 (continued from previous page)

3-month T-bill Rate	Monthly 3-month U.S. Treasury Bill rate, lagged one month. Information is from *St. Louis Fed.* [***tb3ms***]
10-year Treasury Yield	Monthly yield on 10-year U.S. Treasury bonds, lagged one month. Information is from *St. Louis Fed.* [***gs10***]
AAA – 10Y Spread	Yield spread between Moody's AAA corporate bonds and 10-year Treasuries, lagged one month. Information is from *St. Louis Fed.* [***aaa10ym***]
BAA – 10Y Spread	Yield spread between Moody's Baa corporate bonds and 10-year Treasuries, lagged one month. Information is from *St. Louis Fed.* [***baa10ym***]
BAA – AAA Spread	Credit spread between Baa and Aaa corporate bond yields, lagged one month. Information is from *St. Louis Fed* data.