Rev Account Stud (2017) 22:1366–1422
DOI 10.1007/s11142-017-9407-1

Corporate bankruptcy prediction: a high dimensional analysis

Stewart Jones[1]

Published online: 17 June 2017

Abstract Much bankruptcy research has relied on parametric models, such as multiple discriminant analysis and logit, which can only handle a finite number of predictors (Altman in The Journal of Finance 23 (4), 589–609, 1968; Ohlson in Journal of Accounting Research 18 (1), 109–131, 1980). The gradient boosting model is a statistical learning method that overcomes this limitation. The model accommodates very large numbers of predictors which can be rank ordered, from best to worst, based on their overall predictive power (Friedman in The Annals of Statistics 29 (5), 1189–1232, 2001; Hastie et al. 2009). Using a sample of 1115 US bankruptcy filings and 91 predictor variables, the study finds that non-traditional variables, such as ownership structure/concentration and CEO compensation are among the strongest predictors overall. The next best predictors are unscaled market and accounting variables that proxy for size effects. This is followed by market-price measures and financial ratios. The weakest predictors overall included macro-economic variables, analyst recommendations/forecasts and industry variables.

Keywords Corporate bankruptcy modelling · Gradient boosting · Logit · Market prices · Financial ratios

JEL classification C1 · M4

1 Introduction

Corporate bankruptcy prediction dates back at least 50 years (Beaver 1966; Altman 1968). This literature continues to grow, one reason being that distress forecasts are

Data availability From the public sources identified in the paper.

✉ Stewart Jones
stewart.jones@sydney.edu.au

[1] The University of Sydney Business School, The University of Sydney, Building H69 Cnr Codrington and Rose St, Darlington, NSW 2006, Australia

used in many contexts, such as evaluation of loan security by lenders, going-concern evaluations by auditors, and the pricing of assets exposed to credit risk.

Much bankruptcy research has relied on parametric models, such as multiple discriminant analysis (MDA) and logit, which can only handle a finite number of predictors (Altman 1968; Ohlson 1980). Increasing the number of predictors usually leads to over-fitting, which reduces the overall validity of the model. Given these constraints, most bankruptcy studies use a limited set of predictors, typically financial ratios, market-price variables, or both. The purpose of this study is to introduce the gradient boosting model, which is a statistical learning method that overcomes this limitation. The model can accommodate very large numbers of predictors, potentially many thousands, including their interaction effects. The gradient boosting model provides a rank ordering of these predictors, from best to worst, based on their overall predictive power (Friedman 2001; Hastie et al. 2009). Unlike parametric models, it does not rely on the rules of statistical inference, such as significance tests, but is a data analysis approach that uses predictive ability as the basis for variable selection and ranking.

This study uses the gradient boosting model to examine 91 traditional and non-traditional bankruptcy predictors, including financial ratios, market-price indicators, stockholder concentration/structure variables, macroeconomic factors, external ratings, and other variables. The purpose of this study is not to find the best possible bankruptcy model for prediction but to evaluate the relative influence of alternative predictors on the bankruptcy outcome. This is useful to current debates in the literature and is an important practical consideration in the development of any bankruptcy model.

Following Altman (1968) and Ohlson (1980), a number of studies have focused on the predictive value of financial ratios. However, over the years, the literature has examined a broader range of accounting variables, including a number of non-accounting predictors, such as market-price variables (Shumway 2001; Hillegeist et al. 2004; Beaver et al. 2005), corporate governance proxies (Daily and Dalton 1994), and analyst recommendations and forecasts (Clarke et al. 2006; Jones and Johnstone 2012). The gradient boosting model provides a potentially useful method for drawing together many of these predictors and comparing their performance in a single statistical framework.

This study contributes to the bankruptcy prediction literature in at least two ways. First, I introduce the gradient boosting model, which is a technique that combines the predictions of many weak classifiers into a strong classifier (Friedman 2001).[1] It can be contrasted with logit. Logit uses maximum likelihood to find the most parsimonious set of predictors that improves model-fit and discards all other predictors. The model's parameters are optimized jointly, which means that logit cannot handle too many variables before over-fitting and model convergence can become intractable problems. By contrast, the gradient boosting model uses the entire feature space (all predictor variables) to improve prediction outcomes. Rather than finding a small set of strong predictors, it trains many weak predictors where each new predictor learns from the mistakes of the previous predictor. The model keeps adding new predictors until some desired low error rate is achieved. The coefficient for each new predictor is optimized one at a time (a stage-wise process), rather than jointly. Training many weak classifiers

[1] A weak classifier is one that predicts a little better than a random guess.

significantly improves prediction outcomes while remaining highly resilient to model over-fitting (Friedman 2001; Schapire and Freund 2012).[2]

The contribution of different predictors can be assessed through the relative variable importance (RVI) measure, which are rank ordered based on each predictor's weighted classification accuracy averaged across all predictors used in the model (Friedman 2001). The weighted improvements are summed and then scaled relative to the top performing predictor. The predictor with the largest sum of improvements is scored 100, and all other predictors will have lower scores descending towards zero, indicating the variable adds little or nothing to the model's overall predictive power. The gradient boosting model also provides partial dependence plots or marginal effects, which are useful for assessing the direction of the predictor's influence on the dependent variable, whether positive or negative, including the magnitude of this effect. The predictive success of the model is interpreted using a number of traditional outputs, such as receiver operating characteristic (ROC) curves, classification tables and log-likelihood. The gradient boosting model also has particular benefits in terms of implementation, as it requires modest researcher intervention and data preparation.[3] The model can be implemented in many commercial software packages, including freeware such as *R*.

Second, based on a sample of 1115 U.S. bankruptcy filings, the findings show that non-traditional bankruptcy variables such as ownership concentration/structure and CEO compensation are among the strongest predictors overall. Collectively, these variables have an average RVI of 48.18. Ownership variables include level of stockholder concentration, percentage of institutional ownership, and percentage of insider ownership. The percentage of stock owned by the top 5 stockholders is the top ranked variable in the gradient boosting analysis. The next best predictors are unscaled market and accounting variables that proxy for size effects. These variables have an average RVI of 32.24 and include market capitalization, total revenue, total assets, total operating cash flows, and total debt. These are followed by market-price indicators, such as excess returns and stock price volatility, with an average RVI of 21.92. Financial ratios performed comparably well to market-price variables, with an average RVI of 17.43, the best being earnings per share, cash flow per share, interest cover, capital expenditure to total assets, total liabilities to total equity, and the current ratio. Other non-accounting variables, such as auditor type, number of business segments, age of the firm, and credit rating changes, show comparable predictive strength with an average RVI of 17.24.[4] Variables with the weakest predictive power in the model include industry variables (average RVI of 1.93), macroeconomic variables such as inflation and employment rates (average RVI of 3.46), and analyst recommendations/ forecasts (average RVI of 4.67).

[2] Gradient boosting has conceptual similarities with forward stepwise regression. This method starts with an intercept term and sequentially adds variables based on statistical criteria such as model-fit improvement. However, stepwise has many limitations not shared by gradient boosting. For instance, parameter estimates, R-squares and t-values can be severely over-stated and lack interpretability (see Harrell, 2001).

[3] In contrast to conventional models, gradient boosting estimation is less affected by outliers, variable scaling/ transformations, inclusion of irrelevant inputs, missing observations, database errors, and even data snooping—this type of researcher intervention can introduce significant bias to the modeling exercise (Freedman 2010).

[4] The averages discussed above ignore lower impacting variables with RVIs less than 10.

The study also finds that the out-of-sample predictive accuracy of gradient boosting is best in high dimensions (using all predictors and interactions) and significantly outperforms a logit model, irrespective of the number or types of predictors used. The results are found to be robust to different over-fitting and out-of-sample test conditions.

2 Literature and hypothesis development

At least two major themes have emerged from the corporate bankruptcy literature over the past five decades. The first relates to the role and predictive power of alternative bankruptcy predictors. A second theme relates to developments in statistical models to predict corporate failure. I briefly discuss each before introducing the study's hypotheses.

2.1 Alternative bankruptcy predictors

The literature has mainly focused on the role of accounting-based or market-price indicators in bankruptcy prediction modelling (e.g., Altman 1968; Altman et al. 1977; Ohlson 1980; Zmijewski 1984; Shumway 2001; Altman 2002; Duffie and Singleton 2003; Hillegeist et al. 2004; Jones and Hensher 2004; Beaver et al. 2005; Jones and Hensher 2008). A smaller number of studies have investigated other potentially important bankruptcy predictors, including corporate governance proxies (such as stockholder concentration/structure), analyst estimates/forecasts, credit ratings changes, macroeconomic factors, and other industry and firm-specific factors (Jones et al. 2015, 2017).

The use of accounting-based measures (such as financial ratios) to predict bankruptcy has a long history (Beaver 1966; Altman 1968; Zmijewski 1984). Beaver et al. (2005) observe that accounting-based predictors have proven remarkably robust over a 40-year period (1962–2002). Across many empirical studies, a range of accounting-based measures have been shown to have predictive power in corporate bankruptcy. However, measures associated with working capital, cash flow, earnings, and leverage have surfaced as key predictors of firm financial distress in many studies (Altman 2002; Beaver et al. 2005; Jones and Hensher 2008).

As noted by Beaver et al. (2005, 95–96): "The precise combination of ratios used seems to be of minor importance with respect to overall predictive power, because the explanatory variables are correlated." However, conventional bankruptcy models such as logit/probit and LDA are low dimensional models; that is, they are severely limited by the multicollinearity condition (and other statistical problems), which restricts the number of input variables that can be tested in the model. These models have limited capacity to extract signals from other potentially important variables and related interaction effects. The gradient boosting model enables the testing of a much wider range of accounting-based indicators, irrespective of their correlation with other variables and without any cost to model stability or performance. For the purposes of this study, I use several traditional ratios and indicators relating to (1) liquidity and solvency evaluation, (2) capital structure and leverage, (3) profitability and rate of return, (4) turnover and asset efficiency, (5) operating cash flow performance and free cash flows, and (6) change/growth variables, including annual growth in operating cash flow, earnings, revenues and debt. Many of these variables have been tested in prior studies and improvement or deterioration in these ratios and indicators have an intuitive relationship to financial distress. I also use several unscaled measures, such as total revenue, total assets, and total debt, which are useful proxies for firm size effects. Unscaled measures have certain advantages over financial ratios in some situations. For instance, ratios are subject to many types of distortion, such as the proportionality condition, which imposes a strict linearity between the numerator and denominator. However, relationships between financial measures can be highly nonlinear (Foster 1986). While unscaled measures can introduce significant heteroscedasdicity into the dataset, the gradient boosting model is largely insensitive to the shape and structure of data. (For instance, outliers and extreme observations generally do not affect the estimation; see Friedman 2001).

Other accounting-based indicators that might plausibly be associated with corporate failure include total asset write downs and impairments, total investment in intangibles, capital expenditure, and dividends per share. (See Appendix Table 11 for definitions of all variables used in the study.) The level and extent of asset write-downs and impairments is expected to be associated with declining financial performance and therefore a higher likelihood of financial distress. Significant reductions in investments in intangibles, capital expenditure, or dividend payments have also been associated with higher levels of financial distress in prior literature (e.g., DeAngelo and DeAngelo 1990; Asquith et al. 1994; Jones 2011).

The method for selecting accounting measures is to include all important measures used in prior literature and any other plausible indicator that might be associated with corporate distress, even if not tested in prior literature. The reason for including so many accounting-based variables is twofold. First, there is no strong theoretical or empirical evidence pointing to the importance of one financial indicator or ratio class over another (particularly as financial measures tend to be highly correlated), although the bankruptcy literature and experience tells us that some measures should be more important than others in the prediction of financial distress. As the gradient boosting model is effective in extracting a signal from highly correlated predictors, the aim here is to see how well it differentiates among accounting variables involving many correlated and possibly redundant measures. A significant amount of information redundancy is expected from the financial indicators displayed in Appendix Table 11, but where will the redundancy be most concentrated and which particular accounting inputs will feature more prominently in the gradient boosting analysis? By using a large number of accounting inputs, the gradient boosting model can search a wide feature space to determine which accounting inputs have the strongest predictive value overall. The results are expected to foster more theoretical discussion on the role and influence of different financial predictors in bankruptcy prediction. As will be shown in the results section, using a large number of accounting indicators does not destabilize the gradient boosting model nor significantly impact on the predictive power and relative importance of other strong non-accounting-based predictors (such as market price and ownership structure/concentration variables).

Market-price variables There has been significant scholarly interest in the predictive power of market-price indicators. Variables, such as market capitalization, market-to-book, excess (abnormal) stock returns, and price volatility have been used in various studies (Altman et al. 1977; Beaver et al. 2005; Shumway 2001; Bharath and Shumway 2008). The theoretical rationale for market-price measures is compelling. As pointed out by Hillegeist et al. (2004), financial statements have a number of fundamental limitations. For instance, they are backward looking and prepared on the basis of going-concern and conservatism assumptions, which can limit their usefulness. Furthermore, financial statement data does not capture volatility effects, which can have predictive value in corporate bankruptcy models (Beaver et al. 2005).[5] Importantly, the market value of equity represents the equity cushion and reflects the amount by which the value of assets can decline before they become insufficient to cover the present value of the debt payments. As the equity cushion diminishes, the probability of failure is expected to increase (Kealhofer and Kurbat 2001; Chava and Jarrow 2004; Hillegeist et al. 2004; Beaver et al. 2005). A number of empirical studies have shown that bankrupt and distressed firms experience deteriorating excess returns prior to failure (Dichev 1998; Frino et al. 2007, 2014). Excess returns are a useful bankruptcy predictor because stock prices can impound rapidly changes in general market conditions and firm-specific factors that might alter perceptions of credit quality and financial distress (Altman 2002).

Accounting-based vs. market-price indicators Several studies have compared the relative importance of accounting-based ratios and market-price variables. Some empirical studies indicate that market-price variables have stronger predictive power in bankruptcy analysis. However, the evidence is mixed. Shumway (2001) suggests a model based on both accounting-based and market-price variables produces quite accurate forecasts. On the other hand, Hillegeist et al. (2004) evaluate whether Altman's (1968) Z-Score and Ohlson's (1980) O-Score effectively summarize publicly available information about the probability of bankruptcy. They conclude that market-based probability estimates of corporate bankruptcy (based on the Black-Scholes-Merton option-pricing model) are superior, even recommending that future research should focus exclusively on market-price indicators.

Beaver et al. (2005) take a more balanced perspective and argue that market-price indicators are endogenous variables in the sense that they are function, among other things, of the financial statement variables themselves. In this sense, they should not be considered a substitute for accounting-based information "but rather a proxy for the predictive power attainable by capturing the total mix of information, including both financial statement and non-financial statement information" (p.112). Beaver et al. (2005) conclude that market-price variables absorb much of the predictive power of financial statement variables and provide additional explanatory power not reflected in the financial ratios. However, their overall results are consistent with nonfinancial information compensating for a slight loss in predictive power of financial ratios.[6]

[5] Hillegeist et al. (2004) argue that asset volatility is important because it indicates the likelihood that the value of the firm's assets will decline to such a degree that the firm will be unable to repay its debts. In addition to providing direct measures of volatility, Beaver et al. (2005) argue that market price variables are appealing because they reflect a rich and comprehensive mix of information, which includes financial statement data as a subset. Furthermore, market prices can also be measured with a finer partition of time. Whereas most bankruptcy studies use annual data, market prices can exploit the availability of daily prices (Frino et al. 2007).

[6] According to Beaver et al. (2005), this slight loss in predictive power appears to be due to increased discretion in financial reporting or the increase in intangible assets not being offset by improvements due to additional FASB standards.

While market-price indicators have a strong theoretical basis in corporate bankruptcy prediction, they also suffer limitations that could explain some of the mixed findings in the literature. For instance, market-price indicators rely on capital markets, which may not efficiently or accurately impound all publicly available information, including information provided by financial statements (e.g., Sloan 1996). This may be especially true for smaller firms, which predominate in the bankrupt firm sample. Relative to large companies, smaller firms are typically not followed as closely by market participants, such as institutional investors, analysts, ratings agencies, and the financial media.

In addition to market-price variables, such as market capitalization, market capitalization-to-debt, excess returns and stock price volatility, I also test some new market variables that have not been extensively examined in prior bankruptcy research, such as short interest. In an informed market, it is expected that higher levels of short interest will be associated with increased probability of financial distress. Informed traders will short stocks because they are expecting the firm's performance to deteriorate in some way, which will be impounded in declining future stock price returns.

While the literature has focused on accounting-based and market-price variables, many other explanatory variables have plausible theoretical links to corporate bankruptcy. These include ownership concentration and structure; external ratings, including analyst recommendations/forecasts and credit rating changes; macroeconomic factors; executive compensation; and others.

Ownership concentration/structure Ownership concentration/structure variables, including institutional ownership, stockholder concentration, and insider ownership have not been extensively explored in prior bankruptcy research. However, there are good theoretical reasons for expecting an association between these variables and corporate failure. In a practical sense, large stockholders and institutional investors have a voice in all aspects of the Chapter 11 bankruptcy process. As the Chapter 11 process is typically very complex, costly, and ultimately destructive to stockholder value, there are likely to be strong incentives for these stakeholders to consider all available options and remedial actions that avoid Chapter 11.

Prior literature suggests that stockholder concentration is positively correlated with firm value and performance and that large stockholders can influence company performance in various ways. For instance, Shleifer and Vishny (1997) and Ashbaugh et al. (2006) observe the growing evidence that large stockholders can exercise substantial influence on corporate governance practices and financial performance (e.g., CEO turnover and influencing CEOs to rectify poor financial performance and reduce discretionary spending) (Black 1998). Large stockholders also bear excessive risk because they are not diversified. Hence they may have strong incentive to closely monitor the health of the organization and demand changes from management if necessary. According to Shleifer and Vishny (1997), the most direct way for owners to align cash flows and control rights is through stockholder concentration. They argue that substantial minority stockholders have incentives to collect information and monitor management, thereby avoiding the traditional free rider problem. Substantial minority stockholders have enough voting control to exert pressure on management and even remove management in a proxy fight or takeover (Shleifer and Vishny 1997). Thus large stockholders "address the agency problem in that they both have a general interest in profit maximization and enough control over the assets of the firm to have their interests respected" (Shleifer and Vishny 1997, p.754).

There is also some recent evidence that long-horizon institutional investors are better informed when it comes to bankrupt firms (Ramalingegowda 2014).[7] While research has produced little evidence that institutional investors anticipate major corporate events (such as earnings surprises), Ramalingegowda (2014) finds that long-horizon institutions are better informed and sell more stock of impending bankrupt firms well before the bankruptcy announcement. Ramalingegowda (2014) also finds that stock sales are greater in impending bankrupt firms whose stockholders ultimately lose all of their equity. Precourt and Oppenheimer (2013) report stronger herding effects when assessing what motivates institutional investors to start selling off stock in failing enterprises. They conclude that institutional investors sell well in advance of a bankruptcy filing, relative to firms that have smaller institutional shareholdings.[8]

For this study, I include several ownership concentration/structure variables, including percentage of institutional ownership, percentage of stock held by the top five and 10 stockholders, and percentage of insider ownership. From the discussion above, I expect that the level of insider ownership, institutional ownership, and stockholder concentration to be negatively associated with the likelihood of corporate bankruptcy (and vice versa).

Analyst recommendations/estimates Analyst recommendations are typically based on detailed fundamental analysis of the firm and hence should bear on the financial performance and credit quality of firms. There is evidence that the news effects of analyst forecasts can impact on financial distress indicators such as credit ratings (Ederington and Goh 1998). For instance, a sell recommendation or a significant downgrade could lead to lower security prices and higher price volatility. This in turn can impact corporate bankruptcy estimates, because market-based variables have been documented to have predictive power in bankruptcy models (Frino et al. 2007, 2014). Negative analyst reports or downgrades are expected to be based on current or anticipated financial performance of the firm, which is expected to capture impending financial distress as anticipated by analysts. While prior research is sparse, both Clarke et al. (2006) and Jones and Johnstone (2012) have examined analyst recommendations around corporate bankruptcies. Jones and Johnstone (2012) examine a large sample of corporate bankruptcies during the global financial crisis. They find that, while analysts change their recommendations and forecasts slowly in response to the deteriorating health of companies (a positive bias), these factors are nevertheless statistically significant in their bankruptcy forecasting model. The authors also find that analyst intensity is positively associated with nonfailure and vice versa. Here, corporate bankruptcies tend to be more concentrated in smaller companies that have less or no analyst following. Precourt and Oppenheimer (2013) also find that institutional investors react to sell recommendations issued by security analysts. However, consistent with the results of Jones and Johnstone (2012), they find that on average analysts do not

[7] However, the international evidence is somewhat mixed on this issue (e.g., Frino et al. 2014).

[8] Insider ownership is also likely to play a similar role. If the directors and CEO are also stockholders, their interests are expected to be more closely aligned with the longer-term financial performance of the firm. Insiders with higher levels of stock ownership will have stronger incentives to limit the organization's exposure to bankruptcy risks, which inevitably leads to significant destruction in stockholder value.

materially downgrade their recommendations for failing firms until only a few months before a Chapter 11 filing. Consequently, the institutional investor response to these recommendations can be too late.

For this study, I include analyst stock recommendations, consensus EPS forecasts, and analyst intensity as explanatory variables (see Appendix Table 11). I expect downgrades in recommendations and estimates and lower analyst intensity to be associated with a higher likelihood of firm failure (and vice versa).

Credit ratings changes Some studies have used bond ratings as proxies for corporate distress because bond ratings can include information not generally available in the public arena. Bond ratings are based on both public information and private information conveyed to the rating agencies by firms but may not be made generally public (Barth et al. 1998; Billings 1999; Hillegeist et al. 2004). It is expected that credit ratings downgrades will have a positive association with corporate distress. Credit rating agencies evaluate the creditworthiness of a business, and significant deteriorations in financial performance and broader market conditions are expected to be associated with ratings downgrades, implying a higher probability of corporate bankruptcy (see Appendix Table 11).

Macroeconomic variables It is expected that broader macroeconomic conditions and the state of the economy should have an impact on the frequency of corporate bankruptcies. Previous literature has demonstrated that macroeconomic factors can impact on credit ratings changes including corporate bankruptcy (e.g., Keenan et al. 1999; Bangia et al. 2002; Koopman et al. 2009; Koopman et al. 2011; Figlewski et al. 2012; Hensher et al. 2007). Following Jones et al. (2015) and others, we use several broad macroeconomic indicators relating to the overall economic health of the economy and financial market conditions generally. These indicators include real GDP/real GDP growth, the CPI index, interest rate levels, public debt to GDP, unemployment rates, the NBER recession indicator, the Michigan sentiment index, the Moody's AAA and BBB seasoned bond yields, the leading index and other indicators (see Appendix Table 11).[9]

Executive compensation variables Compensation variables, such as total CEO compensation and share-based compensation, are another set of factors that have not been extensively investigated in prior corporate bankruptcy research. The issue of excessive CEO compensation, particularly in distressed and failing companies, has been a

[9] Macroeconomic variables are extracted from a commercial source TradingEconomics.com. Real GDP and real GDP growth are key measures of overall economic health and prosperity. A strong and growing economy is expected to lead to lower default risk and lower overall probability of corporate bankruptcy. Higher interest rates are often associated with general tightness in the economy and increased likelihood of liquidity and debt-servicing pressures on firms. Inflation is a widely cited economic indicator, and the general perception is that high inflation is unhealthy for the economy, although its economic impacts can be ambiguous (Figlewski et al. 2012; Jones et al. 2015). There is also a common perception that a high ratio of public debt to GDP and high unemployment is a sign of broad economic weakness and vulnerability. Hence I expect these indicators to be positively associated with corporate failure. The NBER recession indicator, the Michigan sentiment index, and the leading index provide more direct measures of current economic sentiment and conditions. The Moody's AAA and BBB seasoned bond yields are broad measures of credit risk. For instance, rising bond yields indicate more risk in the economy including default risk.

controversial in the financial press and much contested in public policy since the global financial crisis. However, CEO compensation is used in this study as a proxy for the overall quality of top level management. Total CEO compensation (which includes performance bonuses) is expected to be closely tied to firm financial performance. If top level managers are performing and meeting organizational objectives, their total compensation is expected to be higher. As better management is expected to be associated with better performing companies, I anticipate a negative relationship between executive compensation and firm financial distress. Share-based compensation is also expected to be negatively associated with the failure outcome, as this form of compensation will more closely align the interests of senior managers with firm performance, particularly performance factors rated highly by capital markets, such as earnings stability and growth.

Other variables From prior literature, I use a number of other control variables, including firm size (usually proxied by market capitalization or total assets), age of firm, and industry background (based on the 10 major GICS sector codes—see Appendix Table 11). Much previous research has demonstrated that corporate failures tend to be more highly concentrated in smaller, more recently established companies, with high numbers of bankruptcies occurring in particular industry sectors. Hence it is important to control specifically for these factors. Other variables tested in the gradient boosting analysis are described in Appendix Table 11.

Interaction effects Interaction effects can add significant explanatory and predictive power to bankruptcy models. However, many bankruptcy studies tend to overlook the importance of interaction effects or incorporate them in very limited and ad hoc way. Numerous interaction effects are possible with the explanatory variables discussed above. In high dimensional contexts, there are potentially many thousands of possible interaction effects that conventional models are clearly not designed to detect or handle. A major benefit of a gradient boosting model is that its structure can be set up to automatically detect all important interaction effects across the full set of explanatory variables. The analysis of interactions at this level can display deeper structures in the dataset and reveal more complex and insightful relationships among predictor variables. I now turn briefly to modelling developments in the bankruptcy literature.

2.2 Bankruptcy modelling developments

Much of the corporate bankruptcy literature has relied on logit/probit and MDA models (e.g., Altman 1968; Altman et al. 1977; Ohlson 1980; Zmijewski 1984; Jones and Hensher 2004, 2008). Furthermore, there has been some debate on the merits of survival models vs. discrete choice models such as logit and MDA (e.g., Shumway 2001; Beaver et al. 2005). A smaller number of studies have compared popular statistical learning methods such as neural networks with logit and MDA (Altman et al. 1994; Jones et al. 2015, 2017). While there have been attempts to introduce more sophisticated mixed model approaches, Jones et al. (2015) indicate that the most frequently used model in credit risk and bankruptcy research is the standard logit model.[10]

Studying corporate bankruptcy in a high dimensional setting is generally not possible with conventional modelling frameworks (Hastie et al. 2009). As stated previously, conventional bankruptcy models become unstable as the number of predictor variables increase. This study introduces an advance in statistical learning based on the gradient boosting model. (For this study, I use a recent commercial version of the model called TreeNet®.) The gradient boosting model can be applied to regression and classification problems. (The formal properties of the model are outlined in Section 3.) Hastie et al. (2009) suggests that the model is now among the most powerful off-the-shelf statistical learning methods available. In the context of credit risk and bankruptcy modelling, Jones et al. (2015, 2017) provide evidence that the gradient boosting model and closely related methods (such as random forests and AdaBoost) can significantly outperform more conventional classifiers, including sophisticated statistical learning techniques, such as neural networks and support vector machines (SVMs) (Hastie et al. 2009; Schapire and Freund 2012).

The usefulness of boosting models to bankruptcy prediction has been recognized in fields outside of accounting and finance. Much of this literature compares the performance of boosting techniques with classifiers such as neural networks and SVMs, while other studies attempt to examine whether conventional techniques, such as neural networks, can be enhanced by the boosting approach.[11] Appendix Table 12 provides a tabular summary of this literature. While most of these studies demonstrate the superior predictive power of the boosting technique, relative to conventional models, they vary significantly with respect to sample sizes and sample type (private vs public company samples), reporting jurisdiction, definitions of corporate distress, number input variables used (most are limited to financial ratios), model architecture (such as tree depth), test samples and, not surprisingly, empirical results.[12] However, what emerges most from this literature is that the gradient boosting approach offers a potentially more robust predictive framework for bankruptcy modelling and is more amenable to the characteristics of data typically encountered in corporate bankruptcy datasets (discussed further under Section 3).

From this background, I develop the following specific hypotheses for testing.

> H_1: *A high dimensional analysis of corporate bankruptcy filings will have greater predictive and explanatory power than a low or unidimensional analysis, ceteris paribus.*

[10] This is based on their review of more than 150 empirical studies in the field.

[11] Boosting is a general method for improving the performance of learning algorithms (the idea of combining and weighting many weak classifiers into a powerful 'voting' committee of classifiers). While most applications of boosting use decision trees as the base classifier, some studies have examined how well the boosting technique works when neural networks or logit are used as the base classifier (examples of such studies are included in Appendix Table 12).

[12] This study offers a number of improvements over this literature which is reflected in the generally stronger empirical results. For instance, this study uses a much larger sample size; a much wider range of predictive inputs (and related interaction effects); and applies a consistent definition of corporate failure (Chapter 11 filings). I also use a prize winning commercial gradient boosting package known as TreeNet®, which is well known for its predictive power. I discuss these issues in more detail in Section 6.

H_2: Market-price variables will have greater predictive power on corporate bankruptcy filings, relative to accounting-based variables, ceteris paribus.
H_3: Market-price variables will have greater predictive power on corporate bankruptcy filings relative to non-accounting bankruptcy predictors, including ownership concentration/structure, external ratings, macroeconomic factors, executive compensation, and other variables, ceteris paribus

I test H_1 by estimating a gradient boosting model on the full set of explanatory variables described in Appendix Table 11 and then comparing its performance to several lower unidimensional models. Similar groups or clusters of variables (such as market-price indicators, accounting-based variables, ownership concentration/structure variables, macroeconomic indicators, and so on) are treated as separate dimensions of corporate bankruptcy. There is no precise statistical test or cutoff rule for determining when H_1 should be accepted. However, the evidence is compelling to accept H_1 if most or all of the explanatory variables in Appendix Table 11 are found to have nonzero relative variable importances (RVI).[13] This will also be a test for the multi-dimensional aspect to corporate bankruptcy, as the explanatory variables examined in this study represent several different dimensions of corporate bankruptcy. The explanatory power of the high dimensional model is also tested by examining whether the variables are internally consistency and have a direction with the failure outcome that makes sense. A common criticism of all sophisticated statistical learning models, including gradient boosting, is that they are black boxes and lack interpretability. However, for a gradient boosting model it is possible to understand the direction and behavioral influence of explanatory variables on the failure outcome through analysis of marginal effects. In a gradient boosting analysis, marginal effects can be highly informative because they reveal the strength, direction, and existence of any nonlinear relationships between the predictor variables and the failure outcome.

The same approach is used to assess predictive accuracy. The key test for evaluating predictive power is comparing the out-of-sample classification accuracy of the high dimensional model with a gradient boosting model estimated on each separate dimension of corporate bankruptcy. It is compelling to accept H_1 on the predictive performance criterion if the high dimensional model outperforms all other lower unidimensional models. Also in terms of H_1, the bankruptcy literature has relied heavily low dimensional models such as logistic regression. Low dimensional models can only lend themselves to low dimensional analysis. Given prior research in the statistical learning literature, the gradient boosting model is expected to be a significantly stronger classifier than conventional models such as logit. I also examine this relationship in both high and low dimensions and in multi-dimensional settings.

With respect to H_2, the efficient market literature implies that market prices reflect a rich and comprehensive mix of information, which includes financial statement data as a subset (Beaver et al. 2005). In this sense, market prices are an endogenous source of information to financial statement information, as stock price returns are expected to impound all publicly available information about firm financial performance. Following previous literature, H_2 is tested by comparing market-price indicators (such as excess returns and stock price volatility) with different accounting-based indicators.

[13] An RVI above zero indicates that the variable adds something to the predictive success of the overall model.

However, prior literature has typically compared the predictive power of market-price and accounting-based indicators using a low dimensional framework. For the reasons discussed above, it is important to evaluate the predictive power of these variables in a high-dimensional and multi-dimensional context, that is, after controlling for all other factors and interactions that can impact on corporate failure.

If accounting-based information is a mere subset of the total mix of the information (financial and nonfinancial) impounded in stock price, market-price variables should prove to be stronger predictors of corporate failure than accounting-based indicators (which means they should have stronger RVI's and marginal effects). Beaver et al. (2005) suggests that market-price indicators and accounting-based measures should not be treated as competing but rather complementary sources of information. They conclude that market-based variables absorb much of the predictive power of financial statement variables and provide additional explanatory power not reflected in the financial ratios. However, if accounting and market-price variables are considered *competing* sources of information, it might be expected that accounting-based information will provide additional information not already captured in stock returns.

In the context of the gradient boosting analysis, I test these relationships more formally by examining the interaction effects of excess returns vs. accounting-based variables. If the excess returns variable clearly dominates accounting-based measures (and vice versa), I would not expect to see strong interaction effects in the results, as the presence of interaction effects suggest that one variable (or set of variables) modifies the predictive impact of the other on the failure outcome. Furthermore, if excess returns and accounting-based measures are complementary sources of information (i.e., excess returns absorb the predictive value of accounting-based indicators), I expect that market-price measures will interact in a way that reduces the predictive impact of accounting-based measures on the nonfailure outcome (and vice versa). For instance, higher excess returns should impound the effects of stronger financial performance. Likewise, lower excess returns should reduce the impact of poorer financial performance on the failure outcome. This is consistent with Beaver et al.'s (2005) argument that these variables are complementary rather than competing. If market-price and accounting-based measures are considered competing/exogenous sources of information, I would expect excess returns to interact (if there are observable interactions) with accounting-based measures in a way that increases the impact of accounting-based measures on the failure outcome. If higher excess returns increase the impact of accounting-based measures, this might suggest that accounting-based variables contain predictive information not already impounded in stock prices.

With respect to H_3, prior literature has focused on comparing the predictive performance of market-price and accounting-based indicators. There are a few studies that have examined the performance of market-price indicators, relative to other non-accounting bankruptcy predictors, including corporate governance proxies (ownership concentration/structure), external ratings of analysts and credit agencies, macroeconomic factors, and other indicators. If market prices reflect the total mix of information (financial and nonfinancial), I expect that market-price variables (such as excess returns) will have stronger overall predictive value than other variables tested in this study. As with accounting-based variables, the impact of these indicators is expected to be impounded into stock price returns, assuming capital market efficiency. The remainder of the paper is organized as follows. Section 3 introduces the empirical framework.

Section 4 discusses the sample and data collection, while Section 5 presents the empirical results. Finally, concluding comments are provided.

3 The gradient boosting model

The general idea behind boosting as set out in the work of Schapire and Freund (2012) is to combine the outputs of many weak classifiers (trees) to produce a powerful overall voting committee.[14] All the individual classifiers can be weak, but, as long as the predictive performance of each weak classifier is slightly better than random guessing (i.e., their error rate is smaller than 0.5 for binary classification), the final model can converge to a *very* strong classifier. The weighted voting is based on the quality of the weak classifiers, and every additional weak classifier improves the prediction outcome. The weak learning algorithm is forced to focus on examples where the previous rules of thumbs provided *inaccurate* predictions. The intuition here is straightforward for corporate bankruptcy prediction. The first weak classifier, which might be (taking a one-level decision tree for illustration) the leverage ratio with a cutoff of >70%, is trained on the data where all observations receive equal weights. Some observations will be misclassified by the first weak classifier. A second classifier—say, the current ratio with a cutoff of >1.5—is initiated to focus on the residuals or trainings errors of the first classifier. The second classifier is trained on the same dataset, but misclassified samples receive a higher weighting while correctly classified observations receive less weight. The re-weighting occurs such that the first classifier gives 50% error (random) on the new distribution. Iteratively, each new classifier focuses on ever more difficult samples. The algorithm keeps adding weak classifiers (trees), often many hundreds in a single model, until some desired low error rate is achieved. More formally, Schapire and Freund (2012, p.5) illustrate how the boosting algorithm works in the case of adaptive boosting or AdaBoost.

1. Train weak learner using distribution D_t.
2. Get weak hypothesis or classifier h_t: $X \rightarrow \{-1, +1\}$ which for this study is a binary outcome of bankruptcy vs. nonfailure.
3. Select weak classifier h_t to minimise weighted error.
4. Choose $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$,

[14] The antecedents of modern boosting models come from the classification and regression trees (CART™) technique. However, despite the early popularity of CART™ (particularly in health diagnostics), the technique became associated with a number of limitations, most notably high variance. (It does not generalize well.) More sophisticated techniques, starting with bagging (Breiman 1996), began to develop, which significantly improved on the performance of CART™. Bagging can dramatically reduce the variance of unstable procedures (like trees), leading to improved prediction outcomes. While bagging is a major improvement on CART™, more sophisticated boosting methodologies, such as random forests, adaptive boosting (AdaBoost), and gradient boosting, began to develop. Random forests are essentially a refined form of bagging. The technique improves on bagging by "de-correlating" the trees, which maximizes the reduction in variance. (See Hastie et al. 2009 for details.) There is now an extensive literature devoted to the gradient boosting framework (including related approaches such as AdaBoost), which highlight the many advantages of this approach and, more particularly, the superior forecasting accuracy of the model. Recent applications include biological sciences (such as DNA research), text and speech recognition and processing, satellite imaging analysis (such as oil spill detection), cyber security, geological mapping, and credit risk. Specific applications of boosting to bankruptcy research are provided in Appendix Table 12.

where α_t is the parameter importance assigned to the weak classifier h_t.

5. Update, for $i = 1,..., m$:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if} h_t(x_i) \neq y_i \end{cases}$$

Output the final hypothesis or strong classifier:

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right).$$

where $H(x)$ is the linear combination of weak classifiers computed by the boosting algorithm. The purpose of boosting is to sequentially apply the weak classification algorithm to repeatedly modified versions of the data, thereby producing a sequence of weak classifiers that predicts very accurately. This study uses a commercial version of boosting known as TreeNet®. TreeNet® is based on the boosting approach described above but with some important technical differences and refinements. For instance, the model is based on a gradient boosting approach initially set out by Friedman (2001). Gradient boosting is a generalization of AdaBoost to handle a variety of loss functions. (AdaBoost uses the exponential loss function and is regarded as a special case of gradient boosting in terms of loss function.)[15] The intuition behind gradient boosting is to build a sequence of predictors with the final classifier being the weighted average of these predictors (a multiple additive regression tree approach). At each stage, the gradient boosting algorithm adds a new classifier that improves the performance of the tree ensemble with respect to minimizing some loss function. The basic model is set out by Friedman (2001) and Hastie et al. (2009). It operates as follows.

1. Initialize $F_0(x) = 0$.
2. For $m = 1$ to M:

(a) Compute

$$(\beta_m, \boldsymbol{a}_m) = \arg\min_{\beta, \boldsymbol{a}} \sum_{i=1}^{N} \boldsymbol{L}(y_i, F_{m-1}(x_i) + \beta h(x_i; \boldsymbol{a})) \ ,$$

(b) set $F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \beta_m h(\boldsymbol{x}; \boldsymbol{a}_m)$,

where $\boldsymbol{L}$ is the loss function to be minimized, y_i is the response variable, F_{m-1} represents the current model, x_i are the explanatory variables of the model, $\boldsymbol{a}$ are the parameters of the input variables, h is a function of the explanatory variables and parameters $\boldsymbol{a}$ and a coefficient β, which is the weight assigned to the tree based on

[15] Both AdaBoost and gradient boosting are conceptually similar techniques. Both approaches boost the performance of a base classifier by iteratively focusing attention on observations that are difficult to predict. AdaBoost achieves this by increasing the weight on observations that were incorrectly classified in the previous round. With gradient boosting, difficult observations are identified by large residuals computed in the previous iterations. The idea behind gradient boosting is to build the new base classifiers to be maximally correlated with the negative gradient of the loss function, across the whole tree ensemble (Friedman 2001).

overall model improvement. The expression $\beta h(x_i; \boldsymbol{a})$ represents the improvement to the current model (a weak classifier which is usually a classification tree), which is to be optimized in a stage-wise fashion. The first line initializes to the optimal constant model which is a single terminal node tree. Each new tree is fitted to the generalized residuals of the model, which is the gradient of the loss function.

Specialized software such as TreeNet® includes several features designed to achieve superior accuracy, relative to other boosting techniques such as AdaBoost and more general gradient boosting models (discussed further in Section 6).

Interpretability of the gradient boosting model A frequent criticism of all advanced statistical learning models is their lack of interpretability. Jones et al. (2015) describe different classifiers in terms of an interpretability vs. flexibility trade-off continuum. Sophisticated models such as gradient boosting tend to be very accurate predictors (because they are highly flexible) but at the expense of some interpretability. While the gradient boosting model can be very complex, involving many potential nonlinear relationships and interactions among predictor variables, there are now standard outputs available, which allow the role and influence of different predictors to be more readily interpreted. These outputs include (1) relative variable importances (RVIs) and (2) partial dependencies (marginal effects).

Relative variable importances (RVIs) In statistical learning applications, the predictor variables usually have different impacts on the outcome domain. It is therefore very useful to know the RVIs that show the relative contribution of each input variable on overall model performance. Since these measures are relative, it is customary to assign the largest or most important variable a value of 100 and then scale the all other predictors accordingly. More formally, Breiman et al. (1984) propose the measurement of relative variable importance for a single tree (see Hastie et al. 2009) as follows:

$$\mathcal{I}_\ell^2(T) = \sum_{t=1}^{J-1} \hat{i}_t^2 I\left(v(t) = \ell \right)$$

where $\mathcal{I}_\ell^2$ is the measure of the relevance for each predictor variable X_ℓ. The sum is over J-1 internal nodes of the tree. At each node t, one of the explanatory variables $X_{v(t)}$ is used to partition the region associated with that node into two subregions, and within each subregion, a separate constant is fit to the response values. The particular variable chosen is the one that gives maximal estimated improvement $\hat{i}_t^2$ in squared error risk over that of a constant fit over the entire region.[16] As noted by Hastie et al. (2009), the importance measure is easily generalized to additive tree expansions and is simply averaged over the trees[17]:

$$\mathcal{I}_\ell^2 = \frac{1}{M} \sum_{m=1}^{M} \mathcal{I}_\ell^2 I(T_m).$$

[16] The squared relative importance of variable X_ℓ is the sum of such squared improvements over all internal nodes for which it was chosen as the splitting variable (Hastie et al. 2009).

[17] As shown by Hastie et al. (2009, p.368), it is straightforward to adapt these expressions to classification problems.

Due to the effect of averaging, this measure turns out to be more reliable for a single tree. Because of the effects of shrinkage, the masking of important variables by other highly correlated predictors presents no serious challenge for the model (Hastie et al. 2009).

Partial dependence plots Once the RVIs are determined, it is also useful to understand the nature of the dependence of the approximation $f(X)$ on their joint values. Partial dependence plots are graphical visualizations of the marginal effect of a given variable on the outcome domain. To obtain the univariate partial dependence plot with respect to a variable X_I, all other remaining predictors are fixed at the joint set of values sampled from the dataset, while X_I varies over its range to produce one instance of the dependence curve in terms of model response. This procedure is repeated many times for all available estimation sample observations. Each observation provides another joint set of fixed values and the resulting dependence curve with respect to X_I.[18]

Advantages of GB for bankruptcy research The gradient boosting model has several properties that are particularly appealing for bankruptcy research. For instance, it is demonstrably more suitable for high dimensional, nonlinear analysis, which arguably better reflects the real world context of corporate bankruptcy. The model also appears to be more suitable given the characteristics of bankruptcy datasets. In contrast to conventional models, such as logit, the gradient boosting model is not sensitive to outliers. (The algorithm simply isolates the outliers in a separate node, which does not affect the performance of the final tree). Nor is gradient boosting sensitive to monotonic transformations. For instance, transforming one or several variables to its logarithm or square root will not change the structure of the tree itself. (It only affects the splitting values of the transformed variable.) For conventional models, missing values can also present serious data problems. Typically, missing values are removed case-wise (sacrificing sample), or they must be imputed. However, a gradient boosting model handles missing values by building surrogates from other correlated variables available in the dataset, thus preserving the sample. In this sense, the model exploits multicollinearity to enhance performance, whereas it tends to diminish the performance of conventional models. Gradient boosting can also better handle irrelevant inputs (Friedman 2001).

Furthermore, while multicollinearity among predictors is a serious issue for parametric models, such as logit or MDA (diminishing the interpretability and stability of parameter estimates), it is simply an issue of redundant information in gradient boosting models.[19] Another benefit of the high dimensional gradient boosting analysis is that this approach can eliminate data snooping bias (White 2000). This occurs when a model is continually re-estimated with different iterations of the explanatory variables to arrive at the

[18] At the end of the process, the entire family of dependence curves is averaged and centered to produce the final dependence plot on X_I. (More details are provided in "Introduction to TreeNet" Salford Systems, San Diego, 2015.)

[19] If there are two perfectly correlated predictors in the dataset, the RVI would be 100% for the first one found by gradient boosting and zero for the other (i.e., the second variable is redundant). If two variables are strongly but not perfectly correlated, the gradient boosting algorithm simply partitions the correlated variables, looking for the greatest space between them that can enhance predictive power. A predictive variable will gain a higher RVI, while the other high correlated predictor gets a low RVI, because of redundancy in signal. (It may still have some signal because the correlation is not perfect.)

best model. This approach can result in a good model produced by chance rather than from something innately meaningful in the data. The high dimensional gradient boosting model uses the full set of input variables and automatically detects all important interaction effects. This approach helps remove bias associated with data snooping.

4 Sample data

4.1 Sample characteristics

I use a large sample of U.S. public company bankruptcies extracted from Standard and Poor's Capital IQ service, one of the largest financial and risk database providers in the world. Corresponding annual financial, market, and other input variables are extracted from a customized software application developed with the researcher by Capital IQ technical staff. The sample includes 1115 corporate bankruptcies on the Capital IQ database. The bankruptcy sample spans 27 years (1987 to 2013). Table 1 displays the distribution of bankruptcies over this period. It can be seen from Table 1 that 88.6% of the sample relates to corporate bankruptcies which occurred on or after the year 2000. The greatest concentration of corporate bankruptcies (around 32%) was observed during the global financial crisis and its immediate aftermath (2007–2010).

Around 22.27% of sampled bankruptcies occurred during the height of the dot-com bubble crash (2000–2002), while the remainder are fairly evenly spread over the remaining years. Table 2 shows the bankruptcy vs. active firms across each industry group and reveals a higher frequency of bankruptcies occurring in certain industrial sectors.

Table 1 Distribution of Corporate Bankruptcies over the Sample Period (1987–2013)

Year of Bankruptcy	Number of Bankruptcies	% of Bankruptcies
2013	38	3.86
2012	57	5.79
2011	54	5.49
2010	58	5.90
2009	121	12.30
2008	89	9.05
2007	47	4.78
2006	28	2.84
2005	39	3.96
2004	41	4.17
2003	80	8.13
2002	76	7.73
2001	97	9.86
2000	46	4.67
<2000	112	11.39
Total	983*	100

*Note that 132 firms did not have bankruptcy dates available on Capital IQ

Table 2 Distribution of Bankrupt vs Active Firms across Industry Groups (All Data)

Industry (GICS)	Bankruptcy	Active Firms
Healthcare	544	4,222
Energy	252	2,720
Materials	236	2,200
Industrials	620	4,012
Telecommunications	196	460
Utilities	76	596
Information Technology	708	5,657
Financials	516	5,590
Consumer Staples	212	1,592
Consumer Discretionary	1,100	4,700
Total	4,460	31,749

Table 2 shows that around 24.66% of sampled bankrupt firms are from the consumer discretionary sector, which includes industries such as retail, media, and tourism. A further 15.87% of failures are observable in the information technology sector, which includes industries such as internet companies, software, and technology hardware firms. Around 13.9% of bankruptcies are from the industrials sector, which includes industries such as capital goods, commercial and professional services, and transportation. A further 12.19% of bankrupt firms are from the healthcare sector, represented mainly by pharmaceutical and biotechnology firms. Around 11.56% of bankrupt firms are from the financial sector, which includes banks, insurance companies, diversified financials, and real estate firms. A further 10.94% of bankruptcies are found in the energy and materials sectors, which represent mainly resource and exploration firms. Around 4.75% of bankruptcies are from the consumer staples sector with 4.39% of bankrupt firms coming from the telecommunications sector. The mean market capitalization of bankrupt firms is approximately $184 million versus $2.513 billion for the nonbankrupt sample. This is expected, as smaller more recently established companies tend to have higher failure rates in practice (Altman 2002).

A control group of healthy firms is also collected. The sample of nonfailed firms represents all active U.S. public companies included in the Capital IQ database. I collected up to three years of financial and market data on all bankrupt firms prior to the year of bankruptcy. For bankrupt firms, I use the most recent financial statements available on Capital IQ database immediately prior to the announcement of bankruptcy. Capital IQ provides details on both the bankruptcy filing date and filing date for financial statements. I extracted the same data for the nonfailed control group. Only publicly listed firms are included in the training and test samples. In a small number of cases, firms are deleted from both samples because no financial statement records were available. Following the approach of Ohlson (1980) and Jones and Hensher (2004), no firm is removed from the sample because it is newly or recently listed, and some firms in the sample had only one or two years of financial and market data. The final sample includes 4460 firm year bankruptcies and 31,749 firm year observations for the healthy group.

4.2 Binary response variable

The independent variables of the study are defined in Appendix Table 11. The bankruptcy event is classified as a binary outcome dependent variable, where a firm that is observed to enter bankruptcy is coded 0, and an active or nonfailed firm is coded 1. It is an accepted practice in this literature (at least as it applies to discrete choice models) to code previous years of a bankrupt firm as bankrupt. For instance, if a company failed in 2013 and three preceding years of financial and market data exists on that firm, all years are coded as 0 or bankrupt. This is necessary as researchers often want to examine the performance of a bankruptcy model in any number of years prior to failure (for instance, three or even five years from failure).[20] Corporate bankruptcy is defined in conventional terms as a company that enters Chapter 11 of the United States Bankruptcy Code, which permits an insolvent company to reorganize its affairs under bankruptcy protection but is also a vehicle for liquidation. Most insolvent companies in the United States seeking bankruptcy protection file for Chapter 11 protection (Altman 2002).

4.3 Evaluating predictive performance

Following conventions in the statistical learning literature, the gradient boosting algorithm randomly allocates 70% of the total observations to the training data and 30% of observations to the test sample (Hastie et al. 2009). I test out-of-sample predictive accuracy using several classification statistics. However, the ROC curve is the most ubiquitous measure in the literature for comparing the classification performance of different classifiers (e.g., Swets et al. 2000; Jones et al. 2015). The ROC curve plots the true positive rate (sensitivity) relative to the false positive rate (1 − specificity) with respect to the discretionary cut-off score. (For the binary classifiers, this score is the predicted probability of bankruptcy.)[21] Convention suggests that AUC scores greater than 0.9 demark a very strong classifier, exhibiting an excellent balance between sensitivity and specificity across different probability thresholds, whereas AUCs between 0.8 and 0.9 indicate a very good or strong classifier.

5 Empirical results

The commercial version of gradient boosting model used for this study is interpreted through several outputs, including average log-likelihood, ROC curves, classification accuracy, relative variable importances (RVIs), marginal effects (partial dependence

[20] This approach does involve look-ahead bias. The impact of look-ahead bias on empirical results is examined further in Section 6. Studies by Shumway (2001), Hillegeist et al., (2004) and Beaver et al., (2005) use hazard/duration models to predict corporate bankruptcy. For this type of modeling, it is clearly more appropriate to code a firm as bankrupt only in the year of bankruptcy as hazard models are designed to predict time-to-event using a survival function.

[21] A random guess describes a horizontal curve through the unit interval and has an AUC of exactly 0.5. As a minimum, classifiers are expected to perform >.5 (ie., better than random guessing), whereas an AUC score of 1 represents perfect classification accuracy (zero Type I and Type II errors).

plots), and interaction effects. As a first step in testing the study's hypotheses, a full gradient boosting model is estimated using all input variables described in Appendix Table 11. This model is reported Table 3.

The average log-likelihood, misclassification rates, and ROC curves for both learn (training) and test samples are shown in Figs. 1 and 2 respectively for the full gradient boosting model reported in Table 3. Figure 1 displays the classification accuracy of the full gradient boosting model, while Fig. 2 displays the area under the ROC curve (AUC).

The classification accuracy displayed in Fig. 1 is based on a simple tally of how frequently the model classifies an observation correctly or incorrectly over the tree depth. Figure 1 shows that the gradient boosting model has very impressive out-of-sample predictive accuracy. The model has only misclassified 3.08% of observations for the test sample. The misclassification error is optimized at 680 trees, and the Type I and Type II error rates appear to be significantly lower than results documented in previous bankruptcy studies. (See Jones and Hensher 2008 and Appendix Table 12 studies.) The test error curve in Fig. 1 lies slightly above the learn error curve, and they are in close agreement, which indicates little or no over-fitting on the test sample (Hastie et al. 2009). This result corroborates the findings from other fields that the gradient boosting model is highly resistant to model over-fitting, even when models are estimated on very large numbers of trees (Friedman 2001). Figure 2 reports the AUC, which is the most commonly used predictive performance metric in the statistical learning literature. The AUC is a measure of overall model performance tied closely to the ability of the gradient boosting model to correctly rank records from most likely to least likely to be a 1 (nonfailure) or 0 (bankruptcy).

Figure 2 indicates that the high dimensional gradient boosting model (which includes all explanatory variables in Appendix Table 11) scores a near perfect out-of-sample of AUC of 0.996, which is optimized at 680 trees. This result also corroborates the findings in other literatures that the gradient boosting model is a highly accurate classifier, although the boosting model studies in Appendix Table 12 are quite variable in terms of predictive performance (discussed further in Section 6). Table 4 summarizes overall model performance of the Table 3 model. Table 4 includes the average log-likelihood, areas under the ROC curve (AUC), lift, and baseline classification accuracy.

Panel A of Table 4 reports model performance based on all sampled data (three years of pooled observations). Panels B and C of Table 4 summarize predictive performance of the Table 3 model estimated on *t*-3 data (three years prior to the bankruptcy event) and *t*-1 data (one year prior to the bankruptcy event). Overall, the predictive power of the gradient boosting model is highest when estimated on the full sample (Panel A). The overall classification accuracy (baseline threshold) for the full sample model is 95.78% (and 97.17% for the learn sample). The overall model is 97.83% accurate in predicting failures (Type I error) and 94.98% accurate in predicting nonfailure (Type II error) using baseline threshold as the cut-off score. Table 4 shows that the full gradient boosting model in Table 3 estimated one year from failure misclassifies 6.1% of observations. (The model is around 98% accurate on Type I errors and 90.86% accurate on Type II errors.) Three years from failure, the model misclassifies 8.74% of observations. (The model is around 95% accurate on Type I and 89.2% on Type II errors.) However, the AUCs remain very robust over all time frames.

Table 3 Relative Variable Importances (RVIs) of Full TreeNet (GB) Model - All Data

Explanatory Variable	Relative Variable Importance (RVI) Score	Dimension
Percentage of Stock Owned by Top 5 Stockholders	100.00	Ownership Structure/Concentration
Percentage of Stock Owned by Insiders	61.60	Ownership Structure/Concentration
Market Capitalization	57.60	Market Price
Stock Price Volatility (Two-Year Beta)	36.26	Market Price
Total CEO Compensation	33.59	Executive Compensation
Total Revenues	33.02	Accounting
Percentage of Stock Owned by Institutions	29.43	Ownership Structure/Concentration
Total Assets	27.79	Accounting
Earnings per Share	26.01	Accounting
Cash Flow per Share	24.32	Accounting
Auditor Type	24.18	Other
Net Operating Cash Flows	22.56	Accounting
Excess Returns (Six Months)	22.00	Market Price
Total Debt	20.23	Accounting
Excess Returns (12 Months)	19.98	Market Price
Interest Cover	17.90	Accounting
Number of Business Segments	17.45	Other
Stock Price Volatility (One-Year Beta)	17.38	Market Price
Percentage of Stock Owned by Top 10 Stockholders	16.30	Ownership Structure/Concentration
Capital Expenditure to Total Assets	15.88	Accounting
Age of Firm	15.08	Other
Net Investing Cash Flows	12.88	Accounting
Credit Ratings Change	12.28	External Ratings
Net Financing Cash Flows	11.95	Accounting
Annual Growth in Net Income	11.68	Accounting
Total Short Interest	10.84	Market Price
Total Liabilities to Total Equity	10.38	Accounting
Current Ratio	10.12	Accounting
Stock-Based Compensation	9.81	Executive Compensation
Total Cash and Short-Term Investments/Total Assets	9.34	Accounting
Market Capitalization to Total Debt	9.30	Market Price
Average Inventory Turnover	9.29	Accounting
EBIT Margin	9.20	Accounting
Working Capital to Total Assets	9.19	Accounting
Total Tangible Book Value	9.10	Accounting
Short-Term Debt to Total Liabilities	8.88	Accounting
Change in Working Capital to Total Assets	8.23	Accounting
Sales to Total Assets	7.90	Accounting
Cash Flow Returns	7.76	Accounting

Table 3 (continued)

Explanatory Variable	Relative Variable Importance (RVI) Score	Dimension
Total Debt to Total Assets	7.44	Accounting
Annual Growth in Debt	7.19	Accounting
Annual Growth in Operating Cash Flow	7.00	Accounting
Average Debt Collection Period	6.98	Accounting
Gross Profit Margin	6.64	Accounting
Annual Growth in Capital Expenditure	6.62	Accounting
Cash Flow to Debt	6.46	Accounting
EBIT to Total Assets	6.43	Accounting
Free Cash Flow per Share	6.40	Accounting
Number of Analysts	6.33	External Ratings
Total Debt to Total Equity	6.23	Accounting
Consensus Analyst Forecasts of Target Price	6.21	External Ratings
Total Cash and Cash Equivalents	6.18	Accounting
Annual Growth in Revenue	5.98	Accounting
Identifiable Intangibles to Total Assets	5.93	Accounting
Dividends per Share	5.81	Accounting
CPI Index	5.63	Macroeconomic
Total Bank Debt	5.60	Accounting
Provision for Credit Losses to Total Liabilities	5.56	Accounting
Annual Growth in Leveraged Free Cash Flow	5.56	Accounting
Unemployment Rate	5.31	Macroeconomic
GICS Sector - Industrials	4.77	Other
Analyst Consensus Stock Recommendations	4.76	External Ratings
Total Revolving Credit to Total Debt	4.04	Accounting
Total Revolving Credit Facilities	3.82	Other
TED Spread	3.50	Macroeconomic
Goodwill to Total Assets	3.01	Accounting
GICS Sector – Consumer Discretionary	2.75	Other
Bad Debt Provisions to Total Liabilities	2.73	Accounting
GICS Sector - Materials	2.33	Other
GICS Sector - Financials	2.24	Other
Provision for Credit Losses	2.20	Accounting
GICS Sector – Information Technology	2.02	Other
Moody's Seasoned BBB Bond Yield	1.97	Macroeconomic
Total Asset Write-Downs	1.54	Accounting
Consensus EPS Forecasts (1 YR)	1.41	External Ratings
GICS Sector - Energy	1.32	Other
Provision for Bad Debt Expense	1.15	Accounting
Write-Down to Total Assets	1.08	Accounting
GICS Sector – Consumable Staples	0.93	Other

Table 3 (continued)

Explanatory Variable	Relative Variable Importance (RVI) Score	Dimension
Michigan Sentiment Index	0.89	Macroeconomic
GICS Sector - Telecommunications	0.63	Other
GICS Sector - Healthcare	0.46	Other

Table 3 shows the relative variable importances (RVIs) of the full TreeNet (gradient boosting) model estimated on all explanatory variables displayed in Appendix Table 11. Only nonzero RVIs are displayed. RVIs include the effects of all important two-way interaction effects having predictive value in the model. The RVI is based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and then averaged over all trees. Hence RVIs are calculated relative to all other input variables in the model. The RVIs in Table 3 are ranked according to their contribution to the overall predictive success of the model. Since these measures are relative, it is customary to assign the largest or most important variable a value of 100 and then scale the all other predictors accordingly. Table 3 shows that 82 of the 91 input variables in Appendix Table 11 have nonzero importance. The RVIs are also dispersed across a number of bankruptcy dimensions, such as market-price variables, accounting-based indicators, executive compensation measures, ownership concentration/structure variables, and other variables. However, Table 3 indicates that ownership concentration/structure variables have the strongest overall impact on out-of-sample predictive performance.

It can be seen from Table 3 that 82 of the 91 predictor variables have nonzero RVIs. This means that all 82 variables contributed to out-of-sample predictive success in some way, although the strength of different predictors varies significantly.[22] The RVIs reported in Table 3 are expressed on a scale between 0 and 100, where the most important variable always gets a score of 100 and all other variables are rescaled to reflect their importance relative to the most important variable. RVIs are calculated relative to the predictive power of all other variables in the model. Because gradient boosting uses decision trees as the base learning, the RVI is based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman and Meulman 2003; Hastie et al. 2009).

High dimensional models will typically involve many more variables, thus increasing the number of variables with nonzero importance. Some of the very small RVI values could simply reflect random noise patterns in the data. Hence it is advisable to re-estimate the gradient boosting model excluding variables with close to zero RVI scores. I re-ran the model on all variables with RVIs of at least 5% to eliminate potential noise effects, and the ranking of RVIs did not substantially change. Overall, the results in Table 3 support H_1 regarding the high dimensional nature of corporate bankruptcy. As can be seen by Table 3, a diverse range of bankruptcy predictors dominate the analysis, reflecting several different dimensions of corporate failure. The predictor variables which feature most strongly in the analysis (RVI > 10) include (1) ownership concentration/structure variables, particularly stockholder concentration and institutional ownership, and percentage owned by insiders; (2) a range of accounting variables and ratios including total revenues, total assets, earnings-per-share (EPS), cash flow per share, net operating cash flows, total debt, interest cover, capital expenditure to total

[22] Note that the RVIs reported in Table 3 incorporate all important interaction effects.

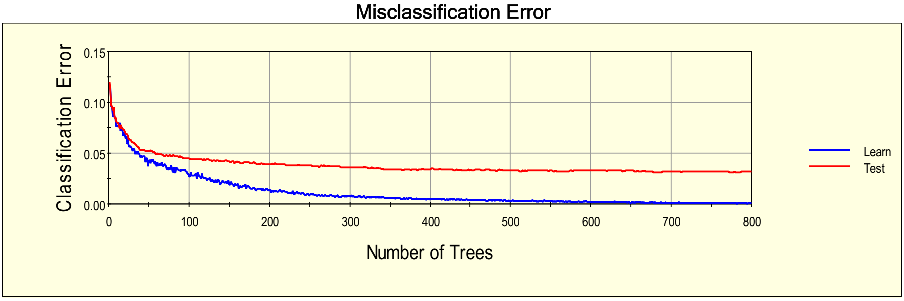

Fig. 1 Misclassification Error for Learn and Test Samples for the TreeNet (Gradient Boosting) Model Reported in Table 3. Figure 1 displays the classification accuracy of the model reported in Table 3. This is based on a simple tally of how frequently the model tags an observation correctly or incorrectly. The misclassification error is optimised at 680 trees, and the Type I and Type II error rates appear to be significantly lower than results documented in previous bankruptcy studies. The test error curve in Fig. 1 lies slightly above the learn error curve, and they are in close agreement, which indicates little or no over-fitting on the test sample. Figure 1 indicates that the model has only misclassified 3.08% of observations for the test sample and 0.01% of observations for the learn sample

assets, net investing cash flow, net financing cash flows, annual growth in net income, total liabilities to total equity, and the current ratio; (3) market-price variables, including market capitalization, excess returns, total short interest, and stock price volatility (measured by beta); (4) total CEO compensation; (5) credit ratings change; and (6) other variables such as auditor type, age of firm, and number of business segments.

Table 3 indicates that the strongest predictor overall is percentage of stock owned by the top 5 stockholders (RVI = 100). The second strongest variable is percentage of stock owned by insiders with an RVI of 61.60, followed by the market capitalization variable with an RVI of 57.60. The next strongest variable is stock price volatility (two-

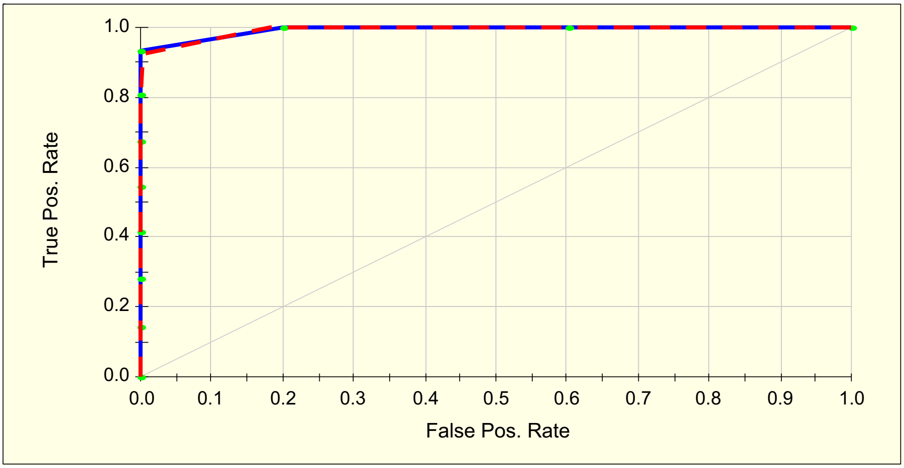

Fig. 2 Out-of-sample AUC for the TreeNet Model Reported in Table 3. Figure 2 reports the out-of-sample AUC performance for the model reported in Table 3. Following conventions in the statistical learning literature, 70% of the sample is randomly allocated to the learn sample, and 30% randomly allocated to the test sample. AUC is the Area under the the receiver operating characteristic (ROC) curve. For a binary classifier, the ROC curve plots the true positive rate (sensitivity), relative to the false positive rate (1 − specificity), as its discrimination threshold or cutoff score is varied. Figure 2 shows the model reported in Table 3 has an out-of-sample AUC of 0.996

Table 4 Summary of Predictive Performance for TreeNet (Gradient Boosting) Model Reported in Table 3

	Learn Sample	Test Sample
Panel A (All Data)		
Average LogLikelihood (Negative)	0.01664	0.04190
ROC (Area Under Curve)	0.99997	0.99686
Lift	1.14245	1.13468
Classification Accuracy (Baseline Threshold)	0.97433	0.95785
Panel B (One Year Prior to Bankruptcy)		
Average LogLikelihood (Negative)	0.02411	0.06297
ROC (Area Under Curve)	.99811	0.99125
Lift	1.14224	1.13629
Classification Accuracy (Baseline Threshold)	0.96525	0.93873
Panel C (Three Years Prior to Bankruptcy)		
Average LogLikelihood (Negative)	0.04880	0.08593
ROC (Area Under Curve)	0.99975	0.98284
Lift	1.14222	1.13617
Classification Accuracy (Baseline Threshold)	0.92721	0.91254

Table 4 summarizes the predictive performance of the full TreeNet (gradient boosting) model reported in Table 3. Panel A displays the results based on all sample data. Panel C summarizes predictive performance three years from the bankruptcy event ($t = -3$), while Panel B displays predictive performance one year prior to bankruptcy ($t = -1$). The ROC curve plots the true positive rate (sensitivity), relative to the false positive rate (1 − specificity), with respect to the discretionary cutoff score. (For the binary classifiers, the score is the predicted probability of bankruptcy.) Average log-likelihood is similar to ROC but stresses a probability interpretation of model predictions. Classification accuracy (baseline threshold) is based on thresholds that reflect the actual distribution of bankruptcy vs. nonfailure firms in the sample. The results show that the Panel A model has the strongest out-of-sample AUC and the lowest misclassification rate overall. However, the full TreeNet model also appears to perform strongly one and three years prior to the bankruptcy event. (Classification accuracy is 93.87% one year from failure and 91.97% three years from failure.) However, there is little evidence of deterioration in AUC performance one and three years from failure.

year beta) with an RVI of 36.26, followed by total CEO compensation with an RVI of 33.59. The excess return variables also show a quite strong impact in the Table 3 model but they do not dominate the analysis. For instance, the six-month excess returns variable has an RVI of 22, while the 12-month excess returns variable has an RVI of 19.98. Other high impacting variables include total revenue (RVI = 33.02), percentage of stock owned by institutions (RVI = 29.43), total assets (RVI = 27.79), earnings-per-share (RVI = 26.01), cash flow per share (RVI = 24.32), auditor type (RVI = 24.18), net operating cash flow (RVI = 22.56), total debt (RVI = 20.23), interest cover (RVI = 17.90), number of business segments (RVI = 17.45), one-year beta (RVI = 17.38), percentage of stock owned by the top 10 stockholders (RVI = 16.30), capital expenditure to total assets (RVI = 15.88), age of firm (RVI = 15.08), net cash flow from investing activities (RVI = 12.88), credit ratings change (RVI = 12.28), net cash flow from financing activities (RVI = 11.95), annual growth in net income (RVI = 11.68), total short interest (RVI = 10.84), total liabilities-to-total equity (RVI = 10.38), and the current ratio (RVI = 10.12).

To get a better sense of what predictors are driving the results, consider the average RVIs across the different predictor categories above. The ownership/concentration and CEO compensation variables have an average RVI of 48.18, the strongest predictors overall. The next best predictors are unscaled market and accounting variables, which proxy for size (such as market capitalization, total revenue, and total assets) and which have an average RVI of 32.24. Next are the market-price indicators, which have an average RVI of 21.92. Financial ratios performed comparably well to market-price variables with an average RVI of 17.43. The RVI difference between financial and market-price variables is only 4.49 overall, which would not translate into significantly better predictive performance. Other non-accounting variables (such as auditor type, number of business segments, firm age, and credit rating changes) show comparable predictive strength with an average RVI of 17.24. Variables with the lowest RVIs in the model include industry variables (average RVI of 1.93), macroeconomic variables such as inflation and employment rates (average RVI of 3.46), and analyst recommendations/forecasts (average RVI of 4.67).

An issue to consider with the Table 3 results is whether having such a large representation of accounting-based variables (and the degree to which they are correlated) can affect the gradient boosting results, particularly the RVIs of strong non-accounting predictors such as stockholder ownership/concentration. However, the gradient boosting model will ignore or assign very small RVIs to variables having little or no predictive power, no matter how many inputs are actually used in the model (or the extent that they are correlated). Only the accounting variables with strong predictive power will receive a high RVI ranking, while others will receive low or nonzero RVIs. In this sense, gradient boosting works very differently from traditional regression procedures. In a logit model, for instance, irrelevant inputs enter the maximum likelihood solution, which affects the overall stability and performance of the model. As stated previously, multicollinearity and irrelevant inputs is just a question of redundant information for a gradient boosting model, which has little impact on overall performance.[23]

Comparisons with logit To illustrate some of the differences between gradient boosting and a traditional logit model, I compare the RVIs in Table 3 results with a logit model. Table 5 below displays a logit model estimated on Table 3 variables, which have the highest RVIs (i.e., RVI scores of at least 15).

Most of the high RVI variables reported in Table 3 are highly correlated with each other, particularly variables such as market capitalization, total revenues, total assets, total debt, age of the firm, percentage of stock owned by the top 5 and 10 stockholders, institutional ownership, insider ownership, auditor type, and total CEO compensation (Most Pearson correlations range from 20% to 95%.) As can be seen from Table 5, several of the logit parameters are not significant and most have a counterintuitive signs. For instance, most of the accounting and market-price variables (such as EPS, total revenues, interest cover, market capitalization, and excess returns) have negative parameters, suggesting they are decreasing of nonfailure (or increasing of bankruptcy).

[23] I re-estimated a gradient boosting model using all the variables in Table 3 but removing all financial variables with RVIs less than 10 (around 30+ variables were dropped as a result). However, I found that the RVIs of the remaining variables did not change significantly.

Table 5 Logit Parameter Estimates Based on Highest RVI Variables Reported in Table 3

Variable	Coefficients	S.E.	t-ratio	*p*-value	RVIs from Table 3	Pearson's Correlation with Bankruptcy Outcome
Constant	−2.166	2.487	−0.870	0.383	n/a	n/a
Percentage of Stock Owned by Top 5 Stockholders	−0.114	0.0112	−10.115**	2.00E-15	100.00	.303**
Percentage of Stock Owned by Insiders	1.907	0.519	3.671**	0.00024	61.60	.176**
Market Capitalization	−0.00092	.000009	−9.613**	2.00E-15	57.60	.05**
Stock Price Volatility (Two-Year Beta)	−0.00136	0.000824	−1.647	0.099555	36.26	.016**
Total CEO Compensation	-0.000003	1.82E-07	-1.649	0.099148	33.59	.093**
Total Revenues	−0.00013	0.0000173	−7.288**	3.12E-13	33.02	.024**
Percentage of Stock Owned by Institutions	0.455	0.204	2.224*	0.026	29.43	.185**
Total Assets	0.00004	.000005	7.0295**	2.07E-12	27.79	.014**
Earnings per Share	−0.00057	0.000231	−2.474**	0.013328	26.01	.024**
Cash Flow per Share	−0.00011	0.000185	−0.568	0.56997	24.32	.014**
Auditor Type	−0.00029	0.0000449	6.3756**	1.82E-10	24.18	−0.0039
Net Operating Cash Flows	.0000025	0.0000446	0.0573	0.95424	22.56	.037**
Excess Returns (Six Months)	−0.27633	0.0817	−3.3801**	0.000724	22.00	−0.007
Total Debt	−.000001	0.0000073	−0.140	0.88845	20.23	0.005
Excess Returns (12 Months)	−0.14075	0.050128	−2.8077**	0.00498	19.98	−0.007
Interest Cover	−.000008	2.01E-05	-0.415	0.67789	17.90	−0.004
Number of Business Segments	1.66242	0.23126	7.188**	6.53E-13	17.45	.089**
Stock Price Volatility (One-Year Beta)	0.0006	0.0020117	0.29948	0.76458	17.38	.016**
Percentage of Stock Owned by Top 10 Stockholders	0.047	0.0082886	5.6707**	1.42E-08	16.30	.314**
Capital Expenditure to Total Assets	0.0087	0.00242	3.58**	0.00034	15.88	0.002
Age of Firm	0.00005	0.00125	0.0405	0.967	15.08	−.23**

*Sig < .05; **Sig < .01 level

Table 5 shows a logit model estimated on the strongest RVI variables reported in Table 3 (RVIs > = 15). The two columns on the far right show the RVIs of each variable as reported in Table 3 and Pearson correlation coefficients of each explanatory variable with the bankruptcy outcome dependent variable. Many of the parameter estimates in Table 5 are not significant, and several of the parameter estimates have a counterintuitive signs. For instance, most of the accounting and market-price variables (such as EPS, total revenues, interest cover, market capitalization, and excess returns) have negative parameters, suggesting they are decreasing of nonfailure (or increasing of bankruptcy). These effects appear to arise from multicollinearity and heteroscedasdicity issues in the dataset, which tend to destabilize parametric models such as logit. This instability has translated into high misclassification error (around 29%) on both learn and test samples. By contrast, the TreeNet model reported in Table 3 appears to be quite effective in extracting signal in the presence of strong correlation and heteroscedasticity in the dataset, as shown by the more even distribution of RVIs across the input variables in Table 3 and high classification accuracy of the overall model.

Another issue in Table 5 is that the logit model has larger parameters for only a small number of variables (such as percentage of stock owned by insiders and institutions and the excess returns variables), while most other variables have very small parameters, suggesting they contribute very little to overall model performance. The logit model in Table 5 has translated into a quite high misclassification error rate (around 29%) on the test sample.[24] By contrast, the gradient boosting model in Table 3 appears to be highly effective in extracting signal from variables exhibiting strong correlation and heteroscedasticity, as shown by the more even distribution or spread of RVIs across the input variables, and the high classification accuracy of the overall model. Gradient boosting also produces marginal effects that broadly make sense in terms of the expected relationship between input variables and the bankruptcy outcome (discussed later in this section).

Multi-dimensionality in corporate bankruptcy Table 3 also provides some evidence of the multi-dimensional nature of corporate bankruptcy, as the RVIs are mostly nonzero and appear reasonably well dispersed across a range of different predictors. However, a more direct test of multi-dimensionality is to compare the Table 3 model with gradient boosting models estimated on each bankruptcy predictor dimension, which includes market-price indicators, ownership structure/concentration variables, external ratings indicators (i.e., analyst forecasts/recommendations and credit ratings changes), macroeconomic factors, executive compensation indicators, and accounting-based variables. If corporate bankruptcy evidences a multi-dimensional aspect, it is expected that each separate dimension will be predictive in its own right. However, it is expected that no one dimension will outperform the classification accuracy of the full gradient boosting model reported in Table 3.

Table 6 summarizes the main findings. It can be seen from Table 6 that the AUCs for each gradient boosting model have quite strong predictive power in their own right. For example, Panel A shows that a gradient boosting model estimated only on market = price variables (i.e., all market-price variables reported in Table 3 above) has an impressive out-of-sample AUC of 0.919. This suggests that market-price variables have strong predictive power in isolation from all other explanatory variables reported in Table 3. However, the out-of-sample AUC performance of a gradient boosting model estimated only on the accounting-based variables of Table 3 is slightly higher at 0.928 (see Panel F of Table 6). Similarly, a gradient boosting model estimated exclusively on ownership concentration/structure variables of Table 3 again shows a strong out-of-sample AUC of 0.926 (see Panel B). A gradient boosting model estimated exclusively on executive compensation variables (Panel E) produces a slightly lower out-of-sample AUC of 0.876. However, the predictive performance of other corporate bankruptcy dimensions was not quite as strong. For instance, Table 6 Panel C indicates that a gradient boosting model estimated only on the external ratings variables of Table 3 has a lower but nevertheless good out-of-sample AUC of 0.802. Finally, Table 6 indicates that the worst performing dimension is macroeconomic variables with an out-of-sample AUC of only 0.563 (see Panel D). Taken together, these results suggest that corporate bankruptcy is not only high dimensional but also multi-dimensional, insofar as each

[24] Based on the same 70/30 random allocation to learn and test samples used for the gradient boosting model in Table 3.

Table 6 Summary of Predictive Performance for TreeNet (Gradient Boosting) Model and Logit Models Estimated on Each Bankruptcy Dimension

	TreeNet Learn Sample	TreeNet Test Sample	Logit Learn Sample	Logit Test Sample
Panel A: Performance Summary of TreeNet vs. Logit on Market Price Variables				
Average Log -Likelihood (Negative)	0.261	0.288	0.340	0.335
ROC (Area Under Curve)	0.951	0.919	0.679	0.662
Lift	1.142	1.135	1.121	1.120
Classification Accuracy (Baseline Threshold)	0.883	0.863	0.515	0.514
Panel B: Performance Summary of TreeNet vs. Logit on Ownership Concentration/Structure Variables				
Average Log -Likelihood (Negative)	0.316	0.331	0.304	0.295
ROC (Area Under Curve)	0.934	0.926	0.799	0.801
Lift	1.142	1.134	1.126	1.124
Classification Accuracy (Baseline Threshold)	0.801	0.792	0.728	0.722
Panel C: Performance Summary of TreeNet vs. Logit on External Ratings Variables				
Average Log- Likelihood (Negative)	0.541	0.543	0.303	0.292
ROC (Area Under Curve)	0.807	0.802	0.695	0.693
Lift	1.309	1.295	1.071	1.072
Classification Accuracy (Baseline Threshold)	0.715	0.715	0.625	0.626
Panel D: Performance Summary of TreeNet vs. Logit on Macroeconomic Variables				
Average Log-Likelihood (Negative)	1.181	1.206	0.36767	0.35914
ROC (Area Under Curve)	0.568	0.563	0.600	0.591
Lift	1.137	1.100	1.048	1.042
Classification Accuracy (Baseline Threshold)	0.607	0.609	0.647	0.649
Panel E: Performance Summary of TreeNet vs. Logit on Executive Compensation Variables				
Average Log-Likelihood (Negative)	0.512	0.510	0.157	0.152
ROC (Area Under Curve)	0.875	0.876	0.678	0.691
Lift	1.138	1.131	1.037	1.036
Classification Accuracy (Baseline Threshold)	0.812	0.801	0.564	0.568
Panel F: Performance Summary of TreeNet vs. Logit on Accounting-Based Variables				
Average Log-Likelihood (Negative)	0.341	0.385	0.313	0.357
ROC (Area Under Curve)	0.962	0.928	0.814	0.787
Lift	1.142	1.134	1.139	1.145
Classification Accuracy (Baseline threshold)	0.846	0.813	0.690	0.730

Table 6 summarizes the predictive performance of each TreeNet (gradient boosting) and logit model estimated on different dimensions of corporate bankruptcy, including market price variables, ownership concentration/structure measures, external ratings indicators, macroeconomic factors, executive compensation, and accounting-based variables. The results show that the model has strong out-of-sample AUC performance across all dimensions of corporate bankruptcy except Panel D (macroeconomic variables). The best AUC performance is on accounting-based variables, followed by ownership concentration/structure, market-price variables, and executive compensation variables, respectively. The results also show that the model has out-of-sample AUC performance substantially higher than logit across all dimensions (except Panel D, where logit slightly outperforms it).

dimension is quite predictive in its own right. However, no one single bankruptcy dimension outperforms the classification accuracy of the full gradient boosting model reported in Table 3.

Explanatory power and marginal effects Having established that a high dimensional gradient boosting model has very strong out-of-sample predictive accuracy, it is also important to assess whether the explanatory variables make sense in terms of their role and influence on the bankruptcy outcome. A major benefit of conventional bankruptcy models (such as logit) is that they are highly interpretable. Highly flexible (nonlinear) models, such as neural networks and support vector machines, often predict well but are widely viewed as black boxes from an interpretative standpoint (Jones et al. 2015).

The gradient boosting model, including commercial versions such as TreeNet®, is a relatively recent statistical learning technology that provides outputs that allow the researcher to see into the black box, particularly through RVIs and partial dependence plots or marginal effects. The RVIs reported in Table 3 are valuable for displaying the strength or influence that a particular explanatory variable has on overall classification performance. However, the RVIs themselves provide no indication of the direction of its relationship with the failure outcome. (For example, RVIs do not show whether the debt-to-equity increases or reduces the probability of failure.) Marginal effects or partial dependence plots reveal both the direction and strength of the relationship between explanatory variables and the failure outcome. While parameter estimates from a conventional model (such as logit) are always linear with respect to the outcome dependent variable,[25] the gradient boosting marginal effects capture all nonlinear impacts, which can be more informative and descriptive of the behavior of bankruptcy predictors in their real world context. In fact, the marginal effects of most of the Table 3 variables have nonlinear relationships with the failure outcome. However, in a number of instances, the broad directions of the underlying relationships are interpretable and seem to make sense. This can be demonstrated with a few examples from the Table 3 variables which are displayed in Figs. 3, 4, 5, 6, 7, 8, 9, and 10 below. To make it easier to interpret the partial dependency plots, the graphs include a first-order single knot spline which smooths out the relationship and reveals the overall direction of the marginal effects on the bankruptcy outcome.

From Figs. 3, 4, and 5, it can be seen that ownership concentration/structure variables have a strong but nonlinear impact on the failure outcome. The direction of the relationship appears to be consistent across all ownership concentration/structure variables. Fig. 3 shows that ownership concentration among the top 5 stockholders sharply impacts the nonfailure outcome up to the 20% ownership level (i.e., if the top 5 stockholders own 20% of the outstanding shares or more, the probability of nonfailure increases significantly). When the top 5 stockholders hold less than 20% of the outstanding shares, the impact on the probability of failure sharply increases. This suggests that lower ownership concentration is an important variable influencing the probability of failure, but this influence changes at different stockholder concentration levels.

[25] The parameter estimate is an "average" effect and has the same value across all observations.

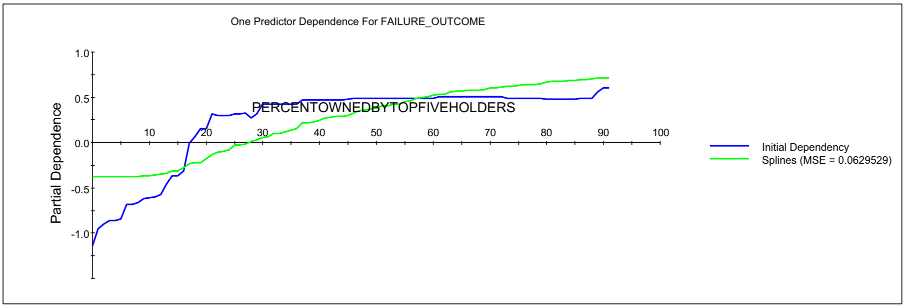

Fig. 3 Partial Dependence Plot for Percentage of Shares Owned by Top Five Stockholders. Figure 3 displays the partial dependence plot of the percentage of shares owned by the top five stockholders on the bankruptcy outcome. The partial dependence plot is a graphical visualization of the marginal effect of this variable on the failure outcome. Its effect is calculated through an iterative averaging process that holds all other variables in Table 3 constant. A first-order single knot spline smooths out the relationship and reveals overall direction. Figure 3 shows that percentage of stock owned by the top five stockholders is strongly increasing (*decreasing*) on the nonfailure (*failure*) outcome

Figure 4 displays a similar story for the institutional ownership variable. The impact on failure is nonlinear and appears to be greatest when there is no or very little institutional ownership in a stock. However, as institutional ownership increases up to the 40% level, there is a quite sharp increase in the probability of nonfailure. Beyond this level, institutional ownership no longer influences the failure outcome. While more extreme concentrations of institutional ownership do not appear to increase the probability of nonfailure, Fig. 4 shows that very low levels of institutional ownership do increase the probability of failure more dramatically. Figure 5 displays the effect of ownership concentration of the top 10 stockholders on the failure outcome. As more stock is owned by the top 10 stockholders (wider dispersion in ownership concentration), the probability of failure actually increases. This reinforces the findings in Figs. 3 and 4 that more dispersion in ownership concentration is strongly related to the failure outcome.

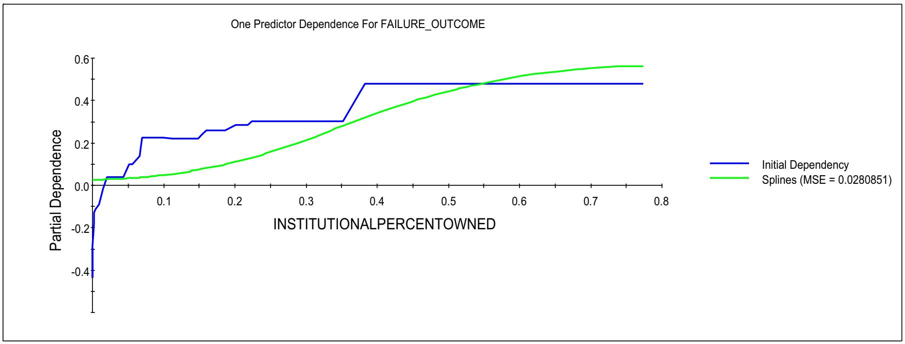

Fig. 4 Partial Dependence Plot for Percentage of Shares Owned by Institutions. Figure 4 shows that percentage of stock owned by institutions is strongly increasing (decreasing) on the nonfailure (failure) outcome. However, lower levels of institutional ownership have a much stronger impact on the failure outcome

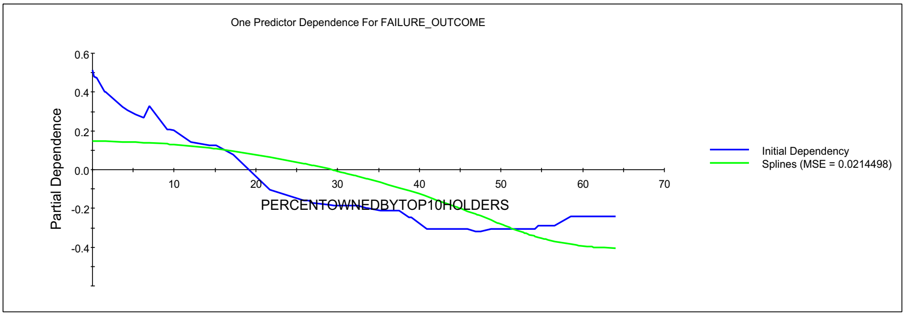

Fig. 5 Partial Dependence Plot for Percentage of Shares Owned by Top 10 Stockholders. Figure 5 shows that percentage of shares owned the top 10 stockholders is strongly increasing (decreasing) on the failure (nonfailure) outcome. This confirms the findings in Figs. 3 and 4 as greater dispersion in ownership appears strongly related to the failure outcome

How well insiders are renumerated also appears to have a nonlinear impact on the failure outcome. Figure 6 indicates that higher levels of total executive compensation increase the probability of nonfailure. When total annual CEO compensation is lower than $200,000, there is a sharper increase in the probability of failure. However, total annual CEO compensation up to the $400,000 level increases the probability of nonfailure. Beyond this level, higher CEO compensation has little effect on the probability of nonfailure. Figure 7 indicates that positive excess return (six months) has a strong impact on nonfailure but only up to the 10% excess return level. Beyond that level, the predictive impact of positive excess returns on the nonfailure outcome starts to flatten out. However, there is a much sharper nonlinear effect in the opposite direction with negative excess returns. Fig. 7 indicates that deteriorating excess returns has a much stronger impact on the probability of failure relative to the nonfailure outcome.[26]

Other variables also appear to make sense but exhibit nonlinear relationships with the failure outcome. For instance, Fig. 8 displays the partial dependence plot for the market capitalization-to-total debt ratio. This ratio has a sharply stronger impact on failure when the ratio falls below 1 (i.e., market capitalization is less than total debt). Similarly, Fig. 9 shows that the current ratio has the strongest impact on the probability of nonfailure up to the three-times level, but beyond that point, the effect of the current ratio on the failure outcome diminishes. Figure 10 indicates that cash flow returns have a strong impact on nonfailure when cash flow is positive. However, the level of cash flow returns does not appear to influence the probability of nonfailure to the same extent as failure. Negative cash flow returns influences failure, but this nonlinear effect is greatest up to the −10% level. Beyond that level, increasingly worse cash flow performance does not noticeably increase the probability of failure.[27]

On balance, there appears to be sufficient empirical evidence to support H_1. The gradient boosting model predicts best in high dimensions. However,

[26] Other market based variables evidence some interesting nonlinear effects. Consistent with previous literature, my analysis of the marginal effects indicates that stock price volatility also exhibits a strong positive relationship with the failure outcome (high stock price volatility increases the probability of failure). Higher levels of short interest are also increase with the failure outcome, but again the impacts are nonlinear and vary over different levels of short interest.

[27] Several other accounting based variables exhibit similar nonlinear relationships with the bankruptcy outcome. For instance, the EBIT to total assets variable displays a similar pattern to cash flow returns.

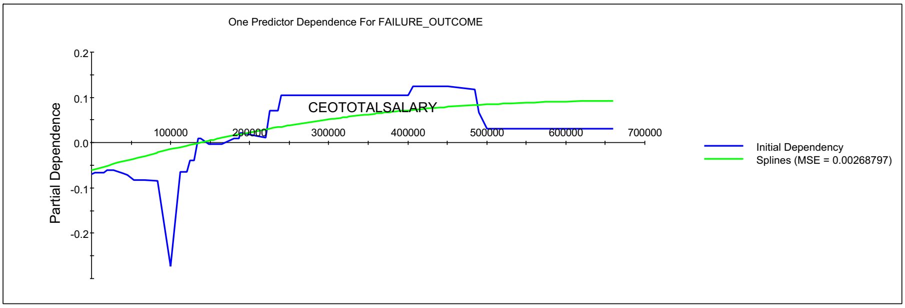

Fig. 6 Partial Dependence Plot for Total CEO Compensation. Figure 6 shows that total CEO compensation is strongly increasing (decreasing) on the nonfailure (failure) outcome up to the $200,000 level

corporate bankruptcy is also multi-dimensional. Separate gradient boosting models estimated on each bankruptcy dimension are quite predictive in their own right, but no single unidimensional model outperforms the full gradient boosting model reported in Table 3. The model has strong explanatory value in terms of the number of variables with nonzero RVIs and the many strong marginal effects that make sense. These marginal effects are associated with many nonlinear effects that arguably better reflect the real world complexity of bankruptcy.

Also implied in H_1 is that a high dimensional model such as gradient boosting should outperform a low dimensional model such as logit. Using identical datasets, variables and model setups, Table 6 shows that gradient boosting outperforms standard logit model on just about every dimension and by very substantial margins.

5.1 Accounting-based variables vs. market-price indicators

It can be seen from Tables 3 and 6 that market-price and accounting-based variables feature quite strongly in the results overall. Table 6 indicates that the predictive performance of accounting-based and market-price indicators are comparable.

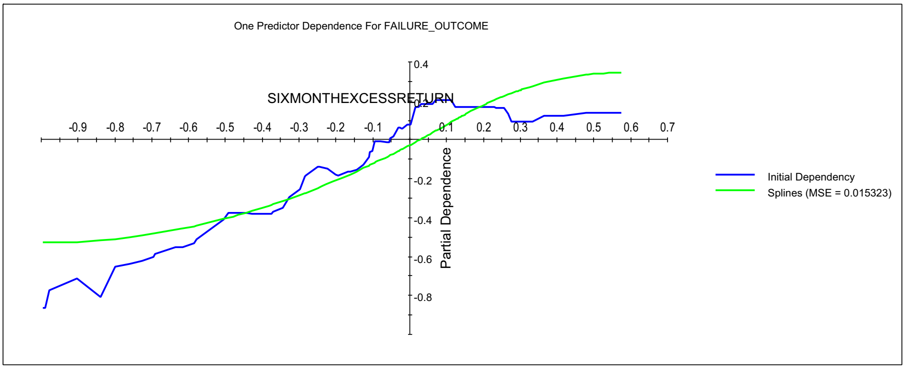

Fig. 7 Partial Dependence Plot for Six-Month Excess Returns. Figure 7 shows that excess returns are strongly increasing on the nonfailure outcome up to the 10%–20% level. However, the relationship is strongly asymmetric. Lower and negative excess returns have a much stronger impact on the failure outcome

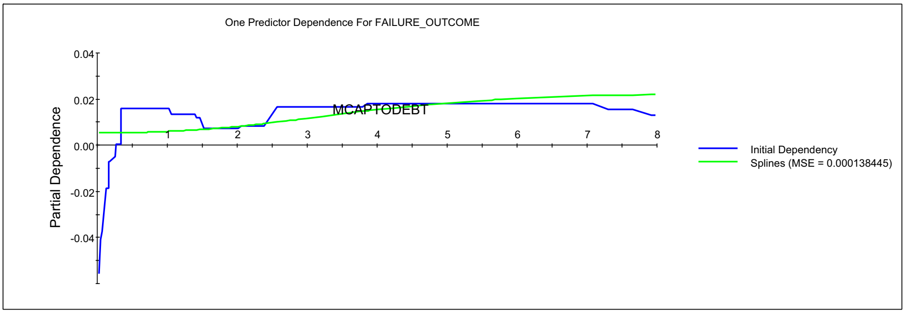

Fig. 8 Partial Dependence Plot for Market Capitalization-to-Total Debt Ratio. Figure 8 shows that the market capitalization-to-debt ratio has a particularly strong impact when less than 1 (market capitalization is less than total debt) but little or no effect above 1

The presence of interaction effects between market-price and accounting-based variables suggests a more subtle and complex relationship.[28] If one variable or dimension clearly dominates the other, strong interaction effects are not expected as this suggests that one variable or dimension can significantly influence/modify the predictive power of the other. Table 7 documents some generally strong interactions effects across several of the high RVI variables reported in Table 3. It can be seen from Table 7 that, among the strongest interacting variables, are several market-price variables, including market capitalization, excess returns, stock price volatility, and short interest. Table 7 displays the whole variable interactions sorted according to their overall interaction strength. Basically, these are the most important interacting variables that contributed most to the predictive performance of the gradient boosting model reported in Table 3 and are ranked according to their interaction weight.

The two-way interaction effects are shown in Table 8 below. Measure 1 in Table 8 shows the impact of the interaction within a pair of variables as a percentage of the overall model prediction (i.e., all variables combined).[29] Measure 2 in Table 8 shows the impact of the interaction within a pair of variables as a percentage of the model prediction attributed *only* to this pair of variables. It displays the effect of the interacting pair normalized to the pair of variables itself (in terms of combined main effects).

Table 8 only reports the strongest Measure 2 interactions across the market-price variables having the strongest influence in Table 7. Unlike standard regression techniques, the gradient boosting model does not provide significance tests or *directions* of interactions as part of the output. The direction of an interaction effect has to be investigated directly by visually examining partial dependence plots. For the purposes of testing H_2, I confine discussion

[28] The commercial version of gradient boosting, TreeNet®, used for this study, allows for automatic detection of all interaction effects. According to Salford Systems, interaction effects are based on comparisons with a genuine bivariate plot (where variables are allowed to interact) and an additive combination of the two corresponding univariate plots. By determining the differences between the two response surfaces, the strength of interaction effect can be measured for a given pair of variables. The core idea behind interaction testing in TreeNet® is that variables can only interact within individual trees (see "Introduction to TreeNet," Salford Systems, San Diego, 2015).

[29] Measure 1 also shows the effect of the interacting pair normalized to the overall response surface (total variation of predicted response).

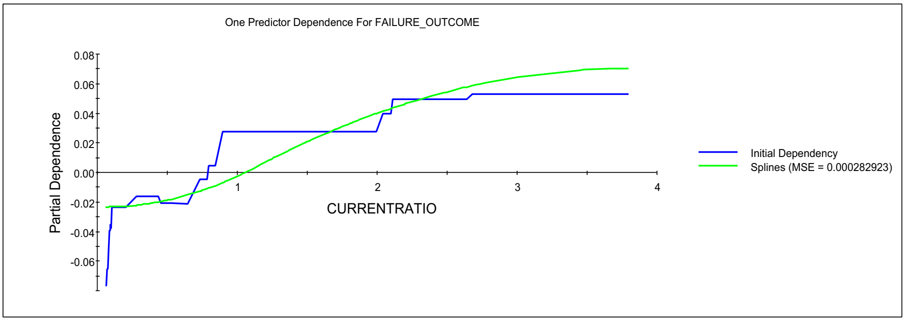

Fig. 9 Partial Dependence Plot of Current Ratio. Figure 9 shows that current ratio has a stronger impact on the nonfailure up to the three-times level but is impact diminishes beyond that level. The current ratio has a strong impact on the failure outcome when less than 1 (or current assets is less than current liabilities)

to the interaction effects of just two market-price variables: excess returns and market capitalization.

Excess returns Table 8 Panel A indicates that the six-month excess returns variable interacts strongly with several accounting-based indicators. The broad direction of the interaction is ascertained from visual analysis of the partial dependence plots. The analysis suggests that across many indicators, higher excess returns interacts with accounting-based indicators so as to *reduce* their predictive power on the failure outcome. Similarly, lower excess returns *increases* the predictive power of accounting based indicators on the failure outcome. In particular, higher excess returns *reduces* the predictive power on the failure outcome of the following accounting-based variables: the current ratio, annual growth in net income, total cash and cash equivalents, annual growth in working capital, net cash flow from financing activities, annual growth in revenues, working capital to total assets, identifiable intangible assets to total assets, EBIT margin, free cash flow per share, annual growth in capital expenditure, earnings per share, capital expenditure to total assets, cash flow returns, sales to total assets, annual growth in operating cash flow, inventory turnover, growth in leverage cash flows, average collection period, and EBIT to total assets. However, higher excess

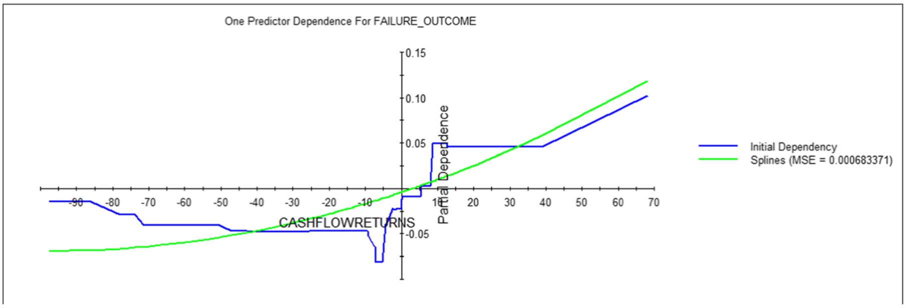

Fig. 10 Partial Dependence Plot for Cash Flow Returns. Figure 10 shows that cash flow returns is strongly increasing (decreasing) on the nonfailure. However, beyond the −10% level, cash flow returns have less of an impact on failure. However, positive cash flow returns have a close to linear relationship on the nonfailure outcome

Table 7 Whole Variable Interactions for Strongest Variable Effects

Measure Predictor	Interaction Weight	Dimension
Percentage of Stock Owned by Institutions	18.92	Ownership concentration/structure
Percentage of Stock Owned by Top Five Stockholders	16.85	Ownership concentration/structure
Percentage of Stock Owned by Top 10 Stockholders	16.44	Ownership concentration/structure
Percentage of Stock Owned by Insiders	16.00	Ownership concentration/structure
Market Capitalization	14.96	Market price
Total CEO Compensation	10.04	Executive compensation
Age of Firm	8.25	Other
Excess Returns (Six Months)	7.85	Market price
Net Operating Cash Flows	7.48	Accounting
Stock Price Volatility (Two-Year Beta)	7.36	Market price
Short Interest	7.26	Market price
Stock Price Volatility (One-Year Beta)	6.37	Market price
Excess Returns (12 Months)	7.09	Market price
Annual Growth in Leverage Cash Flow	6.68	Accounting
Annual Growth in Net Income	6.32	Accounting
Annual Growth in Debt	5.55	Accounting
Average Debt Collection Period	5.15	Accounting
Total Debt-to-Total Assets	5.01	Accounting

Table 7 displays the whole variable interactions sorted according to their overall interaction strength. These are the most important interacting variables, which contributed most to the predictive performance of the TreeNet (gradient boosting) model reported in Table 3. They are ranked according to their interaction weight. Only variables with an interaction weight of greater than 5% are reported in Table 7.

returns *increases* the predictive power on the failure outcome of debt-to-equity, the ratio short-term debt to total liabilities, total debt, total liabilities-to-total equity, and total debt-to-total assets. The interaction effects suggest that accounting-based indicators and market-price measures are complementary sources of information. That is, the presence of one type of information appears to reduce the predictive impact of the other and vice versa. This appears to be consistent with Beaver et al.'s (2005) results, which suggest that market-price variables absorb some of the predictive value of accounting-based measures.

However, the interaction effects reveal some tensions in this interpretation. Lower excess returns also appear to increase the predictive power of accounting-based performance measures on the failure outcome. When excess returns are lower, it is accounting-based measures that seem to absorb the predictive power of excess returns on the failure outcome. One explanation (in a bankruptcy context at least) is that capital markets do not fully impound all accounting information in stock price returns. A part of this explanation is captured in Fig. 7, which shows that the marginal effect of excess returns on the failure outcome is noticeably asymmetric. The impact of excess of returns on the failure outcome is much sharper when excess returns are low or negative, but the impact of higher excess returns on the nonfailure outcome appears less pronounced. This could be explained by some potential market overreaction to negative

Table 8 Interaction Effects with Market-Price and Accounting-Based Variables

Measure 1	Measure 2	Predictor	Direction of Interaction on Failure Outcome
Panel A: Excess Returns (6 Months)			
1.06	8.74	Current Ratio	-
0.972	7.71	Annual Growth in Net Income	-
0.990	7.62	Total Cash and Cash Equivalents	-
0.804	7.11	Annual Growth in Working Capital	-
0.770	6.68	Net Cash Flow from Financing Activities	-
0.620	5.30	Annual Growth in Revenues	-
0.595	5.04	Debt-to-Equity Ratio	+
0.551	4.84	Short-term Debt to Total Liabilities	+
0.537	4.63	Working Capital to Total Assets	-
0.532	4.59	Identifiable Intangible Assets to Total Assets	-
0.510	4.35	EBIT Margin	-
0.504	4.44	Free Cash Flow per Share	-
0.432	3.73	Annual Growth in Capital Expenditure	-
0.413	3.06	Earnings per Share	-
0.403	3.49	Capital Expenditure to Total Assets	+
0.397	3.49	Cash Flow Returns	-
0.395	3.39	Sales to Total Assets	-
0.381	3.25	Total Debt	+
0.353	3.08	Annual Growth in Operating Cash Flow	-
0.324	2.83	Inventory Turnover	-
0.317	2.72	Total Liabilities to Total Equity	+
0.309	2.54	Total Debt to Assets	+
0.304	2.63	Annual Growth in Leverage Cash Flow	-
0.276	2.41	Average Collection Period	-
0.256	2.13	EBIT to Total Assets	-
Panel B: Market Capitalization			
1.53	4.61	Total Debt-to-Total Assets	+
1.27	3.57	Total Assets	-
1.26	3.87	Net Operating Cash Flow	-
1.11	3.30	Capital Expenditure to Total Assets	+
1.02	3.15	Cash Flow per Share	-
0.950	2.93	Net Financing Cash Flows	-
0.903	2.74	Cash Flow Returns	-
0.808	2.48	Free Cash Flow per Share	-
0.737	2.23	Earnings per Share	-
0.680	2.01	Annual Growth in Net Income	-
0.648	1.88	Total Debt	+
0.640	1.82	Total Revenue	-
0.620	1.90	EBIT to Total Assets	-
0.560	1.73	Sales to Total Assets	-

Table 8 reports the strongest overall interactions effects between (1) excess returns and accounting-based variables (Panel A) and (2) market capitalization and accounting-based variables (Panel B). Measure 1 in Table 8 shows the impact of the interaction within a pair of variables as a percentage of the overall model prediction (i.e., all variables combined).· Measure 2 in Table 8 shows the impact of the interaction within a pair of variables as a percentage of the model prediction attributed only to this pair of variables. It displays the effect of the interacting pair normalized to the pair of variables itself (in terms of combined main effects). Unlike standard regression techniques, the TreeNet (gradient boosting) model does not provide significance tests or directions of interactions as part of the output. The direction of an interaction effect has to be investigated directly by examining partial dependence plots. The analysis suggests that higher excess returns interacts with accounting-based indicators so as to reduce their predictive power on the failure outcome. Similarly, lower excess returns increases the predictive power of accounting-based indicators on the failure outcome. A similar effect is observed for the market capitalization variable.

aspects of a firm's performance. However, if accounting-based indicators fail to fully confirm the market's negative assessment (through lower excess returns), the predictive power of accounting-based measures appears to assume more significance than excess returns on the bankruptcy outcome.

Market capitalization Table 8 Panel B indicates that market capitalization also interacts strongly with several accounting-based variables. Consistent with Panel A results, the most noticeable relationship is that higher levels of market capitalization appear to *reduce* the predictive impact of accounting-based variables on the failure outcome, while lower levels *increase* the impact of accounting-based measures on the failure outcome. For instance, lower levels of market capitalization increase the impact of the following accounting-based variables on the failure outcome: total assets, net operating cash flow, capital expenditure to total assets, cash flow per share, net financing cash flows, cash flow returns, free cash flow per share, earnings per share, annual growth in net income, total revenue, EBIT to total assets, and sales to total assets while increasing the impact of the debt-to-assets ratio and total debt. While higher levels of market capitalization appear to absorb some of the predictive power of accounting-based indicators, this relationship works differently for lower levels of market capitalization. Here, accounting-based indicators appear to take away some of the predictive power of market capitalization. This result can be explained by viewing market capitalization as a proxy for firm size and therefore bankruptcy risk. Larger firms are perceived by the market to have lower bankruptcy risk. Hence accounting-based performance variables assume less predictive importance as market capitalization increases. On the hand, smaller firms with lower market capitalization are typically associated with higher bankruptcy risk. Accounting-based measures could be assuming greater predictive power for smaller firms because accounting performance offsets or compensates for small firm risk (i.e., a small firm will be perceived by the market as less risky *only* if its financial performance is stronger).

In summary, there does not appear to be strong evidence supporting H_2 concerning the dominance of mark-price variables over accounting-based variables. They appear to be equally powerful predictors when gradient boosting models are estimated on each dimension separately. However, accounting-based variables have a stronger overall impact in the full (high dimensional) model reported in Table 3. The presence of strong interaction effects between market-price and accounting-based indicators also suggest that these variables can, under different circumstances, act as both complementary and competing sources of information in the prediction of corporate bankruptcy.

5.2 Market-price indicators vs. other non-accounting variables

I assess H_3 from the outputs of Tables 3, 6, and 7. The full gradient boosting model reported in Table 3 indicates that market-price variables are important predictors but appear to be dominated by other variables, such as ownership concentration/structure and, to a lesser extent, executive compensation variables, which have higher overall RVIs. Tables 3 and 6 indicate mixed evidence about the dominance of market-price variables over other non-accounting bankruptcy predictors. Table 6 indicates that market-price variables do not significantly outperform other dimensions, except external ratings and macroeconomic factors and, to a lesser extent, executive compensation variables. However, Table 7 also indicates that market-price variables exhibit strong interaction effects with some non-accounting predictors, such as ownership concentration/structure.[30] This again suggests that market-price variables are not clearly dominant in the gradient boosting analysis, but their predictive influence is moderated in some way by the effects of other strong predictors reported in Table 3.

It is not immediately apparent why variables such as ownership concentration/structure have stronger RVIs than most of the market-price indicators reported in Table 3. As stated by Beaver et al. (2005), market prices are expected to impound the total mix of information, both financial and nonfinancial. One possibility is that ownership concentration/structure is more fundamental than market-price information in the *specific* context of corporate bankruptcies. Not only is Chapter 11 bankruptcy a complex and costly proceeding for the firm, but it usually results in significant diminution in stockholder value. Presumably large stockholders and institutional stockholders who control significant voting rights have strong incentives to monitor the firm financial performance and to rectify deficiencies in financial performance to avoid this outcome. A high percentage of ownership by insiders could lead to similar incentives to avoid the Chapter 11 process.

Large stockholders, institutions, and insiders are also expected to have an important say in all aspects of the Chapter 11 process. The actions of these stakeholders can also have a significant impact on stock price returns. For instance, they could sell off their holdings if they believe they cannot turn the company fortunes around, potentially driving stock returns downward. On the other hand, if they are optimistic about the firm's prospects, they could hold their investments, potentially providing price support. In this sense, stock price returns merely impound value-relevant bankruptcy information, but ownership concentration/structure variables play a more *determining* role in these events.

While it was expected that higher total executive compensation, as a proxy for management quality, will be associated with a lower risk of failure, it is a little surprising that this variable has come out so strongly in the gradient boosting analysis. Some clues appear to come out in the interaction effects. The total CEO compensation variable appears to have one of the strongest interaction weights in the model (see Table 7). This variable interacts most strongly with market capitalization, several accounting-based and ownership concentration/structure variables. Accounting-based indicators appear to interact with total CEO compensation in a way that reduces its predictive influence on the failure outcome. This suggests that this variable could proxy for strong accounting performance and thus act as complementary source of information on the failure outcome. Likewise, higher excess returns also reduce the predictive impact of CEO compensation, suggesting that excess returns absorb the positive impacts of high quality management (which may be reflected in stronger financial performance). Higher market capitalization also reduces the predictive impact of total

[30] The two-way interaction effects suggest that higher levels of ownership concentration and institutional ownership reduce the impact of excess returns on the failure outcome.

CEO compensation, suggesting that the effects of high quality management is most strongly concentrated in and captured by larger firms.

6 Further considerations

The strong predictive performance of gradient boosting in this paper raises a number of additional questions and considerations. First, the empirical results are significantly stronger than many of the related boosting studies reviewed in Appendix 2. Which factors may have contributed to this superior performance? A second issue relates to the stability and robustness of the gradient boosting model under different sampling and data partitioning scenarios.

6.1 Performance of gradient boosting relative to prior literature.

A tabular summary of the boosting literature (as it applies to bankruptcy prediction) is summarized in Appendix Table 12. While predictive performance tends to vary across these studies, with AUCs ranging from 0.75 to 0.95, the average AUC across studies appears to be around the 0.85 level. This is of course very good performance but not as strong as the current results. I attribute the stronger predictive performance to several factors. First, most Appendix Table 12 studies appear to have used freeware software (such as R) or more generalist statistical software for model estimation. This study uses TreeNet®, a prize-winning commercial software for gradient boosting with many specialist enhancements and refinements designed to optimize predictive performance.[31] Second, all of the studies reviewed in Appendix Table 12 have focused primarily on financial variables, and many use private company failure samples (which limits the range of variables which can be tested). The current study generates higher predictive accuracy partly because it includes a wider variety of predictive variables, including market-price variables, stockholder ownership/concentration, executive compensation, macroeconomic variables, and other factors. This study shows that many of these variables have high RVIs, indicating that they contribute quite strongly to overall predictive performance. As can be seen from Table 6, focusing exclusively on accounting variables does significantly reduce AUC performance. Third, none of the studies reviewed in Appendix Table 12 identified or discussed any interaction effects, possibly because of the limited range of input variables tested, limitations with the gradient boosting software used, or both. A useful feature of TreeNet® is the ability to detect all higher order interactions across the entire feature space (i.e., all Appendix Table 11 variables). Interaction effects feature quite strongly in

[31] According to Salford Systems, TreeNet®'s high predictive accuracy comes from the algorithm's capacity to generate thousands of small decision trees built in a "sequential error–correcting process that converges to a highly accurate model." Other reasons include the power of TreeNet®'s interaction detection engine. TreeNet® establishes whether higher order interactions are required in a predictive model. The interaction detection feature not only helps improve model performance (sometimes dramatically) but also assists in the discovery of valuable new variables and patterns. This software has automatic features for detecting higher order interaction effects across any number of variables (not available in most packages or older methods). TreeNet® also has optimization features to help find the analyst find the best tree depth which maximizes predictive performance.

the current results and contribute significantly to overall performance of the model.[32] Fourth, nearly all of the Appendix Table 12 studies (which use public company samples) are based on quite small sample sizes drawn from different reporting jurisdictions (some of the private company samples tend to be larger). While small samples can impair the performance of any machine learning model (or traditional model for that matter), TreeNet® in particular is known to predict more accurately on larger samples.[33]

Fifth, model architecture details such as tree depth can be critical to the overall performance of boosting models (Friedman 2001; Hastie et al. 2009). Most of the studies summarized in Appendix Table 12 provide few details on model setups. For instance, it is quite possible to run a boosting model in R software with (say) a 100 trees, but this may not be the optimal tree depth (or even close to optimal) to achieve the best predictive accuracy. TreeNet® allows the researcher to optimize predictive accuracy through the tree depth functionality. For this study, predictive performance was optimized using a quite large number of trees (680 trees). Finally, some of the studies in Appendix Table 12 are not examining boosting where *decision trees* are used as the base classifier. Recall that boosting is a general method for improving the performance of learning algorithms (the idea of combining and weighting many weak classifiers into a very strong voting committee of classifiers). Other base classifiers can be used, such as neural networks. (Hence several studies in Appendix Table 12 are examining *boosted neural networks.*) However, the vast majority of studies in the broader boosting literature use decision trees as the base classifier, mainly because decision trees are known to predict better overall (Friedman 2001; Hastie et al. 2009; Jones et al. 2015).[34]

To get an indication of how much a gradient boosting model might be impacted by some of the conditions discussed above, I re-estimated the Table 3 gradient boosting model with the following constraints: (1) tree depth set to a maximum of 50–100 trees, (2) no interaction effects permitted, (3) using only financial variables as predictors, and (4) randomly reducing sample size to less than 5000 observations in total. Under these constraints, the model produced Type I & II test error rates and AUCs more consistent with the literature reported in Appendix Table 12. (The best AUC was 0.86.)

6.2 Stability and robustness of gradient boosting model

The high AUCs of Table 3 model raise questions about the stability of the overall model. TreeNet® provides an extensive range of diagnostic checks for model stability and over-fitting. The battery partition is one of the best checks for this purpose. With

[32] The predictive power of interaction effects can readily be isolated by TreeNet® by estimating a model with and without interaction effects and comparing the AUCs accordingly.

[33] A basic issue in any machine learning model is the learn curve—this shows how well the model predicts as the training sample is increased. The smaller the sample size, the higher the expected bias in the model. Model variance is also expected to diminish as the sample size is monotonically increased. A model also needs to be tested out of sample, and the more sample available, the more confidence we can have in the generalization error of the model. It goes without saying that the efficient estimation of more complex models involving many feature vectors and higher order interaction effects requires larger sample sizes, particularly when comparing predicting performance across alternative models.

[34] Jones et al. (2015) reported that gradient boosting and AdaBoost strongly outperformed when decision trees are used as the base classifier. For instance, they examined boosting models where both logit and probit were used as the base classifier, but they did not perform as strongly.

this test, TreeNet® builds a series of models (based on Table 3 inputs and model setups) where the training and test samples are repeatedly drawn at random from the dataset in some proportion. (Basically the training and test samples are reshuffled at random many times and new models estimated on each random reshuffle.) The results provide a more realistic picture of the possible range in predictive performance for the preferred model. If the test error rates are too erratic across models, this will indicate stability and over-fitting problems with the model. Table 9 below provides results of a 20-repeat battery partition test of the Table 3 model based on a 60/30/10 allocation to the training, test, and holdout samples. It shows the performance of each battery test, the optimum tree depth reached, and the misclassification error rates across each model.[35]

As can be seen from Table 9, the misclassification error is quite stable across each model with the highest test error reported in on Model 14 of 3.74% and the lowest test error on Model 17 of 2.66%. I further test the Table 3 model using alternative out-of-sample testing approaches. Table 3 is based on a 70/30 random allocation to the training and test samples. However, this technique can be problematic on small samples (less than 10,000 observations), particularly where there may be insufficient data to obtain a reasonable spread of sample for the training and test sets. A more robust technique for smaller samples is V-fold cross-validation. However, running the same gradient boosting model with a tenfold cross validation made little difference to the results or RVI ranks in Table 3. I also tested the Table 3 model across industry groups as shown in Table 10 below. The performance of the model appears to be fairly stable across industry groups with the highest classification error reported for the telecommunications sector (classification error of 7.3%).

Look-ahead bias For the purposes of this study, a bankrupt firm is coded as bankrupt for all firm year observations available for that firm. While this approach is commonplace for discrete choice models (such as logit and probit), it does involve look-ahead bias.[36] This raises the question as to the extent to which gradient boosting is impacted by look-ahead bias. To test this, I recoded the data so that a bankrupt firm is only coded as bankrupt in the year of bankruptcy (and all years prior to bankruptcy are coded as active or nonfailed). This reduced the bankruptcy sample from 4460 to 1115 firm year observations, while increasing the nonfailed sample to 35,104 firm year observations. Using exactly the same variable inputs and gradient boosting model architecture reported in Table 3, the re-estimated model shows a test error rate of 3.48%, which is very similar to the original model.

7 Conclusion

The empirical results of this study support the proposition that corporate bankruptcy is better explained/predicted in a high dimensional (and multi-dimensional) setting. This

[35] This is double TreeNet®'s battery petition default setting of 10-repeat models.

[36] Bankruptcy researchers often want to predict the success of their models a number of years prior to failure, not just in the year of failure. For instance, testing the predictive power of a model at $t = -3$ (three years from failure) requires knowledge of which companies failed in $t = 0$ (year of failure). Hence look-ahead bias is unavoidable in this situation.

Table 9 Model Stability Tests for Table 3 Model (All Data)

Battery Model	Optimal Trees Count	Misclassification Error on Test Sample
TreeNet Model 1	542	0.0357
TreeNet Model 2	573	0.0280
TreeNet Model 3	598	0.0330
TreeNet Model 4	710	0.0290
TreeNet Model 5	443	0.0365
TreeNet Model 6	385	0.0315
TreeNet Model 7	560	0.0328
TreeNet Model 8	554	0.0313
TreeNet Model 9	783	0.0276
TreeNet Model 10	694	0.0331
TreeNet Model 11	413	0.0349
TreeNet Model 12	513	0.0331
TreeNet Model 13	467	0.0325
TreeNet Model 14	447	0.0374
TreeNet Model 15	610	0.0329
TreeNet Model 16	456	0.0361
TreeNet Model 17	666	0.0266
TreeNet Model 18	636	0.0297
TreeNet Model 19	418	0.0313
TreeNet Model 20	630	0.0346

Table 9 provides results of a 20-repeat battery partition test of the Table 3 model with 60/30/10 allocation for the training, test, and holdout samples. The battery partition is one of the best diagnostic checks for model stability and over-fitting. Under this test, TreeNet builds a series of models (based on Table 3 inputs and model setups) where the training and test samples are repeatedly drawn at random from the dataset in some proportion. (Basically the training and test samples are reshuffled at random many times and new models estimated on each random reshuffle.) It shows the performance of each battery tests, the optimum tree depth reached, and the misclassification error rates across each model. As can be seen from Table 9, the misclassification error is quite stable across each model, with the highest test error reported in on Model 14 of 3.74% and the lowest test error on Model 17 of 2.66%.

is examined using TreeNet®, a commercial version of the gradient boosting model. The analysis indicates that most dimensions of corporate bankruptcy are quite predictive but no single dimension outperforms a gradient boosting model estimated on all available predictors.

The high dimensional context of this study appears to explain a conspicuous pattern in the literature over the past 50 years. Most studies report statistically significant results (and oftentimes impressive classification success), notwithstanding the noticeable variability across studies with respect to sample sizes, time frames, reporting jurisdictions, alternative definitions of corporate failure, and types of explanatory variables tested. A possible explanation is that the literature has fallen into a low dimensional trap. In other words, the literature has tended to rely on restrictive models, such as logit and MDA, which can only handle a finite number of predictors. As a

Table 10 Full TreeNet (Gradient Boosting) Model Estimated Across Industry Groups (All Data)

Industry (GICS)	ROC (test sample)	Classification Accuracy (test sample)
Healthcare	0.998	0.946
Energy	0.998	0.931
Materials	0.996	0.966
Industrials	0.971	0.957
Telecommunications	0.978	0.927
Utilities	0.977	0.983
Information Technology	0.981	0.974
Financials	0.976	0.983
Consumer Staples	0.981	0.940
Consumer Discretionary	0.982	0.975

Table 10 displays the out-of-sample predictive performance of the Table 3 model partitioned across industry groups. The performance of the model appears to be fairly stable across industry groups, with the highest classification error reported for the telecommunications sector (classification error of 7.3%).

result, most bankruptcy studies only capture a small, albeit statistically significant, subset of a much larger (high dimensional) picture.

There are several advantages for using a gradient boosting model in bankruptcy prediction. First, the model provides an effective method for drawing together many different predictors and comparing their performance in a single statistical framework. It can rank order any number of predictors based on the RVI metric. This feature is useful for evaluating the role of alternative bankruptcy predictors, such as accounting-based variables vs. market-price indicators. In contrast to some prior literature, this study rejects H_2 and H_3, relating to the superior predictive performance of market price indicators. While the results clearly show that market-price variables have predictive value, they appear to be no more important than accounting-based indicators in the gradient boosting analysis. There also appears to be a dynamic at play between market-price and accounting-based variables, which surfaces in the interaction effects. In a high dimensional setting, these variables appear to have a moderating influence on each other, and they can act as both competing and complementary sources of information in a bankruptcy prediction context.

The variable ranking feature of gradient boosting can also be used to identify new bankruptcy predictors not widely explored in prior literature. After taking into account the role of all other model influences (including all important interaction effects), the results of this study show that nontraditional variables, such as ownership concentration/structure and CEO compensation variables, are among the best predictors overall. The results also suggest that unscaled market and accounting variables that proxy for size effects tended to perform better than financial ratios and market price variables overall. Other non-accounting variables, such as auditor type, number of business segments, and age of the firm, appear to perform as well as financial ratios.

Second, the gradient boosting model produces noticeable strong out-of-sample predictive performance that is difficult to achieve with conventional models such as logit. The model achieves high predictive accuracy by optimizing many weak classifiers in a stage-wise process. It keeps adding new weak classifiers until some loss function is minimized. This allows gradient boosting to exploit the entire feature space, including all interaction effects, to improve prediction outcomes. The model in Table 3 displays very low Type I and II classification errors and an impressive AUC. The model also appears to be resilient to various model stability and over-fitting tests.

Third, the gradient boosting model requires modest researcher intervention and is largely immune to dirty data issues frequently encountered in bankruptcy datasets. This reduces many forms of bias that can be introduced by the researcher, such as judgments relating to scaling, monotonic transformations of predictor variables, treatment of outliers and missing values, the handling of database errors, and other issues. However, a more pernicious bias relates to data snooping. In a conventional model such as logit, the inclusion of irrelevant predictors affect the maximum likelihood solution, which can undermine the model's stability and performance. Researchers typically weed out irrelevant inputs by re-estimating their models with different explanatory variables, interactions, or both, but this can lead to snooping bias (White 2000). The dangers with data snooping is that an apparently good model can be produced by mere chance and not reflect anything innately meaningful in the data. Any number of predictors can be added to the gradient boosting analysis, irrespective of their assumed importance, without cost to model stability and performance. Variables that are irrelevant or not predictive are simply left out of the tree ensemble. This feature of gradient boosting appears to remove snooping bias.

All bankruptcy models have their strengths and limitations. A weakness of the gradient boosting model is the underlying complexity of the model (the black box effect), which limits interpretability. While outputs such as RVIs and marginal effects go some way to remedying this issue, the logit model still has the undeniable strength of being highly interpretable. On the other hand, the benefits of precise parameter estimates, interpretable signs, and clear-cut significance tests are cold comfort if the underlying model is a poor reflection of reality. Finally, there can be considerable potential for sophisticated statistical learning methods such as gradient boosting to enhance the performance of conventional models such as logit. For instance, the gradient boosting model can function as a useful exploratory and diagnostic tool as well as the bias eliminating framework to identify the most important predictors in any dataset, including important nonlinear relationships and interaction effects. This type of analysis can be harnessed by a logit model to improve predictive and explanatory performance.

Acknowledgements Financial support for this project was provided by the Australian Research Council: ARC A7769.

Appendix 1

Table 11 Variable Definitions

Variable	Definition (based on Standard and Poor's Capital IQ Items)
Accounting-Based Variables:	
Total Short-Term Debt/Total Liabilities	Item 2043/ Item 1276
Total Liabilities/Total Equity	Item 1276/ Item 1275
Total Debt/Total Equity	Item 2161/ Item 1275
Total Debt/Total Assets	Item 2161/ Item 1007
Annual Growth in Total Debt Levels	Item 2161 (*t*) – Item 2161 (*t-1*) / Item 2161 (*t-1*)
Annual Growth in Revenue	Item 28 (*t*) – Item 28 (*t-1*) / Item 28 (*t-1*)
Annual Growth in Operating Cash Flow	Item 2006 (*t*) – Item 2006 (*t-1*) / Item 2006 (*t-1*)
Annual Growth in Net Income	Item 15] (*t*) – Item 15 (*t-1*) / Item 15 (*t-1*)
Annual Growth in Capital Expenditure	Item 2021 (*t*) – Item 2021 (*t-1*) / Item 2021 (*t-1*)
Annual Growth in Leverage Free Cash Flow	Item 4422 (*t*) – Item 4422 (*t-1*) / Item 4422 (*t-1*)
Annual Growth in Working Capital	Item 1311 (*t*) - Item 1311 (*t-1*) / Item 1311 (*t-1*)
Working Capital/Total Assets	Item 1008- Item 1009/ Item 1007
Current Ratio	Item 1008/ Item 1009
Net Operating Cash Flow/Total Debt	Item 2006/ Item 2161
EBIT Margin	Item 400/ Item 28
Net Profit Margin	Item 15/ Item 28
Gross Profit Margin	Item 28 - Item 1/ Item 28
Cash Flow to Revenue	Item 2006/ Item 28
Net Operating Cash Flow to Total Assets (Cash Flow Returns)	Item 2006/ Item 1007
Sales to Total Assets	Item 28/ Item 1007
Inventory Turnover	Item 34/ Item 1043
EBIT/Total Assets	Item 400/ Item 1007
EBIT/Interest Expense	Item 400/ Item 82
Interest Cover	Item 15/ Item 82
EPS (Diluted)	Diluted Net Income [Item 3522] + Earnings Of Discontinued Operations [Item 40] + Extraordinary Item & Accounting Change [Item 42]) / Diluted Weighted Average Shares Outstanding [Item 342]
Free Cash Flow per Share	Cash from Operations [Item 2006] + Preferred Dividends Paid [Item 2116] - Capital Expenditure [Item 2021] + Sale (Purchase) of Intangible Assets) [Item 2029]/ Basic Weighted Average Shares Outstanding [Item 3217]
Cash Flow per Share	Cash from Operations [Item 2006] + Preferred Dividends Paid [Item 2116] + Capital Expenditure [Item 2021] + Sale (Purchase) of Intangible Assets) [Item 2029]/ Basic Weighted Average Shares Outstanding [Item 3217]
Total Capital Expenditure/ Total Assets	Item 2021/ Item 1007
Total Intangible Assets (other than Goodwill)/ Total Assets	Item 1040/ Item 1007

Table 11 (continued)

Variable	Definition (based on Standard and Poor's Capital IQ Items)
Average Collection Peri[illegible]	Item 1021/ Item 28
Total Revenues	Item 28
Total Assets	Item 1007
Net Operating Cash Flows	Item 2006
Net Investing Cash Flows	Item 2007
Net Financing Cash Flows	Item 2004
Total Debt	Item 4173
Total Cash a[illegible]sh Equivalents	Item 1096
Total Bank [illegible]	Item 1046
Total Tangible Book Value	Item 1310
Goodwill to Total Assets	Item 1171/ Item 1007
Total Revolving Credit to Total D[illegible]	Outstanding Balance for Revolving Credit /Item 2161
Total Revolving Credit Facilities	The total represents the total drawn amount of a company's revolving credit facilities.
Provision for Credit Losses to Total Liabilities	Item 2112/Item 1276
Bad Debt Provisions to Total Liabilities	Item 95/Item 1276
Asset Write-Downs	Item 32
Provision for Credit Losses	Item 2112
Provision for Bad Debts Expense	Item 95
Total Write-Downs to Total Assets	Item 32/Item 1007
Dividends per Share	Item 2074/Item 342
Ownership concentration/str[illegible]e	
Percentage of Stock Own[illegible] Top Five Stockholders	Percentage of the total shares outstanding held by the top five holders as aggregated from all S&P Capital IQ sources.
Percentage of Stock Owned by Top 10 Stockholders	Percentage of the total shares outstanding held by the top 10 holders as aggregated from all S&P Capital IQ sources.
Percentage of Stock Owned by Insiders	Percentage of the total shares outstanding held by insiders from S&P Capital IQ Ownership database. Insiders include officer and director ownership as well as non-officer/director individuals.
Percentage of Stock Owned by Institutions	Percentage of the total shares outstanding held by institutions extracted from S&P Capital IQ Ownership database. Institutions include investment managers, private equity, venture capital, and hedge funds.
Executive Compensation Variables:	
Total CEO Compensation	Total CEO compensation is sourced by Capital IQ directly from the company filings based on the actual information reported by companies.
Stock-Based Compensation	Item 24,034
Total CFO Compensation	Total CFO compensation is sourced by Capital IQ directly from the company filings based on the actual information reported by companies.
Market Price Variables:	
Market Capitalization	The market value of shares held in the issuer. This is calculated by multiplying the price by the number of

Table 11 (continued)

Variable	Definition (based on Standard and Poor's Capital IQ Items)
	shares held. Market value is calculated by Capital IQ as of the last available close price.
Market Capitalization to Debt	Market Capitalization /Item 2161
Excess Return (One Year)	Capital IQ calculated 12-month excess return relative to a market benchmark (the S&P 500 for US stocks and the MSCI EAFE for other developed markets).
Excess Return (Six Months)	Capital IQ calculated six-month excess return relative to a market benchmark (the S&P 500 for US stocks and the MSCI EAFE for other developed markets).
Stock Price Volatility (1 Year, 2 Years)	Measured by Beta calculated on a 12 and 24 month basis. Capital IQ estimates beta with a 12 and 24 month regression line of the percentage price change of the stock relative to the percentage price change of its benchmark (the S&P 500 for US stocks).
Shor[illegible]est	Short interest is defined by Capital IQ as the balance of shares outstanding of a publicly traded company that are sold short, but not yet covered or closed out. Capital IQ provides a detailed methodology for the collection and reporting of short interest.
External Rati[illegible]	
Number of Analysts	Number of analysts following the firm
Analyst Consensus Stock Recommendations	Capital IQ provides both the Broker Recommendations and Standardized Recommendations. Capital IQ standardizes all the individual recommendations across the brokers on a 1–5 point scale.
Consensus EPS Forecasts (1 YR)	Consensus analyst estimate of EPS for next 12 months.
Consensus EPS Forecasts (Long Term)	Consensus analyst estimate of long-term EPS forecast.
Credit Ratings Change	Ratings changes are classified as a binary outcome dependent variable, where a ratings upgrade is coded 1 and a ratings downgrade is coded 0. Ratings changes for a company is defined as a rating change from one rating category to another (for example, if a company's rating changes from A to AA or from BBB- to BB).
Macroeconomic Variables:	
CPI Index	This is the official annual inflation rate reported by each country in the sample. For instance, in the United States, it is the official inflation rate reported by the U.S. Bureau of Labour. Accessed from TradingEconomics.com.
Unemployment rate	Each country's official annual unemployment rate. For instance, in the United States, it is the official rate reported by the U.S. Bureau of Labor Statistics.
TED Spread	The price difference between three-month futures con tracts for U.S. Treasuries and three-month contracts for Eurodollars having identical expiration months.
Leading Index USA	An index published monthly by the Conference Board used to predict the direction of the economy's movements in the months to come. The index is broadly made up of 10 economic components.
Michigan Sentiment Index	A consumer confidence index published monthly by the University of Michigan.
Moody's Seasoned BBB Bond Yi[illegible]	

Table 11 (continued)

Variable	Definition (based on Standard and Poor's Capital IQ Items)
	An investment bond that operates as an index for the performance of all bonds assigned a BBB rating by Moody's Investors Service.
Moody's Seasoned AAA Bond Yield	An investment bond that operates as an index of the performance of all bonds assigned a Aaa rating by Moody's Investors Service.
Real GDP and Real GDP Growth	This is the inflation-adjusted figure reflecting the value of all goods and services produced by a nation. It indicates how much a country's production has in creased or decreased compared to the previous year. GDP is reported by the World Bank Group for most countries in the sample.
Public Debt to GDP	The percentage of government debt to GDP.
NBER Recession Indicator	An indicator which defines a period of recession in the economy. NBER defines a recession as "a significant decline in activity spread across the economy, lasting more than a few months, visible in industrial production, employment, real income, and trade. A recession begins just after the economy reaches a peak of output and employment and ends as the economy reaches its trough."
10 Year Treasury Constant Maturity	An index published by the Federal Reserve Board based on the average yield of a range of Treasury securities, all adjusted to the equivalent of a 10-year maturity. This figure is used as a reference point to establish the price of other financial instruments such as corporate bonds.
Other variables	
Firm Size	Proxied by Market Capitalization and Total Assets
Age of firm	Year the firm was first incorporated
Industry/sector variables	Ten major industry groups defined by GICS classification scheme (energy, materials, industrials, consumer discretionary, consumer staples, healthcare, financials, information technology, telecommunications and utilities).
Auditor Type	Dummy coded 1 for a Big Four auditor, 0 otherwise.
Number of Business Segments	Extracted from Capital IQ database.
U.S. vs World Dummy	Dummy coded 1 for a U.S. incorporated public company, 0 otherwise.

Appendix 2

Table 12 Applications of boosting models to bankruptcy prediction

Author/s	Journal	Sample Size	Public/ Private Firm Sample	Definition of distress	Number/Type of Input Variables	Boosting Model Used	Interaction Effects	Model Architecture & software details	Prediction/AUC
Cortés et al. (2007)	Appl Intell (2007) 27:29–37	2730 private firms drawn from BvD's SABI database over the period 2000–2003.	Private firms.	Bankruptcy, temporary receivership, acquired and dissolved firms.	18 predictors with 14 financial variables/-ratios.	AdaBoost vs. single tree	N/A	100 trees for tree depth, R software.	AdaBoost outperforms single tree. Overall test error rate of 6.569%.
Sun et al. (2011).	Expert Systems with Applications 38 (2011) 9305–9312	692 Chinese-listed firms sampled over the period 2000–2008, taken from Shenzhen Stock Exchange and Shanghai Stock Exchange.	Public firms.	Negative net profit in consecutive two years, or net capital lower than the face value.	41 financial variables/-ratios.	AdaBoost with single attribute test vs. single tree and SVMs	N/A	Limited discussio-n, 30-fold holdout tests.	AdaBoost ensemble with SAT outperforms all other models. Test errors are 2.78% (one year from failure), 12.81% (two years from failure) and 27.51% (three years from failure) respectively.
Kim and Kang (2010)	Expert Systems with Applications 37 (2010) 3373–3379	1458 externally audited manufacturing firms, half of which went bankrupt over the period 2002–2005.	Private firms obtained from a Korean commercial bank.	Bankruptcy (presume legal bankruptcy).	32 financial variables/-ratios.	Boosted NNs vs. Bagged NN and NNs.	N/A	Not stated or limited details provided.	Boosting and bagged NN improve performance of NNs. AUCs around 0.75, but best performance appears to be from bagged NNs.
Wang et al. (2014)	Expert Systems with Applications 41 (2014) 2353–2361	Study uses two small datasets, including 240 firms in one dataset and 132 firms in the other.	Not clear from the study (refers to previous studies using the data).	Generally not well described and drawn from previous studies.	Up to 30 financial variables/-ratios.	FS Boosting compared to other tech-niques.	N/A	Limited details provided. WEKA software.	FS-Boosting achieved best overall classification success rates of 81.50 on dataset 1 and 86.70% on dataset 2; followed by boosting and bagging. SVMs performed quite strongly on dataset 2.
Hung and Chen (2009)	Expert Systems with Applications	56 bankrupt companies and 64	Public firms.	Legal bankruptcy.	30 financial variables/-ratios.	Selective ensemble vs.	N/A	Not stated or limited	No conclusion reached for the best bankruptcy

Table 12 (continued)

Author/s	Journal	Sample Size	Public/ Private Firm Sample	Definition of distress	Number/Type of Input Variables	Boosting Model Used	Interaction Effects	Model Architecture & software details	Prediction/AUC
	36 (2009) 5297–5303	nonbankrupt firms sampled between 1997 and 2001.				stacking ensemble.		details provided.	prediction technique. They propose selective ensemble which integrates with the concept of the expected probability.
Chandra et al. (2009)	Expert Systems with Applications 36 (2009) 4830–4837	240 dot-com companies drawn from WRDS database (120 failed vs. 120 nonfailed).	Public firms.	Not explicitly stated.	24 financial variables/-ratios.	Compares MLP, random forests, logit, SVMs, and CART.	N/A	Not stated or limited details provided.	Boosting yielded results superior to those reported in previous studies on the same dataset. Combining boosting with other techniques such as SVM resulted in an AUCs above 0.90.
Fedorova et al. (2013)	Expert Systems with Applications 40 (2013) 7285–7293	888 Russian manufacturing firms sampled over the period 2007–2011.	Private companies from SPARK database.	Bankruptcy according to Russian Federal Law.	23 financial variables/-ratios.	Compares boosted NNs with other tech-niques.	N/A	Not stated or limited details provided.	Boosted NNs appear to achieve the highest accuracy overall on test samples (classification success was 88.8%).
Karas and Režňák-ová (2014)	International Journal of Mathematical Models, Volume 8, 2014, pp214–223.	1908 manufacturing firms within the Czech Republic sampled over the 2004–2011.	Private firms.	Bankruptcy under the laws of the Czech Republic.	34 financial variables/-ratios.	AdaBoost vs. MDA	N/A	Tree depth of 107 trees. Statistical 10 software used.	AdaBoost significantly outperformed the LDA model. AdaBoost had a Type II misclassification rate of 1.48% and Type I misclassification rate of 15.88%. Accuracy differed across models depending on how many input variables were used.
Olson et al. (2012)	Decision Support Systems 52 (2012) 464–473.	A total U.S. sample of 1321 firm years sampled over the period 2005–2011. Study includes 100 bankrupt firms.	Public firms.	Not discussed but assumed to be Chapter 11.	19 financial variables/-ratios.	Decision trees vs. NNs and SVMs.	N/A	Limited discus-sion. WEKA software.	Decision trees performed better than SVMs and NNs. However boosted trees not specifically considered. Best AUC reported was .947.

Table 12 (continued)

Author/s	Journal	Sample Size	Public/ Private Firm Sample	Definition of distress	Number/Type of Input Variables	Boosting Model Used	Interaction Effects	Model Architecture & software details	Prediction/AUC
Kim and Upneja (2014)	Economic Modelling 36 (2014) 354–362.	A U.S. sample of 142 publicly traded restaurant firms sampled over the period 1988–2010.	Public firms.	Uses the Zmijewski score as a dependent variable.	25 mostly financial variables/-ratios.	AdaBoost vs. decision trees	N/A	Not stated or limited details provided.	AdaBoost outperformed decision tree models, with an AUC of 0.988 on the full model, an AUC of 0.969 on the full service model and an AUC of 0.94 on the limited service model.
West et al. (2005)	Computers & Operations Research 32 (2005) 2543–2559.	329 observations (93 bankrupt firm years and 236 healthy firms).	Public firms.	Not clear but assumed to be Chapter 11.	Altman's five key ratios.	Boosting ensemble	N/A	Not stated or limited details provided.	For the bankruptcy dataset, the bagging ensemble had the lowest test error of 0.126 compared to 0.127 for boosting and 0.129 for the CV ensemble.
Alfaro et al. (2008)	Decision Support Systems 45 (2008) 110–122	590 failed and 590 active firms with accounts registered on the Spanish Mercantile Registry.	Private firms.	Bankruptcy and temporary receivership) during the period 2000–2003	16 measures including 13 financial variables/-ratios.	Compares AdaBoost with CART, NNs, and MDA.	N/A	Limited details provided; R software used.	AdaBoost outpeformed NNs, with a test error rate of 8.898%.
Virág and Nyitrai (2014)	Financial & Economic Review, Vol 13 (4), 2014: 178–193.	976 Hungarian firms (51% were solvent and 49% insolvent) sampled over the period 2001 and 2012.	Not clear from the study.	Bankruptcy or winding-up proceedings reported on the Hungari-an Trade Register.	17 financial variables/-ratios.	Comparison of boosting and bagging, using the C4.5 method.	N/A	Not stated or limited details provided.	Bagging slightly out performed boosting (by about 1%). Both models achieved accuracy rates of around 80%.
Heo and Yang (2014)	Applied Soft Computing 24 (2014) 494–499	A total of 1381 Korean construction bankruptcies and 28,481 normal construction firms sampled	Small and medium size private firms.	Firms that went into workout, receivership, or bankruptcy.	12 financial variables/-ratios.	Compares AdaBoost with NNs, SVM, D-Tree, and	N/A	Not stated.	AdaBoost performed the best overall with the highest classification success on the large firm partition of the sample (93.8% accurate). However, AdaBoost performance was more

Table 12 (continued)

Author/s	Journal	Sample Size	Public/ Private Firm Sample	Definition of distress	Number/Type of Input Variables	Boosting Model Used	Interaction Effects	Model Architecture & software details	Prediction/AUC
		over the period 2008 to 2012.				Altman Z scores			modest across the whole sample (78.5% accurate) but still performed better than the other models.
Tsai et al. (2014)	Applied Soft Computing 24 (2014) 977–984	Three datasets with bad vs. good firms in the following ratios: Australian dataset (307/383) German dataset (700/300), and Japanese dataset (307/383)	Not clear from the study.	Not defined.	Up to 20 variables/-ratios.	Compares boosted and bagged SVMs, NNs, and MLPs.	N/A	Limited details provided. WEKA software.	Boosted decision tree ensembles perform best and outperformed other classifier ensembles Including individual classifiers. While predictive accuracy differed across datasets, best performance was achieved on the Australian and Japanese datasets (classification accuracy around 86% overall).
Kim and Kang (2015)	Expert Systems with Applications 42 (2015) 1074–1082	500 bankrupt Korean manufacturing firms and 2500 nonbankrupt manufacturing firms sampled over 2002 –2005.	Private companies obtained from a Korean commercial bank.	Not stated, but assumed to be legal bankruptcy in Korea.	30 financial variables/-ratios.	Compares geometric mean based boosting			(GMBoost) with AdaBoost.
N/A	Not stated or limited details provided.	GMBoost shows monotonically superior performance to AdaBoost when samples are unbalanced.							

References

Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems, 45*, 110–122.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589–609.

Altman, E. I. (2002). *Bankruptcy, credit risk, and high yield junk bonds*. Malden: Blackwell Publishers.

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance, 1*(1), 29–54.

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance, 18*(3), 505–529.

Ashbaugh, H., Collins, D. W., & LaFond, R. (2006). The effects of corporate governance on firms' credit ratings. *Journal of Accounting and Economics, 52*(2), 737–783.

Asquith, P., Gertner, R., & Scharfstein, D. (1994). Anatomy of financial distress: An examination of junk-bond issuers. *The Quarterly Journal of Economics, 109*(3), 625–658.

Bangia, A., Diebold, F., Kronimus, F. X., Schagen, A., Schuermann, T., & T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance, 26*, 445–474.

Barth, M., Beaver, W., & Landsman, W. (1998). Relative valuation roles of equity book value and net income as a function of financial health. *Journal of Accounting and Economics, 25*, 1–34.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 71–111.

Beaver, W. H., McNichols, M. F., & Rhie, J. W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies, 10*, 93–122.

Bharath, S., & Shumway, T. (2008). Forecasting default with the Merton distance to default model. *Review of Financial Studies, 21*(3), 1339–1369.

Billings, B. (1999). Revisiting the relation between the default risk of debt and the earnings response coefficient. *Accounting Review, 74*(4), 509–522.

Black, N. S. (1998). In P. Newman (Ed.), "*shareholder activism and corporate governance in the United States*" *in The New Palgrave Dictionary of Economics and the Law*. New York: Stockton Press.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 26*, 123–140.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth statistics/probability series: CRC Press.

Chandra, D. K., Ravi, R., & Bose, L. (2009). Failure prediction of dotcom companies using hybrid intelligent techniques. *Expert Systems with Applications, 36*, 4830–4837.

Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance, 8*, 537–569.

Clarke, J., Ferris, S. P., Jayaraman, N., & Lee, J. (2006). Are analyst recommendations biased? Evidence from corporate bankruptcies. *Journal of Financial and Quantitative Analysis, 41*, 169–196.

Cortés, E. A., Martínez, M. G., & Rubio, N. G. (2007). A boosting approach for corporate failure prediction. *Applied Intelligence, 27*(1), 29–37.

Daily, C. M., & Dalton, D. R. (1994). Bankruptcy and corporate governance: The impact of Board composition and structure. *The Academy of Management Journal, 37*(6), 1603–1617.

DeAngelo, H., & DeAngelo, L. (1990). Dividend policy and financial distress: An empirical investigation of troubled NYSE firms. *The Journal of Finance, 45*(5), 1415–1431.

Dichev, I. D. (1998). Is the risk of bankruptcy a systematic risk? *The Journal of Finance, 53*(3), 1131–1147.

Duffie, D., & Singleton, K. J. (2003). *Credit risk, pricing, measurements, and management*. Princeton: Princeton University Press.

Ederington, L. H., & Goh, J. C. (1998). The bond agency ratings and stock analysts: Who knows what when? *Journal of Financial and Quantitative Analysis, 33*, 569–585.

Fedorova, E., Gilenko, E., & Dovzhenkon, S. (2013). Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications, 40*, 7285–7293.

Figlewski, S., Frydman, H., & Liang, W. (2012). Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics and Finance, 21*(1), 87–105.

Foster, G. (1986). *Financial statement analysis* (2nd ed.). Englewood Cliffs: Prentice Hall.

Freedman, D.A. (2010). *Statistical Methods and Causal Inference: A Dialogue with the Social Sciences*, edited by collier, D., J.S., Sekhon, and P.B. Stark. Cambridge University Press, New York.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232.

Friedman, J., & Meulman, J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine, 22*(9), 1365–1381.

Frino, A., Jones, S., Wong, J., & J. (2007). Market behaviour around bankruptcy announcements: Evidence from the Australian stock Exchange. *Accounting and Finance, 47*(4), 713–730.

Frino, A., Jones, S., Lepone, A., & Wong, J. (2014). Market behavior of institutional investors around bankruptcy announcements. *Journal of Business Finance and Accounting, 41*, 270–295.

Harrell, F. (2001). *Regression modeling strategies with applications to linear models, logistic and ordinal regression and survival analysis*. New York: Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (2nd ed.). New York: Springer.

Hensher, D. A., Jones, S., & Greene, W. H. (2007). An error component logit analysis of corporate bankruptcy and insolvency. *The Economic Record, 83*(260), 86–103.

Heo, J., & Yang, J. W. (2014). AdaBoost based bankruptcy forecasting of Korean construction companies. *Applied Soft Computing, 24*, 494–499.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies, 9*(1), 5–34.

Hung, C., & Chen, J. H. (2009). A selective ensemble based on expected probabilities for bankruptcy prediction. *Expert Systems with Applications, 36*, 5297–5303.

Jones, S. (2011). Does the capitalization of intangible assets increase the predictability of corporate failure. *Accounting Horizons, 1*(25), 41–70.

Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: A mixed logit model. *Accounting Review, 79*(4), 1011–1038.

Jones, S., & Hensher, D. A. (2008). *Advances in credit risk modelling and corporate bankruptcy prediction*. Cambridge: Cambridge University Press.

Jones, S., & Johnstone, D. (2012). Analyst recommendations, earnings forecasts and corporate bankruptcy: Recent evidence. *Journal of Behavioral Finance, 13*(4), 281–298.

Jones, S., Johnstone, D., & Wilson, R. (2015). An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. *Journal of Banking & Finance, 56*(7), 72–85.

Jones, S., Johnstone, D., & Wilson, R. (2017). Predicting corporate bankruptcy: An evaluation of alternative statistical models. *Journal of Business Finance and Accounting, 44*(1–2), 3–34.

Karas, M., & Režňáková, R. (2014). A parametric or nonparametric approach for creating a new bankruptcy prediction model: The evidence from the Czech Republic. *International Journal of Mathematical Models, 8*, 214-223.

Kealhofer, S., and M. Kurbat. The default prediction power of the Merton approach, relative to debt ratings and accounting variables. Moody's KMV, 2001.

Keenan, S. C., Sobehart, J., & Hamilton, D. T. (1999). *Predicting default rates: A forecasting model for moody's issuer-based default rates*. Special Comment: Moody's Investors Service.

Kim, M. J., & Kang, D. K. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications, 37*, 3373–3379.

Kim, M. J., & Kang, D. K. (2015). Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction. *Expert Systems with Applications, 42*, 1074–1082.

Kim, S. Y., & Upneja, A. (2014). Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Economic Modelling, 36*, 354–362.

Koopman, S. J., Kräussl, R., Lucas, A., & Monteiro, A. B. (2009). Credit cycles and macro fundamentals. *Journal of Empirical Finance, 16*(1), 42–54.

Koopman, S. J., Lucas, A., & Schwaab, B. (2011). Modeling frailty-correlated defaults using many macroeconomic covariates. *Journal of Econometrics, 162*(2), 312–325.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109–131.

Olson, D. L., Delen, D., & Mengm, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems, 52*, 464–473.

Precourt, E., and H.R. Oppenheimer (2013). "What do institutional investors know and act on before almost everyone Else: Evidence from corporate bankruptcies." Seminar Paper at University of Rhode Island, Fall 2013.

Ramalingegowda, S. (2014). Evidence from impending bankrupt firms that long horizon institutional investors are informed about future firm value. *Review of Accounting Studies, 19*(2), 1009–1045.

Salford Systems. (2015). Introduction to TreeNet. San Diego.

Schapire, R., & Freund, Y. (2012). *Boosting: Foundations and algorithms*. USA: MIT Press.
Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. *The Journal of Finance, 52*, 737–783.
Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business, 74*(1), 101–124.
Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review, 71*(3), 289–315.
Sun, J., Jia, M., & Hui, L. (2011). AdaBoost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies. *Expert Systems with Applications, 38*(8), 9305–9312.
Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Better decisions through science. *Scientific American, 283*, 82–87.
Tsai, F., Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing, 24*, 977–984.
Virág, M., & Nyitrai, T. (2014). The application of ensemble methods in forecasting bankruptcy. *Financial & Economic Review, 13*(4), 178–193.
Wang, G., Ma, J., & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications, 41*, 2353–2361.
West, D., Dellan, S., & Qia, J. (2005). Neural network ensemble strategies for financial decision applications. *Computers & Operations Research, 32*, 2543–2559.
White, H. (2000). A reality check for data snooping. *Econometrica, 68*(5), 1097–1126.
Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research, 22*, 59–82.