# Revisiting M&A Premium: New Evidence from Machine Learning

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.

*Key words*: Mergers and Acquisitions, Machine Learning, Explainable Artificial Intelligence

## 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. 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 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

### 2.1. Mergers and Acquisitions

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.

### 2.2. 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.

**2.2.1. 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

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

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 $\lambda$ and mixing parameter $\rho$, 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.

**2.2.2. 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.

**2.2.3. 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.

**2.2.4. 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.

**2.2.5. 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 $T$,

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

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 $R^2_{\text{oos}}$ 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

$$\text{RMSE} = \sqrt{\frac{1}{|T|} \sum_{i \in T} (y_i - \hat{y}_i)^2}, \tag{3}$$

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 $R^2_{\text{oos}}$ 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.

## 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

$$\text{Premium}_{63} = \frac{\text{OfferPrice}_i - P_i(-63)}{P_i(-63)}, \tag{4}$$

where $\text{OfferPrice}_i$ 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):

$$\text{Premium}_{105} = \frac{\text{OfferPrice}_i - P_i(-105)}{P_i(-105)}, \tag{5}$$

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 $\mathrm{Mult}_i$ denote the focal deal's value-to-sales multiple and $\overline{\mathrm{Mult}}_{\mathcal{C}(i)}$ the average multiple of its matched set $\mathcal{C}(i)$. The multiples-based premium is

$$\mathrm{Premium}_i^{\mathrm{CT}} = \frac{\mathrm{Mult}_i - \overline{\mathrm{Mult}}_{\mathcal{C}(i)}}{\overline{\mathrm{Mult}}_{\mathcal{C}(i)}}. \tag{6}$$

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 $\mathrm{Premium}_{63}$ as the outcome, with robustness to $\mathrm{Premium}_{105}$ and the comparable-transactions measure in (6). 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 ($\Sigma$) 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.

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 Table 1 and 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.

[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.[1] 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 can be done, although 70%/30% or 80%/20% splits are typically used in practice.[2] 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. Research Design and Model Selection

We identify a number of machine learning models and methodologies to determine if machine learning can assist in predicting M&A takeover premium. Given our relatively limited sample size, it is not clear that machine learning will be able to accurately identify better or worse deal premium

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

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

for acquirer shareholders. As such, we start with a number of models and standardization methods and then continue tuning the models that appear to be the most promising. We find that non-linear models outperform linear models and the random forest has the overall highest $R^2$ and lowest RMSE.

First, we test the following machine learning models in Python to predict M&A outcomes: Lasso regression, random forest, gradient boost and a neural network model. Our goal in this step is to determine which machine learning model (if any) best predicts takeover premium (i.e., which has the lowest root mean squared error (RMSE) and highest $R^2$).

### 4.3. Prediction Results

The way to evaluate the quality of a model that predicts outcomes is to evaluate the way in which actual premiums increase related to predicted premiums. Table 2 summarizes the ability of the machine learning models, once trained on the earlier portion of the sample, to predict the M&A premiums in the later part. Table 2 indicates that the average actual premium increases across model-predicted premium percentiles for each machine learning model. In contrast, in the OLS model, there is no relation between predicted and actual M&A premiums.

Among the machine learning algorithms, XGBoost and neural network perform best at predicting the premiums of M&A deals. Table 4 compares six models on the same test set using RMSE and out-of-sample $R^2$. The linear benchmark (OLS) performs worst (RMSE $= 0.985$, $R^2 = -0.016$), underperforming even a naïve zero-forecast. Adding regularization yields only modest gains: Lasso and Elastic Net reduce RMSE to about 0.72 and raise $R^2$ to 0.008–0.013, indicating limited linear signal once overfitting is controlled. Nonlinear methods deliver materially better accuracy. Random Forest lowers RMSE to 0.688 with $R^2 = 0.054$; XGBoost and the Neural Network improve further to RMSE $= 0.687/0.684$ and $R^2 = 0.059/0.060$, respectively. Relative to OLS, the Neural Network reduces RMSE by roughly 30.6% and turns $R^2$ positive; relative to the best linear model (Elastic Net), it still trims RMSE by about 4.9% and lifts $R^2$ by 4.7 percentage points. The ranking is consistent across both metrics (OLS < Lasso < Elastic Net < Random Forest < XGBoost < Neural Network), underscoring that capturing nonlinearities and interactions is crucial for predicting M&A premiums. While absolute $R^2$ levels remain modest ($\leq 0.06$)—typical for noisy corporate-finance outcomes—the improvements from tree-based and deep models are economically meaningful and justify the machine-learning approach adopted here.

[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 ranking performance in the tails by reporting realized premiums for deals the models predict to be in very low and very high premium percentiles. At the lower tail, nonlinear models produce sharply lower realized premiums than the linear benchmark. For the bottom 1% bucket, Random Forest, XGBoost, and the Neural Network deliver negative realized premiums, -0.163, -0.173, -0.158, while OLS yields a positive 0.401, indicating poor calibration in identifying genuinely low-premium deals. Even at the $< 5\%$ and 10% thresholds, nonlinear methods keep realized premiums close to zero, whereas OLS remains materially positive, suggesting weak separation of low-premium outcomes under linear specifications.

At the upper tail, all models show rising realized premiums, but the increase is much steeper for nonlinear approaches. For the 100% bucket, the Neural Network attains 1.438 (closely followed by XGBoost at 1.373 and Random Forest at 1.325), compared with only 0.740 for OLS. The implied top–bottom spread is economically large for the best models, about 1.60 for the Neural Network 1.438 to -0.158 versus 0.34 for OLS 0.740 to 0.401, a more than fourfold improvement in cross-sectional discrimination. The monotonic, wider separation at both tails is consistent with nonlinear models capturing interactions and nonlinearities that drive takeover pricing, aligning with their superior out-of-sample accuracy reported earlier.

[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 *Deep Neural Network* 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. Feature Importance

To shed light on the signals behind our predictions, we performed a feature-importance analysis using SHAP. The summary (beeswarm) plot reports *local* contributions of each variable to the predicted premium (right = higher, left = lower), while the companion bar chart ranks variables by their average *magnitude* of impact (mean|SHAP|). This two–step view lets us see both *which* features matter most overall and *how* high vs. low realizations of a feature push the forecast.

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]

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]

## 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 $R^2_{\text{oos}}$) 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.

# Figures and Tables

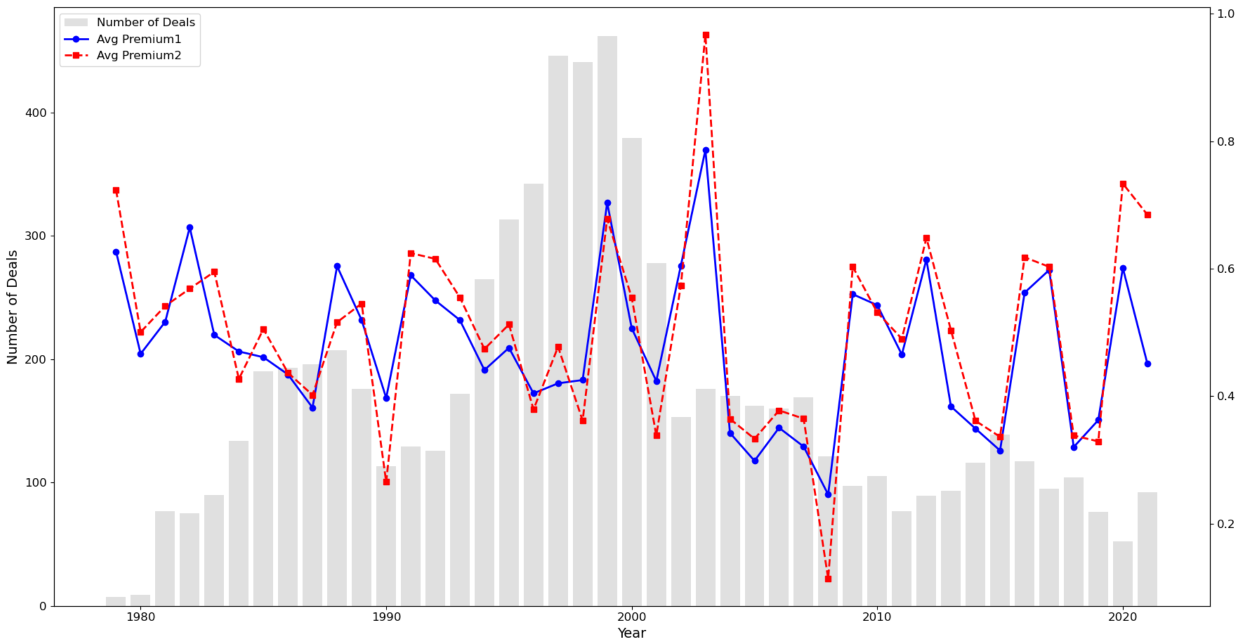
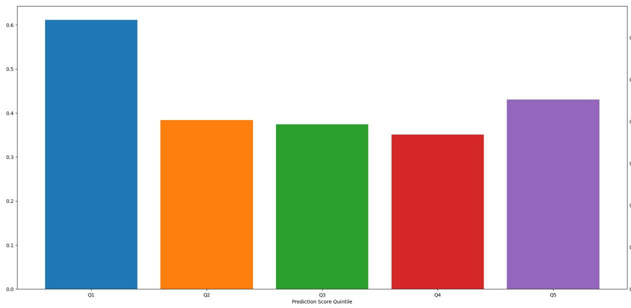
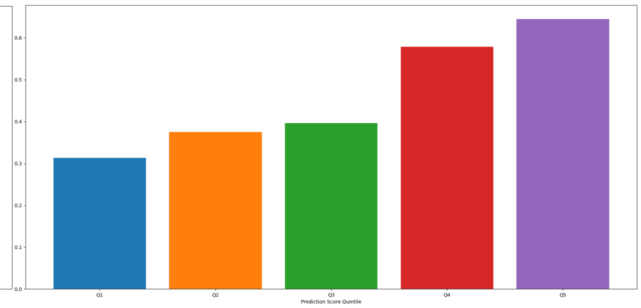


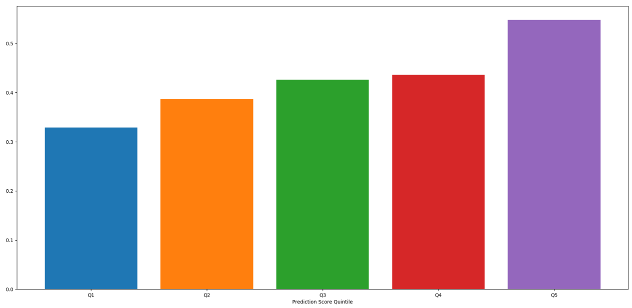

**Figure 1. Annual M&A deal counts and average premiums.** This figure illustrates the annual number of completed M&A deals and the corresponding average deal premiums. The grey bars (left axis) represent the number of unique transactions per year. The blue solid line and red dashed line (right axis) plot the time-series evolution of the average deal premiums, measured respectively by Premium1 and Premium2. Premiums are computed as 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.



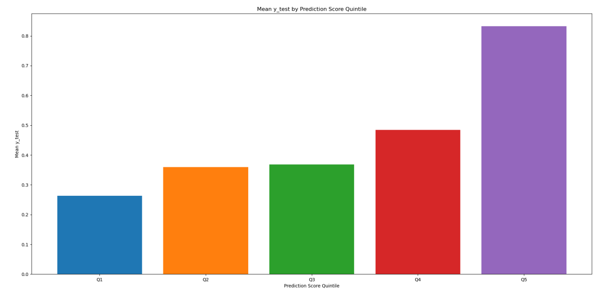

(a) OLS



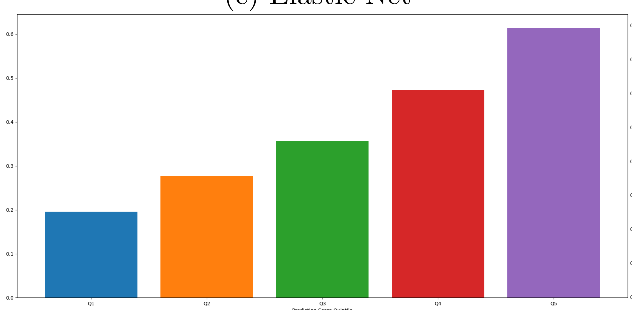

(b) Lasso



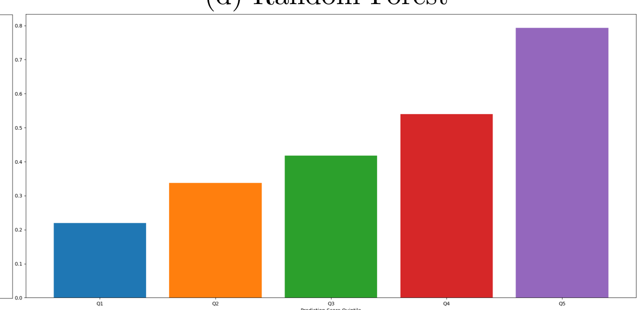

(c) Elastic Net




(d) Random Forest




(e) XGBoost




(f) Neural Network

**Figure 2. Average Predicted M&A Premiums by Quintile, Machine Learning vs. Linear Model.** 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.

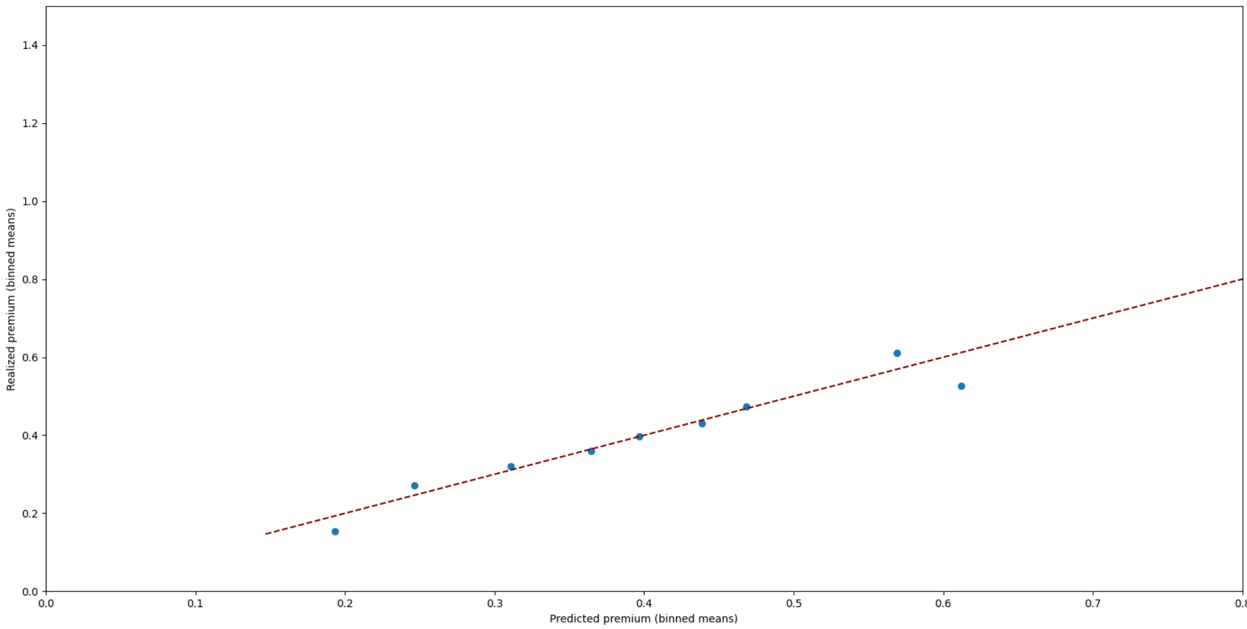



**Figure 3. Algorithm's Predictive Accuracy for M&A Premiums: All Deals in Test Set.** This figure plots average realized M&A premiums (y-axis) against 10 bins of predicted M&A premiums (x-axis) for the 1437 deals in our test set from 2017 to 2021. Both measures are in ratios. The predictions come from 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.

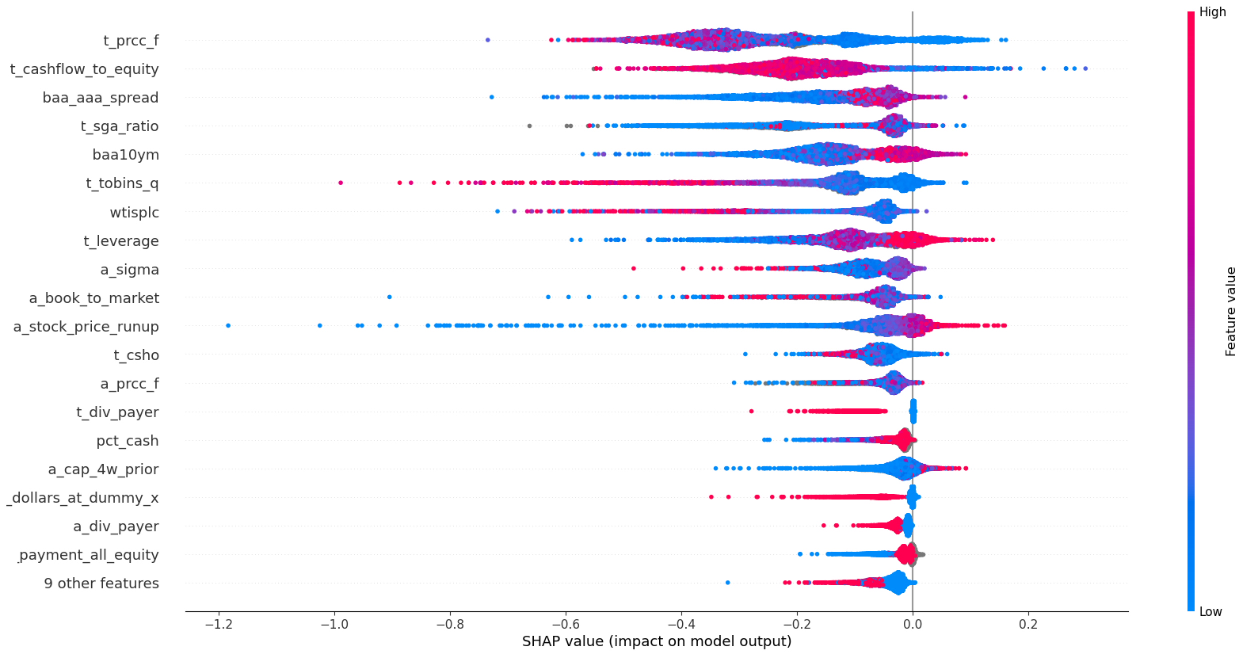



**Figure 4. Machine learning feature importance.** This figure shows the importance of each feature used in the neural network algorithm. Each dot represents a deal in the test sample. The horizontal axis reports the SHAP value (the *local* contribution of a feature to the model's prediction relative 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.

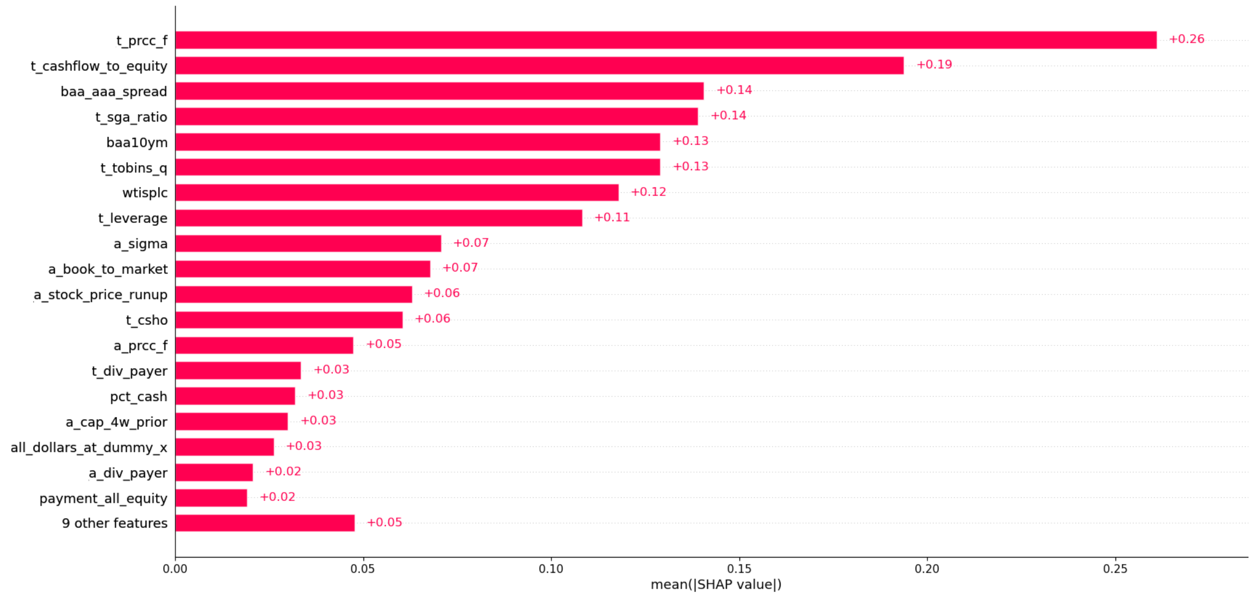



**Figure 5. Machine learning feature importance.** The bars report mean absolute SHAP values (mean |SHAP|) for the same set of predictors, which rank features by their average *magnitude* of influence on predicted premiums. This metric is sign-agnostic: a large bar indicates a strong effect—positive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

| | 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 |
| 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 |

*(continued on next page)*

*Table 1 (continued from previous page)*

| | | | | | |
|---|---|---|---|---|---|
| 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 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*.

| | 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 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 | 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 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.

| **Model** | **RMSE** | $R^2_{\text{oos}}$ |
|---|---|---|
| OLS | 0.985 | -0.016 |
| Lasso | 0.722 | 0.008 |
| Elastic Net | 0.719 | 0.013 |
| Random Forest | 0.688 | 0.054 |
| XGBoost | 0.687 | 0.059 |
| Neural Network | 0.684 | 0.060 |

**Table 4. Out-of-sample Performance by Model.** This table presents the out-of-sample evaluation metrics for the performance of the selected models. The selected candidates include the OLS regression, Lasso, Elastic Net, Random Forest, XGBoost and Neural Network.

| | Predicted Percentile of Premiums | OLS | Lasso | Elastic Net | Random Forest | XGBoost | Neural Network |
|---|---|---|---|---|---|---|---|
| Deals predicted to have low premiums | 1% | 0.401 | -0.117 | -0.082 | -0.163 | -0.173 | -0.158 |
| | $<5\%$ | 0.257 | 0.081 | 0.093 | 0.058 | 0.063 | 0.052 |
| | 10% | 0.255 | 0.130 | 0.225 | 0.121 | 0.128 | 0.113 |
| Deals predicted to have high premiums | 90% | 0.589 | 1.158 | 1.134 | 1.216 | 1.241 | 1.258 |
| | $<95\%$ | 0.605 | 1.280 | 1.301 | 1.339 | 1.358 | 1.419 |
| | 100% | 0.740 | 1.412 | 1.371 | 1.325 | 1.373 | 1.438 |

**Table 5. Realized premiums by predicted percentile and model.** This table reports the average actual premiums of M&A deals in the test sample, which are ranked by their predicted premium values by an OLS model and several machine learning algorithms (Lasso, Elastic Net, Random Forest, XGBoost and Neural Network).

## References

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

Agrawal A, Jaffe JF, Mandelker GN (1992) The post-merger performance of acquiring firms: a re-examination of an anomaly. *The Journal of Finance* 47(4):1605–1621.

Andrade G, Mitchell M, Stafford E (2001) New evidence and perspectives on mergers. *Journal of Economic Perspectives* 15(2):103–120.

Asquith P, Bruner RF, Mullins Jr DW (1983) The gains to bidding firms from merger. *Journal of Financial Economics* 11(1-4):121–139.

Betton S, Eckbo BE, Thorburn KS (2009) Merger negotiations and the toehold puzzle. *Journal of Financial Economics* 91(2):158–178.

Boone AL, Mulherin JH (2011) Do private equity consortiums facilitate collusion in takeover bidding? *Journal of Corporate Finance* 17(5):1475–1495.

Bradley M, Desai A, Kim EH (1988) Synergistic gains from corporate acquisitions and their division between the stockholders of target and acquiring firms. *Journal of Financial Economics* 21(1):3–40.

Breiman L (2001) Random forests. *Machine learning* 45(1):5–32.

Bris A (202) Toeholds, takeover premium, and the probability of being acquired. *Journal of Corporate Finance* 8(3):227–253.

Bubb R, Catan EM (2022) The party structure of mutual funds. *The Review of Financial Studies* 35(6):2839–2878.

Chaudhry AN, 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(2):958–979.

Chemmanur TJ, Fulghieri P (1994) Investment bank reputation, information production, and financial intermediation. *The Journal of Finance* 49(1):57–79.

Cheng P, Li L, Tong WH (2016) Target information asymmetry and acquisition price. *Journal of Business Finance & Accounting* 43(7-8):976–1016.

Cybenko G (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* 2(4):303–314.

Dionne G, La Haye M, Bergerès AS (2015) Does asymmetric information affect the premium in mergers and acquisitions? *Canadian Journal of Economics/Revue canadienne d'économique* 48(3):819–852.

Easley D, Michayluk D, O'Hara M, Putniņš TJ (2021) The active world of passive investing. *Review of Finance* 25(5):1433–1471.

Eaton GW, Liu T, Officer MS (2021) Rethinking measures of mergers & acquisitions deal premiums. *Journal of Financial and Quantitative Analysis* 56(3):1097–1126.

Fich EM, Cai J, Tran AL (2011) Stock option grants to target ceos during private merger negotiations. *Journal of Financial Economics* 101(2):413–430.

Freund Y (1995) Boosting a weak learning algorithm by majority. *Information and Computation* 121(2):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(2):337–407.

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 1189–1232.

Gokkaya S, Liu X, Stulz RM (2023) Do firms with specialized m&a staff make better acquisitions? *Journal of Financial Economics* 147(1):75–105.

Golubov A, Petmezas D, Travlos NG (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(1):271–311.

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. *The Review of Financial Studies* 33(5):2223–2273.

Heitzman S (2011) Equity grants to target ceos during deal negotiations. *Journal of Financial Economics* 102(2):251–271.

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Networks* 2(5):359–366.

Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S (2018) Human decisions and machine predictions. *The Quarterly Journal of Economics* 133(1):237–293.

Li K, Mai F, Shen R, Yan X (2021) Measuring corporate culture using machine learning. *The Review of Financial Studies* 34(7):3265–3315.

Loughran T, Vijh AM (1997) Do long-term shareholders benefit from corporate acquisitions? *The Journal of Finance* 52(5):1765–1790.

Masulis RW, Simsir SA (2018) Deal initiation in mergers and acquisitions. *Journal of Financial and Quantitative Analysis* 53(6):2389–2430.

Moeller SB, Schlingemann FP, Stulz RM (2004) Firm size and the gains from acquisitions. *Journal of Financial Economics* 73(2):201–228.

Mullainathan S, Spiess J (2017) Machine learning: an applied econometric approach. *Journal of Economic Perspectives* 31(2):87–106.

Officer MS (2007) The price of corporate liquidity: Acquisition discounts for unlisted targets. *Journal of Financial Economics* 83(3):571–598.

Rau PR (2000) Investment bank market share, contingent fee payments, and the performance of acquiring firms. *Journal of Financial Economics* 56(2):293–324.

Rau PR, Vermaelen T (1998) Glamour, value and the post-acquisition performance of acquiring firms. *Journal of Financial Economics* 49(2):223–253.

Rossi AG (2018) Predicting stock market returns with machine learning. *Georgetown University* .

Schapire RE (1990) The strength of weak learnability. *Machine learning* 5(2):197–227.

Schwert GW (1996) Markup pricing in mergers and acquisitions. *Journal of Financial Economics* 41(2):153–192.

Schwert GW (2000) Hostility in takeovers: in the eyes of the beholder? *The Journal of Finance* 55(6):2599–2640.

Servaes H, Zenner M (1996) The role of investment banks in acquisitions. *The Review of Financial Studies* 9(3):787–815.

Shleifer A, Vishny RW (2003) Stock market driven acquisitions. *Journal of financial Economics* 70(3):295–311.

Varian HR (2014) Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28(2):3–28.

## Appendix A: 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**] |
| *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**] |

*(continued on next page)*

*Table 2 (continued from previous page)*

| | |
|---|---|
| *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**] |
| *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**] |

*(continued on next page)*

*Table 2 (continued from previous page)*

| | |
|---|---|
| *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**] |
| *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**] |
| *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**] |

*(continued on next page)*

*Table 2 (continued from previous page)*

| | |
|---|---|
| *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* | $Z = \left(\frac{1.2\cdot(\text{act}-\text{lct})-\text{che}}{\text{at}}\right) + \left(\frac{1.4\cdot\text{re}}{\text{at}}\right) + \left(\frac{3.3\cdot\text{oiadp}}{\text{at}}\right) + \left(\frac{0.6\cdot(\text{csho}\cdot\text{prcc\_f})}{\text{lt}}\right) + \left(\frac{\text{sale}}{\text{at}}\right)$. 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*** | |
| *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*** | |

*(continued on next page)*

*Table 2 (continued from previous page)*

| | |
|---|---|
| *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***] |
| *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. |