BACHELOR’S THESIS IN ECONOMICS

Predicting Corporate Takeover Outcomes Using Machine Learning

May 27, 2020

LUND UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT
DEPARTMENT OF ECONOMICS

Author: Gustav Furenmo
Supervisor: Anders Vilhelmsson

Abstract

The aim of this thesis is to investigate if the machine learning based classification procedure, Random Forest, provides superior prediction performance compared to a logistic regression model fitted using the LASSO framework, when predicting outcomes in corporate takeover situations. This is done in the context of merger arbitrage, an event-driven investment strategy. The classification models are fitted using a training data set consisting of 5 922 OECD-domiciled corporate takeover transactions and evaluated on a testing data set consisting of 1 481 observations. Variable selection is based on the extensive research done within the field of takeover prediction. The results suggest that the random forest model outperforms the logistic regression model on all relevant validation measures, such as overall prediction accuracy, sensitivity, and specificity. Given that a vast majority of previous research has been done using logistic regression, this thesis provides cause for considering alternative and complementary classification procedures when attempting takeover prediction.

Keywords – Random Forest, LASSO, Logistic Regression, Merger Arbitrage, Takeover Prediction

Table of Contents

1 Introduction

"Risk arbitrage is not about making money, it's about not losing money. If you can minimize the downside, you get to keep all your earnings and that helps performance"

– John Paulson, Paulson & Co

Event-driven investment strategies constitute an approach employed by a vast range of prominent hedge funds across the asset management industry. Renowned firms such as AQR, Blackrock, and Allianz all have divisions dedicated to applying an event-driven investment scheme to generate returns. Even though this investment style is multifaceted and includes a comprehensive set of subcategories, the general method revolves around capitalising on special situation corporate events (Jetley & Ji, 2010). One such method exploits the effect public merger and acquisition announcements have on stock prices and is commonly referred to as *merger arbitrage*.

In the public equities markets, announcement of a merger or acquisition usually has a significant positive impact on the price of the target company's stock. This occurs since the offer most often includes a substantial bid premium, where the difference between the bid level and the target's share price is referred to as the *merger spread* (also called arbitrage spread). Even though the stock price converges towards the offered level as the announcement dissolves into the market, the spread usually persists during the completion period since there is a non-zero probability that the takeover attempt fails. Given that the takeover attempt is successful, the spread will converge to zero as the deal approaches completion (Branch & Yang, 2003). Hence, there is an opportunity for investors to extract returns from this spread by analysing whether the deal will be completed or not. This investment approach is referred to as merger arbitrage. However, it should be noted that merger arbitrage does not abide by the classical definitions of financial arbitrage in that the strategy does not present a risk-free investment.

The merger arbitrage strategy came about during the 1940s and was developed by Goldman Sachs senior partner Gustave Levy. The approach rose to particular prominence during the fourth big wave of mergers and acquisition that took place in the 1980s. Since inception, the rapidly evolving structures of financial markets have altered the prerequisites for the merger arbitrage strategy. Factors such as increased liquidity, new regulations with respect to market transparency, and integration with international capital flows have provided a basis for merger arbitrageurs to sustain attractive returns (Melka & Shabi, pp. 3-9, 2013). To this background, significant efforts have been made to further analyse the nature of merger arbitrage returns.

Baker and Savasoglu (2002) investigated the return profile of the strategy by constructing merger arbitrage positions for 1 901 transactions that took place over the period 1981 to 1996. The study concludes that a diversified portfolio of merger arbitrage positions can generate an abnormal return of 0.6-0.9% per month, beating the benchmark by 0.3% per month. The authors also make considerable efforts in investigating why these excess returns are not arbitraged away. They conclude that a contributing factor involves undiversified investors selling their positions in the target company to circumvent completion risk, while the amount of available merger arbitrage capital is limited.

The process of executing a merger arbitrage strategy depends on the payment structure of the underlying transaction. If the transaction in question involves a simple *cash payment*, an investor can gain exposure by obtaining a long position in the target's stock. In the case of a *stock-for-stock* transaction, an investor needs to compliment a long position in the target's stock with a short position in the acquiring firm. Additionally, there are more complex payment structures, such as combinations of cash and stock payments, where the execution needs to be altered accordingly (Kirchner, pp. 13-22, 2009).

A central aspect to consider when engaging in merger arbitrage is related to the payoff structure of the investment and this is most apparent in the case of taking a long position in a cash merger. The possible upside gains associated with such a position are usually limited to the merger spread and can therefore be considered fixed. Meanwhile, the downside is usually considerably larger since a failed takeover attempt would allow the target's stock price to resume its pre-announcement level. This payoff structure is clearly asymmetric and highlights the importance for merger arbitrageurs to adequately asses the probability that a takeover attempt will indeed be successful.

Against this background, the subject of using quantitative methods to predict outcomes in corporate takeover situations have become a central aspect when studying merger arbitrage strategies. Previous research efforts rely heavily on conventional classification models, such as *logistic regression*, as a method to determine what variables influence the takeover outcome and to what extent these outcomes can be predicted with accuracy. However, the relatively recent development of machine learning based classification procedures have not been adopted to the same extent. This raises the question if such classification procedures can improve the accuracy when trying to predict the outcome in corporate takeovers.

The purpose of this thesis is to investigate if the machine learning based classification procedure known as *random forest*, is superior to a logistic regression model in predicting outcomes in corporate takeover attempts. Significant amounts of research have previously studied what variables provide the most predictive power when trying to forecast outcomes in such situations. Hence, this thesis focuses on evaluating and comparing different classification models by utilising predictive variables that previous research has shown maintain robust predictive properties. The underlying data sample consists of 7 403 transactions originating from OECD countries during a period stretching from 2000 to 2018.

The underlying data sampling frame enforces several limitations on the conclusions drawn from the results presented in this thesis. Only those bids based on cash payments, stock-for-stock payments, or a combination of the two, are included in the data set. Furthermore, only transactions where the *target company* is listed on an OECD domiciled exchange are considered. In addition to this, transactions in which the target is a financial institution (i.e. banks, insurance companies or investment companies) or a real estate company, are excluded. Intra-group transactions are also excluded since the outcome determinants of these transactions are likely to differ significantly from transactions involving two non-related parties. Finally, minority investments where the acquiring entity does not seek controlling interest in the target company are exempted.

The contents of this thesis are mainly targeting academics with basic knowledge of financial markets and corporate finance. Additionally, professionals in the asset management industry, particularly those employing merger arbitrage strategies, could view the results presented as valuable since the aim is to provide insight into how quantitative strategies can be used to determine outcomes in corporate takeover attempts.

The disposition of the remaining contents of this thesis is as follows. Chapter 2 introduces the institutional setting of merger arbitrage. A general description of mergers and acquisitions is provided along with a formal definition of financial arbitrage and an outline of the merger arbitrage strategy. Chapter 3 provides a literature review containing an overview of relevant research produced within the field of takeover prediction and merger arbitrage. Subsequently, chapter 4 presents a thorough description of the data collection procedure, discusses the variables used in the classification models, and reviews how missing data has been handled. Chapter 5 gives an overview of the applied empirical framework, outlining logistic regression based on the LASSO. In addition, the random forest procedure is introduced. In chapter 6, a range of descriptive statistics for the data sample is presented along with the model fitting results and a comparison of each model's prediction performance. Finally, chapter 7 concludes this thesis and suggestions for further research are made.

2 Institutional Setting

This chapter introduces the institutional setting for merger arbitrage. A general description of mergers and acquisitions is provided along with a formal definition of financial arbitrage and an outline of the merger arbitrage strategy.

2.1 Mergers and Acquisitions

Mergers and acquisitions (M&A) is a broad term referring to a range of corporate transactions in which companies or assets are consolidated to form one entity. Even though they are often used interchangeably, the term *consolidation* typically refers to the act of combining two companies of equal size, while *merger* is used when there is a significant size difference between the two companies (Gaughan, p.14, 2015). Mergers and acquisitions have become a central concept in corporate management and the strategic motives for such transaction are numerous.

Growth is one of the most fundamental reasons for companies to engage in M&A. As opposed to organic growth avenues, M&A sometimes allows businesses to expand more rapidly and capital efficiently. In the eyes of the acquiring firm, M&A can also lead to synergies in which the combination of two firms is believed to be more profitable or strategically sound compared to stand-alone entities. Such synergies can for example take the form of cost reduction or cross-selling opportunities. Furthermore, diversification is another aspect that has been used to motivate M&A activity, particularly during the conglomerate era that elapsed during the 1960s. In addition, vertical and horisontal integration are commonly quoted motivations that corporate management teams rely on to drive an active M&A agenda, but there are many other strategic rationales (Gaughan, pp. 125-169, 2015). While mergers and acquisition are an integral part of many firms' development, transaction volumes tend to be cyclical. For example, transaction volumes are typically positively associated with rising stock prices, cheap financing costs, and a general sense of economic optimism (Goedhart, Koller & Wessels, pp. 445-449, 2010).

2.2 Defining Financial Arbitrage and Merger Arbitrage

The formal definition of financial arbitrage refers to the process of extracting risk-free returns from a self-financed investment scheme. Arbitrage is a fundamental concept in financial theory as it is closely linked to market efficiency. If price discrepancies between two equal financial assets occur, market participants (i.e. arbitrageurs) will seek to exploit the discrepancy to generate returns. This in turn, forces the respective assets' prices to return to their relative equilibrium levels. Hence, if an arbitrage opportunity indeed does occur, they are exploited and eliminated quickly (Byström, pp. 48-51, 2014).

While this description of financial arbitrage adheres to the formal academic definition, practitioners use the term somewhat differently. When participants in the hedge fund universe discuss arbitrage, they usually refer to the act of buying two similar financial assets that they consider as mispriced relative to each other. This assumed imbalance can be exploited by taking a long position in the undervalued asset and a short position in the overvalued asset (BarclayHedge, 2012). In this regard, the trade is only risk-free as long as the assumed price imperfection holds true.

With this in mind, it should be made clear that investments based on a merger arbitrage strategy are by no mean risk-free. However, the risks associated with merger arbitrage are different from those commonly acknowledged by conventional risk models focused on market risk (i.e. beta-risk). Instead, the most severe risk aspect in merger arbitrage is that of *event-risk*, which is a consequence of the uncertainties associated with takeover attempts succeeding. A takeover attempt can fail due to a number of reasons, but insufficient transaction financing, antitrust blockage, and shifting economic fundamentals are a few possible reasons to why a transaction can fall through during the time between announcement and deal completion (Kirchner, p. 10, 2009).

2.3 The General Merger Arbitrage Strategy

The basic idea behind a merger arbitrage strategy is to exploit the often persisting difference between the offered consideration and a target company's stock price in a corporate takeover situation (Branch & Yang, 2003). The execution method and the resulting payoff will however differ depending on the applied payment structure of the underlying transaction. Presented below, is an expose of how a merger strategy can be implemented in the presence of cash payments, stock-for-stock payments, and other more complex structures.

2.3.1 Merger Arbitrage with Cash Payment Transactions

In a corporate takeover, cash payment is the simplest method for settling the financial obligations of the transaction. The acquiring firm assumes control over the target company's stock in exchange for a cash payment. As previously mentioned, the difference between the target's stock price and the offered consideration is referred to as the merger spread. The spread is usually positive but can in some instances become negative (i.e. the stock price is higher than the offered bid per share). As an example, this can occur if market participants believe that rivalry bids, at higher bid premiums, will emerge. Assuming a transaction with cash payment, arbitrageurs seek to exploit the merger spread by taking a long position in the target's stock and hold it until the transaction is completed. Alternatively, the arbitrageur can exit the position before the completion date. Given that a transaction with cash payment is successful and that the target's stock price converges to the offered consideration, an investor receives a return equal to the merger spread obtained at entry, assuming no transaction costs (Kirchner, pp. 13-14, 2009).

2.3.2 Merger Arbitrage with Stock-For-Stock Payment Transactions

Stock-for-stock payment structures entail a more complex transaction profile that warrants a more complicated execution strategy. In such a deal, the owners of the target company receive payment in the form of shares in the acquiring firm, where the offer stipulates a fixed conversion rate of shares. As a consequence, the total consideration of the transaction varies depending on the price of the acquiring firm's stock. Hence, an arbitrageur cannot merely acquire a long position in the target's stock. Instead, such a long position must be complemented with a short position in the acquiring firm's stock. The payoff from this type of trade will be determined by the cash-flow difference between the two positions (Kirchner, pp. 20-22, 2009).

2.3.3 Merger Arbitrage with Other Payment Structures

In addition to the above-mentioned transaction payment structures, several other methods can be applied in corporate takeovers. A mix of cash and stock payment is quite common if the acquiring firm wants to limit dilution or have insufficient funds to finance a pure cash settlement. Furthermore, stock-for-stock methods can be extended to include so called *collars*. These can be applied to impose upper and lower limits to form a range in which the acquiring firm's stock price can fluctuate, hence giving a more certain indication of the dollar value of the transaction (Kirchner, p. 35, 2009).

2.4 The Return Profile of the Merger Arbitrage Strategy

As discussed earlier, the return profile of a merger arbitrage strategy is asymmetric. Assuming that an arbitrageur takes a long position in the underlying stock of a company subject to a takeover attempt with a cash payment structure, the potential gain is usually capped at the given merger spread at entry. Meanwhile, if an arbitrageur engages in such a position and the takeover attempt fails, the price of the target's stock usually reverts to levels close to the pre-announcement level. This in turn would cause a negative return that is substantially greater than the positive outcome. Additionally, at least in theory, the target's stock price could go to zero, which would incur even greater losses for the arbitrageur. In this aspect, the payoff characteristics of a merger arbitrage strategy are similar to that of selling put options on an index (Mitchell & Pulvino, 2001).

Previous research, such as the comprehensive study by Mitchell and Pulvino in 2001, examines the return profile of the merger arbitrage strategy. The authors investigate the return profile by calculating monthly returns for a passive portfolio taking merger arbitrage positions in 4 750 transactions in the United States over a period stretching from 1963 to 1998. They conclude that the strategy generates excess returns amounting to 9.3% per annum compared to their benchmark. When accounting for transaction costs, the excess return decrease to roughly 3.5% per annum. In addition, the authors conclude that returns generated from merger arbitrage are uncorrelated with the overall market during times of flat or expanding market levels. However, during periods of falling markets, the correlation is significant. While these results are interesting in their own right, the underlying reasons are not thoroughly addressed by Mitchell and Pulvino.

With a similar approach, Baker and Savasoglu (2002) investigate returns from a few different merger arbitrage portfolios covering 1 901 transactions from 1981 to 1996 in the United States. Depending on portfolio construction, the authors find excess returns in the range of 0.6-0.9% per month, corresponding to 7-10% per annum. Furthermore, Baker and Savasoglu make considerable effort in trying to explain why these excess returns exists. Using a regressions analysis approach, they find a significant correlation between merger arbitrage returns and deal completion risk as well as target size. Additionally, they find evidence that the excess returns can partly be attributed to undiversified investors selling their positions in the target company to avoid completion risk. Meanwhile, the fairly limited number of arbitrageurs, and their respective capital pools, require a premium for absorbing the completion risk.

3 Literature Review

This chapter provides a literature review containing an overview of relevant research produced within the field of takeover prediction and merger arbitrage.

3.1 Hoffmeister and Dyl (1981) – Predicting Outcomes of Cash Tender Offers

Hoffmeister and Dyl (1981) set out to create a statistical model, based on multivariate discriminant analysis, that is able to predict outcomes in cash tender offers. The examined period stretches from 1976 to 1977 and the analysis focus on the United States. The initial sample contained 313 tender offers, of which 267 utilised a cash payment structure. After removing transactions with insufficient data for the 17 predictive variables used, 84 transactions remained. Included predictive variables measure various financial conditions, vulnerability to takeover attempts, management attitude and the target firms' standing within their respective industries.

After fitting a range of models, the authors find that the management's attitude towards the takeover attempt, along with target size (i.e. market capitalisation), are the most influential factors. At the same time, other variables such as the level of bid premiums, are discarded as insignificant.

3.2 Walkling (1985) – Predicting Tender Offer Success: A Logistic Analysis

Walkling (1985) applies a logistic regression approach with the purpose of predicting outcomes in corporate takeover attempts. The author acknowledges that previous research, such as that of Hoffmeister and Dyl (1981), conclude that management attitude is the primary determinant for the outcome variable, while no support can be found for bid premiums having a meaningful influence. Walkling asserts that the latter conclusion contradicts fundamental principles of economic theory, which in turn begs the question as to why bid premiums exist at all.

The logistic regression relies on a data sample consisting of 158 cash transactions that took place from 1972 to 1977. Furthermore, the sample is divided into a training and testing sample to validate the final model's prediction accuracy. Predictive variables considered by the author are bid premium size, management attitude, percentage of shares owned by the acquiring party, solicitation fees and rivalry bids. In accordance with previous research, the author finds that positive management attitude significantly increases the probability of a successful outcome. In addition, Walkling concludes that there is a positive correlation between the probability of successful takeovers and solicitation fees and acquiring party ownership. In contradiction to the results produced by Hoffmeister and Dyl (1981), Walkling finds that the level of bid premiums in fact play a significant role in determining the outcome variable. Moreover, he attributes this contradiction to failure in accurately defining the bid premium variable, where previous research does not account for leakage of transaction announcement information.

3.3 Mitchell and Pulvino (2001) – Characteristics of Risk and Return in Risk Arbitrage

The 2001 study performed by Mitchell and Pulvino was briefly discussed in section 2.4. This study is comprehensive and has a somewhat different flavour compared to that of Hoffmeister and Dyl (1981) and Walkling (1985). The authors investigate the risk and return profile of the merger arbitrage strategy by analysing 4 750 mergers and acquisition that took place in the United States from 1963 to 1998. While the main focus of the study is on the return profile of merger arbitrage (as outlined in section 2.4), Mitchell and Pulvino also discuss predictive variables other than market risk and transaction costs that can be used to predict outcomes in takeover situations. They find that variables such as payment structure, management attitude and target market capitalisation have a significant impact on the probability that corporate takeover attempts are successful.

3.4 Baker and Savasoglu (2002) – Limited Arbitrage in Mergers and Acquisition

As mentioned in section 2.4, Baker and Savasoglu (2002) make considerable efforts to explain why the merger arbitrage strategy can generate abnormal excess returns by analysing 1 901 transactions from 1981 to 1996. They are able to trace the excess returns to the supply of stock in the underlying target companies, where existing investors tend to sell their positions after a takeover is announced to avoid completion risk. They also conclude that there are only a limited number of arbitrageurs active in the market, and the magnitude of their capital is not sufficient to arbitrage away the excess returns.

The authors also investigate what variables provide most explanatory power when it comes to predicting the outcomes in a corporate takeover attempts. Six predictive variables are included in the most extensive model. Out of the entire set of variables, three variables show statistical significance at the 5% level, namely management attitude, the market capitalisation of the acquiring firm, and the market capitalisation of the target company.

3.5 Branch and Yang (2003) – Predicting Successful Takeovers and Risk Arbitrage

Branch and Yang (2003) use a stepwise logistic regression model to analyse what factors are most relevant when trying to predict the outcome in corporate takeovers. The underlying data sample consists of 1 097 transactions with either cash, stock-for-stock or collar payment structures during the period 1991 to 2000. They cover a range of predictive variables but pay special attention to the importance of payment type. The authors find that payment type, the interaction between management attitude and post-announcement stock price performance, the relative size difference between the acquirer and target company, management attitude, leverage, and percent sought to be significant at the 10% level.

3.6 Branch and Wang (2009) – Takeover Success Prediction and Performance of Risk Arbitrage

Branch and Wang (2009) further investigate the role of different factors in corporate takeover prediction using logistic regression. The authors use a data sample consisting of 1 165 transactions over a period stretching from 1994 to 2003. After analysing eleven predictive variables, they conclude that five variables are significant at the 10% level. These are the target company's stock price performance prior to the transaction announcement, management attitude, merger spread, the relative size difference between the acquiring firm and the target company, and rivalry bids.

In addition, Branch and Wang show that using a logistic regression model on a pair-matched sample, in the setting of takeover prediction, results in biased parameter estimates. As a remedy, they suggest using a weighted logistic regression model to remove the bias and in extension, to improve prediction performance.

3.7 Jetley and Ji (2010) – The Shrinking Merger Arbitrage Spread: Reasons and Implications

In their study from 2010, Jetley and Ji investigate how the merger spread developed over the period 1990 to 2007. The authors base their conclusions on a data set containing 2 182 transactions and find that the average merger spread decreased by roughly 400 basis points (1 basis point = 0.01%) during 1990-1995 and 2002-2007. However, they also assert that out of these 400 basis points, 40-50 might be attributable to a decrease in trading costs.

Since the merger spread is a key determinant in the return profile of the merger arbitrage strategy, the authors investigate the reasons for the observed decline. Jetley and Ji conclude that the shrinking merger spread is caused by increased interest in the merger arbitrage strategy, as well as a reduction in risks associated with corporate takeovers. In addition, they claim that the decreased spread is of permanent nature, suggesting to investors looking to invest in merger arbitrage strategies to focus on returns achieved post 2002 rather than over a more extended period.

Finally, the authors conclude that the variables that have a significant impact on the outcome in corporate takeovers are payment structure, bid premium, management attitude, trading volumes in the target company's stock one day after announcement, and the target company's market capitalisation. All of these variables are statistically significant at the 5% level.

3.8 Summary of Findings in Previous Literature

The previous research covering takeover prediction and merger arbitrage outlined above constitute the basis for variable selection in this thesis, and a thorough expose of considered variables is presented in section 4.3. Existing research is consistent with regards to some explanatory variables, where there is broad agreement concerning the importance of management attitude and the importance of size variables such as target size (either in total target market capitalisation or relative difference to the acquiring party). However, conclusions related to bid premiums as a determining factor are somewhat ambiguous. Hoffmeister and Dyl (1981) assert that bid premiums are not a significant factor when determining the outcome in a corporate takeover attempt. Subsequent research, such as that of Walkling (1985) as well as Jetley and Ji (2010), conclude that bid premium levels are in fact relevant when predicting takeover outcomes. The authors of these studies attribute the contradictive results to failure in accurately specifying the bid premium variable.

As takeover prediction and merger arbitrage has evolved as a research field, a range of intuitively appealing variables have been showed to carry explanatory power. Payment structure is a recurring factor that has been deemed significant in several studies, such as Mitchell and Pulvino (2001) and Branch and Yang (2003). In addition, other variables such as merger spread, rivalry bids, solicitation fees, leverage, percent sought, and acquirer ownership have proven to contribute with predictive value. However, some of these variables are only investigated in single studies and therefore academic consensus is not thoroughly established.

The above outlined studies are to a degree different in nature. While all of them investigate factors relevant for takeover prediction, the approaches applied and the specific areas of analysis vary. Mitchell and Pulvino (2001) and Baker and Savasoglu (2002) particularly focus on the return profile of merger arbitrage, Jetley and Ji (2010) comprehensively investigate the merger spread, and the remaining studies introduce and investigate new predictive variables. It should also be noted that the above outlined studies mostly perform analysis on *in-sample data* with focus on variables selection. Hence, prediction accuracy on *out-of-sample data* is not exhaustively addressed. To this background, the analysis presented in this thesis is of different flavour, where focus is on developing and comparing classification models with the purpose of maximising prediction accuracy. This aspect introduces some difficulties when comparing the results in this thesis to previous research.

4 Corporate Takeover Data Sample

This chapter provides a thorough description of the data collection procedure along with an outline of the variables used in the classification models. In addition, missing data and class imbalance is discussed. The chapter finishes with an overview of the applied data partitioning procedure.

4.1 Data Collection and Inclusion Criteria

The data sample underlying this thesis includes 7 403 transactions that took place during a time period stretching from 2000 to 2018. The interval was set to cover periods of economic expansion and contraction with the purpose of limiting the influence of economic cycles. Furthermore, all transactions contained in the data sample are those where the target company was listed on an OECD domiciled exchange. The decision to include transactions from all OECD countries was made to ensure a sufficient number of observations to allow for dividing the data into distinct training and testing data samples (see section 4.6 for more detail). It can be noted that this is a substantially larger data set compared to those used in most of the studies outlined in chapter 3. Transactions involving target companies categorised as financial institutions (i.e. banks, insurance companies or investment companies) or real estate companies were excluded from the final data sample. Additionally, intra-group transactions were also excluded from the data sample since the outcome determinants in these transactions are expected to differ significantly from transactions involving two non-related parties. Finally, only transactions where the acquiring party sought to buy at least 40% of the target company and opted for total ownership exceeding 50%, were included in the final data sample. This was done since a majority ownership agenda is a prerequisite for a merger arbitrage strategy to be plausible.

All corporate takeover data was obtained using the MA function in the Bloomberg terminal. The following Bloomberg MA search criteria were used to extract the data:

- **Exchange** – the target company must be listed on an OECD domiciled exchange, where all OECD countries as of March 2020 were included (total of 36 countries).

- **Percent sought** – the acquiring firm must attempt to buy at least 40% of the target and have an ambition to gain a majority share.

- **Payment type** – only transactions based on cash payment, stock-for-stock payment, or a combination of the two were extracted.
- **Deal status** – the transaction must be labelled completed, terminated, or withdrawn.
- **Time frame** – all transactions announced during the time period 2000 to 2018.

4.2 Response Variable

The response variable, *outcome*, used in this thesis is binary since a transaction either succeeds or fails. The response variable has been coded to assume the value "1" when the takeover attempt was successful and "0" when the takeover was terminated or withdrawn. These outcome labels were extracted from Bloomberg and are defined in the following way:

- **Completed** – a completed deal has been consummated and no longer needs approvals.
- **Terminated/withdrawn** – a terminated deal has been dissolved and does not continue.

4.3 Predictive Variables

The predictive variables used in the classification models in this thesis rely on the research outlined in chapter 3. An outline of each predictive variable is presented below. Comprehensive definitions and related information regarding these variables can be obtained using the HELP MA function in the Bloomberg terminal.

4.3.1 Completion Time

Completion time describes the number of days that elapsed between the date when the transaction was announced and the date when the transaction was completed or terminated/withdrawn. This is a numeric variable that in theory can take any value greater than zero. The average completion time, measured as the number of days elapsed, for all transactions in the data sample is 126 days. As long completion times can be associated with uncertainty and increased event risk, it seems reasonable to assume that this variable has a significant impact on whether or not a transaction is successful.

4.3.2 Rivalry Bids

This variable indicates if any third-party rivalry bids were made during the course of transaction completion. It is a binary variable that assumes the value "0" if there were no rivalry bids and the value "1" if rivalry bids arose during the completion period. In the total data sample, only a small portion of the transactions, roughly 2%, were countered with rivalry bids.

4.3.3 Total Deal Value

The total deal value measures the total dollar value of the entire offer, including all disclosed payment types and is measured in millions of US dollars. Even though previous research is somewhat inconclusive regarding the impact of deal size, the complexity associated with large transactions could potentially affect the probability of a takeover attempt being successful. Note that total deal values are inflation adjusted.

4.3.4 Payment Type

As outlined in section 2.3, a range of payment structures can be employed in M&A attempts, and cash payments are regarded as the simplest method. The payment variable is binary and assumes the value "0" if the transaction used a stock-for-stock payment structure or a mix of stock-for-stock and cash payment. Meanwhile, the variable assumes the value "1" if the transaction was a pure cash payment offer. Since cash payment structures are easier for investors to understand compared to more complex setups, and because cash structures provide a fixed dollar value offer that can be evaluated, this variable might provide predictive quality to the classification models. Approximately 65% of the transactions in the data sample rely on pure cash payment settlements.

4.3.5 Attitude

A corporate takeover attempt can broadly be classified as *friendly* or *hostile*. The former refers to offers where the target company's board of directors or management team recommend the owners to accept the offer. However, if the board of directors or management team recommend the owners to reject the offer and the acquiring firm keeps pursuing the target, the takeover attempt is considered hostile. In the data sample, 91% of the transactions are considered friendly, indicating that hostile takeovers are relatively uncommon.

4.3.6 Bid Premium

The typical bid premium that an acquiring firm has to pay in order to acquire a public company lie in the range of 30-40% (PwC, 2016). This premium is committed in order to convince the target's owners to sell their respective shares. Meanwhile, the acquiring party can motivate buying the shares at a premium to the current market valuation for any of the reasons outlined in section 2.1. Bid premiums for each transaction have been exported from Bloomberg and are calculated as the difference between the offered consideration and the target's average share price one week before the transaction is announced. In the data sample, the average premium paid over the entire period is roughly 32%.

4.3.7 Financial Leverage

Levels of financial leverage in a target company, and its impact on deal completion, is another aspect that previous research has investigated. Several measurements have been used to quantify the level of financial leverage. This thesis uses the net-debt to equity ratio to quantify financial leverage, but other ratios such as debt to assets or net debt to operating profit could be employed as well.

4.3.8 Percentage Sought

As previously mentioned, only transactions where the acquiring firm sought to acquire at least 40% of the outstanding shares and had an ambition to gain majority ownership, have been included in the data sample. The variable is numeric and can take on any value in the range of 40-100%.

4.3.9 Private Equity

This binary variable indicates if the acquiring party was a strategic buyer (i.e. another company) or a financial sponsor (i.e. private equity firm). The variable assumes the value "0" if the acquiring party was a strategic buyer and the value "1" if the takeover attempt was conducted by a financial sponsor. In the data sample, the vast majority of the transactions (roughly 80%) entails a strategic buyer.

4.4 Missing and Excluded Data

The geographic perimeter of this thesis was set to ensure that a large data sample could be obtained, hence allowing the classification models, in particular the random forest model, sufficient amounts of data to be trained on. The initial data set contained 19 127 transactions. However, this number was reduced down to 7 403 transactions after removing observations where data was missing for one or several variables. Initially, different means of data imputations were considered, but later rejected since 7 403 observations seemed sufficient in relation to the number of predictive variables used. It can be noted that eliminating observations due to missing data for one or several variables has the potential of introducing systematic errors in the data sample. That is, if there is a dependency between the probability of a transaction being successful and missing data for the transactions, observations will be eliminated in a non-random order. This in turn could cause bias in the classification models. Potential bias should be kept in mind when evaluating the results presented in thesis. However, it is the author's opinion that the data set provides support for missing data occurring in a fairly random order.

Transactions where the completion time was less than one day were excluded since these could not provide useful for an arbitrageur. Additionally, all transactions where the target company was classified as a financial institution or real estate company were excluded. Companies belonging to these sectors are quite different compared to operational companies (e.g. retailers or industrial companies) in terms of general accounting practices and particularly in their use of leverage. Hence, including these companies would distort the usage of leverage as a predictive variable. For example, in the transaction where Bank Austria acquired UniCredit Bank, the recorded leverage ratio of the target was over 1 000, which is not comparable to a reasonable leverage ratio for an operational company.

Additionally, intra-group transactions were excluded since the outcome determinants of those transactions are deemed vastly different to transactions including unrelated parties. As an example, the transaction where Canary Wharf Group Plc acquired Canary Wharf Group Investment Holdings Plc was excluded for this reason.

4.5 Data Issues Related to Class Imbalance

The primary purpose of this thesis is to develop classification models that can be used to predict outcomes in corporate takeover attempts. Given the nature of corporate takeovers, where a majority of the attempts are indeed successful, issues related to *class imbalance* in the data are introduced. In binary classification, class imbalance problems occur when the model has vastly fewer observations belonging to one class (i.e. negative outcomes) compared to the other class (i.e. positive outcomes) in the training/fitting procedure. Most classification algorithms aim to maximise the overall prediction accuracy. However, if the classes occur is a skewed ratio, this procedure can lead to incorrect results. Consider the example of extreme class imbalance, where the classes occur in the ratio of 1:100. Given that the algorithm seeks to maximise overall accuracy, it can produce a 99% overall accuracy by simply classifying all observations to the majority class. In many classification domains, such as credit card fraud detection and medical trials, the minority class is usually of much greater interest compared to the majority class (Ali, Ralescu & Shamsuddin, 2015). In these instances, class imbalance can evoke real problems.

To this background, significant efforts have been made to develop various remedies. These solutions can usually be divided into two main approaches, namely *data-level approaches* and *algorithm approaches*. The former includes different sampling methods, such as over-sampling or under-sampling, to rebalance the class ratio distribution to become more even. Meanwhile, the latter approach involves optimising the algorithm to focus on accurately predicting the minority class (Ali, Ralescu & Shamsuddin, 2015).

While class imbalance can be problematic for the scope of this thesis, the ratio between successful and negative corporate takeovers is not that extreme. In the total data sample, the two classes roughly occur in the proportions 82% successful and 18% unsuccessful. Hence, no explicit effort is made to deal with class imbalance.

4.6 Data Partitioning

This thesis will compare the predictive quality of two different classification models: a LASSO logistic regression model and a random forest model (see chapter 5 for detail). To compare the respective models' ability to predict corporate takeovers, each model is trained (i.e. fitted) on a training data set and then evaluated using a testing data set that has not been seen by the models. The data partitioning is done by randomly selecting 80% of the observations from the total data sample to be used in the training stage, while the remaining 20% constitute the testing data set. Since the total data sample is sufficiently large, the random subset selection procedure manages to produce two subsets with similar class proportions. In the training data set, 82.1% of the observations have a positive outcome, and in the testing data set the corresponding quantity is 81.2%.

5 Empirical Framework

This chapter introduces the empirical framework applied throughout the thesis, outlining logistic regression based on the LASSO procedure and the general structure of random forest models.

5.1 Classification

In statistical modelling, *linear regression* is a common approach when the analysed response variable takes the form of a continuous quantitative variable (e.g. income or body height). However, if the response variable dealt with is categorical, such as the outcome in corporate takeovers, linear regression models are insufficient. Instead, models referred to as *classification models* are employed. These models use the information contained in one or several predictive variables (quantitative or categorical) to assign an observation to a specific class. Furthermore, *binary classification* refers to the situation where the response variable belongs to one of two classes, and *multinomial classification* refers to assigning an observation to three or more classes (Hastie et al., p. 127, 2013).

There are a vast number of models used in the domain of classification, all of which are appropriate in different classification settings. This thesis employs two different classification procedures, namely logistic regression and random forest. The logistic regression approach has acted as the backbone in many of the studies described in chapter 3. However, the usage of random forest models in predicting corporate takeovers is rather limited in previous literature.

5.2 Logistic Regression

Logistic regression can be used to model the probability that a response variable belongs to a certain category given a set of predictive variables. In the context of this thesis, the response variable, Y, assumes the value "1" if the transaction is successful and the value "0" if the transaction fails. Logistic regression uses the logistic function to map any real number to a corresponding value in the range [0,1], which in turn can be interpreted as a probability (Hastie et al., pp. 130-133, 2013). The logistic regression model with a binary response variable is given by the probability that the response variable assumes the value "1", indicating the positive outcome (Basto, Pereira & da Silva, 2015):

$$\Pr(\text{Successful takeover}) = \Pr(y_i = 1) = \pi_i = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}} \tag{1}$$

where β is a column vector containing a set of regression coefficients and x_i is the *i*:th row in a matrix containing *n* elements with *p* predictive variables. However, the logistic regression model in (1) is not linear in *x*. To overcome this, we apply a logit transformation, which in terms of π, can be expressed as (Basto, Pereira & da Silva, 2015):

$$\ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = x_i\beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \tag{2}$$

Hence, the logistic regression model described in (1), has a logit that is linear is *x*, and the logit can be viewed as a function that converts a linear combination of predictor variables (that can take on any real number) into the scale of probabilities (i.e. a value between 0 and 1). Finally, the model fitting is done by estimating the parameters in β using maximum likelihood. This procedure involves maximising the following log-likelihood function (Basto, Pereira & da Silva, 2015):

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)] \\ &= \sum_{i=1}^{n} \left[y_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + \log(1 - \pi_i)\right] \\ &= \sum_{i=1}^{n} \left[y_i x_i \beta - \log\left(1 + e^{x_i\beta}\right)\right] \end{aligned} \tag{3}$$

5.2.1 Model Selection Using LASSO

Performing variable selection in linear or logistic regression can be a cumbersome task. In particular, if the analysis uses a large set of predictive variables, a vast range of models can be specified. Furthermore, fitting regression models with many predictive variables introduces problems related to *overfitting*. Overfitting means that the specified model efficiently explains the training data but performs badly when trying to predict data unique to the model (i.e. testing data). This issue arises since adding more predictive variables will improve the model fit, however, at the expense of increased model complexity (Sheather, p. 227, 2009).

The data sample underlying this thesis contains data for nine predictive variables. Additionally, to account for interplay between variables, all interaction terms are considered, bringing the total number of predictive variables to 45. Hence, an efficient strategy is warranted for dealing with model specification, and in extension, overfitting. There are several viable methods for dealing with overfitting, where a common approach involves penalising the model likelihood for having many parameters. This allows for analysing the trade-off between model complexity and model fit. Using this approach, one could in theory fit all possible model combinations and evaluate them based on a quantity such as the Akaike information criteria (AIC) or the Bayesian information criterion (BIC) (Sheather, pp. 230-236, 2009). However, doing this for a large number of models can be inefficient and yield subjective results. Another alternative is to include all predictive variables in the model and then apply procedures that constrain the estimated regression coefficients, shrinking the coefficients towards zero. This is referred to as *regularisation* or *shrinkage* methods, and one well-known such technique is called the *Least Absolute Shrinkage and Selection Operator* (LASSO) (Hastie et al., pp. 214-215, 2013).

The LASSO is a relatively recent regression procedure that performs variable selection and model estimation simultaneously by introducing a shrinkage penalty that forces the regression coefficients towards zero. More specifically, the log-likelihood function being maximised is the following (Basto, Pereira & da Silva, 2015):

$$l_{\lambda}^{L}\ (\beta) = \sum_{i=1}^{n} \left[y_i x_i \beta - \log\left(1 + e^{x_i \beta}\right) \right] - \lambda \sum_{j=1}^{p} \left| \beta_j \right| \qquad (4)$$

where λ is a tuning parameter that is determined individually. Similar to ordinary logistic regression, LASSO regression opts for coefficient estimates that fit the data well. However, the second term in (4), the shrinkage penalty, is small when the coefficients $\beta_1, \dots, \beta_p$ are small. This results in the regression coefficient estimates shrinking towards zero and forcing some coefficients to become exactly zero (given that λ is sufficiently large). In this respect, the LASSO regression procedure performs variable selection (Hastie et al., pp. 219-227, 2013). Note that equation (4) represents the log-likelihood function for the logit LASSO. When the LASSO procedure is applied in the domain of regular linear regression, the first term in equation (4) is replaced by an expression for the *residual sum of squares*.

For LASSO regression to achieve the desired properties described above, the value of λ needs to be determined. When $\lambda = 0$, the shrinkage penalty has no effect and the procedure is equivalent to ordinary regression, and as $\lambda \to \infty$ all coefficients converge to zero. One conventional method for determining the level of the tuning parameter is *k-fold cross-validation*. In short, this method randomly splits the training data sample into k roughly equal sized groups (also called folds), where the first group is used for validation and the remaining $k - 1$ groups are used for fitting. The *misclassification error*, the rate at which the model incorrectly classifies observations, is calculated using the fitted model on the first group (also called held-out fold) that was not used in the fitting process. This operation is performed k times, each time using a new group for validation, resulting in k estimates of the misclassification error. Finally, the *cross-validation error* is obtained by taking the average of the k estimated classification errors. This method can be applied to determine the tuning parameter for LASSO regression by choosing a sequence of different values for λ and then calculate the cross-validation error for each value of λ. The tuning parameter is subsequently set to the sequence value that achieves the lowest cross-validation error. Finally, the LASSO regression model, (4), is fitted using the value of λ that was obtained in the cross-validation stage (Hastie et al., pp. 181-228, 2013).

5.3 Random Forest

Random forest is a machine learning based framework that can be employed in a vast range of classification problems. The mechanics of random forest algorithms are rather intuitive and involve combining a multitude of distinct *decision trees* (i.e. a "forest"), which combined creates a consensus prediction by means of a *majority vote*. The random forest framework was first introduced by Breiman in 2001 and have since risen to prominence due to several convenient features such as being able to handle large data sets with many predictive variables, combined with the ability to recognise non-linear relationships.

5.3.1 Decision Trees

A decision tree contains a sequence of binary splitting rules organised in the structure of a tree. At each split (also called *node*), a test for a certain attribute with a binary outcome is made and each *branch* represents the outcome of that test. In the context of this thesis, the test could be formulated as: "does the bid premium exceed 30%?" (which has a binary answer). This process is continued for an arbitrary number of nodes until the *terminal node*, also called *leaf node*, is reached. In the tree structure, each terminal node represents a classification label (e.g. successful or failed takeover). The process of growing a decision tree is called *recursive binary splitting*, a process that uses the classification error rate as a criterion for making the split at each node (Hastie et al., pp. 303-316, 2013).

An example of a simple decision tree structure and its prediction procedure, in the context of corporate takeover prediction, is presented in Figure 1. First, a new observation is submitted into the *root node* (i.e. the first node in the tree) that tests whether or not the bid premium exceeds 30%. Assuming that the answer to this is "Yes", the left-hand branch is followed down to the first internal node, which in turn tests if the transaction uses a cash payment structure. Once again, assuming that the answer to this test is "Yes", the left-hand branch is chosen. This is procedure is done again for the second internal node. Assuming it is a friendly takeover attempt, a terminal node representing the class "Successful" is reached. Hence, the tree predicts the outcome to be successful given the input data. This example assumes a simple tree structure, but in real applications the structure would usually be more complex.

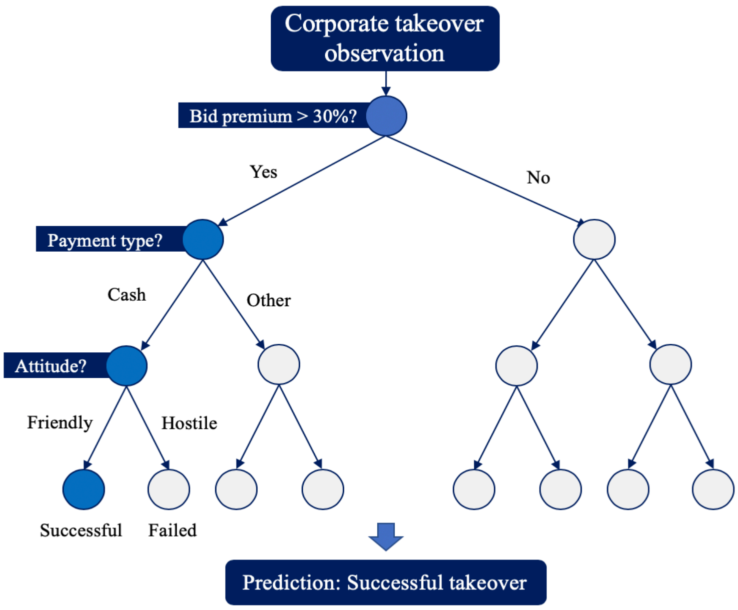

Figure 1: Simplified example of a single decision tree.

The node at the top of the hierarchy is referred to as the root node. Furthermore, the nodes labelled "Payment type?" and "Attitude?" are examples of internal nodes. The nodes at the bottom of the tree are the terminal nodes.

Even though the usage of decision trees in classification is appealing due to its simplicity, it has some major disadvantages. Decision tree structures can become extensively complex, leading to overfitting, and in turn, poor generalisation of the underlying data. Furthermore, decision trees often generate high *variance*, meaning that minor changes in the data result in vastly different tree structures. This is a consequence of the hierarchical nature of decision trees, where errors occurring in splits high up in the structure affects all subsequent splits. Furthermore, recursive binary splitting is an example of a *greedy* approach. This means that at any given level of the decision tree building, the locally optimal split is made at each node, rather than looking ahead and choosing a split that would lead to a globally superior tree (Friedman, Hastie & Tibshirani, p. 312, 2017). However, by generating and combining a multitude of trees using certain procedures, these disadvantages can be handled, for example by using random forests.

5.3.2 Random Forest

Random forest models are based on decision trees but do not suffer from the same level of high variance that is often observed when using single tree structures in classification problems. Consider a data set consisting of n independent observations, $Z_1, \ldots, Z_n$, each with a variance of σ^2. The variance of the arithmetic average, $\bar{Z}$, is then equal to σ^2/n. Hence, taking the average of the set of observations decreases the variance and therefore increases the precision. With this in mind, one could in theory collect several training data sets, construct a prediction model for each training set, and then simply average out the results from each model to form an aggregated prediction. This is not a plausible solution since we usually only have one or a few training sets. Instead, so called *bootstrapping procedures* can be used to take repeated samples from the same training data set. The bootstrap training samples can then be used to fit *N* individual decision trees, which combined constitutes the random forest model. On average, each tree uses roughly two thirds of the observations in the data sample. The observations not used when fitting the trees are referred to as *out-of-bag* (OOB) observations. In extension, the OOB observations can be used to calculate the *OOB error rate*, which is a valid estimate of the overall classification error. The random forest model then makes predictions using a majority vote procedure based on the predictions from each decision tree (Hastie et al., pp. 316-321, 2013).

A simplified random forest is presented in Figure 2. In this example, Tree 1 predicts the class “Successful takeover”, Tree 2 predicts “Failed takeover”, and Tree N predicts “Successful takeover”. Finally, the majority vote, based on the *N* trees in the random forest, makes an aggregate prediction of "Successful takeover". Note that this is a simplified example and in reality, one would usually use a far greater number of trees when constructing a random forest model.

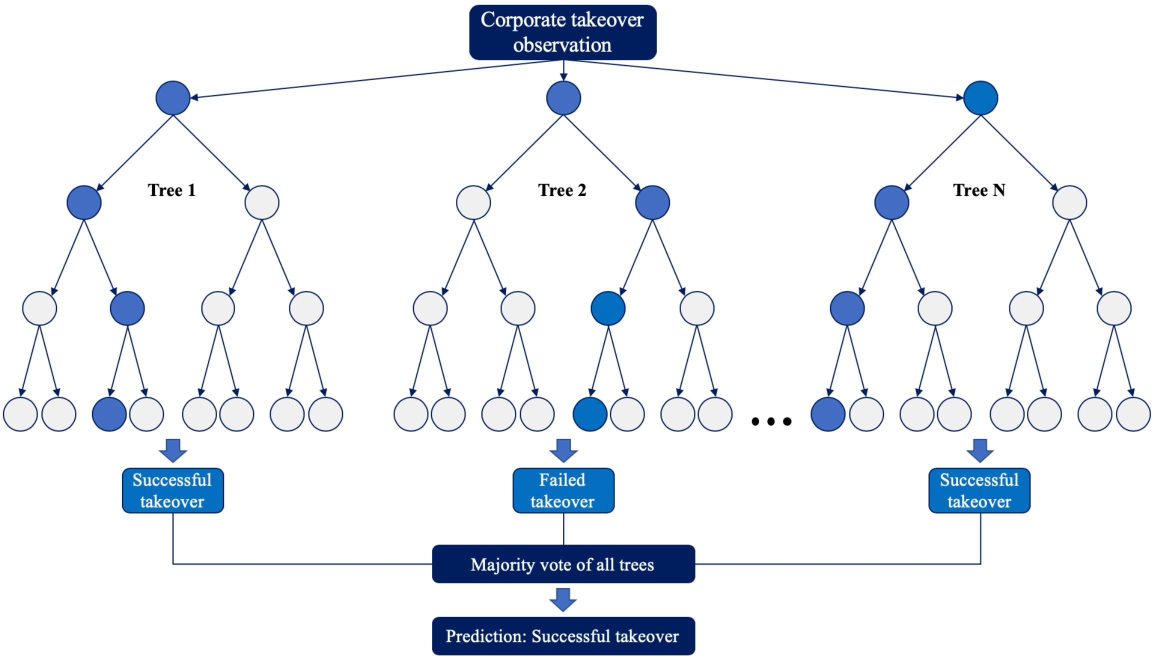

Figure 2: Simplified example of a random forest model consisting of *N* decision trees.

In this simplified example, the first decision tree classifies the observation as "Successful", the second tree makes the prediction "Failed", and the Nth tree classifies the observation as "Successful". A final prediction is made by means of majority vote based on each individual decision tree. In this case, a majority of trees classified the observation as "Successful", leading the random forest to predict "Successful takeover".

Another important aspect of the random forest model is how it builds trees and how the hierarchy is decided. At every instance when a split is considered (i.e. at each node) a random sample of m (also called m_{try}) predictors are chosen from the total set consisting of p predictive variables. This is done at each new split and $m \approx \sqrt{p}$ is commonly used. The reason for only allowing the model to choose between m predictive variables, as opposed to all of them, is to ensure that the distinct trees are uncorrelated. If the model was allowed to choose between all p predictive variables, the variable with most predictive power would occur high up in the hierarchy in most of the trees generated. As a result, many of the trees would look similar and be substantially correlated, which in turn would lower the variance decreasing effect of averaging the predictions from many trees (Hastie et al., pp. 320-321, 2013).

Another parameter that needs to be specified for the random forest is *ntree,* the number of trees generated. In theory, one would ideally use an infinitely large number of trees to increase the model's stability. Breiman (2001) showed that increasing the number of trees does not overfit the data. However, there are practical limitations to growing an infinitely large random forest, particularly with regards to memory requirements and the time it would take to train the model. Meanwhile, using a low number of trees could also present challenges related to the correlation between trees. In practice, common values for *ntree* range between 50-500.

As mentioned in section 5.3.1, decision trees are easy to visualise and interpret using the tree structure diagram. However, when combining a large number of trees into one model, this feature is somewhat lost. Even though the random forest model is more difficult to visualise, an importance summary for the included predictive variables can easily be obtained using a *variable importance plot*. This graphical representation ranks the predictive variables based on *Mean Decrease Gini* or *Mean Decrease Accuracy*. The former is a quantity measuring the reduction in a predictive variable's *Gini impurity*, which referrers to the probability that an observation is misclassified at a specific node. Meanwhile, the latter measures the reduction in prediction accuracy given that a certain predictive variable is removed (Hastie et al., p. 312, 2013). See Figure 3 for an example of a variable importance plot.

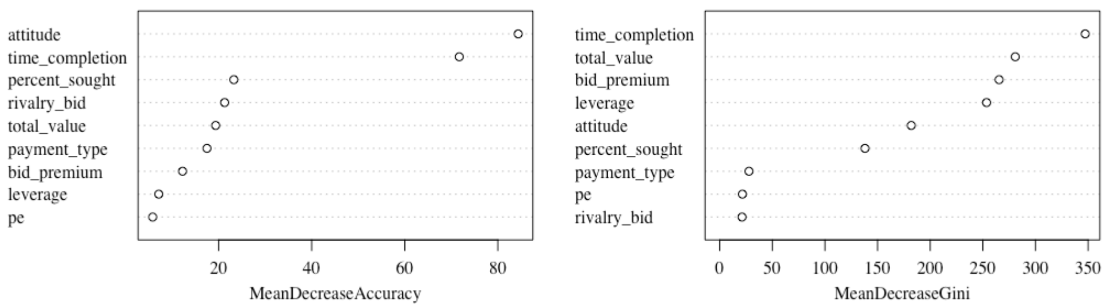

Figure 3: Example of a variable importance plot.

The left panel ranks the predictive variables according to Mean Decrease Accuracy and the right panel ranks each variable with regards to Mean Decrease Gini.

5.4 Evaluating Prediction Accuracy Using the Confusion Matrix

One tool frequently used when evaluating the prediction performance of classification models is the *confusion matrix*. A schematic depiction of the confusion matrix is given in Table 1. This structure is convenient when evaluating classification models because several validation measures can be derived from the data contained in the confusion matrix. These statistics are briefly described below.

Table 1: Schematic depiction of a confusion matrix.

TP = True positive, FN = False negative, FP = False positive, TN = True negative.

		Actual		
		Negative outcome	Positive outcome	**Total**
Prediction	Negative outcome	True negative	False negative	**TN + FN**
	Positive outcome	False positive	True positive	**FP + TP**
	Total	**TN + FP**	**FN + TP**	**TP + FN + FP + TN**

The *overall prediction accuracy* can be calculated according to (5) and is simply the ratio between the total number of correctly classified observations to the total number of observations. This statistic gives an overall measure of the model's ability to classify observations correctly.

$$Overall\ accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{5}$$

While the overall accuracy is interesting, it does not provide insight into the model's ability to classify the respective classes correctly. Therefore, overall accuracy is complemented with two additional statistics, namely *specificity* and *sensitivity*. The former is calculated according to (6) and describes the share of unsuccessful transactions being correctly classified by the model. Meanwhile, the latter is calculated by (7) and indicates the model's ability to classify the successful transactions correctly.

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

6 Descriptive Statistics and Empirical Results

This chapter presents a range of descriptive statistics for the collected data sample. Furthermore, results from the respective model fitting procedures are discussed. The chapter is concluded with a comparison of each model's prediction performance and the results are discussed.

6.1 Descriptive Statistics

The cyclical nature of mergers and acquisitions described in section 2.1 is evident in the data sample. The number of transactions in the data sample per year is presented in Figure 4, where periods of recession (e.g. 2002 and 2008-2009) are associated with lower transaction volumes.

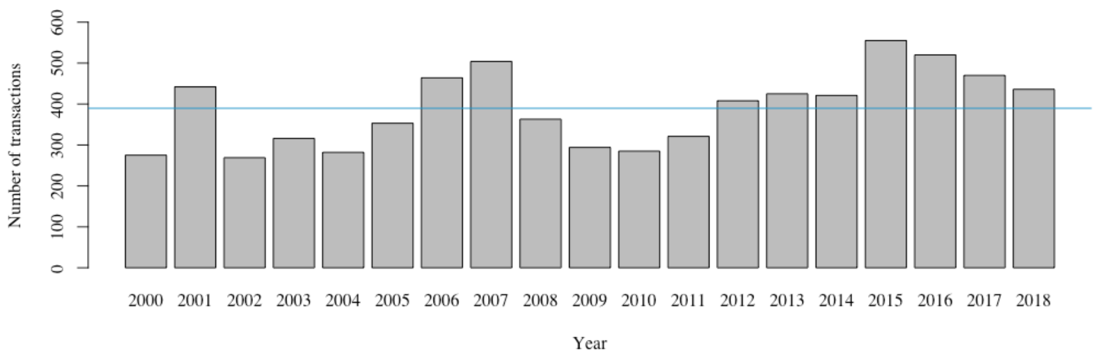

Figure 4: Number of transactions per year in the total data sample with the average number of transactions per year represented by the blue line.

This graph highlights the cyclical nature of mergers and acquisition. The data shows how volumes tend to decrease during times of recession, for example during the economic contraction in the early 2000s and the global financial crisis in 2008-2009.

The median total deal value per year (inflation adjusted) is presented in Figure 5. There is an evident trend of increasing total deal values over the analysed period, with the median over the entire period being roughly 144 million USD. Hence, the data sample suggests that both volumes, in terms of number of transactions, and deal value has increased since the year 2000.

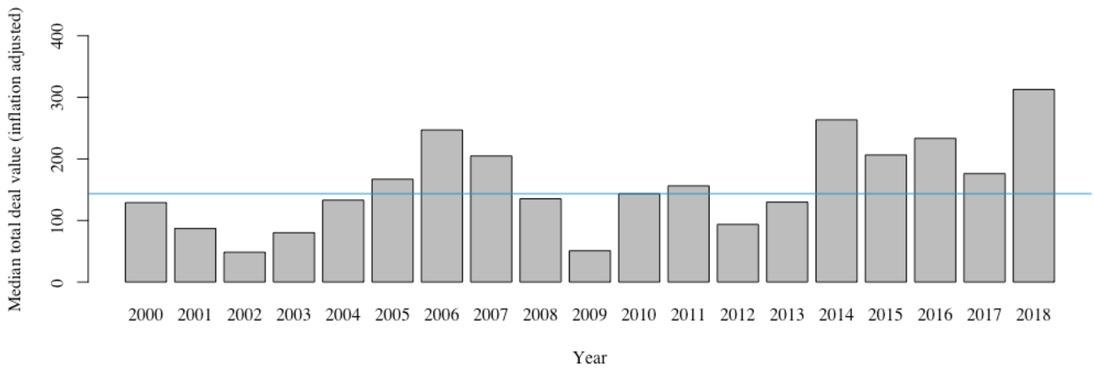

Figure 5: Inflation adjusted median total deal value per year (MUSD) with the median deal value over the entire period being represented by the blue line.

The data suggest a clear trend of transaction getting larger in absolute terms with regards to the sample period stretching from 2000 to 2018. Note that all values are inflation adjusted according to the overall inflation rate for the OECD region.

Furthermore, a boxplot describing how bid premiums have evolved over the measured period is presented in Figure 6 (note that outliers have been excluded from the graph). The data sample suggests that the median bid premium has remained fairly constant since the year 2000. However, it is interesting to note that the variation in bid premium levels, measured by the height of each box, seems to increase during recession year. This is most evident for the years 2002, 2009 and 2012.

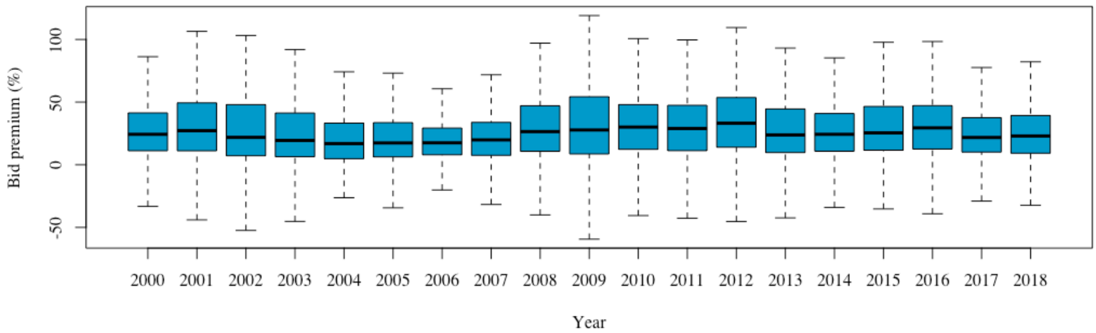

Figure 6: Median bid premium (measured in percent) per year excluding outliers.

The data suggest that bid premiums have remained relatively constant over the sample period, however with increased variation during recession years.

Figure 7 presents how completion time (in days) has developed over the sample period. Once again, there seems to be a correlation between completion time and the general economic cycle, where periods around 2001, 2008 and 2012 exhibit faster completion times.

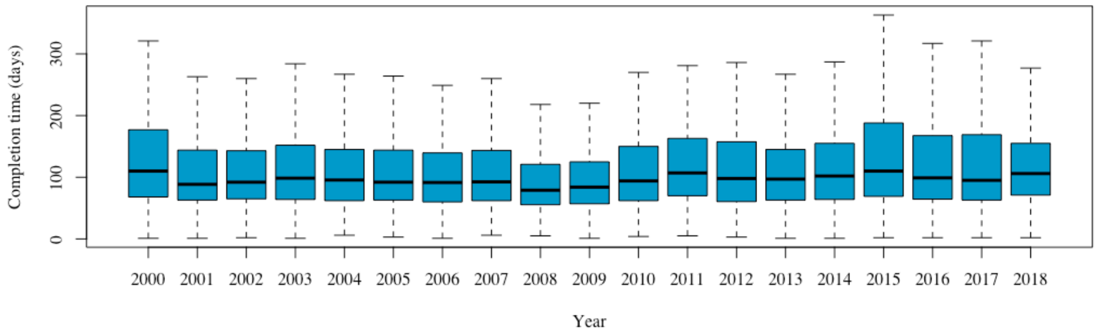

Figure 7: Median completion time (measured in days) per year excluding outliers.

The data indicate fairly constant completion times over the sample period, however with transactions being finalised quicker during recession years.

Table 2 describes the distribution for the response variable (i.e. outcome) for the entire data set. The proportion of successful takeover attempts varies considerably over the sample period and amounts to roughly 82% for the entire data set, corresponding to 6 062 transactions.

Table 3 shows proportions for the categorical variables private equity, payment type and attitude for each year in the sample period. Even though there is some variation, the proportion of friendly takeover attempts seems to be rather consistent over time. However, there is significant variation in the proportion of cash payment transactions carried out per year over the sample period. Additionally, it is interesting to note that private equity sponsored takeovers, in absolute and relative terms, has increased substantially since the year 2000.

Table 2: Distribution of total, completed and failed number of transactions per year. The bottom row refers to the entire sample period.

The proportion of corporate takeover attempts being successful varies significantly from year to year. The lowest share of completed transactions occurred in 2008 and highest share of transactions completed was in 2002.

Year	Total takeovers	Completed	Failed	% Completed	% Failed
2000	275	203	72	73.8%	26.2%
2001	442	375	67	84.8%	15.2%
2002	269	238	31	88.5%	11.5%
2003	316	271	45	85.8%	14.2%
2004	282	235	47	83.3%	16.7%
2005	353	309	44	87.5%	12.5%
2006	464	363	101	78.2%	21.8%
2007	504	399	105	79.2%	20.8%
2008	363	260	103	71.6%	28.4%
2009	294	227	67	77.2%	22.8%
2010	285	227	58	79.6%	20.4%
2011	321	274	47	85.4%	14.6%
2012	408	349	59	85.5%	14.5%
2013	425	351	74	82.6%	17.4%
2014	421	355	66	84.3%	15.7%
2015	555	443	112	79.8%	20.2%
2016	520	437	83	84.0%	16.0%
2017	470	381	89	81.1%	18.9%
2018	436	365	71	83.7%	16.3%
Total	**7 403**	**6 062**	**1 341**	**81.9%**	**18.1%**

Table 3: Proportions for the categorical variables private equity, payment type and attitude per year. The bottom row refers to the entire sample period.

The data suggest fairly constant levels of friendly takeover attempts per year in the data sample. There is a clear trend towards a larger share of private equity sponsored transaction. Additionally, there is considerable variation with regards to payment type over the sample period, however with pure cash deal being most prominent.

Year	% Private equity	% Cash payment	% Other payment	% Friendly	% Hostile
2000	8.4%	55.3%	44.7%	92.0%	8.0%
2001	8.6%	55.0%	45.0%	94.3%	5.7%
2002	13.0%	53.9%	46.1%	94.8%	5.2%
2003	15.8%	57.0%	43.0%	94.6%	5.4%
2004	13.5%	61.3%	38.7%	92.2%	7.8%
2005	19.0%	63.5%	36.5%	91.5%	8.5%
2006	23.9%	70.3%	29.7%	89.0%	11.0%
2007	27.0%	71.4%	28.6%	92.7%	7.3%
2008	17.4%	67.2%	32.8%	89.8%	10.2%
2009	15.3%	50.7%	49.3%	96.3%	3.7%
2010	19.6%	67.0%	33.0%	86.3%	13.7%
2011	19.9%	67.0%	33.0%	87.9%	12.1%
2012	18.4%	68.4%	31.6%	89.7%	10.3%
2013	23.3%	69.4%	30.6%	90.4%	9.6%
2014	20.9%	60.6%	39.4%	90.5%	9.5%
2015	21.4%	61.4%	38.6%	88.5%	11.5%
2016	26.9%	72.1%	27.9%	87.1%	12.9%
2017	27.7%	74.3%	25.7%	91.5%	8.5%
2018	26.1%	66.1%	33.9%	88.8%	11.2%
Total	**20.1%**	**64.6%**	**35.4%**	**90.7%**	**9.3%**

6.2 Results From The Logistic Regression Model Using LASSO

In the subsequent sections, fitting of the logistic regression model on the training data sample using the LASSO procedure is described. Furthermore, the LASSO tuning parameter is assed and variable inclusion is discussed. Lastly, the prediction accuracy of the final model is evaluated by applying it to the testing data set.

6.2.1 Model Fitting and Tuning

As described in section 5.2.1, an important aspect in the LASSO fitting procedure is deciding the value of the tuning parameter, λ, that minimises the cross-validation error. This error rate was calculated for a sequence of values using cross-validation with k = 10. The results are presented in Figure 8. The grey dotted line furthest to the left indicate the value of λ that minimises the cross-validation error and the corresponding value will be used when fitting the final model. Figure 9 provides additional information with regards to the tuning parameters impact on the coefficient estimates, where several of the coefficients are quickly reduced down towards zero as λ increases.

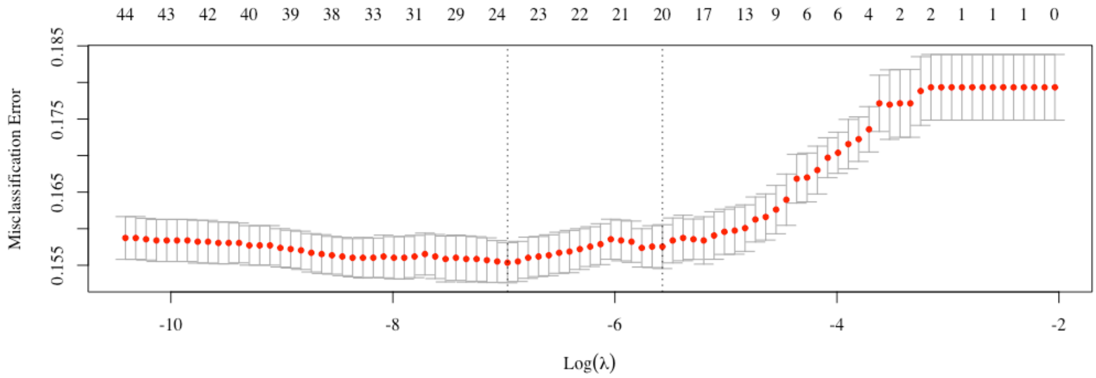

Figure 8: Cross-validation error as a function of λ.

This plot describes the cross-validation error as a function of the tuning parameter. The values in the top row refers to the number of non-zero coefficients in the model given a certain level of the tuning parameter, and the bottom row are the actual values of the tuning parameter. The red dots constitute the cross-validation curve and the grey bars suggest a confidence interval around the point estimates. The dotted vertical line furthest to the left indicates the value of λ that minimises the error rate, and the dotted line to the right suggests the value of λ that shrinks the model the furthest so that the error rate is still within one standard error of the minimum value. The graph suggests that there are at least 24 coefficients that are relevant (including the intercept), and this number is decreased to 20 coefficients given one standard error.

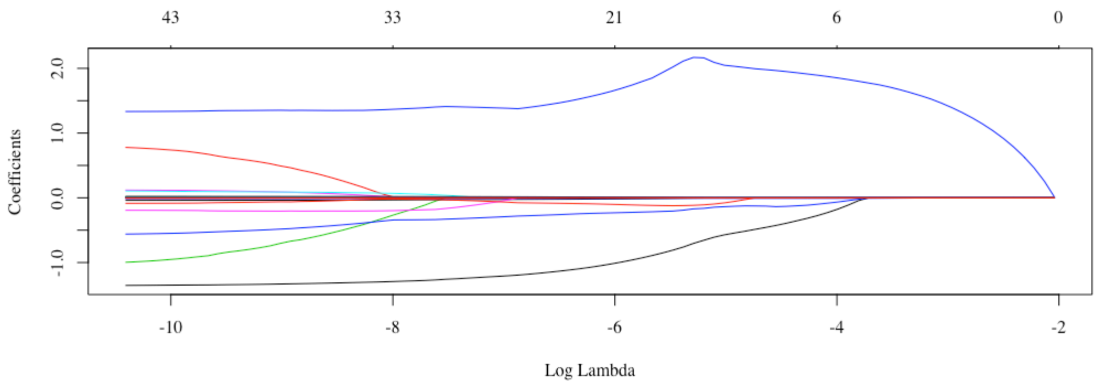

Figure 9: Coefficient estimates plotted as a function of λ.

Each line in the above plot represent an individual coefficient. The majority of all coefficient are reduced down towards zero quickly, while only a few remain far from zero as the tuning parameter is increased.

The ten largest coefficients are presented in Table 4 and a comprehensive table containing estimates for all 46 coefficients can be found in Appendix I. Table 4 suggests that attitude is the most influential variable in determining the outcome in a corporate takeover attempts since this coefficient deviates the furthest from zero (in absolute terms). It is also interesting to note that among the ten largest coefficients, a majority are associated with interactions between variables.

Table 4: The ten largest estimated coefficients from the LASSO procedure.

The notation used in the variables column in this table is the following: variable names written "stand-alone" refers to the actual variable, while variable names separated by a ":" refer to the interaction between the two variables.

Variable	Coefficient
attitude	1.383439000
Intercept	1.164750000
attitude:percent_sought	0.016443220
time_completion:rivalry_bid	0.004947137
payment_type:percent_sought	0.002633868
time_completion	0.002288839
payment_type:bid_premium	0.001120152
bid_premium	0.000807709
leverage:pe	0.000052795
time_completion:percent_sought	0.000045367

6.2.2 Final Model Prediction Performance

The model described in section 6.2.1 is applied to the testing data set to evaluate its prediction performance. Each observation is classified as "successful" if the predicted probability of a successful outcome is larger than 50%. The prediction results, in the form of a confusion matrix and additional statistics, is presented in Table 5. Overall prediction accuracy is 84.2%, which can be compared to the *No Information Rate* value amounting to 81.2%. The latter quantity refers to the largest proportion of the two outcome classes and can be interpreted as the accuracy one would achieve by simply guessing "successful outcome" for each observation. Hence, the model is able to classify the response variable with higher accuracy than what would be obtained by guessing. This conclusion is also confirmed by the 95% confidence interval for the overall accuracy. However, while sensitivity is close to 98%, the specificity is much lower at roughly 26%. This suggest that the model performs well when classifying successful takeover attempts, but less well when classifying failed takeover attempts. Low specificity could be the result of a number of factors, such as a rather substantial class imbalance (see section 4.5).

Meanwhile, in accordance with the interpretation of the No Information Rate above, simply guessing "successful outcome" for each observation would give a specificity of 0%.

Table 5: Confusion matrix and other validation statistics for the LASSO logistic regression model.

The overall accuracy in combination with the 95% confidence interval suggest that the logistic regression model, based on the LASSO procedure, is able to classify the response variable correctly for 84.2% of the observations. While sensitivity is almost 98%, the specificity only reaches about 26%.

		Actual				
		Failed	Successful	**Total**	Overall Accuracy	84.2%
					95% CI	82.3% - 86.0%
Prediction	Failed	73	28	**101**	No Information Rate	81.2%
	Successful	206	1 174	**1 380**	Sensitivity	97.7%
	Total	**279**	**1 202**	**1 481**	Specificity	26.2%

6.3 Results From The Random Forest Model

The sections below outline the model fitting and tuning procedure of the random forest model applied to the training data sample. The parameter controlling the number of predictive variables considered at each split, m, is discussed along with the number of decision trees used in the final model. Additionally, variable importance is examined and prediction performance of the final model is evaluated on the testing data sample.

6.3.1 Model Fitting and Tuning

Following the discussion in section 5.3.2, specifying the number of predictive variables considered at each decision tree split, m, out of the entire set consisting of p predictive variables, is an important aspect when tuning a random forest model. Figure 10 displays the out-of-bag (OOB) error rate as a function of m. The graph suggests that the OOB error is minimised when m is set to three, which is in accordance with the common approach of setting $m \approx \sqrt{p}$.

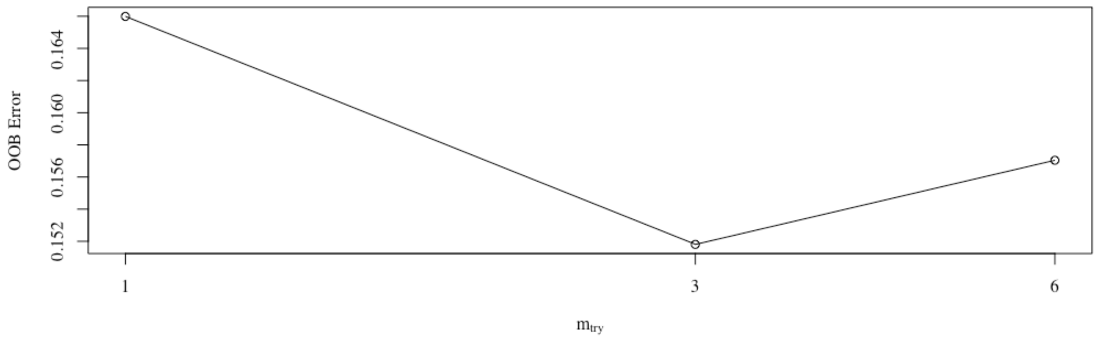

Figure 10: Out-of-bag error as a function of *m*.

This plot describes how the out-of-bag error is impacted by the number of predictive variables considered at each decision tree split. As previously described, the random forest model randomly selects a number of predictive variables from the total set of variables each time a split is considered. This is done to ensure that all the distinct trees are not correlated. The training data sample suggest that $m = 3$ is the optimal choice, which is in accordance with the commonly used $m \approx \sqrt{p}$ rule of thumb.

Figure 11 shows how the random forest model's performance is impacted by the number of distinct decision trees used. The black line represents the total OOB error rate for the model and the green and red dotted lines represent the misclassification error for the classes "completed" and "failed" transactions, respectively. Hence, it can be concluded that the respective error rates stabilise at a fairly low number of trees. As described earlier, one would ideally use an infinite number of trees in the final model. However, using a very large number of trees would not be computationally plausible. Even though Figure 11 suggests that a low number of trees could be used (e.g. 50 trees), the final model is specified to consist of 300 trees since this still allows for quick running times given the size of the sample and number of predictive variables.

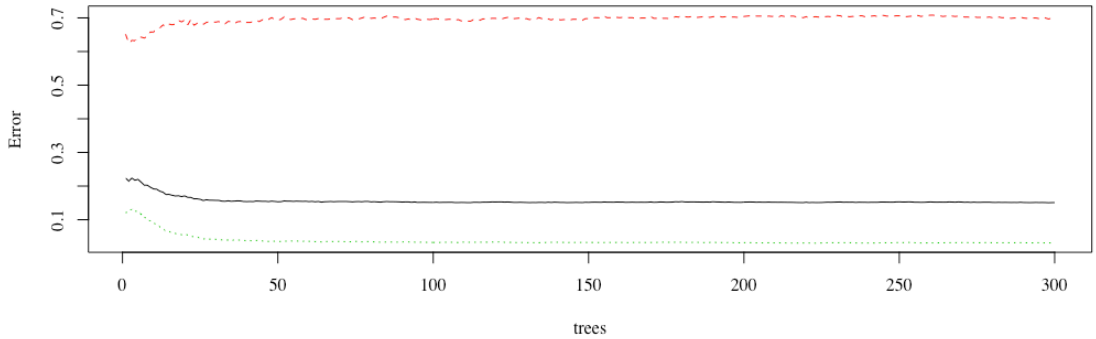

Figure 11: Error rates as a function of the number of decision trees.

The black solid line represents the out-of-bag error for the random forest model as a function of the number of decision trees generated. The dashed green line represents the misclassification error for the class “completed transaction” and the red dashed line represents the equivalent error rate for the class “failed transaction”. The plot suggest that the respective error rates stabilise at roughly 50 trees.

Figure 12 contains a variable importance plot that is based on the fitted random forest model. All predictive variables are ranked in a descending order with regards to Mean Decrease Accuracy and Mean Decrease Gini. Both measurements suggest that the variables attitude, completion time and total value contribute to the model’s performance. Meanwhile, the interpretation for the remaining variables is somewhat ambiguous.

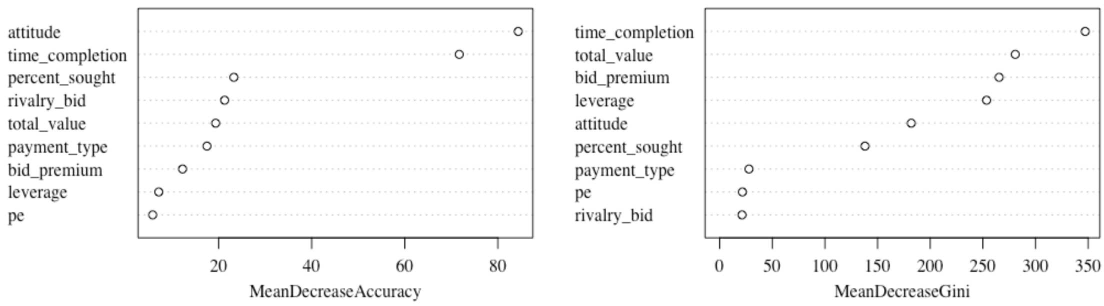

Figure 12: Variable importance plot for the random forest model.

The variable importance plots ranks each predictive variable in a descending order with regards to Mean Decrease Accuracy (left panel) and Mean Decrease Gini (right panel). Both measurements suggest that attitude, completion time and total value are important variables, while there is not a consistent interpretation with regards to the other variables.

Finally, Figure 13 presents a histogram showing the distribution of the number of nodes (i.e. tree size) in each decision tree included in the random forest model. This plot suggests that many of the trees are relatively large, possibly suggesting that there is significant interaction between variables, which in turn leads to complex tree structures. This is in accordance with the results presented in section 6.2.1, where a majority of the most influential predictive variables were interaction terms.

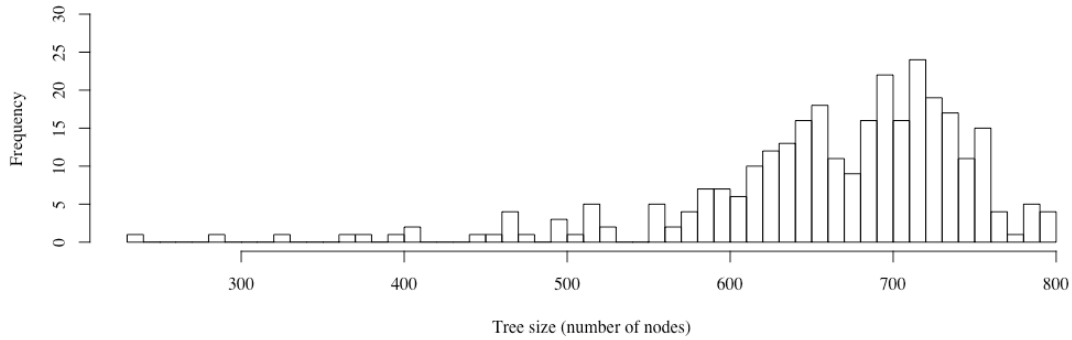

Figure 13: Histogram displaying the number of nodes in the decision trees included in the random forest model.

The average number of nodes included in each tree for the random forest is roughly 660. This suggest fairly complex tree structure, which might be a consequence of significant interaction between predictive variables.

6.3.2 Final Model Prediction Performance

The prediction performance of the random forest model outlined above was evaluate on the testing data set. The prediction results are summarised in a confusion matrix, accompanied with some additional statistics, in Table 6. The overall prediction accuracy is 85.2%, which again can be compared to the No Information Rate value amounting to 81.2%. Additionally, a 95% confidence interval for the overall prediction accuracy is displayed. Similar to the logistic regression model, the sensitivity is high, reaching roughly 98%. However, the specificity is considerably lower at roughly 29%. Hence, the random forest also has problems in classifying the failed transactions accurately.

Table 6: Confusion matrix and prediction quantities.

The random forest is able to classify 85.2% of the observations correctly. While the sensitivity of 98.3% can be considered high, the specificity is considerably lower at 28.7%. Hence, the model experience difficulty in classifying the failed takeover attempts correctly.

		Actual				
		Failed	Successful	**Total**	Overall Accuracy	85.2%
					95% CI	83.2% - 86.9%
Prediction	Failed	80	21	**101**	No Information Rate	81.2%
	Successful	199	1 181	**1 380**	Sensitivity	98.3%
	Total	**279**	**1 202**	**1 481**	Specificity	28.7%

6.4 Model Comparison

A comparison of the two classification models outlined above is presented in Table 7. When subjected to the testing data sample (i.e. data not seen by the models), the random forest achieved better prediction performance on all relevant measures compared to the logistic regression model. Both models have problems related to low specificity. However, it should be noted that even though the random forest only outperformed the logistic regression model by one percentage point on overall prediction accuracy, specificity was almost three percentage points higher. This suggest that the random forest does not only perform better in terms of overall accuracy, but also surpasses the logistic regression model's ability to correctly classify failed corporate takeover attempts. Additionally, these results could indicate that the response variable's two classes are not entirely linearly separable. That is, in the context of the underlying data, the random forest uses a preferred method for dividing the predictor space by more effortlessly acknowledging non-linear and complex data structures. For comparison, Table 7 also contains prediction validation statistics for an ordinary logistic regression model including all predictive variables and the respective interaction terms (see Appendix III for coefficient estimates). As with the other models, the ordinary logistic model was fitted on the training sample and evaluated on the testing sample. Table 7 shows that the ordinary logistic regression model and the LASSO based equivalent produce similar prediction results in terms of overall accuracy, sensitivity and specificity. However, it should be noted that the ordinary model uses 46 predictive variables (including the intercept) even though only 12 coefficients are significant at the 5% level. For comparison, the LASSO based model only uses 25 predictive variables. Hence, the LASSO model and the ordinary model produce similar results in terms of prediction, however, the ordinary model is significantly more complex and therefore more prone to overfitting the data. It can also be noted that the random forest model outperforms the ordinary logistic regression model on all validation statistics.

Table 7: Comparison of prediction performance for the random forest model, the LASSO logistic regression model, and an ordinary logistic regression model including all predictive variables and interaction terms.

The model comparison suggest that the random forest outperforms the LASSO logistic regression model on all relevant measures. It is particularly interesting to note that the specificity obtained by the random forest is considerably higher than that of the LASSO model. This in turn could suggest that the underlying relationships in the data are better described by the predictor space partitioning approach employed by the random forest compared to the logistic regression model. For comparison purposes, an ordinary logistic regression model is included in the table. This model performs similarly to the LASSO version, however, at the expense of higher model complexity.

	Random Forest	**LASSO Regression**	**Logistic Regression**
Overall Accuracy	85.2%	84.2%	84.0%
95% CI	83.2% - 86.9%	82.3% - 86.0%	82.1% - 85.8%
No Information rate	81.2%	81.2%	81.2%
Sensitivity	98.3%	97.7%	97.2%
Specificity	28.7%	26.2%	26.3%

6.5 Analysis and Discussion of Results

There is broad consensus in previous research supporting that management attitude and target size are important determinants of corporate takeover outcomes, as outlined in chapter 3. The results presented in this thesis reinforce this assertion. Table 4 and Figure 12 show that management attitude is one of the most important predictive variables in terms of contributing to prediction accuracy for the LASSO logistic regression model and the random forest model. This interpretation makes logical sense since hostile takeover attempts can trigger various forms of takeover defence strategies that seek to impede the acquiring party from being successful. Furthermore, the variable importance plot for the random forest suggest that total deal value, which can be seen as a proxy for target size, also ranks high in terms of predictive contribution. Concerning the somewhat contradictory results regarding the usage of bid premium as a predictive variable outlined in chapter 3, the above results support the conclusions made by Walkling (1985) and Jetley and Ji (2010), namely that bid premium in fact is a significant determinant of the outcome variable. This in turn makes sense intuitively. If bid premiums do not impact the probability of a takeover attempt being successful, the natural question is why bid premiums exist at all.

Chapter 3 also discuss research on the importance of payment structure as a predictive variable, where Mitchell and Pulvino (2001) and Branch and Yang (2003) show that the underlying payment structure in a takeover has significant impact on the outcome variable. However, this conclusion is not firmly supported by the results presented in this thesis. On the contrary, the variable importance plot in Figure 12 suggests that payment type carries little explanatory power. It could be argued that cash transactions are easier for investors to understand and in turn increase the probability of the takeover being successful. Be that as it may, it could very well be argued that institutional investors possess a level of sophistication that should allow them to accurately evaluate terms of more complex structures. Additionally, the payment type variable occurs among the top ten most influential variables in the LASSO model, however not standalone but in interaction with the variables percent sought and bid premium. This in turn suggests that payment type in combination with other variables could be relevant when predicting the outcome variable.

The random forest model suggest that completion time contributes significantly to the prediction accuracy. This result appears to be reasonable since the variable is closely linked to completion risk and is in line with the results Baker and Savasoglu (2002) report. Furthermore, the LASSO model suggests an interaction between completion time and rivalry bids. Branch and Wang (2009) conclude that rivalry bids have a significant impact on the outcome variable, though it should be noted that the authors use a 10% significance level. The relatively low coefficient estimate for the rivalry bid variable in the LASSO model and moderate importance indicated by the random forest, suggest that the results in this thesis do not firmly support the variable as having a significant impact. Finally, the results above indicate that the variables percent sought and financial leverage contribute to the overall prediction accuracy to some extent. However, there is weak support for claiming that transactions where the acquiring party is a private equity firm would affect the outcome variable in a notable way.

As described in section 3.8, much of the previous literature discussed in this thesis is focused on variable selection and the return profile of merger arbitrage. Only Hoffmeister and Dyl (1981) and Walkling (1985) report out-of-sample validation statistics. The former study performed out-of-sample validation using several model specifications on testing data samples consisting of 7-33 observations. Overall prediction accuracy ranges from 54.5% to 84.8%, sensitivity ranges from 33.3% to 100.0%, and specificity ranges from 28.6% to 100.0%. These statistics are not reliable due to the limited number of out-of-sample observations, which is also highlighted by the authors. The latter study reports an overall accuracy of 76%, a sensitivity of 73.7%, and a specificity of 83.3% on the out-of-sample set consisting of 50 transactions from one isolated year. Once again, the small testing sample limits the reliability of the results and are not entirely comparable to those presented in section 6.2.2 and 6.3.2. This is because validation performed on a small testing sample, stretching over a short period of time, can be heavily influenced by the random composition resulting from the sub-sampling procedure.

Finally, the models fitted in this thesis are optimised to achieve the maximum overall prediction accuracy. However, given the asymmetric return profile of the merger arbitrage strategy, this approach might be faulty if the purpose is to develop an investment strategy. Assuming that an investor takes a long position in a target company and the takeover attempt fails, the potential losses are usually far greater than the potential gains in case of a positive outcome. In this respect, predicting the negative outcomes accurately might be more important than correctly classifying the positive outcomes. Hence, if the models described above are to be used in an applied trading strategy, the models should be optimised to maximise the expected returns of the investment strategy rather than only maximising the overall prediction accuracy.

7 Concluding Remarks

This chapter concludes this thesis by returning to the initial research question posed in chapter 1. Additionally, suggestions for further research related to re-sampling methods and trading applications are made.

7.1 Conclusion

The purpose of this thesis was to investigate if machine learning based classification procedures provide superior prediction performance compared to LASSO logistic regression models when predicting the outcome in corporate takeover situations. The models were fitted using a training data set consisting of 5 922 OECD-domiciled transactions and evaluated on a testing data set consisting of 1 481 observations. At the outset, variable selection was based on the extensive research already done within the field of takeover outcome prediction. The random forest model utilised all predictive variables previously shown to carry explanatory power in the analysed classification setting. Furthermore, variable selection for the logistic regression model was conducted by applying the LASSO procedure. Some of the predictive variables, such as target size and management attitude, were shown to carry significant explanatory power, reinforcing results from previous research. Meanwhile, other variables discussed in earlier literature were somewhat ambiguous in terms of predictive ability when applied to the data sample underlying this thesis.

With a focus on model comparison, the results presented in this thesis suggest that the random forest model outperforms the logistic regression model on all relevant validation measures, particularly with regards to specificity. This conclusion indicates that the outcome variable is not entirely linearly separable and that the random forest, in the context of the underlying data, divides the predictor space more efficiently by more easily acknowledging non-linear data structures. Given that a significant amount of previous research has been done using logistic regression, this thesis might provide cause for considering alternative and complementary classification procedures when analysing and predicting the outcomes in takeover events. Finally, given that both models achieved an overall prediction accuracy higher than the corresponding No Information Rate on out-of-sample data, there is evidence supporting that merger arbitrageurs stand to gain from using quantitative prediction models when evaluating investment opportunities.

7.2 Suggested Further Research

One palpable shortcoming in both the random forest model and the logistic regression model is that of low specificity. Therefore, further research into methods for improving the specificity scores could provide useful in terms of prediction performance. As briefly mentioned in section 4.5, one such method might entail various forms of re-sampling procedures. Additionally, the only machine learning approach applied in this thesis was the random forest framework. Since there is a range of other machine learning based classification methods, it might be of interest to evaluate other such models. For example, one could apply a neural network approach and evaluate how the prediction performance compares to that of the random forest.

Finally, a natural and crucial next step in enabling the models presented in this thesis to be used in an applied investment strategy, would be to consider how to adjust the models to account for the asymmetric merger arbitrage return profile. One approach could entail optimising the respective models to maximise expected returns of the merger arbitrage strategy rather than overall prediction accuracy. That is, optimise overall accuracy in tandem with specificity and sensitivity in a system that accounts for asymmetric returns.

8 References

Ali, A., Ralescu, A.L., & Shamsuddin, S.M. (2015). Classification with class imbalance problem: A review, *International Journal of Advances in Soft Computing and its Applications*, vol. 7, no. 3, pp. 176-204

Baker, M., & Savasoglu, S. (2002). Limited arbitrage in mergers and acquisitions, *Journal of Financial Economics*, vol. 64, no. 1, pp. 91-115

BarclayHedge. (2012). Understanding Merger Arbitrage, Available online: https://www.barclayhedge.com/insider/hedge-fund-strategy-merger-arbitrage [Accessed 11 April 2020]

Basto, M., da Silva, A.M., & Pereira, J.M. (2015). The logistic lasso and ridge regression in predicting corporate failure, *Procedia Economics and Finance*, vol. 39, pp. 634-641

Branch, B., & Wang, J. (2009). Takeover Success Prediction and Performance of Risk Arbitrage, *Journal of Business and Economic Studies*, vol. 15, no. 2, pp. 10-25

Branch, B., & Yang, T. (2003). Predicting Successful Takeovers and Risk Arbitrage, *Quarterly Journal of Business and Economics*, vol. 42, no. 1/2, pp. 3-18

Breiman, L. (2001). Random forests, *Machine Learning*, vol. 45, no. 1, pp. 5-32

Byström, H. (2014). *Finance, Third Edition*, Lund: Studentlitteratur

Gaughan, P.A. (2015). *Mergers, Acquisitions and Corporate Restructurings, Sixth Edition,* New Jersey: Wiley

Goedhart, M., Koller, T., & Wessels, D. (2010). *Valuation - Measuring and managing the value of companies, Fifth Edition*, New Jersey: Wiley

Hastie, T., James, G., Tibshirani, R., & Witten, D. (2013). *An Introduction to Statistical Learning: With applications in R*, New York: Springer

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning - Data mining, inference, and prediction*, New York: Springer

Hoffmeister, J.R., & Dyl, E.A. (1981). Predicting Outcomes of Cash Tender Offers, *Financial Management*, vol. 10, no. 5, pp. 50-58

Jetley, G., & Ji, X. (2010). The Shrinking Merger Arbitrage Spread: Reasons and implications, *Financial Analysts Journal*, vol. 66, no. 2, pp. 54-68

Kirchner, T. (2009). *Merger Arbitrage - How to profit from event-driven arbitrage*, New Jersey: Wiley

Melka, L., & Shabi, A. (2013). *Merger Arbitrage - A fundamental approach to event-driven investing*, Chichester: Wiley

Mitchell, M., & Pulvino, T. (2001). Characteristics of Risk and Return in Risk Arbitrage, *The Journal of Finance*, vol. 56, no. 6, pp. 2135-2175

PwC. (2016). Mergers & Acquisitions, Available online: https://www.pwc.nl/nl/assets/documents/pwc-mergers-acquisitions.pdf [Accessed 14 April 2020]

Sheather, S.J. (2009). *A Modern Approach to Regression with R*, New York: Springer

Walkling, R.A. (1985). Predicting Tender Offer Success: A logistic analysis, *Journal of Financial and Quantitative Analysis*, vol. 20, no. 4, pp. 461-478

9 Appendix

Appendix I – Comprehensive List of LASSO Logistic Regression Coefficients

Table 8: Comprehensive list of coefficient estimates for the logistic regression model based on the LASSO procedure.

The notation used in the variables column in this table is the following: variable names written "stand-alone" refers to the actual variable, while variable names separated by a ":" refer to the interaction between the two variables.

Variable	Coefficient
Intercept	1.164750000
time_completion	0.002288839
rivalry_bid	0.000000000
total_value	-0.000011147
payment_type	0.000000000
Attitude	1.383439000
bid_premium	0.000807709
leverage	0.000000000
percent_sought	-0.025160240
pe	0.000000000
time_completion:rivalry_bid	0.004947137
time_completion:total_value	-0.000000032
time_completion:payment_type	-0.000875017
time_completion:attitude	-0.004881499
time_completion:bid_premium	0.000000000
time_completion:leverage	0.000000000
time_completion:percent_sought	0.000045367
time_completion:pe	-0.001993792
rivalry_bid:total_value	-0.000103726
rivalry_bid:payment_type	0.000000000
rivalry_bid:attitude	-1.204725000
rivalry_bid:bid_premium	0.000000000
rivalry_bid:leverage	0.000000000
rivalry_bid:percent_sought	-0.008192736
rivalry_bid:pe	0.000000000
total_value:payment_type	-0.000007119
total_value:attitude	0.000000000
total_value:bid_premium	0.000000000
total_value:leverage	0.000000000
total_value:percent_sought	0.000000000
total_value:pe	0.000000000
payment_type:attitude	-0.035156000
payment_type:bid_premium	0.001120152
payment_type:leverage	0.000034174
payment_type:percent_sought	0.002633868
payment_type:pe	-0.290252600

Variable	**Coefficient**
attitude:bid_premium	0.000000000
attitude:leverage	0.000000000
attitude:percent_sought	0.016443220
attitude:pe	-0.067609710
bid_premium:leverage	0.000000000
bid_premium:percent_sought	0.000000000
bid_premium:pe	0.000000000
leverage:percent_sought	0.000000026
leverage:pe	0.000052795
percent_sought:pe	0.000000000

Appendix II – Pearson Correlation Matrix

Table 9: Pearson correlation matrix for all predictive variables.

TC = Total to completion, RB = Rivalry bid, TV = Total value, PT = Payment type, A = Attitude, BP = Bid premium, L = Leverage, PS = Percent sought, PE = Private equity.

	TC	RB	TV	PT	A	BP	L	PS	PE
TC	**1.000**	-0.039	0.170	-0.098	-0.054	-0.012	-0.002	-0.022	-0.016
RB	-0.039	**1.000**	0.002	0.059	-0.066	-0.009	0.027	0.051	0.018
TV	0.170	0.002	**1.000**	-0.083	-0.120	-0.027	0.002	0.074	-0.030
PT	-0.098	0.059	-0.083	**1.000**	-0.073	0.085	-0.003	-0.123	0.320
A	-0.054	-0.066	-0.120	-0.073	**1.000**	-0.007	-0.002	-0.019	-0.040
BP	-0.012	-0.009	-0.027	0.085	-0.007	**1.000**	0.001	0.068	-0.021
L	-0.002	0.027	0.002	-0.003	-0.002	0.001	**1.000**	0.008	0.005
PS	-0.022	0.051	0.074	-0.123	-0.019	0.068	0.008	**1.000**	-0.059
PE	-0.016	0.018	-0.030	0.320	-0.040	-0.021	0.005	-0.059	**1.000**

Appendix III – Comprehensive List of Logistic Regression Coefficients

Table 10: Comprehensive list of coefficient estimates for the ordinary logistic regression model.

The notation used in the variables column in this table is the following: variable names written "stand-alone" refers to the actual variable, while variable names separated by a ":" refer to the interaction between the two variables. Note the following notation for the significance levels for each estimated coefficient: *** = 0, ** < 0.001, * < 0.01, ^ < 0.05.

Variable	Coefficient	Std. Error	P-Value	
(Intercept)	2.098000	0.838000	0.012280	*
time_completion	0.003757	0.002088	0.071896	^
rivalry_bid	0.032910	2.399000	0.989054	
total_value	-0.000050	0.000047	0.281988	
payment_type	-1.140000	0.611400	0.062294	^
attitude	1.324000	0.668200	0.047562	*
bid_premium	0.008409	0.006558	0.199772	
leverage	-0.000816	0.001004	0.416634	
percent_sought	-0.040200	0.008456	0.000002	***
pe	0.932100	0.657400	0.156253	
time_completion:rivalry_bid	0.006405	0.003031	0.034574	*
time_completion:total_value	0.000000	0.000000	0.107133	
time_completion:payment_type	-0.001308	0.000896	0.143982	
time_completion:attitude	-0.006246	0.001050	0.000000	***
time_completion:bid_premium	-0.000005	0.000007	0.409337	
time_completion:leverage	-0.000002	0.000001	0.162460	
time_completion:percent_sought	0.000052	0.000018	0.004981	**
time_completion:pe	-0.002455	0.000859	0.004278	**
rivalry_bid:total_value	-0.000145	0.000104	0.162106	
rivalry_bid:payment_type	0.099730	0.570400	0.861197	
rivalry_bid:attitude	-1.392000	0.594000	0.019089	*
rivalry_bid:bid_premium	0.001336	0.008651	0.877287	
rivalry_bid:leverage	-0.000181	0.000346	0.600304	
rivalry_bid:percent_sought	-0.009794	0.023120	0.671786	
rivalry_bid:pe	0.137400	0.496400	0.781928	
total_value:payment_type	-0.000010	0.000011	0.377414	
total_value:attitude	0.000001	0.000011	0.954945	
total_value:bid_premium	0.000000	0.000000	0.779833	
total_value:leverage	0.000000	0.000000	0.988605	
total_value:percent_sought	0.000000	0.000000	0.381414	
total_value:pe	0.000006	0.000025	0.797534	
payment_type:attitude	-0.182700	0.256000	0.475554	
payment_type:bid_premium	0.000573	0.002050	0.780049	
payment_type:leverage	0.000122	0.000306	0.690984	

Variable	Coefficient	Std. Error	P-Value	
payment_type:percent_sought	0.017240	0.005854	0.003237	**
payment_type:pe	-0.603300	0.396400	0.128009	
attitude:bid_premium	-0.001671	0.002848	0.557325	
attitude:leverage	-0.000071	0.000313	0.820813	
attitude:percent_sought	0.021530	0.006536	0.000988	***
attitude:pe	-0.103600	0.249800	0.678422	
bid_premium:leverage	-0.000001	0.000003	0.821287	
bid_premium:percent_sought	-0.000052	0.000060	0.380839	
bid_premium:pe	0.000178	0.002090	0.932286	
leverage:percent_sought	0.000011	0.000008	0.204488	
leverage:pe	0.000179	0.000324	0.580605	
percent_sought:pe	-0.006561	0.005104	0.198671	

Appendix IV – R Code

 # Load packages

```
library(randomForest)
library(readxl)
library(caret)
library(ggplot2)
library(glmnet)
library(Rcpp)
library(e1071)
library(ROSE)
library(tidyverse)
library(stringr)
```

 # Import data

```
master.data <- read_excel("file path", sheet = 5, range = "B3:Q7406", col_names = TRUE)
data <- read_excel(“file path", sheet = 5, range = "G3:P7406", col_names = TRUE)
raw.data <- read_excel("file path", sheet = 5, range = "G3:P7406", col_names = TRUE)
```

 # Data manipulations

```
master.data$year <- substring(master.data$announce_date, 1, 4)
data$rivalry_bid <- as.factor(data$rivalry_bid)
data$payment_type <- as.factor(data$payment_type)
data$outcome <- as.factor(data$outcome)
data$attitude <- as.factor(data$attitude)
data$pe <- as.factor(data$pe)
```

Data partitioning

```
set.seed(123)

data.set.size <- floor(nrow(data) * 0.8)

index <- sample(1:nrow(data), size = data.set.size)

data.training <- data[index, ]

data.testing <- data[-index, ]

data.training <- as.data.frame(data.training)

data.testing <- as.data.frame(data.testing)

data.training.x <- model.matrix(outcome ~ .^2, data.training)[ ,-1]

data.training.y <- data.training$outcome

data.testing.x <- model.matrix(outcome ~ .^2, data.testing)[ ,-1]

data.testing.y <- data.testing$outcome
```

Random forest model

```
set.seed(222)

mtry <- tuneRF(data.training[ ,-5], data.training[ ,5], stepFactor = 0.5, plot = TRUE, ntreeTry
= 300, trace = TRUE, improve = 0.05)

mtry_opt <- mtry[,"mtry"][which.min(mtry[,"OOBError"])]

rf.1 <- randomForest(outcome ~ ., data = data.training, mtry = mtry_opt, ntree = 300,
importance = TRUE)

print(rf.1)

plot(rf.1, family = "serif", main = NULL)

importance(rf.1)

varImpPlot(rf.1, family = "serif", main = NULL)

hist(treesize(rf.1), breaks = 50, ylim = c(0,30), main = NULL, xlab = "Tree size (number of
nodes)")

rf.1.pred <- predict(rf.1, data.testing, type = "response")

confusionMatrix(rf.1.pred, data.testing$outcome, positive = "1")
```

LASSO logistic regression model

```
set.seed(254)

lr.1 <- glmnet(x = data.training.x, y = data.training.y, family = "binomial", alpha = 1)

plot(lr.1, xvar = "lambda")

cv <- cv.glmnet(data.training.x, data.training.y, type.measure = "class", family = "binomial",
alpha = 1)

plot(cv)

cv_opt <- cv$lambda.min

lr.1.pred <- predict(lr.1, s = cv_opt, newx = data.testing.x, type = "response")

lr.1.pred <- as.factor(ifelse(lr.1.pred > 0.5, 1, 0))

confusionMatrix(lr.1.pred, data.testing$outcome, positive = "1")

predict(lr.1, family = "binomial", alpha = 1, s = cv_opt, type = "coef")
```

Standard logistic regression model

```
glm <- glm(outcome ~ .^2, data = data.training, family = binomial)

summary(glm)

glm.probs <- predict(glm, newdata = data.testing, type = "response")

glm.prediction <- as.factor(ifelse(glm.probs > 0.5, 1, 0))

confusionMatrix(glm.prediction, data.testing$outcome, positive = "1")
```