Financial statement fraud detection using supervised learning methods

by

Adrian C B Gepp

June 2015

Submitted in total fulfilment
of the requirements of the
degree of **Doctor of Philosophy**
at the **Faculty of Business, Bond University**.

Abstract

Famous frauds such as at Enron, WorldCom and HealthSouth are potent reminders that the detection of financial statement fraud needs to be improved. Studies estimate the median loss from a single financial statement fraud scheme to be at least one million US dollars. The annual cost of financial statement fraud exceeds 1.2 trillion US dollars worldwide and 377 billion dollars in the US (ACFE, 2014).

Many business decisions rely on the accuracy of financial statements, but resources are not available to comprehensively investigate all of them. Moreover, detection of fraud in financial statements is difficult. Consequently, there is a need for better decision aids such as detection models developed using supervised learning methods. Standard parametric regression-based techniques, particularly logistic regression, have been extensively studied for detecting financial statement fraud. More investigation is needed into non-parametric techniques such as decision trees and ensemble techniques that combine multiple models such as bagging and boosting. Using data about companies listed on US stock exchanges, multiple statistical modelling techniques new to the field are compared with established techniques for detecting this type of fraud. Comparisons are made using a range of ratios for the cost of failing to detect fraud relative to the cost of falsely alleging it, as these costs differ depending on the stakeholder. Newly developed ensemble models that include decision-tree based techniques performed particularly well.

A large number of potential indicators (explanatory variables) of financial statement fraud are investigated in order to study which are the most useful to detection models. These include financial information, non-financial information and comparisons of the two. Empirical support has been found for both financial and non-financial explanatory variables, including new variables. A new framework, the Fraud Detection Triangle, is also developed to assist in the selection of explanatory variables for financial statement fraud detection models. Empirical evidence is provided to support the use of this new framework.

Using models developed in this research, financial statements can be automatically classified as either fraudulent or legitimate, as well as being ranked according to their likelihood of being fraudulent. This information can be used to improve early detection, which would mitigate the costs of fraud and help deter it from occurring by increasing the probability of being detected. Beneficiaries of this information include auditors, investors, financiers, employees, customers, suppliers, regulators, company directors and the financial markets as a whole through improved integrity and allocation of resources.

Declaration

This thesis is submitted to Bond University in fulfilment of the requirements of the degree of Doctor of Philosophy. This thesis represents my own original work towards this research degree and contains no material which has been previously submitted for a degree or diploma at this University or any other institution, except where due acknowledgement is made.

Adrian C B Gepp

Date: 19 June 2015

Acknowledgements

I would first like to thank my primary supervisor Professor Kuldeep Kumar for his advice, guidance and support, all of which has been extremely helpful throughout my Ph.D. studies. Thank you also for being one of the first, along with Steve Sugden and Phil Stocks, to start me on the path of discovery through academic research. I would also like to thank my other supervisor Dr Sukanto Bhattacharya for his valuable feedback and encouragement.

I am grateful for the scholarship awarded to me by Bond University and the support of the people who work there. I am particularly grateful to Professor Keitha Dunstan for her personal support and for providing me with many opportunities to grow as a professional, through attending and presenting at both industry and academic conferences and seminars.

Thank you to James, Becky, Chris and Anne for being such understanding and wonderful friends throughout this adventure. I also want to acknowledge and thank the numerous other people who have greatly influenced who I am today and will become tomorrow; Gary, Tim and Geoffrey, you are at the top of that list.

To Peter and Vesna, no words can accurately express my gratitude for the advice, friendship and support that you have given, and continue to give, me.

To my parents, thank you, now and always, for your unconditional love, honest advice and your unwavering belief in me.

He who is strong conquers others. He who is mighty conquers himself.

–Shawn Michaels, adapted from Lao Tzu

Table of Contents

List of Figures

List of Tables

Abbreviations, Model Codes and Variable Identifiers

Abbreviation	Full Description
AAERs	Accounting and Auditing Enforcement Releases
ACFE	Association of Certified Fraud Examiners
AFDA	Allowance for Doubtful Accounts
AICPA	American Institute of Certified Public Accountants
AMEX	American Stock Exchange
Bagging	Bootstrap Aggregation
CART	Classification and Regression Trees
CEO	Chief Executive Officer
CFO	Chief Financial Officer
DA	Discriminant Analysis
DT	Decision Tree
E-Index	A corporate governance index developed by Bebchuk et al. (2009)
EPS	Earnings Per Share
FIN	Financial Assets, a component of RSST accruals
G-Index	A corporate governance index developed by Gompers et al. (2003)
GDP	Gross Domestic Product
GFC	Global Financial Crisis
I	Pressure/Incentive factor within the new Fraud Detection Triangle framework
ID3	Iterative Dichotomiser 3, a type of decision tree that has led to newer See4.5 and See5 decision trees
LIFO	Last-In, First-Out inventory valuation method
LR	Logistic Regression
LR_MS	Multi-stage Logistic Regression
MARLEDA	Markovian Learning Estimation of Distribution Algorithm
MICE	Money, Ideology, Coercion and Ego or Entitlement
NASDAQ	A stock exchange in the US
NCO	Non-Current Operating assets, a component of RSST accruals
NN	Artificial Neural Network
NYSE	New York Stock Exchange

Abbreviation	Full Description
O	Exploitable Opportunity factor within the new Fraud Detection Triangle framework
OTC markets	Over-The-Counter markets
PARCIF	Prior-adjusted Relative Cost of Missing Fraud
PP&E	Property, Plant and Equipment (a category of asset)
R	Integrity/Attitude/Rationalisation factor within the new Fraud Detection Triangle framework
RF	Random Forests
ROC curve	Receiver Operating Characteristic curve
ROE	Return on Equity
RQ1	Research Question 1 (defined on page 6)
RQ2	Research Question 2 (defined on page 7)
RQ3	Research Question 3 (defined on page 7)
RQ4	Research Question 4 (defined on page 7)
RSST accruals	Unadjusted accruals as developed by Richardson et al. (2005)
S	New Suspicious Information factor within the new Fraud Detection Triangle framework
SAS	Statement on Auditing Standards
SEC	US Securities and Exchange Commission
See4.5	A type of decision tree based on ID3 that has led to the newer See5 decision trees
See5	A type of decision tree, based on See4.5 and ID3
SIC Code	Standard Industrial Classification code
SPM	Salford Predictive Modeler
TN	TreeNet
US	United States of America
WC	Working Capital, a component of RSST accruals
WEC	Weighted Error Cost
Z-score	Financial distress measure developed by Altman (1968)

Model Code	Model Details
Benchmark Models as estimated in prior research	
M-score	Probit analysis model developed by Beneish (1997, 1999a)
F-score	Stepwise logistic regression model developed by Dechow et al. (2011)
Discriminant Analysis (DA) and Logistic Regression (LR)	
DA_All	Inclusion of all variables (except the missing value exclusions)
DA_U15%	Only variables statistically significant at the 15% level in univariate tests
DA_U15%LR	Only variables statistically significant at the 15% level in the LR_U15% model
DA_U1	Only the most statistically significant variable from each main variable
DA_Step	Stepwise variable selection
LR_All	Inclusion of all variables (except the missing value exclusions)
LR_U15%	Only variables statistically significant at the 15% level in univariate tests
LR_U15%LR	Only variables statistically significant at the 15% level in the LR_U15% model
LR_U1	Only the most statistically significant variable from each main variable
LR_Step	Stepwise variable selection
LR_MS_S	Multi-stage LR model using the overall schema of explanatory variables
LR_MS_F	Multi-stage LR model using the new Fraud Detection Triangle framework
Artificial Neural Network (NN)	
NN_BK	Backpropagation neural network with 1 hidden layer containing 4 neurons
NN_GA_1	Backpropagation neural network optimised by a genetic algorithm with a minimum learning rate of 0.1 (NN_GA_1) or 0.5 (NN_GA_5)
NN_GA_5	
Decision Trees (DTs) and Ensembles of them: Random Forests (RF) and TreeNet (TN)	
DT_Suite	Suite of cost-sensitive CART decision trees, one for each value of $PaRC_{IF}$
DT_One	One cost-insensitive CART decision tree used for all values of $PaRC_{IF}$
RF_8	Random Forests with 1000 trees and variable subset size of 8
RF_66	Random Forests with 1000 trees and variable subset size of 66
TN	TreeNet with 0.01 learning rate and maximum number of nodes per tree of 12
Ensembles involving multiple modelling techniques	
NN_LR	NN_BK classifications included as an additional explanatory variable in LR_All
DT_LR	DT_One classifications included as an additional explanatory variable in LR_All
DTnode_LR_Step	Terminal node assignments from DT_One included as additional explanatory variables in the LR_Step model
NN-DTnode_LR_Step	NN_BK classifications included as an additional explanatory variable in the DTnode_LR_Step model

Model Code	**Model Details**
Vote5	Majority Vote between DT_One, TN, RF_8, NN_BK and LR_All
Vote3_RF_TN_DT	Majority Vote between TN, RF_8 and DT_One
Vote3_RF_TN_DA	Majority Vote between TN, RF_8 and DA_All
Vote3_RF_TN_NN	Majority Vote between TN, RF_8 and NN_BK
AV5_NoNN	Average between DT_One, TN, RF_8, LR_All and DA_All
AV2_RF_TN	Average between TN and RF_8
AV3_RF_TN_DT	Average between TN, RF_8 and DT_One
DT_One_DA	Discriminant analysis (based on DA_All) performed as a second stage on the riskier classifications from DT_One
Simpler Models	
TN_90%_6	TreeNet using the six variables with a minimum importance score of 90% in the TN model
TN_4	TreeNet using four variables; the most important (according to the TN model) from each of the four factors of the new Fraud Detection Triangle framework
LR_Step_11	Logistic regression using 11 of the 14 variables in the LR_Step model
LR_Int_11	Logistic regression using 11 variables including an interaction term
LR_TN_6	Logistic regression using the six variables with a minimum importance score of 90% in the TN model

Variable Identifier	Variable Name
V1 V1a V1b V1c	Accounts Receivable Value for the specified year Percentage change Was Percentage change > 10%?
V2	Percentage change in Accounts Receivable to Sales
V3	Percentage change in Accounts Receivable to Total Assets
V4	Percentage change in AFDA
V5	Percentage change in AFDA to Accounts Receivable
V6	Percentage change in AFDA to Sales
V7	Change in Inventory to average Total Assets
V8 V8a V8b	Inventory to Sales Value for the specified year Change
V9	Was Last-In, First-Out (LIFO) inventory valuation used?
V10 V10a V10b V10c V10d V10e	Sales Growth Percentage change V10a minus the Industry Average Previous year's Percentage change Four-year growth rate Previous year's percentage change in total assets
V11 V11a V11b	Sales to Total Assets Value for the specified year Percentage change
V12 V12a V12b	Gross Margin to Sales Percentage change Was percentage change > 10%?
V13 V13a V13b	Cash Sales Percentage change Was change < 0?
V14	Were any sales from acquisitions?
V15	Current Assets to Total Assets
V16	Net Property Plant & Equipment (PP&E) to Total Assets
V17	Soft Assets to Total Assets
V18	Percentage Change in Assets other than Current Assets and Net PP&E to Total Assets
V19	Total Accruals to Total Assets

Variable Identifier	**Variable Name**
V20	Were the specified and the prior year's Total Accruals > 0?
V21	Total Discretionary Accruals
V22	RSST (unadjusted) Accruals
V23	Debt to Total Assets
V24	Debt to Equity
V25	Altman's (1968) financial distress measure (Z-score)
V26	Four-period average of Times Interest Earned
V27 V27a V27b	Return on Equity Value for the specified year Industry Average minus Specific Company
V28 V28a V28b V28c	Return on Average Prior Assets Value for the specified year Previous year Change
V29 V29a V29b	Holding Period Return One-year Previous One-year
V30	Were analyst Earnings Per Share (EPS) forecasts achieved or exceeded?
V31 V31a V31b	Were New Securities issued? Common Stock? Common Stock or Long-term Debt?
V32	Proportion of common stock that is newly issued
V33 V33a V33b V33c	Demand for financing Specific Value (ex ante) Was there demand (ex ante)? Cash from operating and investment activities
V34	Were there operating leases?
V35	Was the auditor a Big Six firm?
V36	Number of auditor changes in the most recent four financial statements
V37 V37a V37b	CEO Tenure Number of changes in the last three years
V38	Has the CFO changed in the last three years?

Variable Identifier	Variable Name
V39 V39a V39b V39c	Composition/Holdings of the Board Number of Directors Percentage of Directors who are also Executives Percentage of Director shares owned by those who are also Executives
V40	Percentage of total shares owned by the CEO
V41	Percentage change in the number of Employees minus percentage change in Total Assets
V42	Percentage change in Sales minus percentage change in the number of Employees
V43	Percentage Change in Sales to Employees: Specific Company minus Industry Average
V44	Company Age: Number of years since foundation
V45	Company Size: natural log of Total Assets
V46	Industry: Standard Industrial Classification (SIC) code starts with a 3?
V47 V47a V47b	Stock Exchange listed on NASDAQ stock exchange? New York Stock Exchange (NYSE)? (No to both V47a and V47b indicates American Stock Exchange [AMEX])
V48 V48a V48b V48c	Macroeconomic indicators Previous year's percentage change in annual real GDP Previous year's percentage change in annual retail sales Previous year's unemployment rate inverted
V49 V49a V49b	Corporate governance indices G-Index E-Index
V50	Accounting complexity of the industry

Nobody, as long as they move about among the chaotic currents of life, is without influence from fraud.

- modified from Carl Jung

Chapter 1 Introduction

Fraud, in general terms, is an act of intentional deception to gain a benefit or to cause a loss to another party. While it is impossible to know the true cost of fraud because of the inherent concealment involved, in a bid to make a reliable estimate the Association of Certified Fraud Examiners (ACFE) recently concluded that fraud in business costs the world nearly 3.7 trillion US dollars each year (ACFE 2014). This was based on an analysis of 1,483 fraud cases that were investigated by experienced Certified Fraud Examiners. It is unquestionably an alarming figure that must not be ignored. It also provides justification for more research on the early detection of fraud before its cost to society rises more.

Financial statement fraud involves the intentional publication of false or misleading information in financial statements. It is also known as *management fraud* most likely because it is commonly perpetrated by managers trying to overstate a business's financial profitability or viability. In their 2014 publication, the ACFE (2014) reported that financial statement fraud now occurs in 9% of the cases they studied, and that this figure has progressively increased from 4.8% in 2010 to 7.6% in 2012. Although it is more common to encounter asset misappropriation or corruption, the cost of a financial statement fraud scheme is much higher. The median loss from an individual financial statement fraud scheme was reported to be one million US dollars by the ACFE (2014). A much higher estimate was found in a ten-year study (Beasley et al. 2010) commissioned by a private sector committee specifically formed to study financial statement fraud in the US. They found that the median loss from financial statement fraud prosecuted by the US Securities and Exchange

Commission (SEC) was 12.1 million dollars, which is nearly three times larger than was found by the same committee ten years earlier. Using the worldwide ACFE figures[1], the annual cost of financial statement fraud is estimated to be more than 1.2 trillion US dollars worldwide and more than 377 billion dollars in the US.

Waste Management, Sunbeam, Tyco, Enron, WorldCom and other cases of financial statement fraud in the US have made news headlines around the world, stark reminders of how costly it can be. The frauds by Harris Scarfe in Australia, Parmalat in Italy, Ahold in The Netherlands, Satyam in India and Vivendi in France further demonstrate that the damage from this problem is felt around the world. Financial statement fraud is not a victimless crime, but rather leaves behind a very real financial loss and victims that include investors, financiers, employees and other stakeholders. There are also wider costs to society such as a reduction in confidence and trust in regulators, and also a reduction in the integrity of financial markets. This results in higher transaction costs and reduced efficiency (Perols 2011).

Multiple high-profile fraud cases over the last decade have left legislators, regulators, practitioners (accountants and auditors) and academics searching for answers (Erickson et al. 2006). This fact probably contributed to a review paper, in which Ngai et al. (2011) identified that research into corporate fraud (which includes financial statement fraud) was prevalent compared with other types of financial fraud. Despite this research, anti-fraud legislation such as the 2002 Sarbanes-Oxley Act in the United States, changes to auditing standards and enforcement efforts, the risk of financial statement fraud remains substantial (Silver et al. 2008) and is a public concern (Deloitte 2009). Furthermore, the magnitude of the problem has increased (Beasley et al. 2010) and it appears that the penalties facing perpetrators of fraud are not a sufficient deterrent (Beasley et al. 2010; Partnoy 2010).

Humpherys et al. (2011) state that there is a need for better decision aids to help detect financial statement fraud because research has shown that human beings have only a slightly better than random chance ability at detecting deception. Worryingly, the authors go on to state that meeting this need for better aids is critical because most external auditors do not have a lot of experience in detecting fraud. The ACFE found that audits are not very

[1] The ACFE (2014) report contains estimates of the median cost and relative frequency per fraud category and per region. The cost of financial statement fraud was then estimated assuming that the relative differences among the mean costs are similar to differences in median costs. This method was also used by Perols (2011).

effective at detecting fraud as less than 19% of occupational fraud[2] was initially detected by either internal or external audit (ACFE 2010, 2012, 2014). This has occurred even after the American Institute of Certified Public Accountants (AICPA) in the US explicitly clarified the responsibility of auditors to provide reasonable assurance that financial statements are free from material misstatement (Perols 2011). Fraud detection decision aids also need to be made available to stakeholders other than auditors, because Beasley et al. (2010) found that external auditors were implicated in the fraud in 23% of the SEC-enforced cases between 1998 and 2007.

Classification models can be used as decision aids to assist in detecting financial statement fraud. These models assign classifications (fraudulent or legitimate) by analysing information such as publicly available financial and accounting ratios derived from data in financial statements. Because the financial information in cases of fraud is, by definition, fraudulent, non-financial information is also routinely incorporated into models. While financial statement fraud is less common than other types of fraud, models that can act as a first test to identify firms warranting more investigation are useful (Dechow et al. 2011). Accurate classification models can reduce the costs of detecting financial statement fraud by facilitating better directed investigations. This would reduce the loss from financial statement fraud by two separate means, first as a result of faster detection and secondly the factor of deterrence to potential fraudsters caused by the faster detection. It is useful to note that fraudsters will generally not be deterred by harsher penalties because they usually believe they will never be caught (ACFE 2013a). Consequently, increasing the perception of the likelihood of being caught is the primary deterrent (ACFE 2013a) and classification models and education about the use of them can assist in increasing this perception.

1.1 Introduction to fraud detection models and their benefits

Specialists in this area, such as forensic accountants, are a scarce and expensive resource so it is impracticable to have them investigate all financial statements. Classification models can be used as a first step to quickly process large amounts of data, find patterns and then produce a predicted classification that can be used to rank cases in terms of the

[2] Occupational frauds occur when an employee abuses the trust of an employer, and consequently they include asset misappropriation, corruption and financial statement fraud (ACFE 2014).

likelihood of fraud. These results can then be used to prioritise and more efficiently allocate human specialists to investigate cases with the highest likelihood of fraud. Unlike computer models that might consider probabilities, human specialists can investigate with a view to establishing proof. As suggested by Belhadji et al. (2000) as a general modelling principle, fraud detection models are to complement, not replace, human specialists.

It is also important to note that static classification models developed using supervised learning methods are inferior to humans at adapting to new events such as new ways to commit fraud. Care must also be taken to avoid fraud models that are too rigid and predictable as they allow fraudsters to learn how to conceal their behaviour over time. Consequently, to maintain initial accuracy levels, models implemented in industry need to be regularly updated with new information.

Fraud detection models must pay particular attention to minimising both types of errors resulting from misclassifying financial statements.

1. Missing fraud, that is misclassifying fraudulent financial statements as legitimate, is costly in terms of stakeholders being misled into making decisions using fraudulent information.
2. Falsely alleging fraud, that is misclassifying legitimate financial statements as fraudulent, is also costly in terms of lost investment opportunities, wasted investigatory and regulatory time and money, and damage to the misclassified businesses.

While it is agreed that the first error is more costly than the second, a quantifiable difference in the error costs has not been, and is unlikely to be, agreed upon since the costs vary depending on the particular conditions and stakeholders (Beneish 1999a). Consequently, research considering a range of error costs is desirable.

The exact proportion of fraudulent statements relative to all financial statements is also unknown, because frauds that have gone undetected remain unidentified. Some studies have estimated the proportion of fraudulent statements to be less than 1% (Beneish 1997; Bell and Carcello 2000; Dechow et al. 2011), but recent estimates are as high as 14.5% (Dyck et al. 2013). All the estimates indicate that the proportions of fraudulent and legitimate statements are drastically unequal, which presents challenges for classification models and is often referred to as the “class imbalance” problem. Consequently, it is important to assess models using metrics that take the proportion of fraud into account.

In addition to the importance of considering class imbalance and unequal error costs, Perols (2011) states that the noisy nature of the explanatory variables (data) and inherent concealment involved means that empirical research using data specific to financial statement fraud is necessary, because results from other fields cannot be assumed to be applicable. And while the similar area of business failure prediction has been well-researched, fewer empirical studies have focused on financial statement fraud detection.

1.2 Introduction to supervised learning methods for classification

Supervised learning methods attempt to use existing data to learn the relationship between input variables and an output variable, a process referred to as training. Thus, supervised learning methods can be used to train classification models on past (or simulated) known data before being used on new data. The name of the method can be better understood when it is viewed as a conceptual supervisor providing the known data for the training process. For fraud detection, the output variable is often a binary variable that indicates either fraudulence or legitimacy. Supervised learning methods train models using both fraudulent and legitimate cases.

There are many different modelling techniques that use supervised learning methods. Discriminant analysis and logistic regression have traditionally been popular modelling techniques for analysing classification problems, such as financial statement fraud detection. These traditional modelling techniques are parametric, and there is now a search for new, more accurate and less distribution-dependent approaches. Potentially suitable non-parametric modelling techniques include artificial neural networks, decision trees and ensembles of multiple decision trees such as Random Forests and TreeNet. To date, decision trees and ensembles of them have been the subject of very few research studies on financial statement fraud detection.

Major advantages of decision trees include that they are non-parametric, immune to outliers, resistant to irrelevant variables, can easily model interactions between explanatory variables and are relatively easy to interpret and simple to develop into automated systems. Decision tree models suffer from being sensitive to small changes in the training data set (Sudjianto et al. 2010) similar to artificial neural networks, but newer ensemble techniques such as TreeNet overcome this issue to provide more stable models that also make better classifications close to region boundaries. Salford Systems are the providers of CART decision trees (Breiman et al. 1984), TreeNet (Friedman 1999) and Random Forests (Breiman 2001a). They have had unparalleled success at data mining competitions and credit TreeNet for the majority of their award-winning results[3].

1.3 Aim of this research

The overall aims of this research are to advance the field of detecting financial statement fraud in terms of:

- Identifying the most appropriate modelling techniques and parameter settings;
- Developing and evaluating new models in comparison with existing models; and,
- Identifying the most appropriate explanatory variables (data) for models to use.

This research also has the potential to develop new theory and methodology in order to achieve these aims. Additionally, it is the aim of this research to produce findings that are widely applicable to investors, regulators, auditors and other stakeholders.

As a consequence of these aims, this research has the potential to uncover new information about the nature of financial statement fraud. Insights from research on detecting financial statement fraud can also be used by others to refine general theories of deception (Humpherys et al. 2011). What happens in the years after financial statement fraud is detected is also an interesting and important research topic, but one that is outside the scope of this research[4].

The four main research questions that this study addresses are introduced below.

Research Question 1 (RQ1) Which supervised-learning modelling techniques are the most accurate at detecting financial statement fraud under varying assumptions about the prior probability of fraud and ratios of misclassification error costs?

As suggested above in Section 1.1, this research considers numerous ratios of misclassification error costs because the cost of missing fraud relative to falsely alleging fraud differs depending on the circumstance and the stakeholder. A range of prior fraud

[3] This information is from Salford Systems' website, https://www.salford-systems.com/en/company/awards and https://www.salford-systems.com/products/treenet, accessed January 2013.

[4] The article by Farber (2005) is a good reference on what happens after fraud occurs.

probabilities is also considered because estimates vary in the literature of the proportion of financial statements that are fraudulent.

The ratio of error costs and the prior fraud probability might not be able to be reliably estimated in some cases. The following research question addresses the issue of which modelling technique to use in such a circumstance.

Research Question 2 (RQ2): Which supervised-learning modelling technique is the best overall at detecting financial statement fraud, considering the entire range of assumptions investigated in RQ1?

It is also of interest to investigate the following related questions.

Research Question 3 (RQ3): Which explanatory variables are the most useful in models that detect financial statement fraud?

Research Question 4 (RQ4): How do simpler models compare with those developed for the first two research questions in their ability to detect financial statement fraud?

1.4 Main contributions of this research

The set of explanatory variables used in this research is more comprehensive than any prior studies. Further to this, the number of fraud cases analysed is often more than double that used in prior studies. Consequently, findings from this research greatly contribute to a better understanding of the variables that are important in models detecting financial statement fraud and how these variables are associated with it. Some of these findings that are new include:

- Empirical support for the use of a corporate governance index in models that detect financial statement fraud;
- Empirical support for increased debt being associated with reduced financial statement fraud, in accordance with agency theory (Jensen and Meckling 1976); and
- Empirical support for the use of an interaction variable between CEO tenure and company size in models that detect financial statement fraud. Theoretical justification for this interaction according to the new Fraud Detection Triangle framework is also provided.

This research also makes a theoretical contribution to the field of financial statement fraud. Despite the fact that the selection of variables is crucial to developing a fraud detection model, their selection in prior financial statement fraud detection research is not standardised by a common overall theory (Perols and Lougee 2011). Consequently, a new theoretical framework has been developed and proposed for this purpose. It is called the Fraud Detection Triangle framework and is based on the famous Fraud Triangle (Cressey 1953). Empirical evidence is also provided to support the use of this new framework to assist in the selection of explanatory variables for models that detect financial statement fraud.

Many supervised-learning modelling techniques, particularly decision trees and ensembles of them, have not been as extensively tested for financial statement fraud detection as they have been in this research. This is the first study to evaluate models both on separate holdout data for accurate evaluation of model performance and under varying assumptions about prior fraud probabilities and ratios of misclassification error costs. Consequently, results from this research greatly contribute to a better understanding of various models' performance in financial statement fraud detection. Some of these results are briefly overviewed below.

More than 30 different models have been developed, after many more were initially analysed in order to choose the best parameters. Some models are based on those used in prior studies, others are modifications of previously used models, and entirely new ones have also been developed. The best model to detect financial statement fraud varies depending on the assumptions about prior fraud probabilities and ratios of misclassification error costs. Overall, models that have been newly developed in this research have outperformed benchmark models and others based on those used in prior research. Newly developed ensemble models performed particularly well. Simpler models have also been developed for stakeholders who prefer models easier to interpret or which use fewer explanatory variables. It is important to note that all explanatory variables used in this research are publicly available, because variables too difficult to obtain are unlikely to be used in a practical context (Perols 2011).

Overall, new contributions are made to the current body of research on financial statement fraud. This information can be used to improve its early detection, which would mitigate its cost to society and help deter its occurrence. There are also specific beneficiaries of this information such as auditors and regulators, who can use detection models to assist them to better assess the risk of fraud having occurred and whether there is a need to investigate further. Additionally, investors, financiers and employees are examples of stakeholders who can use detection models to help them avoid prolonged association with fraudulent companies.

1.5 Dissertation Structure

The next chapter (Chapter 2) introduces and discusses the concepts of financial statements and fraud within them. This includes a discussion of the key entities involved in, and the different types of, financial statement fraud. This information is followed by a discussion of deterrence, detection and prevention of such fraud.

Following this overview of financial statement fraud, Chapter 3 presents a review of prior research into financial statement fraud detection in terms of the modelling techniques used. This includes an explanation of each modelling technique and a discussion of both its advantages and disadvantages.

The new Fraud Detection Triangle framework is developed in Chapter 4. It is also used in this chapter to assist in the selection of the explanatory variables to be used in this research. The justification of this selection incorporates a review of explanatory variables used in prior research, which complements the review of modelling techniques in Chapter 3.

Based on the review of modelling techniques and selection of explanatory variables in the prior two chapters, Chapter 5 presents the research that addresses the first two research questions (RQ1 and RQ2, defined on pages 6 and 7). The data and methodology being used are discussed first, including the development of new models. The chapter then concludes with presentation and discussion of the results from comparing multiple modelling techniques to detect financial statement fraud.

Chapter 6 presents the research that addresses the last two research questions (RQ3 and RQ4, defined on page 7). It begins with an analysis of the explanatory variables in terms of their usefulness in the models developed in Chapter 5. These results are then used, along with those from Chapter 5, to develop and evaluate simpler models that use fewer variables and are easier to interpret.

Finally, Chapter 7 presents the overall conclusions and suggested future work.

Chapter 2 Overview of Financial Statement Fraud

Fraud in general is any act to deceive in order to gain a benefit. It has existed as long as humans have and there are countless forms it can take (Bose et al. 2011); some examples include personal and corporate identity fraud, credit card fraud, insurance fraud and financial statement fraud, which is also known as management fraud. Based on information from the US Federal Bureau of Investigation, Ngai et al. (2011) classify financial fraud into four areas: bank fraud, insurance fraud, securities and commodities fraud and other related financial fraud. Financial statement fraud is classified as a type of corporate fraud, which is a member of the other related financial fraud category.

To better understand financial statement fraud, this chapter begins with a brief introduction into the concept of financial statements in accounting. This is followed by an introduction to fraud within financial statements and that provides the background information needed to discuss the detection, deterrence and prevention of that type of fraud.

2.1 Brief Introduction to Financial Statements

This subsection briefly introduces financial statements from an accounting viewpoint and some of the relevant terms used later in this dissertation. A more detailed introduction into financial statements can be found in the early chapters of most reputable textbooks on financial accounting such as the first three chapters of Trotman and Gibbins (2005) or Anthony et al. (2004). An alternative and even more comprehensive treatment of financial statements is provided in the book by Ittelson (2009), which is exclusively focused on this topic.

2.1.1 Accrual Accounting

Most corporations today use the accrual method of accounting. Accrual accounting recognises events when they occur, rather than cash accounting which recognises events at the time the cash transaction takes place. As an example, consider a retailer who sells and delivers a suite of furniture during a particular accounting period, but the customer has credit terms which means they will not pay until next accounting period. Accrual accounting recognises the sales revenue in the particular period when the sale took place, rather than the next period when the payment was made. Another example is a painting contractor, who completes work during one period but who will not be paid until the next period. This accrual expense is recognised in the period the contractor completed the work rather than in the following period when payment occurs. Apportioning the cost of a long-term asset over its useful life[5] (rather than entirely at the time of purchase) is a component of accrual accounting and is achieved by the concept known as depreciation, further explained in the next subsection. The reasoning behind apportioning the cost is to better match the expense of the asset with the revenue it will generate over multiple accounting periods. While accrual accounting is a more complex concept than cash accounting, it is arguably more economically relevant and more useful for decision-making. However, it will be seen later that this complexity also provides more opportunities for fraud to occur.

2.1.2 Main Types of Financial Statements

The four main financial statements are the:

- Balance Sheet[6],
- Income Statement[7],
- Cash Flow Statement, and the
- Statement of Changes in Equity[8].

These will be discussed briefly in this order. Different accounting standards have slightly different definitions, but the following explanations are general in nature to give a basic

[5] Useful life refers to the period over which an asset is expected to be available for use by a business (Trotman and Gibbins 2005).

[6] The Balance Sheet is called the Statement of Financial Position in some countries such as Australia.

[7] The Income Statement is also known as the Profit and Loss Statement, Statement of Operations and Statement of Earnings.

[8] The Statement of Changes in Equity is also known as the Statement of Retained Earnings.

understanding of the concepts to a reader who is unfamiliar with accounting language and standards.

The *Balance Sheet* describes the financial position of a business at a point in time. It is based on the model that *Assets = Liabilities + Equity*. Brief explanations of the three categories (assets, liabilities and equity) in this model and accounts commonly found within them are provided below.

- *Assets* are essentially useful financial resources, and more precisely assets are future economic benefits that include
    - *Current Assets*, which are expected to be realised (converted to cash) within the next operating cycle (usually one year in the normal course of business), such as
        - *Accounts Receivable*: Amount owing from customers, minus any amounts that are not expected to be collected that are shown in the *Allowance for Doubtful Accounts* contra account[9].
        - *Inventory*: Unsold products on hand (also known as stock).
    - *Noncurrent assets*, which are not expected to be fully realised in the next operating cycle (usually one year), such as
        - *Fixed Assets* or *Property, Plant and Equipment (PP&E)*: Assets such as land, buildings, vehicles, computers and office furniture.
            - PP&E is usually listed as Net PP&E, which is the Gross PP&E minus any *Accumulated Depreciation*. Assets often lose value over time as a result of wear and tear or becoming obsolete and accumulated depreciation is a measure of how much value an asset has lost. Accumulated depreciation can be calculated as the sum of the depreciation expense that occurs in each period (which is discussed more below).
        - *Intangible Assets* such as *Goodwill*, which arises when the price paid for a group of assets (such as in the acquisition of a company) is higher than the sum of the tangible assets.

[9] A contra account is an account used to record certain deductions from an asset

- *Liabilities* are essentially obligations to be paid, or more precisely future economic sacrifices of economic benefits that are obliged to be made by the business. Liabilities can be categorised as current (shorter-term) or noncurrent (longer-term), as was shown with the assets presented above. Examples of liabilities include
  - *Accounts payable (*a *current liability)*, which are amounts owed to suppliers (including the painting contractor from the example above), and
  - Debt agreements such as a *loan* (a *current* or *noncurrent liability*) or *bank overdraft* (a *current liability)*.
- *Equity* is the owners' interest in the business (which is assets minus liabilities), which can be derived from
  - *Retained earnings*, which are accumulated profits that have not been withdrawn from the company and distributed to owners and
  - *Direct financial contributions* from owners.

The *Income Statement* describes the financial performance and profitability over a given period of time (often one year). It is based on the relationship that *Income = Revenue - Expenses*, where income can also be referred to as profit or earnings. Brief explanations follow of the categories of this financial statement (revenues and expenses) and accounts commonly found within them. All of them are accruals that are recognised when they occur, not necessarily when payment is made.

- *Revenues* are inflows excluding contributions by owners and primarily include *Sales* from a product or service.
- *Expenses* are outflows excluding distributions paid to owners and include
  - *Cost of Goods Sold* such as the cost of the inputs to the product or service offered
  - *Interest* payable on debt
  - *Income Taxes*
  - *Salaries* of employees
  - *Depreciation* that recognises a decline in the value of a noncurrent asset[10]

[10] Depreciation is termed Amortisation when it is applied to intangible assets such as goodwill.

The *Cash Flow Statement* shows the changes in cash during a period. The cash flows are broken into the following categories

- *Operating* activities that relate to providing goods and services,
- *Investing* activities that relate to the purchase or sale of noncurrent assets such as PP&E, and
- *Financing* activities that relate to the financial structure of the company such as distributions paid to owners (which are called dividends).

Each category has inflows, outflows and net cash flows that are calculated as the inflows minus the outflows.

The *Statement of Changes in Equity* details any changes in the owners' interest in the company. This statement is influenced by events such as new shares being sold or bought back, profit or losses being made and distributions paid to owners (called dividends).

2.1.3 Double-Entry Accounting

The double-entry accounting used today has its origins many hundreds of years ago (Section 3.11 of Trotman and Gibbins [2005]), and is such that there are two equal and opposite aspects to be recorded for every event. This concept is easier to understand by using some examples, such as those that follow.

The examples use double-entries to maintain the Balance Sheet relationship Assets = Liabilities + Equity.

a. The purchase of a new machine will increase total assets via PP&E and will also cause a corresponding

   a.1. Decrease in another asset such as the cash paid for the machine,

   a.2. Creation of a new liability for future payment if the machine was purchased on credit, or

   a.3. A combination of a.1 and a.2[11].

b. New contributions from owners will increase equity, as well as correspondingly increase assets via increased cash.

[11] It is unlikely that the machine will be directly purchased out of Equity using new funds, because they will be listed as separate transactions as the next dot-point example suggests.

c. Paying for products bought on credit will decrease liabilities by reducing accounts payable and will also reduce assets via decreased cash.

d. Obtaining a loan from a bank will create a new debt liability and also increase assets via the cash provided by the bank.

Double-entry accounting becomes more complicated when the income statement is included, but the concept of two equal and opposite aspects being recorded remains true. For example,

e. When employees are paid, cash (from the balance sheet) decreases and correspondingly the wage expenses increase.

f. When a product is sold for cash or credit, two events are recorded:

  f.1. *Record of the sale:* An asset from the balance sheet (cash or accounts receivable) increases and sales revenue increases.

  f.2. *Record of the cost of the sale:* An asset from the balance sheet (inventory) decreases and an expense from the income statement (cost of goods sold) increases.

The implications of the double-entry system for financial statement fraud is that fraudulent accounting affects two accounts, and typically two different financial statements, commonly the balance sheet and the income statement (ACFE 2013b).

2.2 Fraud within Financial Statements

This section briefly introduces the concept of fraud within financial statements. The following subsections define financial statement fraud, clarify the difference between similar terms, discuss the key entities involved and the different ways it is committed.

2.2.1 Defining Financial Statement Fraud

As mentioned in the previous chapter financial statement fraud involves the intentional publication of false or misleading information in financial statements. The precise definition may vary slightly from country to country, but it is common that fraud involves breaking the law. A formal definition can be found in Section 240 of the Statement on Auditing Standards (SAS) 122 issued by the American Institute of Certified Public Accountants (AICPA) titled *Consideration of Fraud in a Financial Statement Audit*. The definition provided is

An intentional act by one or more individuals among management, those charged with governance, employees, or third parties, involving the use of deception that results in a misstatement in financial statements that are the subject of an audit.

Key terms used in this definition are elaborated in the following dot points.

- **Financial Statements**: These include the Balance Sheet[12], Income Statement[13], Cash Flow Statement, Statement of Changes in Equity[14] and all other sections such as the notes to the statements that provide additional relevant information not contained in the main statements.
- **Misstatement**: Something must be incorrect, which includes something that is fictitious, improper, inappropriate, omitted or false. This includes the manipulation of records or supporting documents, misrepresentation or omission of information or misapplication of principles and regulations with reference to financial statements.
- **Intentional**: This means that the act is done with reason and purpose (TheLawDictionary.org 2015) . A report by the National Commission on Fraudulent Financial Reporting (1987) in the US includes reckless acts in this definition.
- **Material**: There are slight differences between the definitions of what constitutes a *material* misstatement, but a core issue is that it be substantial enough to influence decisions of users of financial statements. The US Securities and Exchange Commission (SEC) refers to it being material if “there is substantial likelihood that a reasonable shareholder would consider it important in making an investment decision” and the US accounting standards board refers to it being material if “it makes it probable that the judgment of a reasonable person relying on the information would have been changed or influenced” (ACFE 2013b).
- **Subject of an Audit**: Financial statements are reviewed by external auditors who are required to be independent and whose task it is to express an opinion on the financial statements.

---

[12] The Balance Sheet is called the Statement of Financial Position in some countries such as Australia.

[13] The Income Statement is also known as the Profit and Loss Statement, Statement of Operations and Statement of Earnings.

[14] The Statement of Changes in Equity is also known as the Statement of Retained Earnings.

Legally, a fraudulent act also requires an entity relying on the misstatement and an injury or loss caused to that entity by the misstatement (Romney et al. 2013). In brief, financial statement fraud is an intentional violation of applicable accounting standards and regulations.

2.2.2 Terms Similar to Financial Statement Fraud

Financial statements can range from true and fair to fraudulent. Stages on a continuum describing this range are extracted and modified from Jones and Library (2011) and described below.

a. **Intended Accounting:** Working within the regulatory framework and accounting rules to produce a “true and fair view” that *best serves the users* of financial statements. This is the ideal scenario and the intention of financial statements (Jones and Library 2011).

b. **Creative Accounting[15]:** Working within the regulatory framework and accounting rules to produce a “creative view” that *best serves the producers* of financial statements. Flexibility when producing financial statements is necessary to produce true and fair statements that are useful to assist the users in making economic decisions, but this same flexibility makes possible creative accounting that benefits the producers (Jones and Library 2011). Creative accounting does not clearly break any rules, but it is not consistent with Intended Accounting.

c. **False Accounting (including financial statement fraud):** Working “outside” the regulatory framework and accounting rules to produce financial statements that *do not comply* with regulations and rules. If the intent and material criteria described above are met, then false accounting can be fraudulent. Fraud needs to be proved in a court of law or by a regulatory body (Jones and Library 2011). Financial statement fraud is the extreme of this continuum and is the focus of this dissertation.

Other commonly used terms are briefly described below.

- **Aggressive accounting** is a term that essentially means the same as creative accounting.
- **Impression management** is similar to creative accounting in that it serves the best interests of the producers of the statements, but it is usually specific to the visual representation of information including the creation of graphs.

[15] There are many different definitions of creative accounting. The preferred definition from Jones and Library (2011) has been used here. Refer to Jones and Library (2011) for more information on creative accounting and a discussion of other definitions.

- **Profit Smoothing** or **Income Smoothing** uses creative accounting to reduce the variability in profits and income over time.
- **Earnings management** involves using flexibility to try to deliver a predetermined objective, such as the profit level expected by analysts[16] (Jones and Library 2011). This term is very commonly found in the academic literature. Depending upon the source, earnings management can either be exclusively creative accounting or it can encompass both creative and false accounting. This differs from financial statement fraud, which invariably involves false accounting.
- **Restatements** occur when regulators require companies to correct false accounting. This usually only occurs if the misstatement is regarded as being material. Restatements differ from fraud in that they do not require the intent to deceive. For example, Palmrose and Scholz (2004) examined 492 restatements and found that only 19% of cases involved fraud.

While earnings management, profit smoothing, restatements and fraud share certain traits and are all phenomena worthy of study, they are not the same. Examining financial statement fraud is important in its own right, as evidenced by the widespread impact of the recent frauds of this century such as Tyco, Enron and WorldCom (Erickson et al. 2006).

2.2.3 Key Entities Involved

The following subsections discuss the victims and perpetrators of financial statement fraud after introducing some of the main entities responsible for protecting against financial statement fraud.

2.2.3.a Responsible Entities – the Regulator

The US Securities and Exchange Commission (SEC) is the primary overseer and regulator of the securities markets in the US, the place of focus in this study. The SEC has a mission to protect investors, maintain fair, orderly, and efficient markets, and facilitate capital formation. This includes protecting investors from fraud and taking enforcement action relating to fraud. The SEC investigates on the basis of many internal and external sources and then can take enforcement action relating to financial statement fraud (Cotter and Young

[16] Analysts are sophisticated users of financial statements (Block 1999) who issue reports on their view of companies to market participants, either for nothing or for a fee.

2007). This enforcement action includes civil lawsuits brought by the SEC in US Federal Court, administrative notices or orders issued by the SEC and information on the settlement of cases involving the SEC. Lane and O'Connell (2009) found that SEC has become more aggressive and that enforcement activities have increased since 2002, when high-profile frauds such as Enron were revealed.

Since 1982 the SEC has issued Accounting and Auditing Enforcement Releases (AAERs) for misconduct related to financial statements during or at the conclusion of an investigation against a company, an auditor or an officer (Dechow et al. 2011), as well as other misconduct such as insider trading unrelated to this research. These AAERs are publicly available on the SEC's website www.sec.gov and contain varying levels of detail. That is, the AAERs contain SEC alleged cases of fraudulent misstatement in financial statements and have been a source of data in many research studies about financial statement fraud.

2.2.3.b Responsible Entities – Management and the Board

Senior management personnel, including the Chief Executive Officer (CEO) and Chief Financial Officer (CFO), are responsible for ensuring quality financial statements free from fraud. However, as is mentioned below they are also the main perpetrators of fraud that conflicts with their responsibility. As overseers of senior management that have a responsibility to act on behalf of the shareholders, the board of directors through its audit committee also has a responsibility for the financial statement's quality and for its being free from fraud (ACFE 2013b).

Companies also have internal auditors who are responsible for continual auditing throughout the year, which includes (but is not limited to) the production of the financial statements. Internal auditors work across all areas of the organisation and should be able to report directly without senior management involvement to the board of directors via the audit committee.

2.2.3.c Responsible Entities – External Auditors

The American Institute of Certified Public Accountants (AICPA) clarified the auditor's responsibility and provided guidance for considering fraud in an audit in SAS 82 issued in 1997 (AICPA 1997), SAS 99 issued in 2002 (AICPA 2002) and more recently with effect from the end of 2012 in Section 240 from SAS 122 (AICPA 2011). Additionally, Section 200 from SAS 122/123 (AICPA 2011) that also came into effect at the end of 2012 specifies that the objective of the external auditor includes the need "to obtain *reasonable assurance* about whether the financial statements as a whole are free from material misstatement, whether due to fraud or error". Reasonable assurance is specified to differentiate it from a guarantee that is essentially impossible to provide, but the exact definition of "reasonable assurance" has been the subject of much debate in the past (Hogan et al. 2008). Despite their responsibility, unfortunately Beasley et al. (2010) found that external auditors were implicated in 23% of the AAERs between 1998 and 2007.

2.2.3.d Perpetrators (Fraudsters)

While perpetrators of fraud in general have been shown to be very diverse, perpetrators of financial statement fraud are most often members of senior management. Nieschwietz et al. (2000) state that financial statement fraud is typically perpetrated by more than one highly motivated and intelligent manager well-versed in persuasion and intimidation. From analysing AAERs from 1998 to 2007, Beasley et al. (2010) found that in 89% of cases the CEO, CFO or both were named as being associated with the fraud. This increased from 83% from the equivalent prior study. The increase is caused by an increase in CFO involvement. The ACFE (2014) also found from its worldwide survey that senior management were the group most responsible for committing financial statement fraud. It is noteworthy that while senior management most often commit fraud, they often pressure other employees such as clerks to perform the data entry to implement the fraud (ACFE 2013a). It is also worth noting that the most costly frauds usually involve collusion (Silver et al. 2008). Silver et al. (2008) reveal that collusion assists with withholding information and providing false information to auditors and the audit committee, as well as overriding traditional internal accounting controls that attempt to prevent fraud being committed by a single person.

From analysing AAERs, Beasley et al. (2010) state that the most common motivators alleged by the SEC are to:

- Meet internal or external earnings expectations,
- Hide deteriorating financial conditions,
- Increase the share price,
- Improve financial performance to assist with raising new equity or debt, and to
- Increase management compensation based on financial results.

The motivations for committing fraud are discussed more in Section 4.2 on page 76. For an analysis of the psychological aspects of fraud, refer to Ramamoorti (2008).

2.2.3.e Victims

Financial statement fraud occurs across a range of industries and company sizes (Beasley et al. 2010). The ACFE (2014) refers to fraud as being ubiquitous, claiming that there is no entity that is immune from the problem. Estimates of the aggregated cost of financial statement fraud were given in Chapter 1, but the cost to individual organisations is estimated to be 5% of revenues according to the ACFE (2014) based on frauds reported from a survey of experts who investigated them. Using data from security class actions that are not limited to financial statement fraud, Dyck et al. (2013) measure the cost to fraudulent companies at 22% of their value and the cost to all firms at 3% of their value. The authors explain their well thought-out methodology for the cost valuation in detail, but in brief they calculated the difference between the value of the company after the fraud and the projected value of the company had the fraud not occurred. This was seen as an improvement to analysing the decline in the stock prices in the days following disclosure of a fraud. Incidentally, this alternative method estimated a large decline in value from a drop in the share price of 16.7% in the two days surrounding the disclosure of a fraud alleged by the SEC in an AAER (Beasley et al. 2010).

In addition to the companies themselves, entities that are negatively affected by financial statement fraud include:

- Investors and other financiers who have provided funds to the company under false pretences, who might not recover their funds. KPMG Forensic (2013) estimates that recovery of proceeds from a major fraud in Australia and New Zealand only occurs in full 8% of the time and in part 49% of the time.
- Employees who have their professional reputation diminished by working for a company involved in fraud, even if they weren’t involved. Additionally, many employees lose their jobs because companies often over-hire during fraud schemes and then reduce employee numbers after the fraud is detected (Kedia and Philippon 2009).
- Suppliers, customers and other stakeholders having their reputation diminished by association and potentially having to find an alternative to the fraudulent company after the fraud is detected.

- Accounting and audit firms that face legal and image problems. An extreme example is that the large accounting and audit firm Arthur Andersen does not exist any more because of the major frauds in financial statements that it audited (Whiting et al. 2012)[17].
- Analysts (Cotter and Young 2007) and regulators (Dechow et al. 2011) who could have their reputation diminished if they missed the fraud.
- All entities involved in the financial markets, because they face higher transaction costs and reduced efficiency (Perols 2011), and reduced liquidity (Dechow et al. 2011), as a result of a loss in confidence and trust in regulators and the integrity of financial markets.
- All entities in the economy; Kedia and Philippon (2009) argue that fraud (and earnings management) causes the misallocation of resources in the economy and the amplification of business cycles.

There is also evidence that suggests fraud increases the likelihood of the company failing (Rosner 2003; Deloitte 2008; Beasley et al. 2010). In fact, six of the ten most costly corporate failures in the US were associated with large fraud schemes, including Enron and WorldCom (Whiting et al. 2012). When financial statement fraud leads to failure then there will be even more damage to the entities listed above.

Whiting et al. (2012) provide a good example of the magnitude of the overall cost of fraud to society, which is presented here. The example is the large frauds that occurred in the US in the early 2000s, such as Enron and WorldCom. In brief, Enron managers profited personally while hiding billions of dollars of liabilities and WorldCom falsely capitalised billions of dollars of expenses. These and other types of financial statement fraud are explained in the next section. Overall these and other frauds caused a loss of confidence in business in the US (Carson 2003) and large declines on sharemarkets worldwide (Whiting et al. 2012). The NASDAQ stock exchange in the US reached a peak of 5049 before a suite of large frauds including Enron and WorldCom were made public, after which the NASDAQ dropped 78% to a low of 1141. There were other factors involved, but the drop in the NASDAQ from the September 11 2001 terrorist acts was only a small fraction of the drop from these frauds. These frauds also caused the accounting profession and the SEC to change basic accounting and internal control procedures and recommendations.

[17] For this statement, Whiting et al. (2012) cited another article by C. Albrecht and W.S. Albrecht (2001), but a copy of this article could not be obtained via the library of the author of this dissertation.

2.2.4 Types of Financial Statement Fraud

There are many different ways to fraudulently manipulate financial statements and many different ways to categorise them. Examples of financial statement fraud are presented in five categories based on those in ACFE (2013b): improper revenue recognition, improper treatment of expenses and costs, improper asset valuation, improper recording of liabilities and inadequate disclosures. It is worth noting that manipulating cash flows does not obviously fall into any of these categories, as cash is not commonly fraudulently manipulated because it is more easily verified through the analysis of bank statements[18] (ACFE 2013a). It is also important to note that there are other ways to categorise financial statement fraud; for example, Gao and Srivastava (2011) presented an alternative classification of fraud types, as well as the most relevant evidence schemes available to auditors for each type of fraud. Example evidence types include fake documents, false responses from companies (management) and hidden documents. This taxonomy is exclusively focused on being used within an audit, and so it is not the preferred choice for this study that has a broader scope.

Known warning signs (commonly referred to as red flags and referred to as such hereafter) for each type of fraud are listed for each category of fraud discussed below[19]. The Association of Certified Fraud Examiners (2013) has been used as the primary reference for these red flags, because it is the world's largest anti-fraud organization. Many of these red flags use information available to the auditor but not easily available to the public, such as complex transactions being recorded near the end of a reporting period, sales being recorded by corporate headquarters (rather than the appropriate accounting department) or recurring attempts by management to justify unusual accounting practices to the auditor. However, this dissertation focuses on using publicly available information to enable the findings to be more widely applicable. Consequently red flags that use public information will be the focus of this section and this dissertation in general. Part of the contribution of this research is to identify indicators to assist in detection of financial statement fraud. In addition to the red flags presented in this section, the development of potential indicators of fraud for modelling purposes (which are influenced by these red flags) will be discussed in detail in Chapter 4.

[18] It is possible to fraudulently manipulate cash and it has been done in the past by forging bank statements, but it is not very common.

[19] For readers who are interested in how the red flags for financial statement fraud differ from asset misappropriation please refer to the paper by Gullkvist and Jokipii (2013).

There is some overlap between the different types of fraud listed in the following subsections. Fraud schemes commonly involve more than one type of fraud because of the double-entry nature of modern accounting. This means that there are usually at least two ways to detect any given fraud scheme.

2.2.4.a Improper Revenue Recognition

In simple terms, revenue is recognised and recorded for financial statement purposes when the sale occurs, which usually occurs on delivery of the product or provision of the service. Thus it is different from cash flows that are recorded when the cash is received. In reality, determining when to recognise revenue in all of the scenarios that happen in the real world is very complex. There are accounting rules to guide the revenue recognition process, which can be broken when committing fraud. In fact, improper revenue recognition is the most common type of financial statement fraud (Dechow et al. 1996; Beasley et al. 2010; Dechow et al. 2011; Gao and Srivastava 2011).

The usual reason revenue is improperly recognised is to fraudulently increase revenue and consequently raise income, in order to improve the financial performance as shown in the income statement. It can also be used to decrease these figures, but that is comparatively rare. Ways to fraudulently manipulate revenue are briefly discussed below.

- Premature revenue recognition is the most common method (Gao and Srivastava 2011; ACFE 2013a) and it involves recording a sale before all the conditions for recognising revenue have been met. An example is recording a conditional or consignment sale as a standard sale. A conditional or consignment sale can be returned by the buyer at a later date without payment. For example, buyers usually return the items they do not successfully on-sell to their customers. This means that there is a chance that the sale might not ever occur, and so revenue should not be recorded until the sale is confirmed and payment is made or unconditionally expected to be made.
- Delayed revenue recognition involves the opposite strategy whereby revenue that has occurred in a specific reporting period is fraudulently moved into the next period. This is usually done if the current period has been very profitable, but there are concerns about the next period. This technique will consequently assist in income smoothing.
- The creation of fictitious revenue is also common (Brennan and McGrath 2007), which usually results in a fictitious asset because of the double-entry accounting.

- Incorrectly classified revenue includes manoeuvres such as the sale of an asset being booked as general operating revenue.
- Indirect manipulation involves methods such as fraudulently changing discounts, sales returns and bad receivables accounts.

There are many known red flags for improperly recognising revenue. Some examples are:

- Noticeably higher growth in profitability, particularly compared to that of competitors;
- Positive income and negative operating cash flows;
- Rapidly increasing gross margin (which is sales minus cost of goods sold);
- A substantial increase in the time between sales being recorded and the payment being collected, as measured by the ratio of accounts receivable to sales;
- Financial trouble or failures in the industry; and
- A change in auditors.

2.2.4.b Improper Treatment of Expenses and Costs

Improper treatment of expenses and costs usually involves fraudulently reducing them to improve the financial appearance of the company. This can be easier to do than revenue manipulation because it can be perpetrated by simply omitting costs or expenses. Ways to fraudulently manipulate expenses are listed below.

- Omitting expenses to increase income, which can be easier than creating fictitious revenue.
- Capitalizing expenditures that should be recorded as expenses. Capitalized expenditures are recorded as a long-term asset, which allows their cost to be allocated over a long period using depreciation without an immediate reduction in income, which is what happens if the item is recorded as an expense. Capitalizing expenditures is only allowed in certain situations as defined in accounting reporting standards. Inappropriately capitalising expenses is commonly referred to as deferring the inevitable, because the cost will eventually be seen in the statements as depreciation for tangible assets.
- Shifting current period expenses to prior or future reporting periods, and thus avoiding reporting them in the current period. This can be done by improperly changing accounting policies or manipulating dates close to the end of reporting periods.
- Not accounting for costs such as increases in doubtful debts, obsolete inventory and increases in accumulated depreciation.

The red flags for improper treatment of costs and expenses include:

- A high proportion of assets derived from capitalised expenses;
- High levels of inventory relative to other types of assets; and,
- Low levels of or lack of growth in allowance for doubtful debts and accumulated depreciation.

2.2.4.c Improper Asset Valuation

Improperly valuing assets usually involves the value of an asset being artificially inflated to improve the financial situation as shown in the balance sheet. These schemes all have ties to either improper treatment of revenues or expenses on the income statement as per the double-entry accounting system. Most schemes fall into one of the following categories:

- Illegitimate valuation of inventory, such as fraudulently valuing obsolete inventory as if it were saleable inventory;
- Illegitimate manipulation of accounts receivable, which is particularly prevalent because (1) it directly influences income, (2) it allows for influential changes because accounts receivable commonly has a relatively high value and (3) it can be relatively easily done. The two most common ways to do this are creating fictitious receivables and not recording receivables that are doubtful of being collected.
- Incorrectly recording fixed assets including fictitious assets, misrepresenting asset value and incorrect capitalisation of unrelated costs in order to artificially inflate the value of an asset.
- Not recognising impairments on assets. Assets are often required to be decreased in value on the financial statements if their economic value falls below the net amount[20] on the financial statements, and this is known as impairment. Impairments are also recognised as expenses on the income statement[21].

The red flags for improperly valuing assets include:

- Increased levels of inventory, relative to other assets and particularly relative to sales;
- High levels of accounts receivable, or low levels of allowance for doubtful debts, relative to other assets and amounts in the previous year(s);

[20] The net amount refers to the fact that depreciation has already been deducted.
[21] Impairments on revalued assets are not treated as expenses, but rather a decrease in the revaluation.

- Noticeably higher growth or profitability, particularly compared to those of competitors;
- Positive income and negative operating cash flows; and,
- Unusual increases in gross margin (which is sales minus cost of goods sold).

2.2.4.d Improper Recording of Liabilities

Similar to improper asset valuation, improperly recording liabilities is usually done to improve the financial situation by understating liabilities that appear on the balance sheet. Common ways to do this include:

- Omitting liabilities, which is the easiest way to manipulate liabilities. Examples include omitting court judgements, misplacing and destroying invoices from suppliers and omitting liabilities for product warranty claims or product returns.
- Improperly removing liabilities such as allowances for restructuring costs or severance costs.
- Inappropriately changing accounting assumptions – for example, reducing the number of expected loyalty program claims without appropriate justification when revenue drops. This was done by Tasmanian Airlines who reduced their estimated frequent flyer liability from $7 million to $2 million because of a lack of paying passengers (ACFE 2013a).

The red flags for improperly recording liabilities include:

- Positive income and negative operating cash flows;
- Unusual increases in gross margins;
- Decreases in allowances for product returns or warranty claims; and,
- Unusual changes in accounts payable.

2.2.4.e Inadequate Disclosures

Inadequate disclosures can result in fraudulent financial statements because knowledge of the information important to the financial statements is required to be able to interpret them accurately and meaningfully. The main ways fraud can occur are by

- Incorrectly explaining the key accounting principles and methods used to produce the financial statements,
- Grouping items together into single line items, when they should be presented separately, and
- Not adequately disclosing contingencies. For example, when a liability such as losses from ongoing litigation is not recorded in the balance sheet because the value cannot be

reasonably estimated, it must still be disclosed in the notes to the financial statements if it is probable that it will have to be paid.

The red flags for inadequate disclosures are limited when using publicly available information, but include rapid growth in profitability particularly relative to competitors and also the control of management decision-making by a small number of people (as much as this can be observed from public information).

2.3 Deterrence, Detection and Prevention

This section discusses the deterrence, detection and prevention of financial statement fraud and concludes with the benefits that fraud detection models can provide.

2.3.1 Red Flags

In addition to the red flags associated with the types of financial statement fraud listed above, the ACFE (2014) compiled a list of red flag behaviours that perpetrators[22] share from analysing 1,483 fraud cases as reported by the Certified Fraud Examiners who investigated them. The top red-flag behaviours of perpetrators are:

1. Living beyond their means,
2. Control issues including not being willing to share duties,
3. Excessive pressure on them from within the organisation,
4. A Wheeler-dealer[23] attitude involving shrewd or unscrupulous behaviour, and
5. Financial difficulties.

These behaviours can be useful for organisations in identifying potential fraudsters. In fact, the ACFE (2014) found that at least one red-flag behaviour was present before the fraud was detected in 92% of cases they analysed. Ramamoorti (2008) also lists in his appendix five behavioural-oriented solutions to financial statement fraud that include cultivating a culture of integrity with a tone from management that does not encourage fraud, decisive response to any instances of fraud, background checks for new employees, fraud awareness

[22] These traits are not limited to perpetrators of financial statement fraud, but also asset misappropriation and corruption.

[23] Wheeler-dealer is defined in the Merriam-Webster online dictionary (http://www.merriam-webster.com/dictionary) as "a person who makes deals in business or politics in a skillful and sometimes dishonest way" (accessed 23 January 2015).

training and hotlines for reporting suspicious activity. The ACFE (2013a) also suggested a culture where it is acceptable to ask questions if something appears strange to an employee, even if it is about a relatively small issue. KPMG Forensic (2013) noted two strategies that still remain relatively unused; these are enforced job rotation and screening of employees due for promotion or transfer into positions that have a high opportunity to commit fraud. However, the ACFE (2014) cautioned that background checks might not perform well at predicting fraudulent behaviour, stating that only 5% of perpetrators have been convicted of a prior fraud-related offence and only 18% had been punished by an employer for a prior fraud-related offence. Consequently, the ACFE suggested continuing monitoring of employees and understanding the red flags to be alert for were better approaches.

While the red flags described in this and the previous section are observed in many fraud cases, it is important to note that the presence of a red flag does not guarantee fraud has occurred. Red flags also occur many times in legitimate cases, which makes minimising the error of falsely alleging fraud a difficult task for the investigators or models detecting financial statement fraud (Hogan et al. 2008).

2.3.2 Main Methods of Detection

The ACFE (2014, 2012, 2010) found the most common way financial statement fraud and asset misappropriation fraud schemes are detected is by tips from people. Internal audits and management reviews are the next two most common methods which have been found to follow the tips. Nearly half of the tips come from employees, as well as nearly 22% from customers and almost 10% from suppliers. Furthermore, companies that implemented an anonymous hotline for reporting tips experienced even more detection of fraud from tips. However, more could be done to encourage tips as only 54% of companies had a hotline in place and too many hotlines that were provided were limited to employees when customers and suppliers should also be included (ACFE 2014). Only 11% of companies offered rewards for whistle-blowers[24] (ACFE 2014), which is a particular concern because employees lose by whistle-blowing without them (Dyck et al. 2010). The SEC has created a division called the Office of the Whistleblower (https://www.sec.gov/whistleblower) that receives tips as well as pays whistle-blowers between 10% and 30% of the money collected. This acknowledges the

[24] Whistle-blowers in this context are people who provide useful and actionable tips and information about a fraud.

effectiveness of tips in detecting fraud and the importance of offering rewards to incentivise them. It also suggests that the SEC might not have the resources to investigate and prosecute all the financial statement fraud that occurs, a likelihood also acknowledged throughout the academic literature.

It is notable that there is a gap between the perceived and actual problem of fraud. Although the study was not limited to financial statement fraud, KPMG's (2012) fraud survey found that while 43% of organisations surveyed have been victims of fraud, only 15% of respondents thought fraud was a problem for their organisation. Addressing this gap between perception and reality might assist people to be more alert to the red flags that have been described above that might in turn generate more tips and whistle-blowing.

In addition to the responsible entities (regulators, management, boards of directors and external auditors), there are also other entities that might offer a layer of protection against financial statement fraud, which is important because Dyck et al. (2010) find that it takes a wide range of entities to detect fraud. These other entities include employees, customers and suppliers as listed above, as well as analysts who report on the company, as there is some evidence they signal fraud in advance of its occurring (Cotter and Young 2007). In fact, all the entities listed above as victims have incentives of varying degrees to stop financial statement fraud. Private litigation lawyers also have large incentives to bring cases against companies committing large financial statement fraud and media journalists gain reputation benefits from exposing financial statement fraud.

The ACFE (2014) empirically confirmed the intuitive result that longer fraud schemes cost more. They extend this analysis to discover that frauds detected passively (such as by external audit, law enforcement, accident or confession) have been going on longer and cost more than frauds proactively detected (such as by hotlines, internal audit and controls and management reviews). Effective detection models could assist to reduce the length of time until detection and consequently reduce the cost for both passive (such as external audit) and proactive (such as management review and internal audit) detection.

2.3.3 Detection As a Means of Deterrence and Prevention

Becker (1968) presented an economic theory of crime that states people commit crime because the expected utility of the payoff exceeds the expected disutility of getting caught and being punished. Johnson et al. (2009) found evidence to suggest this is true for financial statement fraud because it is committed more at companies with a lower likelihood (and consequently lower expected disutility) of being caught. This implies that better detection of financial statement fraud, which will increase the perception of the likelihood of being caught, is an important deterrent to potential fraudsters. While it is difficult to change the expected payoff from fraud, another aspect of Becker's theory concerns the expected punishment. Although there are consequences, sometimes severe, for perpetrators of financial statement fraud, the severity of these consequences may not be a sufficient deterrent (Beasley et al. 2010; Partnoy 2010). Moreover, perpetrators are difficult to prosecute because they are usually well connected and often first-time offenders (Ramamoorti 2008; ACFE 2014). Consequently, increasing the perception of the likelihood of being caught is likely to be more effective at deterring and preventing fraud than harsher punishments (ACFE 2013a)

2.3.4 External Auditors Are Not Enough

Although stealing or embezzling company assets occurs more often, management and auditors are more concerned with financial statement fraud because its loss is higher (Romney et al. 2013). Although it is usually large and of great interest to auditors, research finds they perform poorly at detecting financial statement fraud[25].

- Financial statement and asset misappropriation fraud are only detected 3% of the time, reduced from 3.3% two years prior and 4.6% another two years prior (ACFE 2014). This is put into perspective when approximately 7% were detected by accident. The use of external audits did reduce the loss involved, but it was amongst the least effective methods analysed.
- The Treadway Commission in the US, which was set up to study financial statement fraud, found undetected fraud to be a factor in half of the 450 lawsuits against auditors they studied (Romney et al. 2013).

Some possible reasons for this poor performance are listed below.

- Most external auditors do not have a lot of experience in detecting fraud (Humpherys et al. 2011). This has been the case for many years as a survey by Loebbecke et al. (1989) found that only about half of audit partners had ever encountered fraud.

[25] Francis (2004) provides a review of the overall quality of audits, rather than specifically related to fraud detection.

- Managers who commit fraud understand the limitations of an audit and deliberately try to deceive inexperienced auditors (Fanning and Cogger 1998).
- Auditors struggle to effectively combine fraud risk factors to assess the overall likelihood of fraud having occurred (Patterson and Noel 2003), and it is easy for auditors of large companies to be overwhelmed by the volume of information (Eppler and Mengis 2004).
- Audits are not typically designed to reveal the collusion and forgery often involved in financial statement fraud (Whiting et al. 2012).

While external audits are important, they should not be relied upon to detect fraud (ACFE 2014). Given that knowledgeable managers deliberately try to deceive inexperienced auditors, auditors need better decision aids to help them in detection (Fanning and Cogger 1998; Humpherys et al. 2011). In fact, the use of analytical procedures as decision aids for auditors was suggested back in 1988 (Coglitore and Berryman 1988). Some prior research has shown that decision aids such as checklists and questionnaires reduce the effectiveness of an audit (Hogan et al. 2008), possibly because they cannot model the complexity required to effectively detect fraud. However, auditors' ability to detect fraud improves as their understanding of the probability (risk) of fraud existing improves (Bernardi 1994). These facts support the creation of more powerful fraud detection models that can assist auditors in estimating the probability of fraud and assist in detecting it, particularly given the adverse consequences to auditors when it goes undiscovered[26] (Bell and Carcello 2000; Lin et al. 2003). Detection models might also reduce the error associated with people trying to combine multiple risk factors (Bell and Carcello 2000). Additionally, detection models might assist in uncovering costly fraud schemes that involve collusion as Silver et al. (2008) state that they generally cannot be prevented using traditional controls. Furthermore, correctly identifying statements with a low risk of fraud increases the efficiency of an audit by reducing unnecessary tests (Lin et al. 2003).

2.3.5 Benefits of Fraud Detection Models

It is clear that the current situation regarding financial statement fraud detection needs to be improved. This is highlighted by the fact that a large public company in the US, HealthSouth, was able to falsify its financial statements for eleven years without discovery

[26] Refer to Burton et al. (2011) for a discussion on what type of penalties increase auditors' effort to detect fraud.

(Whiting et al. 2012). Detection is difficult, so better tools and aids are needed to assist in making it more effective. Financial statement fraud detection models are an example of such an aid.

Poor as it is, auditors' current capacity to detect fraud has acted as a deterrent (Hogan et al. 2008). Consequently, fraud detection models that assist auditors to improve their ability to detect it will also act as an increased deterrent and possibly reduce the occurrence rates. Given that longer fraud schemes are more costly, models that speed up the process will also assist in mitigating the cost of fraud schemes that do occur.

In addition to their value to auditors, fraud detection models can also be used to benefit:

- Other responsible entities (that are not perpetrators of the fraud) including regulators, management, the board of directors and its audit committee. The models can assist them to better assess the risk of fraud having occurred and whether there is a need to investigate further.
- Victim entities including investors, financiers, employees, customers, suppliers and other stakeholders. The models can assist them to better assess the likelihood of fraudulence with a view to avoiding association with fraudulent companies. In addition, models can be used by analysts to better assess the true value of a company.

Unfortunately, potential fraudsters can use detection models to help them conceal their fraud and avoid detection. Hence, it is important that models are regularly updated to incorporate new information so that their accuracy level does not decline over time. Ideally, this means that before potential fraudsters can exploit what they have learnt from the model, a changed (updated) version will be developed and implemented. Additionally, given the double-entry nature of accounting, it might not be possible for a fraudster to avoid being detected by a sophisticated model even if they know about the model. Furthermore, models are only to complement, not replace, experts such as auditors, and so there will also always be a human component that can add flexibility and unpredictability to the fraud detection process.

Fraud detection models can make an important contribution to the field of financial statement fraud. The next chapter introduces techniques that both have been and could be used to build these models.

Chapter 3 Review of Modelling Techniques

The importance of models to detect financial statement fraud and the contribution they can make was described in the previous chapter. This chapter follows on by introducing and explaining the relevant terminology and classifications of the detection models. Issues identified from other review papers will then be discussed before a review of prior research into financial statement fraud detection from the perspective of the modelling technique used.

This review of modelling techniques complements the following chapter that discusses the types of data or variables used in these models. This review also assists in validating the importance of the research presented later in this dissertation. Within the field of financial statement fraud detection, the focus is on:

- Data from the US, particularly SEC issued AAERs. Despite studies outside the US not being the focus, many are briefly mentioned for their modelling methodology.
- Data that are publicly available and assessable (discussed more in the next chapter). Accurate models that only use this data are referred to as the "Holy Grail" by Whiting et al. (2012). Modelling can also be used in many other ways to assist auditors who have access to private, disaggregated data. For example, research has been undertaken on models analysing journal entries and internal accounting databases (Debreceny and Gray 2010; Argyrou and Andreev 2011), emails (Debreceny and Gray 2011) and event logs from company's information systems (Jans et al. 2013). Gray and Debreceny (2014) discuss this type of research further. Models can also be used to analyse and combine data from auditors' opinions or working papers (Loebbecke et al. 1989; Fanning et al. 1995; Hansen et al. 1996; Deshmukh et al. 1997; Deshmukh and Talluru 1998; Chen et al. 2009;

Krambia-Kapardis et al. 2010), and to study characteristics of auditors (Welch et al. 1998) and how they relate to their ability to detect fraud (Bernardi 1994).

- Multivariate modelling techniques. While there have been univariate studies (Beasley et al. 2000), fraud is like many other phenomena, too complex to be able to be described by one variable, and so modelling techniques that utilise multiple variables are called for.
- Supervised learning methods, which are discussed in the next section. As the title of this dissertation suggests, these methods will be a focus of this research.

3.1 Introduction to Modelling Terminology and Classifications

There are a large number of interrelated terms used in the modelling literature, partially because of its cross-disciplinary nature. There is also a lack of widespread agreement on the definitions of each term and how much they overlap with each other. A brief comment on some of the more common terms is given here to assist the reader unfamiliar with modelling. More information can be found in the introductory sections of Rokach and Maimon (2015), Sathya and Abraham (2013) or reputable textbooks on data mining or data science.

Data science refers to the entire process involving data, from generation or collection through to all the processing and analysis. It is a very similar definition to the discipline of *statistics*, but much debate occurs on whether statistics has or is growing to include *data mining*[27]. *Data mining*, *knowledge mining*, *knowledge discovery* and *machine learning* are overlapping parts of data science that, broadly speaking, involve discovering useful information, patterns, rules or models from one or more samples (or populations) of data. It is possible to categorise the types of techniques used in data mining as Classification, Clustering, Outlier Detection, Prediction, Regression and Visualisation (Ngai et al. 2011; Sharma and Panigrahi 2012). Many of these techniques involve models that learn from data (hence the name, machine learning). This learning can happen in a variety of ways including supervised, unsupervised, reinforcement or stochastic learning (Sathya and Abraham 2013). Supervised learning is the method most relevant to this study and is explained in the context of fraud detection below.

[27] Breiman (2001b) provides an excellent argument for including data mining, termed algorithmic modelling in the paper, as a part of statistics.

3.1.1 Supervised versus Unsupervised Learning Methods

Supervised learning methods attempt to use existing data to learn the relationship between input variables (also known as independent or explanatory variables) and an output variable (also known as a dependent variable). This process is referred to as training a model. For fraud detection, the dependent variable is often a binary variable that indicates whether it is either fraudulent or legitimate. When the dependent variable takes on a predefined number of values (such as two, fraudulent and legitimate), the resulting model is often termed a classification model.

Supervised learning methods learn from data that comprise explanatory variables for a range of known values of the dependent variable. The name of the method can be understood by thinking that a conceptual supervisor provides the value of the dependent variable for each training case. In the case of fraud detection, models are trained on both fraudulent and legitimate cases. Unlike supervised learning that requires training data with specified values for the dependent variable (fraudulent or legitimate), unsupervised learning is a heuristic process that does not require any information about the dependent variable (from a supervisor)[28]. Consequently, it does not rely on past data with classifications of either fraudulence or legitimacy from human investigations that could be misclassified, which can be an advantage of this type of learning. As mentioned in the previous chapter, with fraud data from SEC-issued AAERs the likelihood of incorrectly alleged frauds is extremely low, but there almost certainly are missed cases of fraud.

As opposed to a specific relationship between inputs and output, unsupervised learning methods look for any relationships in the data. Clustering is an example of unsupervised learning and anomaly detection can be a mixture of both supervised and unsupervised. Furthermore, some techniques such as neural networks can be trained using either supervised or unsupervised learning (Sathya and Abraham 2013). Sudjianto et al. (2010) provide more information about these methods as used in a financial crime setting.

[28] Some papers refer to supervised and unsupervised learning respectively as directed and undirected.

3.2 Models to Detect Financial Fraud: Lessons from Review Papers

The purpose of fraud detection models is to accurately identify signals of fraudulent reporting. Phua et al. (2010) provide a concise review of fraud detection modelling that includes references to other review papers such as Bolton and Hand (2002) who review credit card fraud, telecommunications and medical fraud, computer intrusion detection and money laundering. Sudjianto et al. (2010) provide an excellent review that includes the challenges faced when modelling, but it is focused on money laundering and retail banking fraud. More recently, Ngai et al. (2011) review financial fraud detection that encompasses bank fraud, insurance fraud, securities and commodities fraud and other, which includes financial statement fraud (termed corporate fraud in the paper). They also provide references to other review papers on the topic of financial fraud detection. Also there have been specific reviews of financial statement fraud detection with comments on the modelling techniques used (Apparao et al. 2009; Shiguo 2010; Sharma and Panigrahi 2012), but they are brief.

Relevant issues and lessons for producing models to detect fraud that arise from these review papers are listed below.

- Detection of insurance fraud and credit card fraud have been the most researched areas, but financial statement fraud detection has also received considerable focus;
- Bridging the gap between researchers and practitioners is important (Ngai et al. 2011), which indicates that presenting results in a clear, usable way is important, too.
- Prevention activities are important to complement detection (Sudjianto et al. 2010), and as presented in the previous chapter detection also helps deter and prevent fraud.
- Accessing large and good quality data sets is a challenge, particularly with regard to financial statement fraud that has some of the smallest data sets (Phua et al. 2010; Ngai et al. 2011). This issue is important as the size of the data set has implications for the ability of models to distinguish between fraudulence and legitimacy, and the accuracy of performance measurements. Consequently, research using larger data sets would be more valuable.
- Model performance is overestimated if it is tested on the same data used to develop the model (Shiguo 2010). If there are sufficient data, the best method is to keep a portion of the data separate from the training phase (known as a holdout sample) that will be used to obtain an unbiased estimate of accuracy (Sutton 2005). To best assess real-world performance, the data can be partitioned such that the holdout sample occurs

chronologically after the other data. This more closely models reality where models would be developed using currently available data and then implemented on as yet unknown future data. When the data set is too small to have a separate holdout sample, cross-validation is preferable (Sutton 2005). Cross-validation with *k*-folds partitions the data into *k* non-overlapping sets, and then trains the model on *k-1* sets and tests it on the remaining set. This is performed so that every set is used for testing once and then the results are aggregated. In the extreme case, usually for the smallest sample sizes, the number of data items in each set size can be set to one (so that *k* equals the total number of data items), which is called leave-one-out or jackknife cross-validation. Five- or ten-fold (k = 5 or 10) is common for cross-validation in the modelling literature.

- It is very important to appropriately handle the difference between the cost of the two types of errors, missing fraud and falsely alleging fraud (as discussed earlier in Section 1.1) (Phua et al. 2010; Ngai et al. 2011). Consequently, assessing models using a performance measure such as the cost of errors is superior to studying the percentage accuracy (Phua et al. 2010). While this will be taken into account in this study, comments about the performance of the models used in prior studies are presented in the next section using the performance measure used in each particular study.
- Regression-based techniques have been widely used and neural networks are very appropriate techniques for financial statement fraud detection (Sharma and Panigrahi 2012). These two techniques will be discussed below in the first two subsections of the next section.

3.3 Review of Modelling Techniques for Financial Statement Fraud Detection

Purely accounting-based accrual models have been used to detect earnings management (Jones 1991; Dechow et al. 1995; Peasnell et al. 2000), but much has been done using accrual models and some researchers (Bayley and Taylor 2007; Dechow et al. 2011) now believe research should move away from refining accrual models and towards utilising other information in financial statements. The modelling techniques presented in this section are ways that additional information from financial statements can be utilised to assist in detecting financial statement fraud.

This section first covers studies that use or introduce a single modelling technique, and includes an explanation of the relevant techniques. Comparative studies are then reviewed, which are more relevant as this research also compares multiple modelling techniques. The modelling techniques presented in this section are more comprehensive than those in the review papers mentioned in the previous section.

Standard regression-based techniques have been extremely popular in financial statement fraud detection as well as in other classification problems in business. There is also a search for new, more accurate and less distribution-dependent modelling techniques, such as neural networks, survival analysis, decision trees and decision tree ensembles (multiple decision trees). These will be discussed in the following subsections. A summary table of the most relevant studies is presented at the end of each subsection, except for those subsections that do not contain any studies focused on financial statement fraud detection that assess model performance using publicly available US data. An overall summary table covering all techniques is also presented at the end of this section.

3.3.1 Standard Regression-based Techniques

Discriminant analysis is a linear model that has been around for many decades (Fisher 1938) and can be thought of as multiple regression applied to classification problems. A standard discriminant function can be represented as $D = a_1x_1 + a_2x_2 + \ldots + a_nx_n + c$, where x_i are the independent explanatory variables, and a_i and c are the estimated parameters. N cut-off (or critical) values for D can then be established to separate the data into $N+1$ groups. For example, values of $D \leq R$ might be classified as legitimate, while values of $D > R$ are classified as fraudulent, where R is a real number. Although not designed for calculating probabilities associated with classifications they can be calculated, usually using Bayes's theorem.

Two notable assumptions include the x_is being multivariate normal and the covariance of the two classification groups being equal, both of which are rarely satisfied with data from financial statements (Skogsvik 2005). The first is often violated, for example with binary explanatory variables, but the practical significance of this violation in terms of modelling performance is questionable (Laitinen and Kankaanpää 1999). The second assumption is also a common problem for fraud data sets as they are unbalanced, which means that the probability of fraud significantly differs from 50%. Squaring the explanatory variables and conducting quadratic discriminant analysis overcomes the second assumption, but it is questionable whether it has any practical significance (Altman et al. 1977). Overall, prior research suggests that both linear and quadratic discriminant analysis are robust even when their assumptions are not met (Fanning and Cogger 1998).

Unlike discriminant analysis, logistic regression or logit analysis is designed for two (or more) class classification problems such as classifying data points as legitimate or fraudulent. It is theoretically more suitable to these classification problems as it does not suffer from the same assumptions as discriminant analysis. A logistic regression when applied to fraud detection can be represented as

$$Probability\ that\ fraud\ occurs = \frac{e^{a_1x_1 + a_2x_2 + \dots + a_nx_n + c}}{1 + e^{a_1x_1 + a_2x_2 + \dots + a_nx_n + c}},$$

and the estimates are usually estimated using the maximum likelihood procedure. Unlike discriminant analysis the probability of fraud is modelled directly and cut-off probabilities can then be used on the probability for classification purposes. The default cut-off rule is to classify as fraudulent if the probability is greater than 0.5, but this cut-off can be changed to cater for different error costs. Probit analysis is very similar to logistic regression, but it differs by being based on the cumulative standard normal distribution instead of the cumulative logistic distribution. Although the logistic distribution has fatter tails, both distributions are very similar. Consequently, both techniques usually produce the same classifications for the same data. One advantage for logistic regression is that it is computationally more efficient, but this is only an advantage with very large data sets.

As a result of consistently good performance, discriminant analysis and logistic regression (or logit analysis) have become standard benchmark techniques in most fields of classification, and probit analysis has also been used as an alternative to logit analysis. Multicollinearity is an issue for all these regression methods. One way to help address this problem is to use stepwise techniques that either add one variable at a time or remove one variable at a time, and assess the fit of the model at each stage to try to use the best subset of explanatory variables. However, better handling of multicollinearity is one advantage of other techniques that will be presented. The shape of the underlying distributions not being representative for fraud detection, the inability to incorporate error costs into the model estimation stage with these regression techniques, and the desire for greater classification accuracy are all reasons to consider other modelling techniques that might overcome some or all of these issues.

3.3.1.a Standard Regression-based Techniques applied to Financial Statement Fraud Detection

Logistic regression is the most frequently used technique in the financial statement fraud detection literature. Many studies have used logistic regression as a means to empirically study what variables might explain or indicate financial statement fraud or a specific type of financial statement fraud (Beasley 1996; Dechow et al. 1996; Abbott et al. 2000; Owusu-Ansah et al. 2002; Carcello and Nagy 2004; Uzun et al. 2004; Erickson et al. 2006; O'Connor et al. 2006; Ettredge et al. 2008; Yuan et al. 2008; Brazel et al. 2009; Johnson et al. 2009; Perols and Lougee 2011; Wang et al. 2011). Beneish (1999b) used the closely related probit analysis for the same purpose. None of these studies considered their models from a detection viewpoint and so do not provide estimates of classification accuracy. The number of fraud data points in these studies is rather small with the vast majority being fewer than one hundred.

Standard regression-based techniques have also been used to study financial statement fraud in other ways. For example, Bonner et al. (1998) used logistic regression to analyse the relationship between the type of financial statement fraud and the likelihood of litigation against external auditors, and Stice (1991) used probit to model whether litigation would be brought against external auditors. Gao and Srivastava (2011) also used logistic regression to analyse how the methods used by perpetrators to commit financial statement fraud differ according to the type of fraud (such as fictitious revenue compared with fictitious assets). Logistic regression was also used to study whether fraud helps to explain cumulative abnormal returns (Feroz et al. 1991), the effect revenue-related fraud has on cash flows (Scott 2012) and the influence of restatements on management turnover (Desai et al. 2006).

Standard-regression-based techniques are now primarily used as a benchmark with which to compare other techniques, but there are also studies that solely used these techniques and analysed them from a classification accuracy perspective. As is the case with the studies above, the number of fraud cases in the data sets remains small. Bayley and Taylor (2007) trained a logistic regression model on data comprising 129 fraud cases from AAERs and found that it outperformed traditional accounting and accrual models. The study considered varying error costs, but did not have holdout sample or perform cross-validation and so the model performance estimates are not reliable. Similarly, Summers and Sweeney (1998) used logistic regression to analyse data comprising 102 fraud cases, but did not include holdout or cross-validated results. In contrast though, they used a cascaded or multi-stage logistic regression, which appears to be a development based on the idea in Bell et al. (1991). First, the explanatory variables were partitioned into two groups, insider trading and financial statement variables, and separate logistic models were estimated using each group of variables. Then in the second stage, a third logistic regression was estimated using the probability outputs of the first two models. Interestingly, analysis on a sample comprising 51 fraud cases showed that the second-stage cascaded logistic model had superior classification accuracy compared with the first-stage model. Using Greek data, Spathis (2002) had success using a logistic regression and a stepwise procedure to select variables, but without testing the model using cross-validation or holdout data.

Kaminski et al. (2004) found a discriminant analysis model to have impractically high error rates using jackknife cross-validation on a sample comprising 79 fraud cases from AAERs. Better results have also been attained using logistic regression or univariate analysis to select the variables to be used in a discriminant analysis model (Skousen and Wright 2008; Skousen et al. 2009). Specifically, using Jackknife cross-validation on a sample comprising 86 fraud cases, 65-70% and 72-77% accuracy was attained for classifying fraudulent and legitimate cases respectively.

Some studies also considered varying error costs by varying the cut-off probability of fraud for logistic regression. Using 36-fold (one fold for each industry group) cross-validation on data comprising 56 fraud cases, Lee et al. (1999) respectively achieved classification accuracy ranging from 73% for fraudulent and 90% legitimate cases with a 0.1 cut-off to 43% for fraudulent and 98% for legitimate cases with a 0.3 cut-off. It was noted that the drop in classification accuracy for fraud is very large relative to the small gain for the legitimate cases. Bell and Carcello (2000) tried hundreds of different model variations, including using univariate tests to select combinations of explanatory variables, and used cut-off values ranging from 0.05 to 0.95 in 0.05 intervals. The models were trained on data with 37 fraud cases and then tested on holdout data with 40 fraud cases. Overall, their model was found to outperform auditors and was recommended for use as an aid to auditors. Using Taiwanese data, Lou and Wang (2009) found promising results using jackknife cross-validation with the same range of cut-off values.

Beneish (1997, 1999a) and Persons (1995) addressed varying error costs further by using a weighted error cost measure to assess accuracy as well as empirically optimising (during the training phase) the cut-off value to minimise the error cost for each cost ratio. Persons (1995) used stepwise logistic regression trained on data comprising 103 fraud cases from AAERs and tested it using jackknife cross-validation. They also considered cost ratios of missing fraud relative to falsely alleging fraud that ranged from 1:1 to 30:1 and showed their model improved on the naïve strategy of classifying everything as legitimate. In contrast, Beneish (1997, 1999a) used probit analysis and considered cost ratios that ranged from 1:1 to 100:1. The data for Beneish's studies comprised 43/50 fraud cases for training and 21/24 cases that occurred in the years after the training cases in the holdout sample. Overall, they demonstrated that their model outperformed simple accrual models. The latter study has led to the now well-known M-score model, which is defined later in Section 5.2.2 as it is used as a benchmark model in this current study. Classification accuracy of the M-score model on holdout data is presented in Table 3-1.

Table 3-1. Classification accuracy of the M-score model on holdout data for three different cost ratios of missing fraud relative to falsely alleging fraud.

Cost ratio	**Classification Accuracy for**	
	Fraud cases	**Legitimate cases**
10:1	37.5%	96.5%
20:1	50%	92.8%
40:1	54.2%	90.9%

Once again the classification accuracy for fraud cases is worse and more sensitive than for legitimate cases. These percentages are an improvement over Persons (1995); for example, with a ratio of 20:1 the accuracy is 47% and 86% respectively for fraud and legitimate cases.

Dechow et al. (2011) also used logistic regression to produce a probability of fraud that was converted into their F-score by considering the prior probability that financial statement fraud occurs. The model selected a small subset of explanatory variables from a large initial set by using all the variables and then using a stepwise removal procedure to remove statistically insignificant variables. The resulting F-score model is defined later in Section 5.2.2. Testing was conducted on a holdout sample comprising 107 fraud cases from AAERs that occurred after the 247 training cases, which is also similar to Beneish's work but using a much larger data set. In contrast to other research, the F-score is better at classifying fraud with 73.8% accuracy compared with 61.7% correctly classified legitimate cases on the holdout sample. Dechow et al. (2011) demonstrated that the cut-off values could be changed to represent different cost ratios, but they did not optimise the cut-off values for specific cost ratios as was done by Beneish.

The most relevant studies presented in this section are summarised in Table 3-2.

Table 3-2. Summary of studies that exclusively use standard regression-based techniques to detect financial statement fraud, and that assess model performance using publicly available US data. The last column refers to whether the study considers different values for the cost ratio of falsely missing fraud relative to falsely alleging fraud.

Study	**Modelling Technique**	**Test Data**	**Consideration of Costs**
Bayley and Taylor (2007)	Logistic Regression	No	Cost ratios 20-50:1, optimised cut-off values
Summers and Sweeney (1998)	Multi-stage Logistic Regression	No	No, 0.5 cut-off used
Kaminski et al. (2004)	Discriminant Analysis	Jackknife Cross-validated	No
Skousen and Wright (2008); Skousen et al. (2009)	Discriminant Analysis: (Variables selected by Logistic/Univariate)	Jackknife Cross-validated	No
Lee et al. (1999)	Logistic Regression	No	Cut-offs 0.1,0.2,0.3
Bell and Carcello (2000)	Logistic Regression (Stepwise Entry)	Random Holdout	Cut-offs 0.05-0.95 in 0.05 intervals
Persons (1995)	Logistic Regression (Stepwise Entry and Removal)	Jackknife Cross-validated	Cost ratios 1-30:1, optimised cut-off values
Beneish (1997, 1999a) M-score	Probit Analysis	Holdout chronologically after training	Cost ratios 1-100:1, optimised cut-off values
Dechow et al. (2011) F-score	Logistic Regression (Stepwise removal)	Holdout chronologically after training	Varied F-score cut-off to model different cost ratios

3.3.2 Survival Analysis

Survival analysis (also known as duration analysis) techniques analyse the time until a certain event. Survival analysis techniques model problems as a timeline of a life, which is most commonly described by the survival or hazard function (where each is derivable from the other). The survival function $S(t)$ indicates the probability that an individual survives until time t. Contrastingly, the hazard function $h(t)$ indicates the instantaneous rate of death at a certain time t.

There are many different survival analysis techniques including regression-based models that define relationships between one of the descriptor functions (usually survival or hazard) and a set of explanatory variables. The most prominent is the semi-parametric proportional hazards (PH) model proposed by Cox (1972), but there are alternatives such as fully-parametric PH models, accelerated failure time models and Aalen's additive model. Cox's PH model is defined as $h(t) = h_0(t)\ e^{X'\beta+c}$, where:

- $h_0(t)$ is the non-parametric baseline hazards function that describes the change in the hazard function over time. The flexibility from not having to specify the hazard distribution is one of the key reasons for the model's popularity; and,
- $e^{X'\beta+c}$ describes how the hazard function relates to the explanatory variables (X) and is the parametric part of the model, where β is a vector of variable coefficients and c a constant estimated by a method very similar to the maximum likelihood procedure.

The survival function is then computed as $S(t) = e^{-H(t)}$, where $H(t)$ is the cumulative hazard function from time 0 to t. Survival probabilities can then be compared with cut-off values as is performed when using discriminant analysis and logistic regression.

3.3.2.a Survival Analysis applied to Fraud Detection

Survival analysis techniques have been widely and successfully used in biomedical sciences, where the concepts of lifetime, survival and death can be directly applied to a person. However, it is less obvious how to use these techniques in business problems. The Cox and other models have been used with moderate success to predict the failure of a business (Lane et al. 1986; Crapp and Stevenson 1987; Laitinen and Kankaanpää 1999; Shumway 2001; Gepp and Kumar 2014), where a business and its failure were respectively modelled as an individual and death. The Cox model has also been used on a very small data set for the detection of car insurance fraud (Gepp et al. 2012), where a policy and fraud on the policy were respectively modelled as an individual and death. However, while death and company failure are terminal states, fraud is not, as it can occur, then stop, then start again and then stop again and so on. As the survival analysis model considers fraud as death without the possibility of coming back to life, an individual case must be re-entered back into the model every time a company changes from having fraudulent statements to legitimate statements. This could cause substantial impediments to implementing the model (Gepp et al. 2012) and is a particular problem for continuing detection of financial statement fraud as companies can produce statements that vary between legitimate and fraudulent over time.

Even if financial statement fraud occurs within a company, that same company usually continues to produce financial statements in future years that could be either legitimate or fraudulent, and so modelling the fraud as a terminal event is not representative. Nevertheless, survival analysis is a very useful technique in related research that is concerned only with changes leading up to the first instance of a fraud occurring, because the first instance of fraud only happens once and so can be treated as a terminal event. Furthermore, survival analysis is particularly useful when an inherent time dimension is being studied. For example, Cotter and Young (2007) analysed AAERs from 1995 to 2002 using the Cox model and found that companies committing fraud were more likely to have analysts stop reporting on them earlier in the period preceding the public announcement of the fraud. Additionally, Yu and Yu (2011) used the Cox model as well as two other methods, the Kaplan-Meier method and Weilbull distribution regression. They produced findings such as businesses that engaged in lobbying activities evaded fraud detection 117 days longer than those that did not. Recently, Cao et al. (2015) have used survival analysis along with multiple regression to study the way firms change in terms of corporate governance and default risk after they are named in an AAER.

Overall, while it can lead to failure, financial statement fraud is not a terminal event like death or company failure, and so financial statement fraud detection as a classification problem does not naturally lend itself to be modelled as a single lifetime. Considering that this study focuses on the classification problem of detecting fraud, other techniques will be used in preference to survival analysis.

3.3.3 Artificial Neural Networks

Artificial neural networks are a set of interconnected neurons designed on the inner-processes of the human brain, primarily with respect to pattern learning and problem-solving tasks. Artificial neural networks can be trained using supervised methods such as backpropagation, unsupervised methods such as self-organising maps, stochastic learning or reinforcement learning[29]. These different learning methods are discussed further in Sathya and Abraham (2013), but in all cases the model is trained by adjusting the interconnections of neurons in the model. Neural networks trained using backpropagation are by far the most widely applied in business and social science applications. (Bhattacharya et al. 2011b).

[29] Reinforcement learning can be considered a special case of supervised learning (Koskivaara 2004).

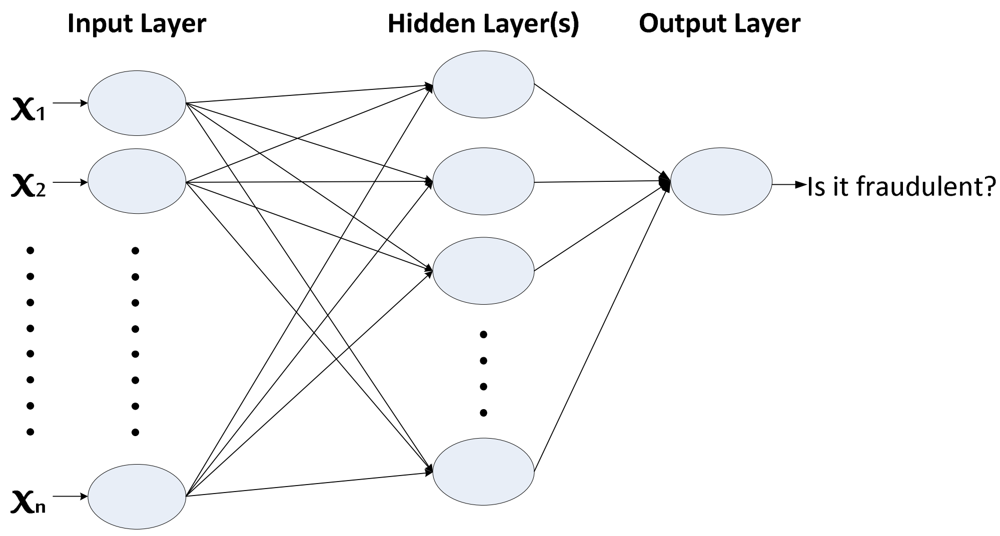
Koskivaara (2004) and Chapter 6 of Negnevitsky (2011) respectively provide a brief and more detailed introduction into artificial neural networks, including an explanation of the backpropagation algorithm.

An example neural network is shown in Figure 3-1 and comprises directional, weighted connections between

- One layer of input neurons for the independent explanatory variables,
- One (or more) hidden layer(s) of interconnected neurons[30], and
- One output layer such as a binary neuron representing either a fraudulent or legitimate case.

The structure of the network is often described using the number of neurons in each layer. For example, a 6-4-1 neural network would have 6 input neurons, 4 neurons in the hidden layer and 1 output neuron. During training, connections of neurons that lead to correct classifications are rewarded by increasing their weighting and those that lead to incorrect classifications are reduced in weighting. It is also possible that the output layer can produce a continuous score that can be compared against different cut-off values to model different error costs, as is done with standard regression techniques.

Figure 3-1. Example neural network for fraud detection.



[30] One hidden layer is common, partially because multiple hidden layers using linear functions can always be transformed into one hidden layer using a non-linear function (Feroz et al. 2000).

Artificial neural networks generally perform well on large, complex data sets in comparison to standard regression-based techniques (Gepp et al. 2012). One advantage of the use of artificial neural networks for financial statement fraud detection is that they are non-parametric and can model non-linear relationships that are likely to be present (Fanning and Cogger 1998). Elliot and Kennedy (1988) provide a review of the statistical assumptions made in standard parametric regression approaches that are not applicable to artificial neural networks. Because of large-scale redundancies in their architecture, artificial neural networks are also better than standard regression-based techniques at choosing among explanatory variables in situations where multicollinearity is present, which is often the case in financial statement fraud detection (Bhattacharya et al. 2011b). Artificial neural networks are also better able to handle outliers and have a high tolerance for ambiguity or noise in the data (Feroz et al. 2000), all of which are common in financial statement fraud data.

The performance of an artificial neural network varies greatly depending on how it was trained. The training phrase requires setting the levels of multiple parameters such as the number of hidden layers and the number of neurons in each layer to match the complexity of the problem. A simple rule for setting these parameters does not exist, so the values can greatly vary with the person training the network. Automatic techniques to train neural networks have consequently been proposed to save time and remove bias from the trainer (Fanning and Cogger 1998). Genetic algorithms are likely the best way to optimise the parameters associated with training an artificial neural network for fraud detection (Bhattacharya et al. 2011b). Genetic algorithms are a type of evolutionary algorithm, heavily based on Darwin’s survival of the fittest principle to evolve better individuals. Genetic algorithms make the training phase more efficient by removing the need to compare alternative configurations, because only the most optimal configuration is expected to survive the optimisation process (Bhattacharya et al. 2011b).

Artificial neural networks are black-box in nature with hidden internal logic. For complex networks such as those needed to detect fraud, it is not possible to gain a full understanding of the importance of each explanatory variable and the interactions between them. However, it is possible to extract decision rules from an artificial neural network in the form of a tree (similar to those discussed in Section 3.3.4) that can be easily analysed (Schmitz et al. 1999).

3.3.3.a Artificial Neural Networks applied to Financial Statement Fraud Detection

Artificial neural networks have been extensively used for classification problems in recent decades and have had success in related fields of auditing (reviewed by Koskivaara [2004]), credit card fraud (Ngai et al. 2011; Khan et al. 2014), fraud involving mobile phone SIM cards (Sallehuddin et al. 2014), energy consumption fraud (Ford et al. 2014), accounting, economics and finance (Fanning and Cogger 1998) and handwritten word recognition (Blumenstein et al. 2007). They have also been used by a public accounting firm to assist in detecting financial statement fraud (Cerullo and Cerullo 1999a, 1999b). Academic studies that focus on the use of artificial neural networks for detecting financial statement fraud are discussed below.

Green and Choi (1997) compared three back-propagation networks with one hidden layer on a randomly selected holdout sample. The network was trained using variables calculated as a one-year percentage change and outperformed networks trained using variables based on other trend calculations. After testing between two and thirty nodes in a layer, an 8-4-1 structure was chosen. The best model was trained and tested on data comprising 27 and 19 fraud cases from AAERs respectively. It achieved a 74% classification accuracy for fraudulent cases and 68% for legitimate cases, which is promising, albeit on a very small data set.

Fanning and Cogger (1998) found that an artificial neural network optimised by an evolutionary algorithm outperformed stepwise regression techniques, specifically logistic regression and both linear and quadratic discriminant analysis. Their network was trained on data comprising 75 fraud cases from AAERs using a program called AutoNet that adds extra hidden layers as required according to the evolutionary algorithm that guides the training. Two stages of variable selection were undertaken, the first with univariate tests and the second with AutoNet. The testing was performed assuming equal error costs on a holdout sample that chronologically occurred after the training sample and comprised 27 fraud cases. The neural network achieved an overall classification percentage of 63% compared with 50% for logistic regression and quadratic discriminant analysis and 52% for standard discriminant analysis. While the other techniques had similar or superior ability to classify fraud, the neural network was the only technique able to effectively classify legitimate statements. It is also interesting to note that the quadratic discriminant analysis did perform slightly worse than standard discriminant analysis, despite its theoretical advantages mentioned above. Using private auditor data, Fanning et al. (1995) also found that AutoNet and another evolutionary neural network performed well relative to cascaded logistic regression on a holdout sample and considering different cut-off values.

Feroz et al. (2000) also considered error costs and found an artificial neural network performed well compared with logistic regression using data comprising 28 training and 14 holdout cases of fraud from AAERs. Both models, but particularly logistic regression, were better at detecting legitimate cases compared with fraudulent cases. Their 7-14-1 feed-forward network clearly outperformed logistic regression in terms of overall accuracy (81% versus 70%) with equal error costs. Earlier, the same authors (Kwon and Feroz 1996; Feroz and Kwon 1999) found that a similar neural network had higher overall classification accuracy than both logistic regression and a self-organising fuzzy set model, but Feroz et al. (2000) also considered varying error costs. For cost ratios of missing fraud relative to falsely alleging fraud that ranged from 10:1 to 40:1, logistic regression was often superior using a weighted error cost measure, but no model was superior in all cases. It appears that the cut-off values were not manipulated or optimised as the error cost ratios changes, although this is not clear in the paper. It is also important to note that prior to final testing, performance on the holdout sample was used to determine model parameters such as the number of hidden nodes. This reduces the real-world applicability of the accuracy estimates, but it was probably necessary given the small size of the data set.

Lin et al. (2003) considered cost ratios that ranged from 1:1 to 100:1, and found that a fuzzy neural network outperformed logistic regression for cost ratios greater than 30:1 using a weighted error cost measure. The fuzzy neural network model was better with larger cost ratios because it was better at detecting fraud, while logistic regression was better at detecting legitimate cases. Overall, logistic regression had the higher classification accuracy with 79% compared with 76% for the fuzzy neural network. This study used a test sample of equal size to the training sample comprising 20 fraud cases. Similar to Feroz et al. (2000), the cut-off values appear not to have been manipulated or optimised for different error cost ratios, and the holdout sample was used to determine training parameters such as the number of training cycles. The fuzzy neural network model utilises a backpropagation network to train a fuzzy logic clustering system of IF-THEN rules. It is called an Adaptive Neuro-Fuzzy Inference System and is explained further in Section 8.4 of Negnevitsky (2011). The theoretical advantage of fuzzy logic is that it better handles the uncertainty and imprecision of the real world by behaving with less analytical precision. However, fuzzy models do not learn, which is overcome by combining them with neural networks.

Unsupervised neural networks have been used to cluster similar cases together, extract features comprising multiple variables, and visualise those clusters to help understand financial statement fraud (Huang et al. 2012; Huang et al. 2014). Using the extracted features and inputs, discriminant analysis has also been used to detect fraud on Taiwan data without a holdout sample or consideration for error costs (Huang et al. 2014). Logistic regression has also been found to outperform a backpropagation neural network on Taiwanese data with the same limitations (Shih et al. 2011).

The most relevant studies presented in this section are summarised in Table 3-3. In addition, some more recent studies have used artificial neural network models as a comparison technique. These studies are discussed below in Section 3.3.7.

Table 3-3. Summary of studies that focus on artificial neural network models to detect financial statement fraud, and that assess model performance using publicly available US data. The "Consideration of Costs" column refers to whether the study considers different values for the cost ratio of falsely missing fraud relative to falsely alleging fraud.

Study	Neural Network	Test Data	Consideration of Costs	Comparison
Green and Choi (1997)	Backpropagation	Random Holdout	No, 0.5 cut-off	No
Fanning and Cogger (1998)	AutoNet: Optimised by Evolutionary Algorithm (Initial variable selection by univariate)	Holdout Chronologically after training	No, assumed equal	Outperformed logistic regression & discriminant analysis
Feroz et al. (2000)	Supervised Learning[31]	Random Holdout	Cost ratios 1-40:1	Comparable to logistic regression
Lin et al. (2003)	Backpropagation using Fuzzy Logic	Holdout[32]	Cost ratios 1-100:1	Outperformed logistic regression for cost ratios >30:1

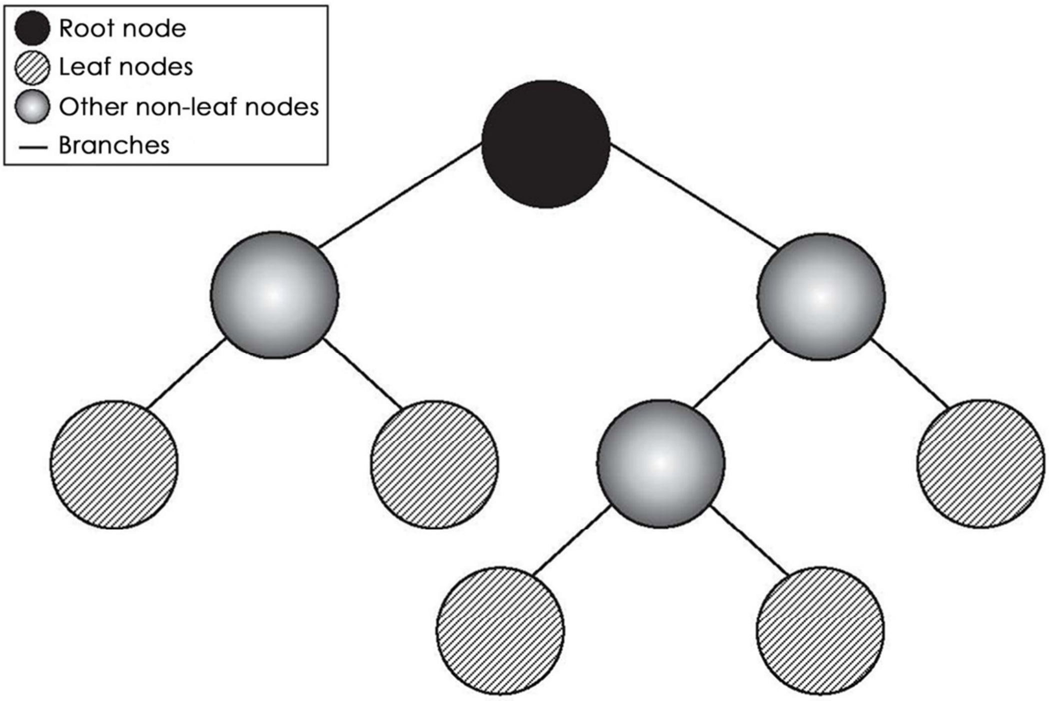
3.3.4 Decision Trees

Decision trees can be used for both regression and classification, and when used for each purpose they are commonly referred to as regression trees and classification trees respectively. Decision trees are most commonly binary trees that consist of a root node, other

[31] The authors did not specify whether the supervised learning algorithm was backpropagation.
[32] It is assumed that the holdout sample was randomly formed, although the authors did not specify.

non-leaf nodes and leaf nodes connected by branches, whereby each non-leaf node has two branches leading to two distinct nodes, as shown in Figure 3-2 below.

Figure 3-2. Basic structure of a binary tree[33].



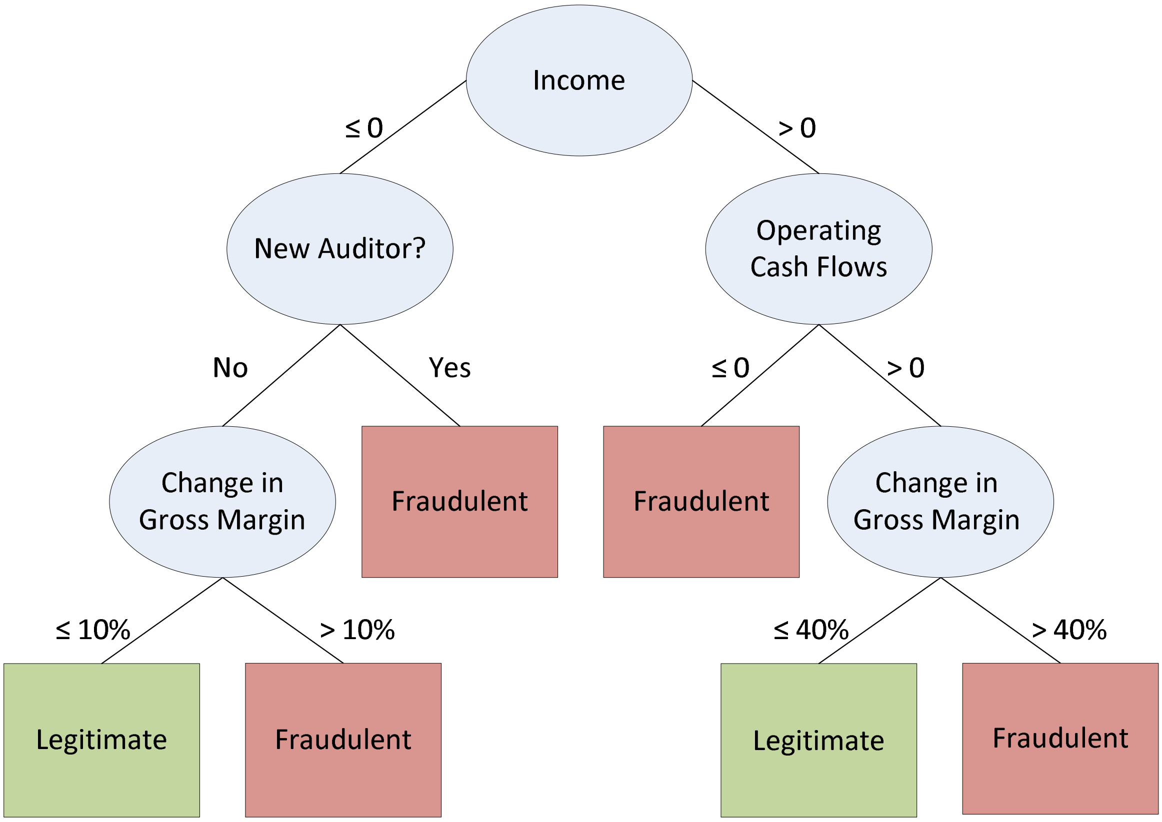
The tree is built by a recursive process of splitting the data when moving from a higher to a lower level of the tree. This process is described by splitting (or decision) rules that are associated with every non-leaf node including the root node decision. While decision trees have all the power of a multivariate technique, the splitting rules are often univariate. The fact that variables are considered individually is more akin to stepwise techniques and assists them in being able to handle irrelevant variables in the data. Multi-way splits have been proposed instead of binary (two-way) splits, but all multi-way splits can be modelled as multiple binary splits. Further to this, binary splits are preferred to reduce complexity and computation time, and increase the potential number of explanatory variables.

When applied to classification problems such as fraud detection, leaf nodes (also known as terminal or decision nodes) represent classification groups such as fraud and

[33] This figure has been reproduced from Figure 1 in Gepp et al. (2010).

legitimate as shown in Figure 3-3. A terminal node is assigned fraud or legitimate according to which classification produces the lowest error cost on the training sample.

Figure 3-3. An example decision tree for detecting financial statement fraud.



The tree structure automatically models interactions between multiple explanatory variables as demonstrated by the model in Figure 3-3 that classifies statements as fraudulent if one of the following occurs:

- "Income > 0" and "Operating Cash Flows ≤ 0", or
- "Income > 0" and "Change in Gross Margin > 40%", or
- "Income ≤ 0" and "New Auditor? = Yes", or
- "Income > 0" and "New Auditor? = No" and "Change in Gross Margin > 10%".

This is an advantage over standard regression-based techniques that can only model interactions that are predefined, which is difficult and time-consuming when there are many variables and many or unknown interactions. Decision tree interactions are more flexible and can correspond to specific regions of data compared with traditional $x_i \times x_j$ interactions in regression equations. Variables can be repeated in a decision tree to allow for more complex interactions, such as the case with Change in Gross Margin in the example above. The dot points above also illustrate that categorical variables such as *New Auditor* are easily modelled and that only the ranking of numerical data is important because only ≤ and > operators are used. Treating numerical variables as ordinal (ranked) means that decision trees are immune to outliers and that monotonically transforming variables (such as taking the natural logarithm) is unnecessary and irrelevant. This is an advantage over standard techniques because the appropriate transformation is not always clear in the real world (Derrig and Francis 2008). It is important to note that if multivariate splitting rules are allowed then monotonic transformations can make a difference (Sutton 2005). Although linear combination splits are not common because they increase the complexity of interpretation, sometimes they can offset this increase in complexity by reducing the number of nodes required in the tree (Sutton 2005).

The major advantages of decision trees include that they are non-parametric and are able to model interactions similar to artificial neural networks, but in contrast they are also transparent, easy to represent visually and simple to interpret as seen in Figure 3-3[34]. Decision trees are also immune to outliers, resistant to irrelevant variables, do not require variable transformations, can incorporate varying error costs into model building (not just model assessment) and are easy to develop into automated systems. However, they also have disadvantages. A major one is that they do not produce an accurate probability of classification, because they do not differentiate between cases classified into the same leaf node[35]. Decision trees also suffer from their construction being sensitive to small changes in the training data set (Sudjianto et al. 2010) similar to artificial neural networks, and therefore perform better when trained on larger data sets. Consequently, the choice of the decision tree building technique is important (Derrig and Francis 2008) as there can be a large variation in accuracy between them.

Sophisticated decision tree software called CART (Classification and Regression Trees) based on the seminal work of Breiman et al. (1984) is sold exclusively by Salford Systems. Decision trees similar to those underlying the CART program can also be created in

[34] It should be noted that extremely large trees with many nodes are difficult to interpret and display visually.

[35] This disadvantage can be addressed by combining decision trees with other techniques. This will be done in the research presented later in Section 5.2.3.f.

the software packages S-Plus and Matlab (Sutton 2005). Other alternatives include Quinlan's See5 based on See4.5 (Quinlan 1993) and ID3 (Iterative Dichotomiser 3), and CHAID based on THAID (Morgan and Messenger 1973) that is available in the SPSS software package. In addition to choosing the best splitting rule at each non-leaf node, the tree building techniques determine how many nodes to include in the tree. Having a rule to stop adding nodes to a tree does not work well, because a split at a higher level that only offers a small improvement might facilitate a latter split that significantly improves the classification ability (Sutton 2005). Consequently, most trees are built using a two-stage process to:

1. Create an overly complex tree that has many nodes, and then
2. "Prune" the tree to the desired complexity by multiple node sub-trees with single leaf nodes[36]. This is important for accuracy on holdout data (Breiman et al. 1984).

Some techniques allow "pre-pruning" during stage 1 to improve computational efficiency, but this is at the risk of lowering classification accuracy. Growing an overly complex tree and then pruning it is a clearly promoted feature of CART.

Salford Systems claim that CART outperforms all competitors in terms of features, accuracy, reliability and robustness (Steinberg 2015). This claim is supported by CART having a better record than any other decision tree software in data mining competitions. This record includes winning multiple first prizes in a very competitive data mining competition, the KDDCup. One of the features of CART in addition to classification accuracy is its ability to assess the relative importance of explanatory variables. Accurately assessing variable importance is complicated when correlations between explanatory variables exist, as is the case with financial statement fraud data. A basic assessment of variable importance can be gained by considering variables that appear in higher levels of the tree to be more important. However, consider the situation when an excluded variable such as sales could have replaced income as the root node in Figure 3-3 without a large loss in accuracy. It is arguable that sales (called a *surrogate* in CART) would then be an important variable for detecting fraud even though it is not present in the model. CART provides a variable importance score ranging from 0 to 100 that considers the contribution of variables in the model as well as the

[36] Pages 313 and 314 of Sutton (2005) explain how minimum cost-complexity pruning works, which is the most common algorithm used for pruning and is the one used by CART.

contributions that variables would have had if they were included. Additionally, CART provides a sophisticated means of handling missing values using surrogates.

3.3.4.a Decision Trees applied to Fraud Detection

Decision trees have been frequently used in finance, marketing, engineering and medicine contexts (Rokach and Maimon 2015). They have also shown promise in detecting insurance fraud (Derrig and Francis 2008; Gepp et al. 2012). In a study (Gepp et al. 2010) that used a weighted error cost measure to assess performance over a range of error cost ratios, CART was found to outperform See5 decision trees at predicting company failure using data from financial statements (similar to financial statement fraud detection models).

In terms of detecting financial statement fraud, CART decision trees were assessed using a holdout sample of Chinese data and were found to outperform logistic regression (Bai et al. 2008). However, no study has focused on decision trees to detect financial statement fraud using US data. Consequently, no summary table is provided for this section. Decision trees have been used in comparative studies that compare multiple techniques. These studies are presented in Section 3.3.7.

3.3.5 Ensembles of Decision Trees

Ensemble techniques aggregate the results from multiple models (called base classifiers) that include, but are not limited to, decision trees. One previously mentioned disadvantage of decision trees is that they are sensitive to changes in the training data. However, newer ensemble techniques that use multiple trees overcome this issue and provide more stable and robust models with reduced variance. Two main ensemble techniques are bagging and boosting, both of which are designed to improve accuracy by reducing the associated variance (Sutton 2005) and will be explained in the following sections. For ensembles to be effective the underlying base classifiers need to be both accurate and diverse. Consequently, trees are excellent underlying base classifiers because their instability results in diverse models given small changes in the training process.

Tree ensembles have the potential to be more accurate than single trees and have fewer parameters to be set, because the final decision is aggregated from a large number of trees. They also include more explanatory variables and make better classifications close to region boundaries as a consequence of using a large number of trees. Tree ensemble techniques are also better at handling missing values and noise or ambiguity in the data.

Conversely, ensembles of trees suffer from a disadvantage common to all ensemble techniques, which is that they become more difficult to interpret and more similar to neural networks with hidden internal logic because of the increased complexity that results from so many individual trees. This disadvantage is weighed against the advantage of being both more accurate and more stable than single trees.

No financial statement fraud detection studies have exclusively focused on ensembles of decision trees, but they have been used in comparative studies that are presented in Section 3.3.7. Ensembles of trees have also been used for related classification problems such as the detection of credit card fraud (Bhattacharya et al. 2011a) and money laundering (Sudjianto et al. 2010) with favourable results.

3.3.5.a Bagging and Random Forests

Bagging (Breiman 1996a) is short for bootstrap aggregation and is designed to improve stability and accuracy of individual models. It is based on repeated sampling with replacement. Many samples are taken from the data, models are then fitted to each sample and the results are aggregated to provide a classification (Sutton 2005). The aggregated classification is usually made by Majority Vote (or mode) from the individual models[37], but the average of the probability of fraud could be used. Average probability has been shown to be better with a small number of trees, but Majority Vote is better with a larger number of trees (Sutton 2005).

Random Forests (Breiman 2001a) is an example of bagging that uses decision trees. Random Forests using CART decision trees is exclusively available through Salford Systems, but other similar implementations are available. Random Forests uses complex trees that have not been pruned and that are built from random samples with replacement. Usually unpruned trees would find patterns specific to the training data that do not generalise and consequently perform poorly on holdout data, but this problem is avoided because of the large number of trees built using random samples that never contain all the data. Random Forests also incorporates more randomness by using the random subspace method that randomly selects a different subset of explanatory variables for each individual node in each decision tree. The size of the subset is held constant and previous research has shown that the square root of the

[37] Majority Vote determines the final classification as that which has been assigned the most times (the mode) by a set of underlying models. For example, if two trees classified a statement as fraudulent and three classified it as legitimate, then legitimate would be the final classification because three is bigger than two.

total number of variables is a good choice (Bhattacharya et al. 2011a). The reduction in the number of variables available at each node provides an opportunity for the second and third (and so on) variables to be used in some nodes. This is desirable because it creates diversity by reducing the correlation between base classifiers, which subsequently increases accuracy (Breiman 2001a).

In summary, for each tree, Random Forests draws a random sample from the entire training data set and uses it to grow an unpruned tree using a different randomly chosen subset of explanatory variables at each node. The result is a form of dynamically constructed nearest neighbour classifier (Whiting et al. 2012).

The advantages of Random Forests over single decision trees are increased stability through reduced variance and reduced sensitivity to the training data, and commonly increased accuracy. They are also easy to use as only two main parameters need to be set: the number of trees in the forest and the size of the variable subset. Their accuracy has been demonstrated by success in data mining competitions (Aldhous 2012). Random Forests have also recently become popular in the academic modelling literature with excellent results when compared with other techniques (Bhattacharya et al. 2011a). The drawback is the black box nature of Random Forests caused by the complexity of a large number of trees, but the Random Forests program by Salford Systems does still provide a variable importance score ranking the explanatory variables.

3.3.5.b Boosting and TreeNet

The classification power of decision trees can be "boosted" by iteratively applying the classification function and combining (with weights) the output so that the classification error is minimised. Unlike bagging that builds models independently of one another, the boosting process is iterative as the incorrect predictions from the current model are given higher weighting (probability) of being selected in the data used to grow the next tree. This process iteratively improves the classification accuracy. Unlike bagging that uses complex models (unpruned trees), boosting uses simple classifiers (small trees) that are poor classifiers on their own. This results in a slow learning process over many iterations, which avoids the problem of largely reduced performance on holdout data.

Stochastic gradient boosting (Friedman 1999; Friedman 2002) is a leading method of boosting. TreeNet, also sold exclusively by Salford Systems, is the commercial product based on Friedman's work. This process also incorporates random sampling and uses the next tree to model errors from the current tree (instead of reweighting data) to improve results (Friedman 1999).

Similar to Random Forests, the advantages of stochastic gradient boosting over single decision trees are increased stability through reduced variance and reduced sensitivity to the training data, and commonly increased accuracy. In addition, as errors are specifically modelled by the next tree, stochastic gradient boosting is particularly good at classifications near region boundaries with smooth decision boundaries (Derrig and Francis 2008; Whiting et al. 2012). Stochastic gradient boosting is also noted for being able to handle unbalanced data sets (Whiting et al. 2012), which is the case with financial statement fraud data sets.

The nature of decision trees forces newly entered variables to interact with variables at higher levels. In contrast, the level of interaction can be controlled in TreeNet by varying the parameter for the size of the trees. This is possible because the size of the tree is fixed by a predefined parameter, unlike the unpruned trees in Random Forests that vary in size. If the size of individual trees in TreeNet is limited to three nodes (one root and two terminal nodes) then no interactions are possible. The level of interactions that are able to be modelled then increase as the size of the trees increases.

It has been suggested that boosting is one of the most powerful data modelling concepts to have been introduced in the 1990s, and furthermore boosted decision trees are generally competitive with any other classifier (Sutton 2005)[38]. Salford Systems credit TreeNet for most of their outstanding results at data mining competitions and claim that stochastic gradient boosting is widely regarded as the most powerful technique available for most cases. Once again, the drawback is the black box nature of TreeNet, but it does provide a variable importance score that ranks the explanatory variables. This ranking is very useful because variables have many opportunities to be used (as a result of many trees), and their importance is built up slowly one iteration at a time.

3.3.6 Other Studies

Expert systems that incorporate fuzzy logic have been proposed for detecting financial statement fraud, and the best results were obtained when the results were fed into a logistic regression (Lenard and Alam 2004; Lenard et al. 2007). However, data comprising

[38] See Section 1.2 of Sutton (2005) for a brief history of boosting and references to more detailed histories.

only 15 cases of fraud in service-based computer and technology firms were analysed from information provided from surveying auditors, which severely limits the generalizability of the results. Comunale et al. (2010) also used fuzzy logic to aggregate survey responses from auditors. Expert systems that incorporate some fuzzy logic have also been used to model insurance fraud (Pathak et al. 2005) and litigation against public accountants (Ragothaman et al. 1995). In the latter example, the fuzzy expert system outperformed discriminant analysis, but the current applicability of the results is questionable because it used data prior to 1984.

In addition to being used to optimise parameters in artificial neural networks, genetic algorithms[39] have been used to detect time-based patterns that identify fraudulent companies, rather than specific financial statements (Kiehl et al. 2005; Hoogs et al. 2007). This research used quarterly, rather than annual, data comprising 51 firms accused of fraud in AAERs. They also used a holdout sample approximately 30% of the size of the main data. However, the authors did not present results that incorporated the detection of multiple patterns. This is unfortunate as single pattern detection only achieved accuracy of 27% or less in detecting fraud, although it is good at detecting legitimacy. The results were however similar to a logistic regression model tested on the same data. Chai et al. (2006) complement this genetic algorithm approach by assigning to each pattern detected a fuzzy logic score representing the degree to which it is present in a company's financial statement (Alden et al. 2012). The advantages of genetic algorithms include being both nonparametric and transparent, and their ability to handle small data sets, missing values and interactions between explanatory variables (Hoogs et al. 2007). However, they also have drawbacks that include being sensitive to changes in the large number of parameters that need to be set, as well as assuming that fraud cases are located "near" each other in the search space (Alden et al. 2012). In response to these and other limitations of genetic algorithms, Alden et al. (2012) also tested a technique called the Markovian Learning Estimation of Distribution Algorithm (MARLEDA) that incorporates learning from a Markov random field probability model with evolutionary algorithms. Using ten-fold cross-validation on data comprising 229 fraud cases alleged in AAERs, both models were found to have similar promising results without considering different error costs. The genetic algorithm and MARLEDA models were able to

[39] Genetic algorithms and other evolutionary algorithms are explained in more detail in Chapter 7 of Negnevitsky (2011).

respectively correctly detect 66.4% and 68.1% of fraud cases, and 61.1% and 60.7% of legitimate cases.

Amershi and Feroz (2000) researched-whether numbers or sequences of numbers with mathematical significance (Golden Ratio, Golden Mean, Fibonacci sequence), and the frequency of them, might be able to distinguish between fraud and legitimate companies. They found some early promising results and concluded that more research needed to be done, including into Silver Means.

Research has also investigated whether deviations from particular statistical distributions are a good indicator of fraud. Zipf's Law describes the expected frequency of the occurrence of various complex patterns or natural languages, and has been tested using simulation for its ability to assist auditors by highlighting potentially fraudulent accounting records (Huang et al. 2008). They found that there are limitations; in particular, the accounting data that were used might not conform to the usual distribution described by Zipf's Law. Benford's Law, which some consider to be a special case of Zipf's Law, describes the expected frequency of individual digits within sets of numbers. This has been applied to disaggregated transaction level data (Durtschi et al. 2004), as well as on aggregated financial statement data (Reed and Pence 2005). Incorporating Benford's Law into an artificial neural network has also shown promise using simulated data (Busta and Weinberg 1998; Bhattacharya et al. 2011b). Hogan et al. (2008) also discusses other studies that use Benford's Law.

Other approaches have also been used to model financial statement fraud. An evidential reasoning approach using a Bayesian framework to assess fraud risk is presented by Srivastava et al. (2009), which needs further empirical analysis to determine probabilities for the Bayesian formulas. Using Greek data, Spathis et al. (2002) used discriminant analysis and logistic regression as benchmark techniques to show that their multi-criteria decision aid model has good detection accuracy. A cutting plan formulation using mathematical programming has been tested on Turkish manufacturing data and found to compare favourably to a probit model (Dikmen and Küçükkocaoğlu 2010). Using Taiwanese financial statement fraud data, Pai et al. (2011) used a support vector machine with some additional processing from CART and nearest neighbour techniques that successfully assisted in properly allocating audit resources. In recent years, studies (Goel et al. 2010; Glancy and Yadav 2011; Gupta and Gill 2013) have begun to investigate analysing the text in financial statements, such as in the Management Discussion and Analysis section. An excellent summary of this work is given by Gray and Debreceny (2014). Finally, Zhou and Kapoor (2011) propose a self-adaptive framework (based on a response surface model) with domain knowledge without any empirical analysis.

The most relevant studies presented in this section are summarised in Table 3-4.

Table 3-4. Summary of other studies to detect financial statement fraud that assess model performance using publicly available US data. The "Consideration of Costs" column refers to whether the study considers different values for the cost ratio of falsely missing fraud relative to falsely alleging fraud.

Study	Technique	Test Data	Consideration of costs	Note
Kiehl et al. (2005); Hoogs et al. (2007)	Genetic Algorithm	Random Holdout	No	Poor accuracy
Alden et al. (2012)	Genetic Algorithm, MARLEDA (both with fuzzy logic)	Ten-fold Cross-validation	No	Both have similar good accuracy

3.3.7 Comparative Studies

The following studies compare multiple supervised learning methods for their utility in detecting financial statement fraud.

Liou (2008) found that stepwise logistic regression outperformed a neural network and decision tree using a Taiwanese data set, without considering costs or separate test data. More recently, a probabilistic neural network and genetic programming (a type of evolutionary algorithm) technique were found to outperform logistic regression, a backpropagation neural network and another type of neural network, and a support vector machine using a Chinese data set (Ravisankar et al. 2011). It is noteworthy that they trialled using only variables that were significant according to a simple t-statistic test, but this reduced accuracy. Using Greek data, Kirkos et al. (2007) found a Bayesian network to outperform an artificial neural network that subsequently outperformed an ID3 decision tree. In contrast, another Greek study (Kotsiantis et al. 2007) found that See4.5 decision trees had the highest accuracy for a single technique outperforming a Bayesian network, neural network, logistic regression, support vector machine and a nearest neighbour and rule-based technique. Interestingly, this study also showed that ensemble techniques improved accuracy. Specifically, combining the individual model results using a decision tree with linear (not univariate) splitting rules was found to be better than simple voting (mode), a nearest neighbour technique and aggregating probabilities of fraud. Combining multiple methods has also more recently been found to be useful on a Chinese data set (Song et al. 2014). A voting system with varying weights for each base classifier outperformed individual models from a backpropagation artificial neural network, See5 decision tree, support vector machine and logistic regression. Overall, the lesson from the studies using data from outside the US is that ensemble techniques that combine multiple models are worth investigating.

McKee (2009) also concluded that combining multiple models warranted further research from analysing US data comprising 50 fraud cases. He used information from prior to the occurrence of fraud (lagged variables), and so his model predicted, rather than detected, fraud. The results, however, are still of interest to this study. He found that feeding the results of one model into another achieved results superior to any of the individual models. Specifically, the ensemble model (referred to as meta-learning model or stacking[40] by McKee) used was produced by the following steps.

1. Feed the original data[41] into a backpropagation neural network with one hidden layer. This resulted in accuracy of classifying fraud of 47.8% and classifying legitimate statements of 95.6%. Overall, this represents an accuracy of 71.4%.
2. Feed the original data and the binary classification from the neural network into a logistic regression. This notably improved the ability to classify fraud by raising it to 70.7%, while the ability to classify legitimate statements dropped to 82.5%. Overall, this represented an increase to 76.5% accuracy.
3. Feed the original data and the binary classification from logistic regression into a See5 decision tree. This increased the ability to classify fraud further to 92.7%, while the ability to classify legitimate statements decreased to 72.5%. Overall, this represented an increase to 82.7% accuracy.

Overall, the results improve with each step. These results also surpassed individual logistic regression and decision tree models that achieved a maximum overall accuracy of 70% and 69% respectively. It is notable that the artificial neural network was the most accurate single technique model. McKee also considered cost ratios of missing fraud relative to falsely

---

[40] Stacking usually refers to ensemble models aggregated from multiple types of underlying modelling techniques.

[41] The original data only contained variables that were statistically significant according to a univariate t-test.

alleging fraud that ranged from 1:1 to 50:1 and found similar superior performance from the ensemble models from steps 2 and 3 as measured by a weighted error cost measure. However, the reliability of these accuracy estimates and comparisons is unclear. The authors state that half the data are partitioned into a holdout sample and it is clear that the individual logistic regression is tested using this holdout sample. However, the individual decision trees are assessed differently using ten-fold cross-validation. Further to this, the tables reporting the ensemble model results include the full data set and so it appears that the accuracy of the ensemble methods might be upwardly biased from including results from the training data.

Using ten-fold cross-validation on data comprising only 51 fraud cases alleged in AAERs, Perols (2011)[42] found that logistic regression and support vector machines performed well compared with artificial neural networks[43], See4.5 decision trees[44] and two ensemble techniques, bagging and stacking. Bagging was performed using 50 See4.5 decision trees as the base classifiers, which is far fewer than the number of trees usually used in Random Forests and also does not include random subsets of variables. Stacking was performed by using a Bayesian classifier to aggregate the results from the other five techniques including bagging. It was noted that CART decision trees were a second preference to the Bayesian classifier for aggregation. All models were compared using a weighted error cost measure for cost ratios of missing fraud relative to falsely alleging fraud that ranged from 1:1 to 1:100. Furthermore, Perols (2011) considered varying cost ratios at both the model building and assessment stage; additionally, the cut-off values were empirically optimised for each cost ratio as in previously presented research such as by Beneish. Overall, Perols acknowledges that these results are surprisingly contrary to many previous studies. For example, most other studies of fraud detection and other classification problems found neural networks to be at least as good as logistic regression. The fact that incorporating varying error costs were incorporated into both the model building and assessment stages is cited as one possible reason for the different results. Given that more complex methods such as artificial neural networks and decision trees require more training data than simpler methods (such as logistic regression), perhaps the small number of fraud

[42] This paper is based on the earlier dissertation by Perols (2008), which had to be examined for some of the details not included in the paper.

[43] The exact artificial neural network used is unclear, but from the discussion in the dissertation it appears that it is a standard backpropagation network with one hidden layer.

[44] J48 was the actual decision tree package used, which was developed based on See4.5 version 8.

cases in the data set is part of the reason for the unusually poor performance of neural networks and decision trees. Consequently, it would be useful to re-test the utility of artificial neural networks and decision trees on another data set with more fraud cases, while still using varying error costs.

Using ten-fold cross-validation on a larger sample of 114 fraud companies, Whiting et al. (2012) found ensemble techniques to be superior to single technique models. The single technique models included probit analysis, logit analysis and partially adaptive estimators (using custom-written computer code written by the authors) that are generalisations of probit and logit analyses. The ensemble models included Random Forests, stochastic gradient boosting and RuleFit, which employs an ensemble of easily understandable rules derived from random sampling with replacement (Whiting et al. 2012). The stochastic gradient boosting used a learning rate of 0.001 and Random Forests used a variable subset size the square root of the number of variables, and both models were aggregated from 1000 trees. The three ensemble methods had comparable accuracy, but Random Forests was the best overall (82%) and also the best at classifying fraud cases with an accuracy of 81.5%. Interestingly, all models were better at classifying legitimate cases compared with fraudulent cases, but Random Forests had the most similar performance between the two with 82.5% accuracy at classifying legitimate statements. The authors noted that the three ensemble methods they chose, particularly the ensemble of rules, were easier to interpret than many other ensemble methods. While Whiting et al. (2012) did not consider specific ratios of error costs they did assess performance using the area under the receiver operating characteristic (ROC) curve, which is a popular measure that provides a type of performance averaged over possible error cost ratios. Despite its advantage of not requiring knowledge of the actual ratio of error costs, Perols (2011) noted that the averaging includes ratios that are not of interest. Moreover, Hand (2009b) mathematically demonstrated that using the area under the ROC curve is akin to stating that the ratio of error costs depends on the classification technique that is used and not the underlying problem, a concept which he claims is absurd (Hand 2009a). Consequently, it would be useful to re-test some of these models in a study that considers a range of error cost ratios.

The most relevant studies presented in this section are summarised in Table 3-5 on the next page.

Table 3-5. Summary of comparison studies that focus on detecting financial statement fraud using US data. The last column refers to whether the study considers different values for the cost ratio of falsely missing fraud relative to falsely alleging fraud. Information about model accuracy is not included because the studies have not presented it in a way that permits comparison between them all.

Study	Techniques Used	Test Data	Consideration of Costs
McKee (2009)	Backpropagation Neural Network, Logistic Regression, See5 Decision Tree, Ensemble of the above (best)	Inconsistent, Unclear	Cost ratios 1-50:1
Perols (2011)	Logistic Regression (best), Support Vector Machine (best), Artificial Neural Network, See4.5 Decision Trees, Bagging using See4.5, Ensemble of the above	Ten-fold Cross-validation	Considered in both model building and evaluation, Cost ratios 1-100:1, Optimised cut-off values
Whiting et al. (2012)	Probit Analysis, Logistic Regression, Partially Adaptive Estimators, Stochastic Gradient Boosting, Random Forests (best), Rule Ensemble	Ten-fold Cross-validation	Performance measured by area under the ROC curve

3.3.8 Summary

The most relevant studies presented in this chapter are summarised in Table 3-6 below with their characteristics for ease of comparison.

Standard regression-based techniques, particularly logistic regression, have been widely studied for detecting financial statement fraud and are therefore excellent benchmark techniques for future research. The most well-known specific models are the M-score (Beneish 1997, 1999a) and F-score (Dechow et al. 2011), both of which are good benchmarks for future models to be compared against. Backpropagation artificial neural networks with one hidden layer have also been applied to the detection of financial statement fraud with encouraging empirical results, and are another technique well-suited for comparing new models against. Integrating fuzzy logic into neural networks (and other techniques) has also shown promise, as has optimising neural network parameters with evolutionary algorithms. However, there is a shortage of studies that assess neural networks over a range of error cost ratios and that empirically optimise the cut-offs. The research presented in this dissertation will address this issue, as well as use the benchmarks identified and suggested above.

Evolutionary algorithms such as genetic algorithms have shown promise, but need further study. There would be benefit from comparing them to other techniques in a study that considers varying ratios of error costs. There is also a lack of research that uses decision tree techniques, particularly CART that has been shown to be superior to other decision tree techniques when applied to other classification problems in business. Consequently, CART will be used in this research. There are also other techniques that have been proposed for financial statement fraud detection that require further empirical analysis, such as Silver Means and Bayesian frameworks and text analysis. Overall, there are a lack of studies that have more than 100 fraud cases. The research presented in this dissertation will use a larger number of fraud cases.

The comparative studies clearly indicate that ensembles of models, including Random Forests and stochastic gradient boosting, perform relatively well at detecting financial statement fraud. As a result of the small number of fraud cases, cross-validation has been used, so there is an opportunity to test the ensembles further by using data with a larger number of fraud cases using a holdout sample that occurs chronologically after the training data to give more real-world applicable results. This would also be an opportunity to re-test the surprisingly poor results of ensemble methods, decision trees and artificial neural networks from the work of Perols (2011), on a data with a larger number of fraud cases and a holdout sample, while still considering optimised cut-off values for a wide range of error cost ratios. Stochastic gradient boosting (TreeNet) and Random Forests are two ensemble techniques that will be used in this research because of their advantages (presented in earlier sections), as well as their promising empirical performance from limited application to financial statement fraud detection. Other ensemble models are also tested in this research, both newly developed and those mentioned earlier in this review including TreeNet, Random Forests and feeding the results of one model into another model. Applying a method to reduce the number of variables before the final modelling stage was found to be useful in many studies, and this finding will be considered when developing the models used in this research. For example, using TreeNet to select variables for use in other techniques will be trialled because TreeNet provides an excellent ranking of variable importance (as mentioned above). Details on the models developed in this research are presented in Chapter 5.

Table 3-6. Summary of studies that focus on detecting financial statement fraud detection and that assess model performance using publicly available US data. The last column refers to whether the study considers different values for the cost ratio of falsely missing fraud relative to falsely alleging fraud. Information about model accuracy is not included because the studies have not presented it in a way that permits comparison between them all.

Study	Modelling Technique	Test Data	Consideration of Costs
Persons (1995)	Logistic Regression	Jackknife Cross-validated	Cost ratios 1-30:1, optimised cut-off values
Green and Choi (1997)	Backpropagation Neural Network	Random Holdout	No, 0.5 cut-off
Beneish (1997, 1999a) M-score	Probit Analysis	Holdout chronologically after training	Cost ratios 1-100:1, optimised cut-off values
Fanning and Cogger (1998)	Neural Network optimised by Evolutionary Algorithm, Logistic Regression, Discriminant Analysis	Holdout chronologically after training	No, assumed equal
Summers and Sweeney (1998)	Multi-stage Logistic Regression	No	No, 0.5 cut-off used
Lee et al. (1999)	Logistic Regression	No	Cut-offs 0.1,0.2,0.3
Bell and Carcello (2000)	Logistic Regression	Random Holdout	Cut-offs 0.05-0.95 in 0.05 intervals
Feroz et al. (2000)	Neural Network	Random Holdout	Cost ratios 1-40:1
Lin et al. (2003)	Backpropagation Neural Network using Fuzzy Logic, Logistic Regression	Holdout[45]	Cost ratios 1-100:1
Kaminski et al. (2004)	Discriminant Analysis	Jackknife Cross-validated	No
Kiehl et al. (2005); Hoogs et al. (2007)	Genetic Algorithm	Random Holdout	No
Bayley and Taylor (2007)	Logistic Regression	No	Cost ratios 20-50:1, optimised cut-off values
Skousen and Wright (2008); Skousen et al. (2009)	Discriminant Analysis	Jackknife Cross-validated	No

[45] It is assumed that the holdout sample was randomly formed, although the authors did not specify.

Study	Modelling Technique	Test Data	Consideration of Costs
McKee (2009)	Backpropagation Neural Network, Logistic Regression, See5 Decision Tree, Ensemble of the above	Inconsistent	Cost ratios 1-50:1
Dechow et al. (2011) F-score	Logistic Regression	Holdout chronologically after training	Varied F-score cut-off to model different cost ratios
Perols (2011)	Logistic Regression, Support Vector Machine, Artificial Neural Network, See4.5 Decision Trees, Bagging using See4.5, Ensemble of the above	Ten-fold Cross-validation	Considered in both model building and evaluation, Cost ratios 1-100:1, Optimised cut-off values
Alden et al. (2012)	Genetic Algorithm, MARLEDA (both with fuzzy logic)	Ten-fold Cross-validation	No
Whiting et al. (2012)	Probit Analysis, Logistic Regression, Partially Adaptive Estimators, Stochastic Gradient Boosting, Random Forests, Rule Ensemble	Ten-fold Cross-validation	Performance measured by area under the ROC curve

Chapter 4 Selection of Explanatory Variables

Fraud detection models process data, called explanatory or independent variables, from both fraudulent and legitimate cases in an attempt to differentiate between the two cases. With supervised learning models, the same explanatory variables (data) are collected for both fraudulent and legitimate cases. The variables are then analysed to determine patterns that indicate whether fraud has likely occurred before a formal investigation has occurred to determine with certainty.

Despite the fact that the selection of variables is crucial to developing a fraud detection model, the use of explanatory variables in prior financial statement fraud detection research is not standardised by a common overall theory or schema (Perols and Lougee 2011). In addition to the variables themselves differing between studies, the categories of variables also vary, which is likely because there is no overarching theory to guide variable selection (Beneish 1997; Kaminski et al. 2004).

Past studies have evaluated a large number of potential indicators of financial statement fraud. Using this past research, only variables that have empirically been found to be useful in financial statement fraud detection models will be included in this study. Further to this requirement for demonstrated empirical usefulness, variables that have been included must also have an underlying theoretical rationale for their association to financial statement fraud. That is, variables that have been theoretically proposed as indicators of financial statement fraud but have not been supported empirically are not included. The underlying rationale or justification for including each variable is derived from prior research, and auditing practice and standards.

The research aims to produce findings that are widely applicable to investors, regulators, auditors and other stakeholders (as discussed in Section 2.3.5), and consequently only variables that are publicly available and relatively easy to obtain have been included. This is consistent with prior research stating that variables too difficult to obtain are unlikely to be used in a practical context (Perols 2011). However, it does introduce a challenge, because publicly available financial statements are highly summarised and aggregated.

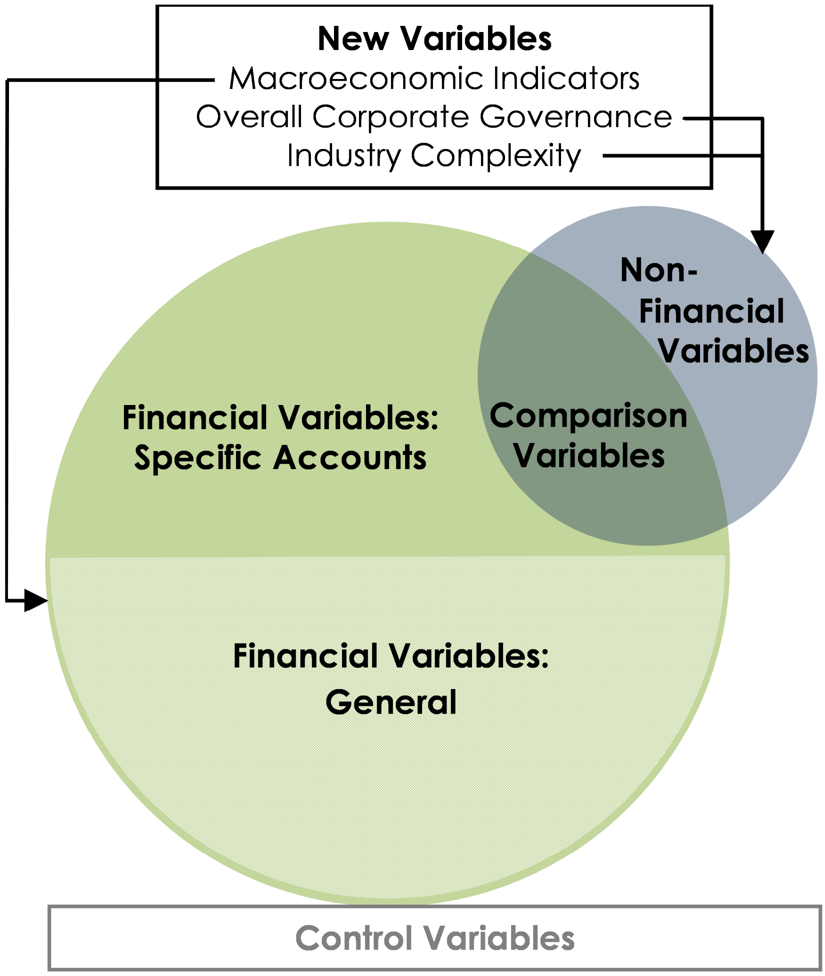
The overall schema of the explanatory variables is presented first, followed by a new theoretical framework called the Fraud Detection Triangle that is based on the famous Fraud Triangle (Cressey 1953). The link between the overall schema and the Fraud Detection Triangle is also described, before a list of all the explanatory variables is presented. An analysis of each individual explanatory variable including the rationale for its inclusion is then given. A summary of all the explanatory variables is then provided, together with their position within the overall schema and the theoretical Fraud Detection Triangle framework. Finally, a summary of the data sources is provided.

4.1 Overall Schema of Explanatory Variables

In the absence of a standardised approach or representation, an overall schema has been constructed from an extensive review of the variables found to be both empirically useful in financial statement fraud detection models and that also have an underlying logical reasoning for their inclusion.

The schema of the explanatory variables used in this study to develop financial statement fraud detection models is shown in Figure 4-1 below. The schema depicts the categories of explanatory variables, and the interrelationships within and between the categories. The largest category of explanatory variables is financial, made up of general financial variables and those that focus on specific accounts within the financial statements. There are also non-financial variables as well as variables that compare the growth of non-financial variables with the growth of financial variables that measure specific accounts. Finally, new variables are added and others are included to control for the matching in the data selection process (described more in Section 5.1.2). While not exactly to scale, the size of the shapes in Figure 4-1 provides an indication of the relative contribution of each category in terms of the number of explanatory variables, except for the new variables category that is disproportionately large to highlight them. The exact size of each category is shown in Table 4-1 on the next page and each category is described more below.

Figure 4-1. Overall schema of explanatory variables, including the categories of variables (financial, non-financial, comparison, new and control) and the interrelationships within and between the categories.



The categories of explanatory variables follow:

> **Financial variables** measure **specific accounts** in the financial statements that are common targets of fraud (accounts receivable, allowance for doubtful accounts, inventory and sales), as well as **general** measures that span multiple accounts comprising measures of asset composition, accruals, debt and financial distress, performance and profitability, and the financing of the company. These general measures include accounting information from the financial statements themselves, market-based information such as share prices and analyst expectations, and comparisons with industry averages. This is the largest category of explanatory variables, as has been the case with most previous research.

Non-Financial variables consider the key roles and positions both within and external to an organisation, including the auditors, board of directors, CEO, CFO and other executives.

Comparison variables measure discrepancies between non-financial variables and financial variables measuring specific accounts, because largely different growth rates between the two are uncommon and might indicate fraud.

There are also control variables and new variables as described below.

Control variables allow for differences in company size, company age, industry membership and the exchange on which it is listed. These are control variables because they are influential in the sample selection process, specifically the selection of legitimate companies that match each fraudulent company. This process is described in more detail in Sections 4.5.13 and 5.1.2.

New variables measure the macro-economy (financial), overall corporate governance (non-financial) and the accounting and organisational complexity of the industry (non-financial). These variables require more testing to assess their usefulness in financial statement fraud detection models and so have been included in this study. The justification for the inclusion of these variables is provided in Section 4.5.14 to Section 4.5.16.

Table 4-1. The number of explanatory variables in each category of the overall schema, and the relative contribution of each category.

Category of Explanatory Variable	Number of Variables[46]	Relative Contribution
Financial: Specific Account	14	28%
Financial: General	20	40%
Non-Financial	3	6%
Comparison: Financial and Non-Financial	6	12%
Control	4	8%
New	3	6%
Total	50	100%

[46] These numbers exclude sub-types; for example, there are multiple ways of measuring variable V1 that are referred to later as V1a, V1b, V1c and so on, but this only counts as one variable in this list.

4.1.1 Temporal Nature of Explanatory Variables

Variables that incorporate information from prior years as well as the year of investigation (hereafter referred to as the specified year) are very important to financial statement fraud detection models. Information from the year being analysed for fraud allows models to look for unusual patterns that occur relative to other companies, which might then indicate fraud. The addition of prior information allows models also to find unusual changes in patterns that occur over time, which also might indicate fraud. For readers familiar with accounting, this means that models are able to perform horizontal analysis (in addition to vertical analysis). For readers more familiar with modelling domains, this represents the inclusion of temporal information. The inclusion of temporal information is consistent with both other research into fraud detection modelling and financial classification problems such as predicting the failure of a company. Table 4-2 shows that 58% of the variables in the overall schema incorporate temporal information. Additionally, more than 50% of the variables in every category of the overall schema incorporate temporal information, except for the control and new variables.

Table 4-2. The temporal nature of explanatory variables in each category of the overall schema.

	Number of Variables[47] with	
Category of Explanatory Variable	**No Temporal Information**	**Temporal Information**
Financial: Specific Account	2 (14%)	12 (86%)
Financial: General	10 (50%)	10 (50%)
Non-Financial	3 (50%)	3 (50%)
Comparison: Financial and Non-Financial	0 (0%)	3 (100%)
Control	4 (100%)	0 (0%)
New	2 (67%)	1 (33%)
Total	**21 (42%)**	**29 (58%)**

The 29 (58%) variables with temporal information are made up of 21 variables (42%) that incorporate information from the specified year and one year prior and 8 variables (16%)

[47] These numbers exclude sub-types, for example, there are multiple ways of measuring variable V1 that are referred to later as V1a, V1b, V1c and so on, but this only counts as one variable in this list.

utilise information from the specified year, one year prior and from even further back. The majority of variables with temporal information compare the specified year with the previous year.

It is notable that the concentration of variables with temporal information is higher within the Specific Account section, compared to the General section, of the Financial Variables. This could indicate when focusing on individual accounts in the financial statements it is more important to study changes over time rather than absolute or relative values from the specified year.

4.1.2 Explanatory Variables with Exogenous Information

Senior management, who are the most common perpetrators of fraud, are in a position to try to conceal fraud in information produced by the company. Consequently, variables that incorporate information that comes from outside the company (exogenous) are important (Hogan et al. 2008), because senior management do not have direct influence over this information.

Eight (16%) variables within the overall schema incorporate exogenous information. These variables comprise

- One (7%) **Specific Account** financial variable measuring the difference between the company and the industry average,
- Four (20%) **General Financial** variables incorporating the share price (two variables), analyst forecasts and the difference between the company's performance and the industry average,
- One (33%) **Comparison** variable measuring the difference between the company and the industry average, and
- Two (67%) **New** variables measuring information about the company's industry and the economy.

The Non-Financial variables that consider or involve decisions influenced by the board of directors (that includes some outside directors) might be arguably exogenous to management (although not the company itself). However, this was not listed above because senior management could have substantial influence, possibly control, over the board of directors in some cases.

4.2 Fraud Detection Triangle Framework Development

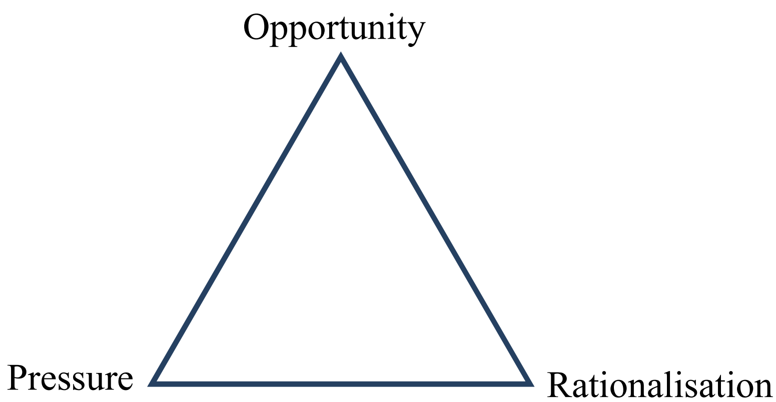
4.2.1 The Original Fraud Triangle

Based on interviewing convicted fraudsters in prison, Cressey (1953)[48] proposed that instances of fraud share three common factors, namely

- An *opportunity* to commit fraud,
- A *pressure* to commit fraud, and
- A *rationalisation* for committing the fraud that is consistent with the perpetrator's personal ethics, sometimes called a willingness to commit fraud.

This theory has since become famous and is referred to as the Fraud Triangle as shown in Figure 4-2 and has been used in numerous prior research and audit standards (Brazel et al. 2009; Kassem and Higson 2012; Lou and Wang 2009; Skousen and Wright 2008; Skousen et al. 2009).

Figure 4-2. Original Fraud Triangle by Cressey (1953) showing the three factors necessary for fraud to occur.



Cressey stated that all three factors had to be present (to some extent) simultaneously for a fraud to occur, but professional accounting bodies contend that the presence of only one factor is enough for fraud to occur (Skousen et al. 2009). Dorminey et al. (2012) provide evidence to support that the presence of one factor is enough, because they point out that fraudsters with a predatory nature only require an opportunity to commit a fraud.

[48] At the time, Cressey was being mentored by Sutherland (1940) who is credited with being the first to integrate the areas of white-collar crime (such as fraud) and business.

Furthermore, Boulter et al. (2013) have mathematically demonstrated that the three factors are inherently inter-linked and consequently the presence of one factor can predispose the presence of the other two factors. This is further evidence that one factor alone is enough to increase the likelihood of a fraudulent act.

Over the years since Cressey's (1953) publication, there have also been several different definitions of the three factors, and different examples given for each (Kassem and Higson 2012). One such definition with a set of examples for each factor is given below.

4.2.1.a The Opportunity Factor

The opportunity factor requires the fraudster to perceive that there is an opportunity to commit fraud. An opportunity arises from situations that provide (Romney et al. 2013):

1. An opportunity to commit fraud,
2. An opportunity to conceal or attempt to conceal the fraud, thereby reducing the chance of being caught, and
3. An opportunity to then convert that fraud into gain such as by receiving a bonus.

The following are examples of conditions that present an opportunity for financial statement fraud to occur:

- New accounting standards that are not yet fully understood by all accountants and auditors;
- Complex or unusual transactions that are not well understood or require subjective judgement;
- Unchallenged management who are dominant over subordinates or auditors;
- Weak internal control systems, such as a lack of segregation of duties or unjustified management overrides to control systems;
- A lack of monitoring internal controls or response to internal control flags;
- Ineffective internal auditors;
- Ineffective board of directors and weak corporate governance; and,
- Frequent changes of external auditors.

4.2.1.b The Pressure Factor

The pressure factor requires the fraudster to perceive that there is a pressure or incentive to commit fraud[49]. It was originally proposed that the fraudster also needs to perceive that they cannot seek help or share their problem, which can be caused by a strong sense of ego or pride (Dorminey et al. 2012). The pressure factor can be thought of as another triangle of three sub-factors (Romney et al. 2013) that include

1. Financial pressure; for example,
    - Pressure to meet or exceed earnings expectations of analysts,
    - Cash flow problems and possible difficulty paying interest and accounts payable,
    - Restrictive debt covenants,
    - Unusually rapid growth with the implied pressure for the growth to continue, and
    - Poor overall economic conditions.
2. Pressures relating to management characteristics; for example,
    - Unrealistic budgets and growth targets,
    - Questionable management ethics and management style,
    - Aggressive earnings forecasts, and
    - Substantial component of compensation linked to performance.
3. Industry condition; for example,
    - Adverse new industry regulations or tax changes,
    - Declining industry, and
    - Poor performance relative to industry competitors.

An alternative breakdown of the pressure factor (Kassem and Higson 2012) also considers external pressures such as negative publicity as well as personal pressure, both financial such as a gambling addiction and non-financial such as greed.

4.2.1.c The Rationalisation Factor

The rationalisation factor is about the fraudster being able to commit the fraud whilst remaining within their personal moral comfort zone (Dorminey et al. 2012). This factor can also take three forms (Romney et al. 2013):

[49] This factor has also been referred to as "perceived need" (Ramamoorti 2008).

1. *Justifying* that the fraud is not actually an ethical wrongdoing; for example, "I only took what I was owed", "It was for a good cause" or "It is best for the company". Those implicated in the financial statement fraud at HealthSouth Corporation used the rationalisation that it was for a good cause because the company was involved in the manufacture of medical equipment that saved lives;
2. An *attitude* that allows the fraud to be rationalised; for example, "The rules don't apply to me";
3. A *lack of personal integrity*; for example, "What I want is more important than honesty".

4.2.2 Extending The Fraud Triangle

The Fraud Triangle is a useful conceptual model for studying and understanding the conditions under which fraud occurs and it is explained in almost all industry and academic education on fraud (Dorminey et al. 2012). Although it has been widely used, the Fraud Triangle has also been criticised for being inadequate (Kassem and Higson 2012). One such criticism is that the pressure and rationalisation factors cannot be directly observed and so solely relying on the Fraud Triangle is not sufficient in fraud investigations (Dorminey et al. 2012). Dorminey et al. (2012) go further to explain that there is a Triangle of Fraud Action that looks at the required actions needed to commit fraud, which is useful for investigators who require proof because all the elements can be directly observed and documented. As this research is looking at identifying cases with higher likelihoods of fraud and not proving fraud occurred, the original Fraud Triangle is the most relevant.

Dorminey et al. (2012) point out that fraud has grown in complexity since the 1940s, which might mean that the Fraud Triangle does not capture all of the precursors to fraud. Additional models have been proposed that can be considered as extensions to the Fraud Triangle (Kassem and Higson 2012). The key contributions of each of the well-known alternative models are described below, with a view to incorporating them as incremental improvements into the pre-existing Fraud Triangle as a basis for this research. It is also important to note that this complexity is also a reason for incorporating the assistance of computer models to identify complex patterns and thereby assist in detecting fraud.

4.2.2.a Extending the Opportunity Factor

After analysing past fraud cases that revealed that the worst are committed by intelligent and experienced people, Wolfe and Hermanson (2004) contend that even if there were an opportunity, a pressure (or incentive) and a rationalisation to commit fraud, the fraudster also needs the necessary capabilities for the fraud to occur. Consequently, they propose a fourth factor, capability, in addition to the three from the Fraud Triangle, thereby constructing The Fraud Diamond. Kassem and Higson (2012) also added capability of committing fraud as a fourth factor to their New Fraud Triangle model, while Dorminey et al. (2012) incorporated it within the existing opportunity factor. It has also been pointed out that a fraudster needs to be capable of concealing the fraud (Dorminey et al. 2012), or at least perceives that they are capable of concealing it (Hurley and Boyd 2007).

Opportunities to commit fraud only lead to the committing of it if there are people with the capability of exploiting them (Dorminey et al. 2012). Consequently, the opportunity factor in the framework used in this research will be referred to as an *Exploitable Opportunity*, which is an opportunity in the presence of people with the capability of committing the fraud. Additionally, the concept of concealment (and the perception of it) is already incorporated into the broader definition of the opportunity factor presented in the previous section based on Romney et al. (2013).

Senior management are the primary perpetrators of financial statement fraud and it would be valuable future research to study how many senior managers have the necessary capabilities to commit financial statement fraud (given the opportunity). It would not be surprising if the vast majority of senior management were capable of committing fraud given the extensive skills and knowledge required to be appointed to such positions. If this were the case, limiting the opportunity factor to exploitable opportunities would not be a substantial change to the model. However, it is still valuable to acknowledge that capability is an important precursor to financial statement fraud and this will be done in the framework in this research by focusing on opportunities that are exploitable.

4.2.2.b Extending the Pressure Factor

The famous Tyco financial statement fraud case was missing a pressure factor, but there was a strong incentive to commit the fraud as evidenced by fraudulent executives making $430 million by inflating the share price based on publishing fraudulent information (Dorminey et al. 2012). Recent research (Dorminey et al. 2012; Kassem and Higson 2012) has also expanded the pressure factor to include a broader set of motivators to commit fraud according to the acronym MICE as presented by Kranacher et al. (2011). Inclusion of a wider set of motivations also has support in previous research such as done by Beasley et al. (2010). MICE stands for Money, Ideology, Coercion and Ego or Entitlement. More money (or financial wealth) and boosting ego are the most common motivators and were present in high-profile cases such as Tyco, Enron and WorldCom (Dorminey et al. 2012), all of which are part of the sample used in this research. An example of coercion is a mid-level accountant in WorldCom being ordered to make false accounting entries (Dorminey et al. 2012). A less-frequent motivation is ideology. An example of this was the HealthSouth case mentioned above where senior management considered that falsification of financial statements helped them provide life-saving equipment to hospitals. Overall, MICE is an incomplete explanation of fraud motivations (Dorminey et al. 2012), but it does provide some additional and useful considerations to be incorporated into the framework used in this research.

A broader definition of the pressure factor as described above includes incentives and in the framework used in this research this factor will be referred to as *Pressure/Incentive*, which incorporates MICE similar to Dorminey et al. (2012) and Kassem and Higson (2012), as well as the categories mentioned in the previous section (such as financial and management characteristics). *Money* is captured by financial pressure and incentives, while *coercion* is captured by management characteristic related pressures. *Ego* and *ideology* are both captured by including incentives as well as pressures, and it is important to note that both also play a role in the rationalisation factor.

4.2.2.c Extending the Rationalisation Factor

Based on the analysis of information from internal auditors of companies that were victims of fraud, The Fraud Scale was proposed by Albrecht et al. (1984). This model replaced the *rationalisation* factor with *personal integrity*, where the latter is defined as "the personal code of ethical behaviour each person adopts". The importance of personal integrity is also stated by Rezaee and Riley (2010).

When an opportunity occurs along with a pressure or incentive and a rationalisation, a generally law-abiding person might succumb to the temptation to commit fraud. However, predatory fraudsters only require an opportunity (Dorminey et al. 2012). Consequently, in the framework used in this research the presence of one factor is enough to be concerned about fraud having occurred. For predators, the pressure and rationalisation factors are replaced by arrogance and criminal mindset, which are issues of attitude and personal integrity. Furthermore, a broader definition of the rationalisation factor as described above based on Romney et al. (2013) already includes the concept of attitude and personal integrity. Attitude was also considered as part of rationalisation by previous financial statement fraud modelling research (Brazel et al. 2009; Lou and Wang 2009). Consequently, in the framework used in this research the rationalisation factor will be referred to as *Integrity/Attitude/Rationalisation.*

4.2.2.d An Additional Model

Rezaee and Riley (2010) used a different model for studying financial statement fraud; it is referred to as the 3Cs model. This model states financial statement fraud will occur if there are:

1. Favourable conditions such as pressures and incentives (financial and non-financial), and opportunities and motives to commit and conceal fraud,
2. A corporate culture that provides the opportunity and motivations for senior management to commit fraud, and
3. Senior management that makes the choice to commit fraud and rationalises their decision.

This model is a different grouping of similar factors to those in the original Fraud Triangle, such that conditions and culture are addressed in the pressure and opportunity factors, and then choice is addressed by the rationalisation factor.

4.2.2.e New Suspicious Information Factor

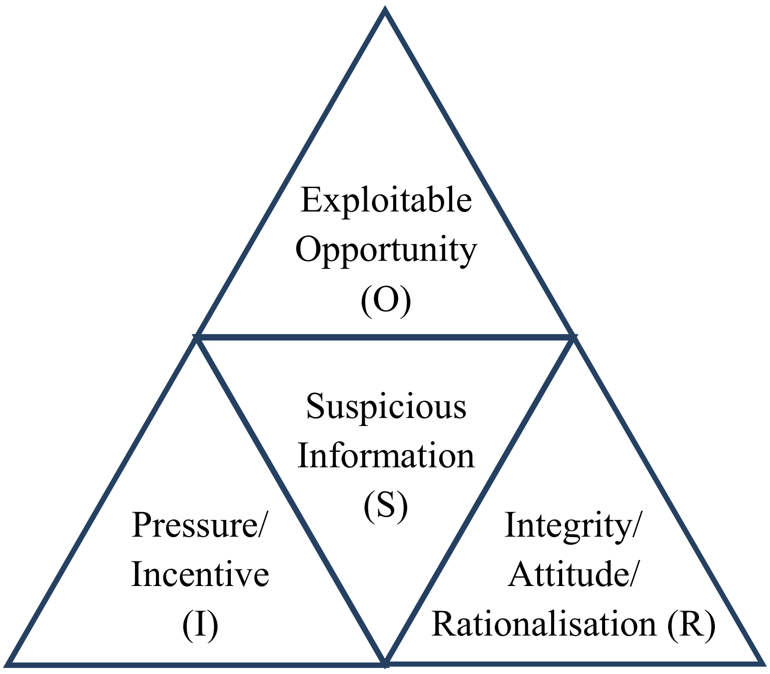
This study focuses on detecting financial statement fraud and one technique is to search for precursors that can also be used as indicators that it might have occurred. This has been the focus of the three framework factors presented so far. However, it is also possible to detect fraud by finding unusual patterns in figures that often occur as a result of it. Consequently, an additional factor called Suspicious Information is proposed. Unlike the other factors that are concerned with whether the precursors to fraud exist, Suspicious Information occurs as a consequence of fraud. This additional factor allows for the inclusion of variables that are not directly related to the precursor factors, but might simply reveal suspicious patterns in them when fraud actually occurs. This additional factor also provides a theoretical basis for the inclusion of complex interactions between variables that empirically might be excellent at detecting fraud even though they might not be clearly related to one of the precursor factors to fraud.

Brazel et al. (2009) included a suspicious accounting factor in their study, but they only viewed accounting information that, arguably, could also fall into one of the precursor factors. In fact, if a simple standard definition of the word suspicious is used then all precursor factors would automatically be included in this category. The reason for this is that the existence precursors of fraud are suspicious in terms of fraud having potentially occurred, which is why this framework includes precursor factors. For example, sales growth is related to both opportunity and incentive factors (as discussed in detail later), but sales growth could also be considered suspicious if it were unusually high. Consequently, to make Suspicious Information a meaningful addition it has been defined as occurring as a consequence of fraud.

4.2.3 The New Fraud Detection Triangle

There are four factors that will form part of the new Fraud Detection Triangle framework. As shown in Figure 4-3 and as defined in the previous section, the factors are Exploitable Opportunity (O), Pressure/Incentive (I), Integrity/Attitude/Rationalisation (R) and Suspicious Information (S). This framework incorporates the extensions of the O, I and R factors suggested in previous research, as well as adapting the Fraud Triangle for use in fraud detection (rather than explaining the precursors) by the addition of the S factor. It is important to note that only one factor in this framework needs to be present for there to be a concern that fraud might have occurred. While developed for this study into financial statement fraud detection, the framework is not limited to this specific type of fraud and is applicable more broadly to fraud detection in general.

Figure 4-3. The new Fraud Detection Triangle framework[50].



[50] It is acknowledged that the layout of this diagram is very similar to the model presented by Kassem and Higson (2012).

There have been a few financial statement fraud detection modelling studies that have used or incorporated all or part of the original Fraud Triangle. Brazel et al. (2009) considered incentive, opportunity, suspicious accounting and other factors, but did not include rationalisation and their definition of suspicious factor is very different from the S factor in this framework (as described in the previous section). Pressure, opportunity and rationalisation factors, but not suspicious information factors, were considered by Lou and Wang (2009) using Taiwanese data, and by Skousen and Wright (2008) and Skousen et al. (2009) using US data. The use of the new Fraud Detection Triangle in this study represents an extension to these prior research studies.

The number of explanatory variables that are associated with each factor of this framework are shown in Table 4-3. The relatively few variables associated with the R factor are not an indication that it is unimportant, but rather an indication that less focus has been placed on it in prior research, probably because it is the most difficult factor to measure (Skousen et al. 2009). For example, Gillett and Uddin (2005) found that while the attitude of the Chief Financial Officer (CFO) was an important factor, the compensation of the CFO (publicly available information in many cases) was not a useful proxy[51]. The relative contribution of the rationalisation factor was also low in a study by Skousen et al. (2009) whose initial set of explanatory variables (totalling 29) comprised 38% measuring opportunity, 45% measuring pressure and 10% measuring rationalisation. Furthermore, no variables measuring rationalisation were found to be statistically significant at a 15% level. It is also expected that the Suspicious Information category is relatively small, because it is a new factor to a fraud triangle.

[51] It should be noted that Gillett and Uddin (2005) also called for more research to be done on the link between the compensation of the CFO and financial statement fraud.

Table 4-3. The number of explanatory variables that are associated with each factor in the new Fraud Detection Triangle framework, and the relative contribution of each category. Variables associated with multiple factors are counted in all associated factors.

Fraud Detection Triangle Factor	**Number of Variables**[52]	**Relative Contribution**
Exploitable Opportunity (O)	31	49%
Incentive/Pressure (I)	24	38%
Integrity/Attitude/Rationalisation (R)	3	5%
Suspicious Information (S)	5	8%

4.3 The Link between the Framework and the Schema

The factors from the new Fraud Detection Triangle framework that are associated with each category of variables within the overall schema are described below and summarised in Table 4-4. In addition, the theoretical rationales for the individual variables from the overall schema are presented in Section 4.5 with reference to the Fraud Detection Triangle framework for the theoretical justification.

Table 4-4. The factors of the new Fraud Detection Triangle framework that are associated with each category of the overall schema.

Category of Explanatory Variable	**Fraud Detection Triangle Factor**
Financial: Specific Account	O, I
Financial: General	O, I, R
Non-Financial	O, I, R, S
Comparison: Financial and Non-Financial	S
Control	O, I
New	O, I

Financial Variables: Specific Accounts (O, I)

These are accounts commonly targeted by fraud schemes, because they are difficult to audit and involve subjective judgement, and so high values represent an increased

[52] These numbers exclude sub-types; for example, there are multiple ways of measuring variable V1 that are referred to later as V1a, V1b, V1c and so on, but this only counts as one variable in this list.

opportunity to commit fraud. Sales are also a key performance indicator for management and so this is strongly linked to incentives or pressures to commit fraud. Additionally, not using the Last-In, First-Out (LIFO) inventory valuation method is also linked to the incentive factor.

Financial Variables: General – Asset Composition (O)

These variables measure opportunity factors because they primarily focus on aggregate measures of the difficult to audit and subjective judgement accounts considered in the specific accounts section.

Financial Variables: General – General Accrual Measures (O, I, R)

Accruals are comparatively easier to audit compared with cash items and so represent an opportunity to commit fraud. Additionally, positive or increasing prior accruals indicate that management has fewer ways to legitimately manage earnings and so there are pressures or incentives to turn to fraud. Finally, increased use of discretionary accruals might reflect poorly on the integrity of management, who might be more able to rationalise fraud.

Financial Variables: General - Level of Debt and Financial Distress (O, I)

Increased debt raises the incentive and pressure to commit fraud. Financial distress might also increase the opportunity to commit fraud through lack of controls or lack of monitoring or enforcing controls during financially difficult times.

Financial Variables: General – Performance and Profitability (I)

Poor performance and profitability can be considered to increase the pressure and incentive to commit fraud to hide poor results.

Financial Variables: General – Financing (I, R)

New funding or the need for additional funding increases the incentive and pressure to commit fraud. Additionally, the use of operating leases as off-balance sheet funding could be associated with managers who are more short-term focused and more likely to be able to rationalise fraud.

Non-Financial Variables (O, I, R, S)

Non-financial variables about the auditor, Chief Executive Officer (CEO), CFO and board of directors predominantly measure opportunity factors, including the capability of the CEO or CFO to exploit the opportunity to commit fraud. However, there is a competing theory for changes in CEO and CFO that is linked to the new Suspicious Information factor. Additionally, the share ownership of the CEO is related to their incentive and pressure to commit fraud, and the number of changes of audit firm is related to the ability to rationalise fraud.

Comparison Variables: Financial and Non-Financial (S)

Increases in the financial values without corresponding increases in the non-financial raises the suspicion of fraud.

Control Variables (O, I)

The control variables are also associated with the opportunity and incentive factors.

New Variables (O, I)

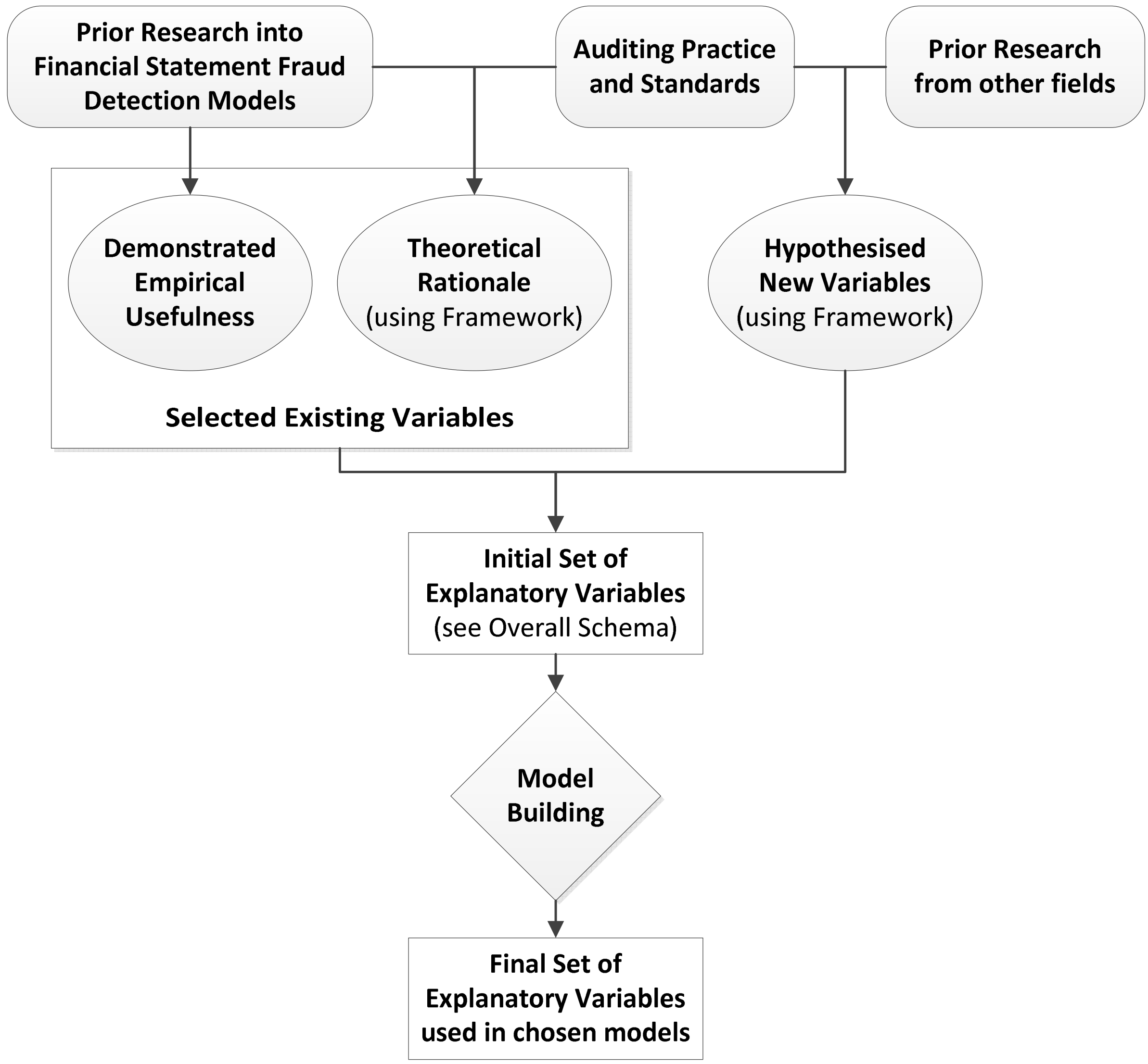
Variations in the economic conditions, as measured by macroeconomic indicators, are hypothesised to change the opportunity and incentive factors associated with fraud. Lower overall corporate governance as measured by an index is hypothesised to indicate reduced monitoring that correspondingly indicates a greater opportunity to commit fraud. Finally, a more complex industry is hypothesised to increase the opportunity to commit and conceal fraud as a result of there being complexities that are understood by fewer people.

4.4 Explanatory Variable Selection Process

In addition to the theoretical rationale with reference to the new Fraud Detection Triangle framework and guided by prior research and auditing practice and standards, the empirical usefulness in prior models was a requirement for variable selection as shown in Figure 4-4. New variables are also proposed based on auditing practice and research from the broader discipline of accounting. The new variables are also justified with reference to the Fraud Detection Triangle framework.

Overall, a large initial set of explanatory variables is used, with variables excluded (noted in Section 4.5.17) based on similar definitions. This was done in preference to using correlation analysis to exclude similar variables as done by Spathis (2002). The reason for this preference is because stepwise procedures and other model building techniques that will be used in this research can handle empirical similarity when selecting the final set of variables to go in the chosen financial statement fraud detection models. This process utilises the strengths of people (researchers) to propose many alternative hypotheses and then uses the strength of statistics and computers to analyse data to determine what hypotheses are supported empirically. This process is summarised below in Figure 4-4.

Figure 4-4. Explanatory Variable Selection Process.



4.5 Detailed Analysis of Each Explanatory Variable

A summary table is presented in the first subsection (4.5.1), followed by an analysis one-by-one of each explanatory variable in the overall schema. The analysis of each variable includes both empirical and theoretical justification for its inclusion, with reference to the new Fraud Detection Triangle framework.

The analysis of each variable also includes stating the expected direction (positive or negative) of association of each variable with financial statement fraud. It is also important to note that linear relationships as well as non-linear relationships and interactions between variables will be analysed. This approach allows for some of the complexity involved in financial statement fraud to be modelled, which could not be done without the use of computer assisted modelling or with only linear computer models.

The theoretical justification for some new variables that have yet to be comprehensively tested in a cost-sensitive manner will also be discussed before the excluded variables are covered. The final subsections cover the excluded variables and the definitions for the terms used that should be referred to for clarification of any terms used.

4.5.1 Summary of Explanatory Variables

A complete list of the explanatory variables used in this research along with their identifiers (IDs) that will be used in future sections is presented in Table 4-5. The variables are listed (and discussed in the following sections) within the categories of the overall schema, and with reference to their expected direction of association with financial statement fraud. The factor or factors of the new Fraud Detection Triangle framework that each variable is associated with are also listed.

Table 4-5. Summary of Explanatory Variables. Definitions of terms are provided in Section 4.5.18 on page 126.

ID	Variable Name	Expected Direction of Association with Fraud	Framework Factor(s) (O/I/R/S)
Specific Account - Accounts Receivable (see Section 4.5.2 on page 93)			
V1 V1a V1b V1c	Accounts Receivable Value for the specified year Percentage change Was Percentage change > 10%?	Positive	O
V2	Percentage change in Accounts Receivable to Sales	Positive	O
V3	Percentage change in Accounts Receivable to Total Assets	Positive	O
Specific Account - Allowance for Doubtful Accounts (AFDA) (see Section 4.5.3 on page 94)			
V4	Percentage change in AFDA	Negative	O
V5	Percentage change in AFDA to Accounts Receivable	Negative	O
V6	Percentage change in AFDA to Sales	Negative	O
Specific Account – Inventory (see Section 4.5.4 on page 95)			
V7	Change in Inventory to average Total Assets	Positive	O

ID	**Variable Name**	**Expected Direction of Association with Fraud**	**Framework Factor(s) (O/I/R/S)**
V8 V8a V8b	Inventory to Sales Value for the specified year Change	Positive	O
V9	Was Last-In, First-Out (LIFO) inventory valuation used?	Negative	I
Specific Account – Sales (see Section 4.5.5 on page 97)			
V10 V10a V10b V10c V10d V10e	Sales Growth Percentage change V10a minus the Industry Average Previous year's Percentage change Four-year growth rate Previous year's percentage change in total assets	Uncertain	O, I
V11 V11a V11b	Sales to Total Assets Value for the specified year Percentage change	Uncertain	O, I
V12 V12a V12b	Gross Margin to Sales Percentage change Was percentage change > 10%?	Uncertain	I
V13 V13a V13b	Cash Sales Percentage change Was change < 0?	Uncertain	I
V14	Were any sales from acquisitions?	Positive	I
General Financial - Asset Composition (see Section 4.5.6 on page 101)			
V15	Current Assets to Total Assets	Positive	O
V16	Net Property Plant & Equipment (PP&E) to Total Assets	Negative	O
V17	Soft Assets to Total Assets	Positive	O
V18	Percentage Change in Assets other than Current Assets and Net PP&E to Total Assets	Positive	O
General Financial - General Accrual Measures (see Section 4.5.7 on page 102)			
V19	Total Accruals to Total Assets	Uncertain	O
V20	Were the specified and the prior year's Total Accruals > 0?	Positive	O, I
V21	Total Discretionary Accruals	Positive	O, I, R
V22	RSST (unadjusted) Accruals	Positive	O, I

ID	**Variable Name**	**Expected Direction of Association with Fraud**	**Framework Factor(s) (O/I/R/S)**
General Financial - Level of Debt and Financial Distress (see Section 4.5.8 on page 104)			
V23	Debt to Total Assets	Positive	O, I
V24	Debt to Equity	Positive	O, I
V25	Altman’s (1968) financial distress measure (Z-score)	Positive	O, I
V26	Four-period average of Times Interest Earned	Positive	O, I
General Financial - Performance and Profitability (see Section 4.5.9 on page 106)			
V27 V27a V27b	Return on Equity Value for the specified year Industry Average minus Specific Company	Uncertain	I
V28 V28a V28b V28c	Return on Average Prior Assets Value for the specified year Previous year Change	Uncertain	I
V29 V29a V29b	Holding Period Return One-year Previous One-year	Uncertain, probably positive	I
V30	Were analyst Earnings Per Share (EPS) forecasts achieved or exceeded?	Uncertain, probably positive	I
General Financial – Financing (see Section 4.5.10 on page 109)			
V31 V31a V31b	Were New Securities issued? Common Stock? Common Stock or Long-term Debt?	Positive	I
V32	Proportion of common stock that is newly issued	Positive	I
V33 V33a V33b V33c	Demand for financing Specific Value (ex ante) Was there demand (ex ante)? Cash from operating and investment activities	 Negative Positive Positive	I
V34	Were there operating leases?	Positive	R
Non-financial - Key Roles and Positions (see Section 4.5.11 on page 110)			
V35	Was the auditor a Big Six firm?	Negative	O
V36	Number of changes of audit firm in the most recent four financial statements	Uncertain	O, R
V37 V37a V37b	CEO Tenure Number of changes in the last three years	Uncertain	O or S

ID	Variable Name	Expected Direction of Association with Fraud	Framework Factor(s) (O/I/R/S)
V38	Has the CFO changed in the last three years?	Uncertain	O or S
V39 V39a V39b V39c	Composition/Holdings of the Board Number of Directors Percentage of Directors who are also Executives Percentage of Director shares owned by those who are also Executives	Positive	O
V40	Percentage of total shares owned by the CEO	Uncertain	I
Comparing Financial and Non-financial (see Section 4.5.12 on page 116)			
V41	Percentage change in the number of Employees minus percentage change in Total Assets	Negative	S
V42	Percentage change in Sales minus percentage change in the number of Employees	Positive	S
V43	Percentage Change in Sales to Employees: Specific Company minus Industry Average	Positive	S
Control variables (see Section 4.5.13 on page 117)			
V44	Company Age: Number of years since foundation	Negative	I
V45	Company Size: natural log of Total Assets	Uncertain	O
V46	Industry: Standard Industrial Classification (SIC) code starts with a 3?	Positive	O, I
V47 V47a V47b	Stock Exchange listed on NASDAQ stock exchange? New York Stock Exchange (NYSE)? (No to both V47a and V47b indicates American Stock Exchange [AMEX])	Negative	O
New variables - Macroeconomic indicators (see Section 4.5.14 on page 119)			
V48 V48a V48b V48c	Macroeconomic indicators Previous year's percentage change in annual real GDP Previous year's percentage change in annual retail sales Previous year's unemployment rate inverted	 Negative Negative Negative	O, I
New variables - Corporate governance indices (see Section 4.5.15 on page 120)			
V49 V49a V49b	Corporate governance indices G-Index E-Index	Positive	O
New variables - Industry complexity measure (see Section 4.5.16 on page 122)			
V50	Accounting complexity of the industry	Positive	O

4.5.2 Financial Variables: Specific Account - Accounts Receivable

Accounts receivable is a very difficult area to audit because of the subjective judgements involved as acknowledged in the Statement on Auditing Standards (SAS) 27 issued by the American Institute of Certified Public Accountants (AICPA). This difficulty to audit creates an opportunity to commit fraud. Prior academic research has also shown that accounts receivable is one of the accounts more likely to be fraudulently manipulated, particularly by premature revenue recognition that overstates accounts receivable (Fanning and Cogger 1998).

The variables found to be empirically significant along with the studies that presented these findings are:

V1 Accounts Receivable:

- V1a. Value for the specified year (Perols 2011);
- V1b. Percentage change (Green and Choi 1997; Lin et al. 2003);
- V1c. Binary variable indicating whether the percentage change is greater than 10%, because analysts and auditors commonly use a 10% change as a threshold for material change (Fanning and Cogger 1998; Perols and Lougee 2011).

V2 Percentage change in Accounts Receivable to Sales (Beneish 1997; Green and Choi 1997; Beneish 1999a; Feroz et al. 2000; Lin et al. 2003; Alden et al. 2012)[53]:

- Feroz et al. (2000) and Alden et al. (2012) used absolute change instead of percentage change and Green and Choi (1997) used Sales to Accounts Receivable (the inverse of this ratio), but these are not material differences and so are not considered separate variables, as is consistent with more recent research (Perols 2011).
- Feroz et al. (2000) also considered the four-period average and Fanning and Cogger (1998) considered the simple value, but these were not considered to be separate variables given the number of other studies supporting the percentage change.

[53] Beneish (1997, 1999a) actually used accounts receivable to sales this year divided by last year, but this only differs from percentage change by a constant of one; that is, a subtraction of one results in a percentage change. Consequently, this has not been considered as an additional variable.

- Bayley and Taylor (2007) use a variable that tries to capture the amount of manipulated revenue, which in essence analyses the change in sales relative to accounts receivable. Again, this variable was not chosen because of the number of other studies supporting the simpler percentage change variable.

V3 Percentage change in Accounts Receivable to Total Assets (Green and Choi 1997; Lin et al. 2003; Dechow et al. 2011)

- Dechow et al. (2011) actually used change in accounts receivable divided by average total assets, but this is considered an immaterial difference as both variables measure the change relative to total assets, which was the theoretical intention in all cases. Consequently, this is not considered as a separate variable.

Lin et al. (2003) and Green and Choi (1997) also stated that prior empirical research and auditing practice support the use of V1,V2 and V3. Feroz et al. (2000) used the ratio of accounts receivable to sales as a proxy for transactions that are difficult to audit because revenue recognition issues are inherent in the measurement of receivables, which is an extremely difficult auditing task. The importance of considering difficult to audit transactions as an explanatory variable is reinforced in the auditor-oriented red flags from SAS 53. Furthermore, as the size of accounts receivable increases relative to total assets (and also to sales) the likelihood of reaching the materiality threshold increases (Stice 1991) and consequently then so might the likelihood of fraud which requires materiality.

Expected direction (positive or negative) of association with financial statement fraud

It is expected that increases in accounts receivable compared to the prior period (V1), compared to sales (V2) and compared to assets (V3) indicate a higher likelihood of fraudulent manipulation to inflate revenue. For example, an increase in the accounts receivable to sales ratio could be the result of a relaxing of credit terms to customers, but it could also be the result of fraudulently overstated earnings (Beneish 1999a) as per the improper revenue recognition red flag stated above in Section 2.2.4.a.

4.5.3 Financial Variables: Specific Account - Allowance for Doubtful Accounts (AFDA)

There is an opportunity to commit fraud by manipulating the amount of allowance for doubtful accounts, which is a contra account to accounts receivable. The variables found to be empirically significant along with the studies that presented these findings are:

V4 Percentage change in AFDA (Green and Choi 1997; Lin et al. 2003)

V5 Percentage change in AFDA to Accounts Receivable (Lin et al. 2003; Green and Choi 1997)

V6 Percentage change in AFDA to Sales (Green and Choi 1997; Lin et al. 2003)

Lin et al. (2003) and Green and Choi (1997) found prior empirical research and auditing practice support the use of V4 through V5. V4 is included because of its involvement in the revenue cycle, which is very susceptible to fraud as established in the previous chapter. V5 and V5 measure the contra-asset account (AFDA) against its main account (sales and accounts receivable) to standardise any changes in AFDA so that only unusual changes are featured.

Expected direction (positive or negative) of association with financial statement fraud

Understating the true value of doubtful accounts falsely increases the net value of accounts receivable, and consequently it is expected that (unusual) decreases in AFDA are associated with a higher likelihood of fraud. This theory is supported by the red flag for low levels or lack of growth in AFDA being listed as red flags for multiple types of fraud in Section 2.2.4. Thus, the expectation is that lower values of V4-V5 will indicate a higher likelihood of fraudulent manipulation.

4.5.4 Financial Variables: Specific Account - Inventory

Inventory is another account more likely to be fraudulently manipulated because of the increased opportunity as a result of its being difficult to audit (Fanning and Cogger 1998). The variables found to be empirically significant along with the studies that presented these findings are:

V7 Change in Inventory over average Total Assets (Dechow et al. 2011; Whiting et al. 2012)

- Whiting et al. (2012) actually used a ratio of the specified year's inventory divided by the previous year's. The justification for this variable was solely about measuring the change in inventory, which V7 does. Consequently this is not a material difference and so is not considered a separate variable.

V8 Inventory to Sales

V8a. Value for the specified year (Kaminski et al. 2004)[54]

V8b. Change (Summers and Sweeney 1998)

- Inventory to current assets (Kaminski et al. 2004) is excluded because of its similarity to this variable, which is consistent with recent research by Perols (2011).

V9 A binary variable indicating whether the inventory method is Last-In, First-Out (LIFO) (Fanning and Cogger 1998)

Dechow et al. (2011) explain that the change in inventory is worth analysing because it is an accrual component of earnings, which are the components most used to fraudulently misstate earnings. Moreover analysing the change in inventory relative to total assets (V7) reduces bias from company size. The ratio of inventory to sales, and any change in this ratio, provides additional insight into whether the level of inventory is an accurate valuation or incorrectly includes obsolete inventory (Spathis 2002; Kaminski et al. 2004), which is a common type of fraud as listed above in Section 2.2.4. This is important because estimating obsolete inventory involves subjective judgement and consequently presents an opportunity to commit fraud.

In times of rising prices, using LIFO results in a higher cost of goods sold, and consequently reduces earnings. Using LIFO also results in lower values of inventory and consequently lower reported assets. There is an incentive for companies that commit fraud not to use LIFO, because the fraud is usually being committed to increase earnings and assets. Thus, using LIFO is a part of the new Suspicious Information factor. This theory was supported empirically by Fanning and Cogger (1998) and assumes prices are generally increasing, which was the case during their study and also during this research[55].

Expected direction (positive or negative) of association with financial statement fraud

Consistent with inventory red flags listed above in Section 2.2.4, it is expected that increases in inventory and so increases in V7 and V8 indicate a higher likelihood of financial

[54] There is also evidence from Greek data to support the use of this variable (Spathis 2002).

[55] Given enough data from periods of deflation, future research could study how financial statement fraud differs during times of falling prices, such as in 2009.

statement fraud. In addition, using LIFO (V9) is expected to decrease the likelihood of fraudulent manipulation.

4.5.5 Financial Variables: Specific Account - Sales

Sales is another account more likely to be fraudulently manipulated because of the increased opportunity as a result of its being difficult to audit (Fanning and Cogger 1998). Furthermore, the sales account is a key performance indicator for management and so it is also a pressure factor in fraud. The variables found to be empirically significant along with the studies that presented these findings are:

V10 Sales Growth

- Sales growth was chosen in preference to asset growth (with the exception of V10e), because of the large number of supporting studies. This was also the reason why sales growth was also chosen in preference to the ratio of market value to book value, which was used by Carcello and Nagy (2004) to measure company growth.

V10a. Percentage change (Beneish 1997; Green and Choi 1997; Beneish 1999a; Lin et al. 2003; Brazel et al. 2009)[56]

V10b. Percentage change: Specific Company minus Industry Average (Alden et al. 2012)

- Absolute change was used in prior research, but percentage change has been used here for consistency with other variables.

V10c. Previous year's percentage change (Erickson et al. 2006; Perols and Lougee 2011)

- Erickson et al. (2006) used the previous year's growth for fraudulent companies and the specified year's growth for legitimate companies, but as done by Perols and Lougee (2011) this research will be consistent across both types of firms because of its matched pair design.

[56] Beneish (1997, 1999a) actually used the ratio of this year divided by last year, but this only differs from percentage change by a constant of one; that is, a subtraction of one results in a percentage change. Consequently, this has not been considered as an additional variable. Additionally, Brazel et al. (2009) only found this variable significant as a control variable in one of their two models.

V10d. Four-year growth rate (Fanning and Cogger 1998)

- Rapid company growth (Bell and Carcello 2000) is excluded because of its similarity to this variable, which is consistent with recent research by Perols (2011). The three-year growth rate (Johnson et al. 2009) is also excluded because of its similarity to this variable.

V10e. Previous year's percentage change in total assets (Abbott et al. 2000; Skousen et al. 2009)[57]

- This variable measures assets growth instead of sales growth, but is still included in this section as the underlying rationale for its association to fraud is the same as for sales growth.

V11 Sales to Total Assets

V11a. Value for the specified year (Persons 1995; Fanning and Cogger 1998; Kaminski et al. 2004; Whiting et al. 2012)[58]

V11b. Percentage change[59]

V12 Gross Margin to Sales, also known as Gross Margin Percentage[60]

V12a. Percentage change (Green and Choi 1997; Beneish 1999a; Lin et al. 2003)[61]

V12b. Binary variable indicating whether the percentage change is greater than 10% (Fanning and Cogger 1998)[62]

---

[57] Abbott et al. (2000) actually measured the average percentage change over the two years prior, but it was not considered a different variable because the authors used V10e when data were not available for this variable.

[58] There is also evidence from Greek data to support the use of this variable (Kirkos et al. 2007).

[59] Percentage change has been included even though it was not used in prior studies because the justification still applies (as explained later in this subsection). Percentage change calculations also remove the size bias and have been useful for many other variables in the literature. Indeed, many other variables in this study involve them.

[60] Perols (2011) also stated that a paper by Chen and Sennetti published in 2005 supported the use of this variable, but this paper could not be obtained to verify it.

[61] Beneish (1997, 1999a) actually used the ratio of last year divided by this year, but this only differs from percentage change by a constant of one when the ratio is inverted; that is, a subtraction of one from the inverted ratio results in a percentage change. Consequently, this has not been considered as an additional variable.

[62] As was the case for V1c, this variable was chosen because analysts and auditors commonly use a 10% change as a threshold for material change (Fanning and Cogger 1998).

V13 Cash Sales

V13a. Percentage change (Dechow et al. 2011)

V13b. Binary variable indicating whether the change is less than 0 (Beneish 1997)

V14 A binary variable indicating whether a portion of sales is from an acquisition (Erickson et al. 2006; Brazel et al. 2009)

Blocher and Cooper (1988) found that the sales trend is useful for detecting fraud in the revenue cycle. Lin et al. (2003) and Green and Choi (1997) also found auditing practice and prior empirical research support for the use of V10 (Sales Growth). While sales growth does not directly imply fraud, it can result in more pressure on managers to maintain consistent growth and reach growth targets, and thereby increase the incentives to commit fraud (Beneish 1997; Summers and Sweeney 1998; Beneish 1999a; Erickson et al. 2006). Backed by findings from a government report and prior research, Beneish (1999a) states that controls in a company lag behind operations in times of high growth, which might increase the opportunity to commit fraud. This finding is also supported by Summers and Sweeney (1998) and Bell and Carcello (2000) who also found weak internal controls associated with fraudulent financial statements. More recently Johnson et al. (2009) also implied that it is more difficult to monitor firms with higher growth as measured by sales growth. They also found empirically that higher sales growth was associated with financial statement fraud and postulate that it was because of increased opportunity to commit and conceal fraud.

Higher sales growth is commonly expected to increase the likelihood of fraud, but lower sales growth can also indicate management that is under pressure to perform better and therefore might consider committing it (Fanning and Cogger 1998; Erickson et al. 2006). As a result of the varying hypotheses and prior results with sales growth, three variables measuring sales growth (V10a-d) in different ways are considered. V10e (asset growth) is also included here because the same growth-related opportunity and pressure factors just mentioned in this paragraph apply, even though it measures asset growth, as opposed to sales growth.

Fanning and Cogger (1998) found that companies operating with lower efficiency in terms of sales to total assets were more likely to commit fraud. An earlier study by Persons (1995) supports this hypothesis as she found that a lower ratio indicates management are less able to cope with competitive situations (as measured by their sales generating ability from assets), and consequently they might turn to fraud as a response to the pressure. As this hypothesis is equally applicable if the "lower ratio" is considered as relative to the year prior, the percentage change in sales to total assets has also been included in this study.

Changes in gross margin were listed above as red flags for multiple types of financial statement fraud in Section 2.2.4 because it is generally expected by industry professionals that the gross margin percentage remains fairly constant over time (ACFE 2013a). The reason for this is that while sales fluctuate with business cycles, the margin will stay fairly constant as costs fluctuate in a highly correlated way to sales. Citing prior research for further evidence, Beneish (1999a) goes further by saying declining gross margin is a signal of poor performance that might be associated with pressure to manipulate accounts.

It was expected that decreases in cash sales would indicate poor performance that would then increase the pressure to fraudulently inflate credit sales (Beneish 1997; Dechow et al. 2011). Beneish (1997) empirically supported this expectation by finding that a declining cash sales binary variable indicated a higher likelihood of fraud. However, more recently Dechow et al. (2011) found unexpected contrary results. They found the statistically significant result that positive changes in cash sales increased the likelihood of fraud. After further investigation, they found that the fraudulent companies were increasing in size overall and so cash sales increased simply as a flow-on effect. Furthermore, many companies were prematurely recognising revenue (particularly just prior to the end of reporting periods), which increased their cash sales, but without properly considering the ability for customers to return products.

Additionally, Erickson et al. (2006) and Brazel et al. (2009) both cite the same prior research which states that firms have additional incentives to overstate earnings (using sales) before acquisitions. The reason for this is to increase their share price prior to the acquisition in order to negotiate more favourable terms of acquisition.

Expected direction (positive or negative) of association with financial statement fraud

The expected direction of association is unclear between (cash) sales (V10 and V13) and the likelihood of fraudulent manipulation. It is initially expected that low or declining efficiency and declining gross margin percentage, and so lower values of V11, V12a and a true value of V12b, will indicate a higher likelihood of fraud. However, fraudulently inflated sales would result in higher values of these variables and so the direction of their association is also uncertain. Finally, V14 measures the presence of an incentive to commit financial statement fraud and so is expected to be associated with a higher likelihood of it.

4.5.6 Financial Variables: General - Asset Composition

In addition to variables and ratios involving specific accounts affected by fraud, the broad composition of assets on the balance sheet has been shown to be useful in fraud detection models. The variables found to be empirically significant along with the studies that presented these findings are:

V15 Current Assets to Total Assets (Persons 1995)

V16 Net Property, Plant and Equipment (PP&E)[63] to Total Assets (Fanning and Cogger 1998; Kaminski et al. 2004)

- Kaminski et al. (2004) actually used Gross, rather than Net, PP&E. However, this is not considered a material difference and a specific rationale was not supplied by the authors. Consequently, it is not considered a separate variable. The choice of Net PP&E in preference to Gross PP&E was made because it was found significant in a multivariate study, while Gross PP&E was only found significant in a univariate study.

V17 Assets other than Cash and Net PP&E (termed Soft Assets) to Total Assets (Dechow et al. 2011)

V18 Percentage Change in Assets other than Current Assets and Net PP&E to Total Assets (Beneish 1999a)[64]

As previously mentioned, accounts receivable and inventory are accounts commonly used in cases of financial statement fraud, and so larger proportions of these accounts relative to other accounts are thought to be associated with fraudulent manipulation of accounts (Persons 1995), the result of increased opportunity. V15 is included to test whether a combined measure of accounts receivable and inventory is better able to discriminate between fraudulent and legitimate cases. Although PP&E can be manipulated by over-capitalising costs, assets other than PP&E (such as accounts receivable and inventory) are

[63] Net PP&E is calculated as Gross PP&E minus Accumulated Depreciation.

[64] Beneish (1999a) actually used the ratio of last year divided by this year (termed Asset Quality Index), but this only differs from percentage change by a constant of one; that is, if one is subtracted from the ratio then the result is percentage change. Consequently, percentage change is used for consistency with many other percentage change variables.

more commonly manipulated and so lower proportions of PP&E (V16) might indicate a higher likelihood of fraudulently manipulated accounts.

When firms have more soft assets on their balance sheet, there is more discretion for management to change assumptions to meet short-term earnings targets (Dechow et al. 2011). However, there is also more opportunity to fraudulently manipulate figures to meet targets.

Assets other than current assets and Net PP&E are considered to have less certain future benefits, and an increased proportion of these assets (as measured by V18) indicate more of a disposition towards capitalising and consequently deferring costs (Beneish 1999a), which can be done fraudulently and so there is increased opportunity to commit fraud. High levels of capitalised expenses were also listed as a red flag for fraud in Section 2.2.4.b. Increases in V18 could also be a result of goodwill from acquiring a company, but Beneish (1999a) found similar results after removing the influence of goodwill.

Expected direction (positive or negative) of association with financial statement fraud

While increases in V15, V17 and V18 are expected to increase the likelihood of financial statement fraud, the opposite association is expected for V16.

4.5.7 Financial Variables: General - General Accrual Measures

There is a large body of literature going back to Healy (1985) that theorises earnings are primarily misstated by manipulating accruals (Dechow et al. 2011). While accrual components such as inventory and accounts receivable are represented in other sections, this section specifically focuses on variables involving multiple accounts that are specifically derived to measure accruals. The variables found to be empirically significant along with the studies that presented these findings are:

V19 Total Accruals to Total Assets (Beneish 1997, 1999a; Brazel et al. 2009; Whiting et al. 2012)

- Total accruals to last year's total assets or revenue (Lee et al. 1999; McKee 2009; Alden et al. 2012) are excluded because of their similarity to this variable. As Beneish (1997) found similar results when using discretionary accruals and total accruals (without dividing by total assets), discretionary accruals (Dechow et al. 1996; Beneish 1999b) were also excluded because of their similarity to this variable, which is again consistent with recent research by Perols (2011). For the same reasons, total accruals (Whiting et al. 2012) was not considered a separate variable. Similarly, working capital accruals to total assets (Bayley and Taylor 2007) was not considered a separate variable with the additional reason that Dechow et al. (2011) empirically found V22 below preferable. Finally, V19 and V20 were chosen in preference to change in total accruals (McKee 2009), because no theoretical justification rationale was provided for its inclusion.

V20 Binary variable indicating whether Total Accruals are positive in both the specified year and the year prior (Beneish 1997)

V21 Sum of the three prior years of Discretionary Accruals, also known as Total Discretionary Accruals (Perols 2011; Perols and Lougee 2011)

V22 RSST (unadjusted) Accruals as developed by Richardson et al. (2005) (Dechow et al. 2011)

Accruals can increase non-cash income, which should then result in future cash inflows for legitimate entries. However, being non-cash items, accruals are easier to fraudulently manipulate compared with cash accounts that are easier to audit and so higher accruals represent a greater opportunity to commit fraud. The ratio of total accruals[65] to total assets measures the extent that income is derived from anything other than cash from normal operations, and so a positive association is expected between it and fraudulent manipulation. The red flag *positive income and negative operating cash flows* listed for multiple types of financial statement fraud in Section 2.2.4 will also be captured in higher values of this variable, because total accruals measure the difference between income and operating cash flows.

Companies with fraudulent accounts have been found to have positive accruals leading up to and including the year the fraud was committed. Prior positive accruals reduce ways to legitimately manage earnings and so may increase the pressure to commit fraud if management attempts to avoid accrual reversals or maintain accrual growth as a means to improve earnings (Beneish 1997; Perols and Lougee 2011). V20 and V21 both consider prior accruals, the former measuring total and the latter measuring discretionary accruals. Furthermore, discretionary accruals might also indirectly measure management character and consequently cover some of the rationalisation aspect of the Fraud Detection Triangle. The

[65] Bayley and Taylor (2007) found that total accruals were better than various measures of unexpected accruals at identifying financial statement fraud.

theory behind this is that management's attitude towards discretionary accruals, which can be thought of as a legitimate method to manage earnings and performance, might also be an indication of their attitude towards fraud (Perols and Lougee 2011). The concept of using accruals to provide insight into the rationalisation factor is also supported by other studies (Skousen et al. 2009). In other words, high amounts of discretionary accruals might indicate poor management character that also predisposes them to be able to rationalise fraud, which makes them more likely to commit fraud.

Unadjusted RSST Accruals developed by (Richardson et al. 2005) consider prior periods in a different way, by focusing on the changes in the past year. It is a complex measure of accruals that considers the change in non-cash working capital (WC), the change in net non-current operating assets (NCO) and the change in net financial assets (FIN). In contrast to Perols (2011), Dechow et al. (2011) found unadjusted measures of accruals had more discriminatory power than discretionary accruals. By including both variables in this research, a contribution will be made by offering evidence towards clarifying these contrary findings.

Expected direction (positive or negative) of association with financial statement fraud

Prior positive or increasing accruals indicate that management has fewer ways to legitimately manage earnings, and so it is expected that V20 and higher values of V21 and V22 will indicate a higher likelihood of fraudulently manipulated accounts. However, Beneish (1997) casts doubt on the directional expectation of V19 because it found the unexpected empirical result that total accruals in the specified year were lower for fraudulent companies. This could possibly indicate accrual reversals occurring in the year the fraud is committed, rather than after the fraud as originally expected.

4.5.8 Financial Variables: General - Level of Debt and Financial Distress

Higher levels of debt increase the incentive and pressure to commit fraud. The variables found to be empirically significant along with the studies that presented these findings are:

V23 Debt to Total Assets (Persons 1995; Dechow et al. 1996; Beneish 1997; Lee et al. 1999; Brazel et al. 2009; Alden et al. 2012)

V24 Debt to Equity (Fanning and Cogger 1998; Erickson et al. 2006)[66]

V25 Altman's (1968) financial distress measure (Z-score) (Feroz et al. 2000; Erickson et al. 2006) [67]

V26 Four-period average of Times Interest Earned (Feroz et al. 2000)

More debt is positively associated with income-increasing accounting policies, but if this is not sufficient to avoid breaching debt covenants then there is pressure to fraudulently overstate assets or understate liabilities (Persons 1995). This hypothesis is supported by prior research that has used debt levels as a proxy for the existence and restrictiveness of covenants (Dechow et al. 1996). This all suggests that higher debt might be associated with higher likelihood of fraudulent misstatements (Fanning and Cogger 1998; Spathis 2002; Perols and Lougee 2011). Debt relative to both assets (V23) and equity (V24) have shown empirical promise in the past and so have both been included. Debt ratios are also positively related to the demand for equity funding (V33) (Dechow et al. 1996) and financial distress (V25) (Erickson et al. 2006) and so also share the rationale for including these variables.

When a company is performing poorly and under more financial distress there is increased motivation to commit fraud (Fanning and Cogger 1998; Spathis 2002; Beasley et al. 2010) to falsely improve the financial situation as presented in the financial statements. Previous research also proposes that a poor financial position might indicate weak controls that will make it easier for fraud to be perpetrated (Spathis 2002). Furthermore, a critique of an early financial statement fraud model suggested adding a financial distress variable (Jiambalvo 1996). The Z-score (Altman 1968) is the most widely used and cited financial distress model, and although individual components are already included as variables elsewhere (such as sales to total assets), including this composite variable (V25) allows for testing the relative contribution of an overall distress measure. Another way to consider

[66] Erickson et al. (2006) used different time periods for fraud companies and non-fraud companies, but just like Perols and Lougee (2011) this research will be consistent across both types of firms because of its matched pair design. There is also evidence from Greek data to support the use of this variable (Kirkos et al. 2007).

[67] Erickson et al. (2006) actually used an updated calculation of the Z-score (Begley et al. 1996), and Feroz et al. (2000) used a four-period average, but the standard Z-score was chosen here as it has long been the preference over newer updated versions. It is used in recent research by Perols (2011). Its use is supported by Greek data as presented by Spathis (2002); Kirkos et al. (2007), and it is also linked to litigation against auditors (Stice 1991). Erickson et al. (2006) also tried a newer financial distress model (Shumway, 2001) without finding any difference over the Z-score model. McKee (2009) also used his own bankruptcy model and found it to be a useful variable in their financial statement fraud detection models.

financial distress is to look at the difficulty changes in interest rates would cause in terms of servicing debt, which can be assessed using the times interest earned variable (V26)[68].

Expected direction (positive or negative) of association with financial statement fraud

More debt and financial distress increases the likelihood of fraud and so higher values of V23, V24, V25 and V26 are expected to indicate a higher likelihood of fraudulent manipulation. Although previous research has not found the opposite direction of association, it is possible. Higher debt levels might be accompanied by increased monitoring and scrutiny from creditors (those who are owed money), which would be consistent with agency theory (Jensen and Meckling 1976). This increased scrutiny would then likely reduce the opportunity to commit and conceal fraud.

4.5.9 Financial Variables: General - Performance and Profitability

Poor performance and profitability have been shown to be a useful factor in detecting financial statement fraud as it can be perceived as pressure to fraudulently improve financial statements to hide the poor results (Perols and Lougee 2011). The variables found to be empirically significant along with the studies that presented these findings are:

V27 Return on Equity ($\frac{\text{Net income}}{\text{Total common equity}}$)

V27a. Value for the specified year (Alden et al. 2012)

V27b. Industry Average minus Specific Company (Feroz et al. 2000; Alden et al. 2012)

- Feroz et al. (2000) actually used a four-period average, but consistent with more recent research by Perols (2011) the specified year is being used because obtaining historical information on return on equity figures for entire industries is considered to be difficult to obtain (without access to financial databases, which, for example, many investors do not have). Additionally, Alden et al. (2012) actually used the ratio of industry to specific company, but this was not considered a separate variable because of its similarity.

---

[68] The inclusion of variable V26 is also supported by Efendi et al. (2007) who found that the ratio of interest to income is empirically associated with the likelihood of financial statements being restated (which is related to, but different from, fraud as previously described in Section 2.2.2).

V28 Return on Average Prior Assets ($\frac{\text{Net income}}{\text{Average total assets}_{t-1}}$)

- Past research used either net income (Erickson et al. 2006; Perols and Lougee 2011) or income before extraordinary items (Summers and Sweeney 1998; Lee et al. 1999; Brazel et al. 2009; Dechow et al. 2011) in the numerator of this ratio. Net income was chosen as it is also used as the measure of return for V27 above.
- Past research has used a variety of measurements of assets in the denominator, including specified year (Erickson et al. 2006; Perols and Lougee 2011; Ndofor et al. 2013), previous year (Lee et al. 1999; Brazel et al. 2009) and average of the specified and previous year (Dechow et al. 2011). These two latter concepts of previous year and averaging over two years were combined to form an average of the two prior years of assets, which is the denominator used in this research. This creates a ratio between a current variable (in the numerator) and a past variable containing information from two prior years (in the denominator). This comparison across multiple years offers more information that might assist in detecting fraud (as discussed in Section 4.1.1), and is therefore the reason for the choice.

V28a. Value for the specified year (Lee et al. 1999; Brazel et al. 2009; Perols and Lougee 2011)

V28b. Previous Year (Summers and Sweeney 1998; Erickson et al. 2006)

V28c. Change (Dechow et al. 2011; Ndofor et al. 2013)

- Studying restatements[69], Ndofor et al. (2013) actually used the specified year subtracted from the year prior (rather than specified year minus year prior), but this only changes the interpretation and because there is no material difference this was not considered a separate variable.

V29 Holding Period Return

V29a. One-year (Beneish 1999b; Dechow et al. 2011)

V29b. Previous One-year (Dechow et al. 2011)

- Dechow et al. (2011) actually measured return in excess of the market return in the specified year and the year prior, but this was not considered a separate

[69] Restatements are related to, but different from, fraud as previously described in Section 2.2.2.

variable because this research focuses solely on US public companies that can be considered to be part of the same market.

- Beneish (1997) also measured return in excess of a size-matched, value-weighted portfolio, but this was also not considered to be a separate variable because of a preference for simplicity, and later research by the same person (Beneish 1999b) changed to use a standard holding period return.

V30 Binary variable indicating whether analyst forecasts were achieved or exceeded for earnings per share (Perols 2011; Perols and Lougee 2011)

Performance can be measured by accounting ratios as well as variables that consider market-based performance. Referring to the SAS 53 red flags and documentation from the Association of Certified Fraud Examiners, Feroz et al. (2000) and Alden et al. (2012) used return on equity (V27) as an accounting measure of performance, while many other studies successfully used variations on return on assets (V28). Both of these are standard accounting measures of profitability and performance. In terms of market-based performance, holding period return (V29) provides a measure of performance based on the share price. Market-based performance is often directly linked to management's performance indicators and remuneration. Management can be overly concerned with stock price performance, and the analyst forecasts that influence them. Undue emphasis on meeting earnings projections is an important factor in misstatements (Bell and Carcello 2000; Brennan and McGrath 2007) and whether analyst forecasts are achieved or exceeded (V30) might be a suitable publicly available proxy for this information.

Expected direction (positive or negative) of association with financial statement fraud

Poor performance has been shown to increase fraud on financial statements. However, particularly with market-based measures of performance, improved performance can actually be an indicator that fraud has occurred to cover up the truth of poor performance as suggested by the profitability growth red flags listed for multiple types of financial statement fraud in Section 2.2.4. Hence, the expected direction of influence is not certain for these variables, but based on prior findings it is more likely that increased market performance (V30 and increased V29) likely indicates a higher likelihood of fraudulent manipulation.

4.5.10 Financial Variables: General - Financing

Variables about financing in addition to the debt ratios already covered are considered. The variables found to be empirically significant along with the studies that presented these findings are:

V31 Binary variable indicating whether New Securities were issued (in the specified year)

- V31a. Common Stock (Dechow et al. 1996; Lee et al. 1999)
- V31b. Common Stock or Long-term Debt (Dechow et al. 2011)[70]

V32 Proportion of common stock that is newly issued (Dechow et al. 1996)

V33 Demand for financing

- V33a. Specific Value (ex ante) (Skousen et al. 2009)
- V33b. Binary variable indicating whether there is demand for financing (ex ante) (Dechow et al. 1996; Erickson et al. 2006)
- V33c. Cash from operating and investment activities (Skousen et al. 2009; Alden et al. 2012)[71]
  - This variable is related to the above demand for financing variables, and so is included using the same underlying rationale, which is consistent with prior research (Skousen et al. 2009).

V34 Binary variable indicating whether there are operating leases (Dechow et al. 2011)

A higher stock price reduces the cost of raising new funds. In the case of equity, it directly reduces the cost of raising funds, and in the case of debt it can result in more favourable contractual terms. To assist in attaining this higher stock price, management can boost earnings and consequently there is also increased incentive to fraudulently increase earnings. Raising new funds can be measured by assessing whether there was issuance during the year (V31) or the proportion of funds that are new (V32) with the view that the larger

[70] The inclusion of variable V31b is also supported by Efendi et al. (2007) who found that it was associated with a higher likelihood of financial statements being restated (which is related to, but different from, fraud as previously described in Section 2.2.2).

[71] Alden et al. (2012) actually used a ratio of this variable to total assets. However, the rationale for its inclusion was the study by Skousen et al. (2009) and so the version used by Skousen et al. (2009) was chosen for this study.

proportions increase the incentives. There might additionally be firms that wanted to raise funds, but chose not to as favourable terms could not be secured. V33 captures this issue as it measures whether firms have the incentive to raise new funds in the next two years because they are close to exhausting their internal funds (Dechow et al. 1996). However, it is generally considered that the SEC is more focused on companies actually issuing securities (Dechow et al. 2010) and consequently it is possible that this variable might not be significant as observed in Dechow et al. (2011).

Operating leases are the main source of off balance sheet finance and as opposed to capital leases their use lowers expenses in the earlier years of the lease (Dechow et al. 2011). Lower expenses result in higher earnings, and Dechow et al. (2011) postulate that the use of operating leases (V34) could be associated with managers who are more focused on short-term window dressing. The implication is then that managers with this disposition are more likely to commit fraud for the same short-term window dressing incentives and also because they are more likely to be able to rationalise the fraud to themselves by focusing on the short-term.

Expected direction (positive or negative) of association with financial statement fraud

Issuing or needing to issue new securities and the use of operating leases are associated with higher incentives to fraudulently inflate income and so higher values of V32 and V31, V33 (except for V33a for which lower values indicate more need for financing) and V34 will indicate a higher likelihood of fraudulently manipulated financial statements.

4.5.11 Non-Financial Variables: Key Roles and Positions

The non-financial variables consider the key roles and positions both within and external to the organisation. The variables found to be empirically significant, along with the studies that presented these findings and the theories behind their inclusion, are presented in this section.

V35 Binary variable indicating whether the auditor was a Big Six[72] firm (Fanning and Cogger 1998; Carcello and Nagy 2004; Perols 2011)

[72] The Big Six (or Four or Five) audit firms are the largest such international professional services firms that offer an auditing service.

- The Big Six Audit firms became the Big Five in 1998 when Price Waterhouse merged with Coopers & Lybrand, and subsequently the Big Four in 2002 with Arthur Anderson's demise. As my study spans both of these key dates, any firm in the Big Six is considered as a positive value in this binary variable.

Researchers have argued that Big Six auditors limit the opportunities to perform earnings management (Hogan et al. 2008) and it is possible that they might also limit the opportunities to commit financial statement fraud. Previous research also shows that the Big Six Audit firms have smaller under-pricings of initial public offerings and command higher fees for audits that possibly indicate they offer a higher quality of audit (Fanning and Cogger 1998). Consequently, Fanning and Cogger (1998) proposed that fraud might be less prevalent in statements audited by any of the Big Six as a result of a decreased opportunity to commit it.

V36 Number of changes of audit firm in the most recent four financial statements (Feroz et al. 2000; Perols 2011)

- This variable also captures (as a zero value) whether the auditor has been with the company for three or fewer years (Carcello and Nagy 2004).
- This variable was chosen in preference to the number of years the auditor has been the same (McKee 2009), because it could lead to biases related to the age of the company and multiple studies found this variable to be statistically significant.

Auditor firm turnover is listed as a red flag in the SAS 53 and above in Section 2.2.4.a, and Beasley et al. (2010) found more changes of audit firm in companies that had committed financial statement fraud relative to those that did not. Stice (1991) also found that litigation against auditors decreases after they had the same account for more than three years, with the theory being that there is a costly learning curve when auditing a new set of accounts and that results in lower quality audits. Myers et al. (2003) made supporting findings that the quality of the audit improves as the auditor tenure increases, because the longer-term auditors reduce extreme management decisions regarding production of financial statements. This theory is further supported by Carcello and Nagy (2004) who found more fraud occurs in accounts that have had the same audit firm for three or fewer year when analysing an eleven year period from 1990 to 2001 in the US. In addition to a lower quality audit, changing audit firm could be related to the opportunity-seeking behaviour known as "opinion shopping", which is searching for auditors that will give favourable findings in preference to auditors who may have concerns of fraudulent manipulation[73]. Management who is engaged in opinion shopping might also indicate an attitude and level of personal integrity prone to make it easier for them to rationalise fraud. However, there is also the contrary argument that a lack of change leads to complacency and overconfidence, which might result in lower quality audits that provide more opportunity to commit fraud (Fanning and Cogger 1998). As a result of the contrary theories, the expected direction of association is uncertain.

V37 Chief Executive Officer (CEO)

V37a. Tenure (Ndofor et al. 2013)[74]

V37b. Number of changes in the last three years (Feroz et al. 2000)

- These variables were chosen in preference to a similar variable indicating whether the CEO was the founder (Dechow et al. 1996; Erickson et al. 2006), which is consistent with recent research by Perols (2011).

Companies in which the founder is the CEO are more prone to financial statement fraud, because company founders are likely to have more of an opportunity as they have greater influence without scrutiny and are less accountable to the board of directors (Dechow et al. 1996). Finkelstein (1992) also states that longer serving CEOs have more power and influence. These findings actually support each other because founding CEOs are likely to be longer serving CEOs, because there have been no changes in CEO. Because these rationales are not directly concerned with founding CEOs but rather about entrenched CEOs (some of whom happen also to be founding CEOs), variables assessing CEO tenure and changes in CEO are chosen in preference to a variable indicating whether the CEO is a founding CEO. Furthermore, it is possible that more entrenched CEOs also have additional knowledge about the company that will make them more capable of committing fraud.

[73] Lennox (2000) provides a discussion of whether companies successfully conduct opinion shopping using data from the UK.

[74] Erickson et al. (2006) also found a variable indicating whether there were missing data for CEO tenure to be useful, but that variable is reliant on the way the databases collect information, rather than on publicly available information. In addition, Ndofor et al. (2013) study restatements, which are related to but different from financial statement fraud.

Fanning and Cogger (1998) present the contrary argument that CEO changes might indicate suspicious behaviour that increase the risk of financial statement fraud, because a board of directors might dismiss a CEO if they suspect fraud before it is publicly discovered with the view to keep it from public knowledge. This hypothesis is also supported by Land (2010) who finds a significant association between CEO changes and companies being accused of committing fraud. CEO change has been studied over a one-year and three-year prior period, but the three-year variable was found to have the greater discriminatory power (Fanning and Cogger 1998).

As a result of the contrary theories, the expected direction of association is not certain.

V38 Binary variable indicating whether the Chief Financial Officer (CFO) has changed in the last three years (Fanning and Cogger 1998)

- The number of CFO changes (Feroz et al. 2000) is excluded because of its similarity to this variable, which is consistent with recent research by Perols (2011).

As is the case with CEO rationales, a longer-term, more entrenched CFO might have more power without scrutiny and consequently a greater opportunity to commit fraud. Again as is the case with CEO rationales, a more entrenched CFO might also have additional knowledge of the company that makes them more capable of committing fraud. Contrastingly, a CFO change could also be the consequence of suspicions that the CFO has been committing fraud, with the idea that if the CFO is removed perhaps this will stop the fraud and keep it from ever becoming public knowledge. Similarly, the CFO might choose to leave if concerned the fraud will soon be discovered. Therefore, as with CEO variables, the expected direction of association is uncertain.

V39 Composition and holdings of the board of directors

V39a. Number of Directors (Beasley 1996; Carcello and Nagy 2004)

V39b. Percentage of the Directors who are also Executives (Dechow et al. 1996)[75]

- Percentage of outside directors (Beasley 1996; Fanning and Cogger 1998; Carcello and Nagy 2004; Uzun et al. 2004) and whether there are more than 50% insiders on the board (Dechow et al. 1996) are excluded because of similarity to

[75] A person holding multiple executive positions is only counted once, and this is the same for directorships.

this variable, which is consistent with recent research by Perols (2011). The number of outside directors (Ndofor et al. 2013), the percentage of directors with any business relationship with management (Beasley 1996; Uzun et al. 2004) and the percentage of independent directors (Beasley 1996; Uzun et al. 2004) are also excluded because of their similarity[76]. Furthermore, whether the CEO is also the chairman of the board of directors (Carcello and Nagy 2004; Erickson et al. 2006; O'Connor et al. 2006; Skousen et al. 2009) has been excluded because the reasoning is based on CEO influence or CEO power, which is covered by this variable and the previous CEO variables.

V39c. Percentage of Director shares[77] owned by those who are also Executives (Dechow et al. 1996)

- Percentage held by outside directors (Beasley 1996) is excluded because of its similarity to this variable, which is consistent with recent research by Perols (2011). Similarly, percentage held by insiders and percentage held by insiders with more than 5% ownership (Skousen et al. 2009) is also excluded because of its similarity to this variable and also V40 (that is discussed next).

- Data from the specified year will be used as is consistent with all the research cited above in support of V39 except Dechow et al. (1996) who used the previous year's data.

The results from prior research are inconclusive when it comes to the influence of board size (that is, the number of directors), and this includes a study that has linked smaller boards to company failure and a study that finds board size irrelevant to company performance (Fanning and Cogger 1998). A study of restatements found that larger boards increased the chance of restatement of the financial statements (Abbott et al. (2004). Additionally, the most relevant study showed that increases in the board size increase the likelihood of financial statement fraud (Beasley 1996), which is consistent with the theoretical view of Jensen (1993) that smaller boards provide a better controlling function,

[76] Determining whether directors are truly independent or have a business relationship with management is also too complex to fit with the widely applicable aim for this model.

[77] The total number of shares was used because it is more easily accessed compared with the number of unrestricted shares that were found by Johnson et al. (2009) to be more appropriate measures of incentive to commit fraud. Unrestricted, as opposed to restricted, shares do not have any restrictions on their ability to be sold.

reducing the opportunity to commit fraud. This is the directional association that is expected in this research.

Monitoring company management is important because without costs being associated with behaviour that is detrimental to the company, management act more in their personal interests (Fama and Jensen 1983). A specific example of this is that they might commit financial statement fraud (Fanning and Cogger 1998) to increase their annual bonus or other forms of remuneration. Research has shown that boards with more outsiders are more effective at monitoring management (Weisbach 1988; Beasley 1996). The rationale is that people who are not executives and are more independent will increase the monitoring function of the board and consequently reduce the opportunities for fraud to be committed. Also, Dechow et al. (1996) refer to prior research that claims directors who are not executives, but do have a large equity holding in the company, play a notable monitoring role. Thus, these governance variables (V39b and V39c) are an inverse proxy for board vigilance. Larger proportions of, and holdings by, directors who are also executives are expected to be associated with more financial statement fraud.

V40 Percentage of total shares[77] owned by the CEO (Ndofor et al. 2013)

Ndofor et al. (2013) state that the primary method of aligning the desires of management and shareholders is through ownership of the company by management. This alignment of desires should decrease the incentive to commit fraud. Consistent with agency theory, both academics and industry professionals state that increasing the level of ownership by top management improves the alignment between the incentives of shareholders (the owners or principals) and management (the agents), and this consequently decreases the likelihood of financial misconduct, such as restatements (Ndofor et al. 2013) or financial statement fraud. However, there are also contrary findings that suggests that higher concentration of ownership by top management might unintentionally increase aggressive and possibly fraudulent accounting practices (Loebbecke et al. 1989; Desai et al. 2006; O'Connor et al. 2006; Ndofor et al. 2013). The reason hypothesised for this is by fraudulently improving performance, CEOs can increase the short-term value of their ownership-based remuneration. Consequently there is no clear directional expectation associated with this variable.

4.5.12 Comparison Variables: Financial and Non-Financial

It is common that financial figures (such as sales) and non-financial figures (such as number of employees) will grow or shrink together. However, because it is often difficult to fraudulently manipulate both categories simultaneously, discrepancies between the growth rates of these two categories can be considered suspicious and indicate fraud. The variables found to be empirically significant along with the studies that presented these findings are:

V41 Percentage change in the number of Employees – percentage change in Total Assets (Dechow et al. 2011)

V42 Percentage change in Sales – percentage change in the number of Employees (Brazel et al. 2009)

V43 Percentage Change in Sales to Employees: Specific Company minus Industry Average (Perols and Lougee 2011; Perols 2011)

Basic economics tells us that inputs into production such as the number of employees are correlated with the outputs from production such as the level of sales (Perols and Lougee 2011). Brazel et al. (2009) support this theory using a case study in which they point out it is unlikely for substantial decreases in employee numbers to correspond to increases in sales. Because the number of employees is easier to audit and consequently more difficult to fraudulently manipulate (Brazel et al. 2009), changes in the difference between employees and revenue indicate a higher likelihood of fraud. Dechow et al. (2011) also point out that the argument also applies to the employees and company assets, not just employees and sales, as managers fraudulently overstating assets are likely not to be correspondingly manipulating employee numbers.

Expected direction (positive or negative) of association with financial statement fraud

Increases in the financial values (sales, assets) without corresponding increases in the non-financial (employees) raises the suspicion of fraud, and so higher V42 and lower V41 are expected to indicate higher likelihood of fraudulent manipulation. Furthermore, increases above the industry average as measured by V43 are expected to indicate a higher likelihood of fraud.

4.5.13 Control Variables

When developing the data set, companies will be selected by a *matched pair* process that matches legitimate companies to known fraud companies. The matching will be done on their age, size, industry and the stock exchange on which they are listed. These features will also be included as explanatory variables as has been done in other studies, for example, both Perols and Lougee (2011) and Erickson et al. (2006) included total assets in their analysis even though they matched companies based on their size in terms of assets. While these are unlikely to be statistically significant because they will not be good differentiators given the matching process, they are still included for two reasons: (1) the matching process is not exact and more importantly (2) their inclusion allows for the possibility that these control variables might be significant when they interact with other variables. For example, Dechow et al. (2011) found that the influence of the variable V17 (Assets other than Cash and Net PP&E [termed Soft Assets] to Total Assets) varied with the industry.

Some justification for the importance of these variables is included here with more justification in Section 5.1.2 on page 136 that further details the matching process.

V44 Company Age, measured as the number of years since foundation (Lee et al. 1999; Carcello and Nagy 2004; Erickson et al. 2006; Brazel et al. 2009)[78]

This variable controls for the financial incentive to commit fraud in younger firms around the period of the initial stock offering (Erickson et al. 2006). Additionally, Beasley (1996) suggests that younger firms might not have the controls established to adhere to reporting requirements, a point supported by Abbott et al. (2004) who empirically found that younger firms were associated with more restatements.

V45 Company Size, measured by the natural log of Total Assets (Persons 1995; McKee 2009)

V46 McKee (2009) actually used base 10 log, but this was not considered to be a significant difference.

[78] Lee et al. (1999) actually calculated age as the number of years the company had data in the Compustat database, but this was not considered a significant difference and was also not chosen because it cannot be used by stakeholders who do not have access to Compustat. Additionally, Erickson et al. (2006) and Carcello and Nagy (2004) used the length of time the company has been publicly traded.

Smaller companies usually have weaker internal controls (Lou and Wang 2009) that might increase the opportunity to commit fraud. Correlation between company size and fraud-related litigation has been found (Bonner et al. 1998) and fraud has been found to be more prevalent in smaller firms (Persons 1995). However, the SEC may have a bias towards prosecuting larger companies.

V47 Industry, specifically a binary variable indicating whether the Standard Industrial Classification (SIC) code starts with a 3 (Lee et al. 1999)

- SIC codes starting with a 3 represent a subset of the manufacturing sector that comprises the manufacture of rubber and plastic products; leather and leather products; stone, clay, glass and concrete products; metals and metal products; industrial and commercial machinery, and computer equipment; electronic and electrical products; transportation equipment; instruments of measuring, analysing and control; and miscellaneous products.

This variable was included to control for the fact that industry membership is a factor affecting the likelihood of fraud (Summers and Sweeney 1998), likely because of differences in the opportunities and incentives to commit fraud. This variable was chosen in preference to other industry membership variables as it was found to produce superior results by Lee et al. (1999) who found fraud more prominent in the industry group of SIC codes starting with a three, which are primarily computer companies.

V48 Stock Exchange (Lee et al. 1999)

V48a. Binary variable indicating whether the company was listed on the NASDAQ stock exchange

V48b. Binary variable indicating whether the company was listed on the New York Stock Exchange (NYSE)

- A zero in both of these variables indicates the company was listed on the only other exchange in the sample, the American Stock Exchange (AMEX). These variables will actually be combined into one categorical variable for those techniques that can handle categorical variables.

This variable controls for the fact that audit requirements affecting the opportunity to commit fraud vary across stock exchanges. It will also pick up any differences in the type of companies that decide to list on each exchange, possibly based on the differences in the opportunity factor. For example, Lee et al. (1999) found less fraud amongst firms listed on the AMEX. Jiambalvo (1996) provides further support for the inclusion of these variables as binary variables were used to indicate the listing exchange, but this study did not conduct any significance tests on the variables.

4.5.14 New Variables: Macroeconomic Indicators

The ACFE (2009) revealed that most industry experts believe fraud increased during the recent economic slump, a belief that was supported by a Brisbane (Australia) partner of Deloitte in a question and answer session[79]. Furthermore, Rezaee and Riley (2010) imply that more financial statement fraud occurs in economic recessions (which can be measured by macroeconomic indicators). It is intuitive that poor economic conditions might increase the opportunity to commit fraud because control systems are weaker as a result of more emphasis upon the economic downturn. Consequently, the inclusion of macroeconomic indicators is proposed to allow probability of fraud to vary depending on macroeconomic conditions. This will allow for overall macroeconomic conditions mentioned as an opportunity factor by Romney et al. (2013) to be incorporated into the model. However, this variable is also hypothesised to be involved with the pressure factor as varying economic conditions will influence the amount of pressure managers perceive they are under. For example, when the economy is in a poor state managers might feel more pressure in line with the more pessimistic view that people hold.

This research and most previous studies use matched observations by year, which means that a matched pair of fraud and legitimate companies will always have the same macroeconomic indicator and so cannot be differentiated by a macroeconomic indicator. However, the inclusion of the macroeconomic indicators does allow for the probability of fraud to differ with economic conditions, which could be important when classifying companies based on the probability that fraud has occurred.

The following widely reported co-incident indicators of macroeconomic health will be included as explanatory variables[80]:

[79] Graham Newton was a speaker at the ACFE's Brisbane chapter's fraud lunch on November 14, 2012.

[80] As this research only deals with data from the US, it does not need to be concerned with differences in the quality and accuracy of macroeconomic data from country to country.

V49 Macroeconomic Indicators

V49a. Previous year's percentage change in annual real GDP

V49b. Previous year's percentage change in annual retail sales

V49c. Previous year's unemployment rate inverted

The previous year is used because at the time of release of the current year's financial statement, the current year's macroeconomic indicators might not be available. While the unemployment rate is usually considered a lagging indicator, it can also be used as a co-incident indicator because people lose or gain confidence at the time when higher or lower unemployment rates are published (Layton et al. 2012).

Lower values in any of these variables indicate deteriorating macroeconomic conditions and so this is expected to be associated with more financial statement fraud, a concept supported by Warren Buffet's famous statement, "Only when the tide goes out do you discover who's been swimming naked"[81]. The nature of fraud might also change with the cycle of the economy (Dorminey et al. 2012), in which case variable V48 would contribute when it interacts with other variables. Consequently, decision-tree techniques that intrinsically model interactions between explanatory variables will be analysed and if interactions are found between V48 and other variables, then new interaction variables will be added to the initial set of explanatory variables available to other model building techniques.

4.5.15 New Variables: Corporate Governance Indices

V50 Corporate Governance Index

V50a. G-Index

V50b. E-Index

Past research has shown that weak corporate governance increases the likelihood of financial statement fraud (Hogan et al. 2008). The underlying proposition is that less monitoring means more opportunity to commit fraud. While a variety of individual corporate governance variables such as the percentage of executives on the board have been tested,

[81] It is possible that the opposite direction of association will be found as Wang et al. (2010) suggest that we need to be vigilant against fraud in economic booms as well.

there has been little empirical testing of corporate governance indices for use in fraud detection models. However, Kent et al. (2012) point out that there is support for both individual variables and indices for measuring corporate governance in the accounting literature.

The hypothesis to be tested is that financial statement fraud is related to the aggregation of many corporate governance initiatives, more than the presence or absence of any particular one. The most prominent index is the G-index (Gompers et al. 2003), which aggregates the existence of 24 governance provisions[82]. The G-index will be added as an explanatory variable along with the more recently developed E-index (Bebchuk et al. 2009) that uses the six (out of the 24) provisions that were found to be the most important[83]. Specifically, the 18 unselected provisions were found to be uncorrelated with either reduced firm valuation or negative abnormal returns index (Bebchuk et al. 2009).

Both indices are calculated simply as the sum of the individual provisions. The existence of each provision that restricts shareholder rights (and increases managerial power) is represented as 1, while the absence of that provision is represented as 0. While this method of calculation does not reflect the relative influence of each provision, it has advantages of transparency and objectivity. Higher scores in either index indicate lower governance and monitoring, and weaker shareholder rights. A G-index value less than 6 is referred to as a Democracy, compared with a Dictatorship if it is greater than 13 (Gompers et al. 2003).

One study (Johnson et al. 2009) used logistic regression on a matched-pair sample and did not find either index statistically significant, but the sample was only 49 matched pairs and this result is yet to be confirmed or contrasted using other supervised learning methods. Erickson et al. (2006) also found the G-index not to be significant in a logistic regression model, but again it was only one model and only used a limited number (50) of fraud cases.

[82] The 24 provisions are Antigreenmail, Antigreenmail laws, Blank check preferred stock, Business combination laws, Limits to shareholder bylaw amendments, Limits to charter amendments, Control share cash-out laws, Staggered boards, Compensation plans with changes-in-control provisions, Director indemnification contracts, Cumulative voting, Directors' duties provisions, Directors' duties laws, Fair-price provisions, Fair price laws, Golden parachutes, Director indemnification, Limitations of director liability, Pension parachutes, Poison pills, Secret ballot (or confidential voting), Executive severance agreements, Silver parachutes, Special meeting limitations, Supermajority requirements for mergers and charter amendments, Control-share acquisition laws, Unequal voting rights and Limitations on action by written consent.

[83] The E-index is made up of the following subset of 6 provisions: Staggered boards, Limits to shareholder bylaw amendments, Poison pills, Golden parachutes, and Supermajority requirements for mergers and charter amendments.

Because these indices have not been comprehensively tested, both indices will be tested more thoroughly by their inclusion as explanatory variables in this study. Furthermore, higher values of V49a and V49b represent increased opportunity to commit fraud and so are expected to be associated with more fraud.

4.5.16 New Variables: Industry Complexity Measure

V51 Accounting complexity of the industry

The hypothesis being tested is that financial statement fraud is different and more prevalent in a more complex industry, presumably because there are more ways to attempt to cover up the fraud and fewer people who understand the details. A recent measure of an industry's accounting complexity (Seavey 2011; Francis and Seavey 2012) will be included and tested as an explanatory variable, in addition to industry control variable V47.

Quantifying the complexity of an industry is a difficult task and variable V51 has been chosen as the desirable measure of complexity because it is calculated as the mean of ten different complexity measures. V51 was also chosen as it attempts to specifically measure the accounting complexity of an industry as specified by its two-digit SIC code, which is preferential to using an industry concentration measure as a proxy for complexity as done by Ndofor et al. (2013)[84]. Table 4-6 below provides a mapping between 65 two-digit SIC codes and V51 complexity values, where a higher value indicates increased complexity. Further details about the construction of V47 are provided by Seavey (2011).

It is also possible that the complexity of the industry will change the other fraud indicators in which case the V50 variable will be significant as it interacts with other variables. Consequently as with the macroeconomic variables, decision tree techniques that intrinsically model interactions will be analysed and if interactions are found between V50 and other variables, then new interaction variables will be added to the initial set of explanatory variables available to other model building techniques.

[84] Ndofor et al. (2013) also included the number of firms in the industry as an explanatory variable for explaining restatements, but the only reason they did this was as a control for the industry concentration variable. However, as I am not using this proxy for industry complexity, I have excluded the number of firms in an industry.

Table 4-6. The accounting complexity measure for each industry based on its two-digit SIC codes as per Seavey (2011).

2-digit SIC code	Complexity Measure	2-digit SIC code	Complexity Measure
01	35.9	41	25.4
02	39.3	42	21.8
07	33.7	44	36
08	29.3	45	29.9
09	32.9	46	34.5
10	40	47	25.3
12	40.3	48	40
13	39.6	49	32.9
14	33.8	50	23.6
15	36.9	51	28.9
16	29.2	52	20.5
17	30.7	53	17.3
20	35.1	54	15.7
21	44.3	55	25.7
22	31.2	56	18.2
23	29.4	57	17.6
24	26.3	58	20.9
25	31.3	59	27.5
26	36.1	70	30.2
27	33.6	72	33
28	45.4	73	42.6
29	40.2	75	32.6
30	35.4	76	36.2
31	29.7	78	35.5
32	38.5	79	33.9
33	40.9	80	31.9
34	35	81	30.4
35	41.6	82	33
36	42	83	23
37	40.1	87	36.1
38	41.2	89	42.9
39	39	99	40.3
40	26.8		

4.5.17 Excluded Variables

This subsection presents and discusses variables that have been excluded from this study. If stakeholders such as regulators or auditors have information in addition to the public information included in this study, then refinements to the models developed in this study could be made for particular stakeholders with additional information, with a view to enhancing the ability to detect financial statement fraud. Consistent with the statements by Feroz et al. (2000) about their models, the models developed in this research represent a minimum level of performance that is obtainable to all users, because refinements given additional data would only be made if they improved performance.

As stated at the beginning of this chapter this research aims to produce findings that are widely applicable to investors, regulators, auditors and other stakeholders, and consequently variables that are relatively difficult to obtain are excluded as they are less likely to be used in a practical context. The following variables have been found to be significant in prior research, but are excluded primarily because data on them is difficult to obtain in practice, which is consistent with the aim of this research and similar recent research by Perols (2011) and Dechow et al. (2011):

- Insider trading factors (Summers and Sweeney 1998; Beneish 1999b) that are not easily obtainable in practice;
- Retained earnings to assets and Market value of Equity to Assets (Lee et al. 1999) are also excluded because they were found to be significant in a test that used delisted companies[85] as the comparison to fraudulent companies, and this is not the focus of this research;
- Whether management lies to or is evasive with auditors and whether there is a weak internal control environment (Bell and Carcello 2000), which requires information from internal audit reports that are not publicly available;
- The standard deviation of the volatility of returns over the past 60 months (Erickson et al. 2006), which requires a large amount of data in addition to the financial statements;
- The difference between revenue growth and the growth of a complex non-financial measure that Brazel et al. (2009) estimated using university students to imitate the auditing practice;

[85] A delisted company is a company that is removed from the exchange on which it trades.

- Corporate lobbying factors (Yu and Yu 2011) which are not easily obtainable information in practice;
- Information from analysts in addition to their estimates, specifically whether and when coverage or recommendations are dropped for firms, which were related to specific types of financial statement fraud (Cotter and Young 2007);
- Text analysis (Glancy and Yadav 2011) and linguistic cues, for example, the use of more pleasantness and less lexical diversity being associated with fraud (Humpherys et al. 2011);
- The proportion of the proceeds from newly issued equity that went to insiders in the company, whether the company has a bonus plan and whether managers recently redeemed their stock appreciation rights (Beneish 1999b), as well as the unexercised options (Ndofor et al. 2013) or average stock options (O'Connor et al. 2006) the CEO owns[86].
- Information on committees under the board of directors and additional information about the directors[87], specifically variables indicating the:
    - Existence of an audit committee (Dechow et al. 1996; Uzun et al. 2004) and the compensation committee (Dechow et al. 1996)
    - Composition of the audit committee (Abbott et al. 2000; Abbott et al. 2004; Uzun et al. 2004; Johnson et al. 2009; Skousen et al. 2009; Ndofor et al. 2013) and the compensation and nominating committees (Uzun et al. 2004)[88]
    - Number of meetings held by the audit committee (Abbott et al. 2000; O'Connor et al. 2006; Ndofor et al. 2013), as well as the number of meetings held by the overall board of directors as this information is not reliably made public (Johnson et al. 2009; Ndofor et al. 2013)
    - The number of additional directorships held by outside directors and the average number of years outside directors have spent on the board (Beasley 1996)

[86] Burns and Kedia (2006) also found the sensitivity of the CEO's option portfolio to stock price and Efendi et al. (2007) found CEO compensation measures using salary and options important variables in explaining restatements, which are related to, but different from, fraud as previously described in Section 2.2.2.

[87] It is worth stating that one high profile study (Beasley et al. 2010) claims that there is little evidence that audit committee characteristics are associated with financial statement fraud.

[88] Beasley (2000) finds variables such as audit committee composition to vary between fraud and legitimate companies in certain industries, but no financial statement fraud detection model was constructed.

- The age of the CEO (O'Connor et al. 2006);
- Deferred tax variables (Ettredge et al. 2008);
- Company complexity as measured by a diversification index that requires the analysis of market share for each industry segment the company operates within (Ndofor et al. 2013), which is information not easily obtainable in practice.

4.5.18 Definition of Terms

The first table defines general terms, while the second defines terms specific to particular variables.

Table 4-7. Definitions of general terms.

General Term	**Definition**
Subscript $_{t}$, $_{t-1}$…	The specified year, previous year…
Change (absolute)	$X_t - X_{t-1}$
Percentage change	$(X_t - X_{t-1})/ X_{t-1}$
Average	$(X_t + X_{t-1})/2$
Four-year growth rate	(X_t/ X_{t-3})^0.25 -1
Binary variable	Coded 1 for 'yes', 0 for 'no'
Four-period average	$(X_t + X_{t-1} + X_{t-2} + X_{t-3})/4$
Industry average	Calculated as the average of the constituents of the S&P500[89] at the end of the specified year that have the same two-digit industry Standard Industrial Classification (SIC) code.
X inverted	1 / X

Table 4-8. Definitions of terms specific to particular explanatory variables. The data sources are listed below in Section 4.6.

Specific Term	**Definition**	**Variable ID**
Gross Margin	Sales – Cost of Goods Sold	V12
Cash Sales	Sales – change in Accounts Receivable	V13
Were any Sales from Acquisitions?	IF (Mergers & Acquisitions Transactions during previous year > 0) THEN “yes” ELSE “no”	V14
Soft Assets	Assets other than Cash and Net PP&E	V17

[89] The constituents of the Wiltshire 5000 were preferred, but this research was not able to acquire this data historically.

Specific Term	**Definition**	**Variable ID**
Total Accruals	Income before extraordinary items – Net cash flow from operating activities ($\mathrm{NCF_{OA}}$)[90]	V19
Discretionary Accruals	$\frac{\text{(Total Accruals + Change in Accounts Receivable − Changes in Sales − Change in } \mathrm{NCF_{OA}} \text{ − Gross Property Plant and Equipment − 1)}}{\text{prior year's Total Assets}}$	V21
Total Discretionary Accruals	Discretionary Accruals$_{t-1}$ + Discretionary Accruals$_{t-2}$ + Discretionary Accruals $_{t-3}$	V21
RSST Accruals	$\frac{\text{Change in WC + Change in NCO + Change in FIN}}{\text{Average Total Assets}}$ (Richardson et al. 2005)	V22
…WC	(Current Assets – Cash & Short-term Investments) – (Current Liabilities – Debt in Current Liabilities)	
…NCO	(Total Assets – Current Assets – Investments and Advances) – (Total Liabilities – Current Liabilities – Long-term Debt)	
…FIN	(Short-term Investments + Long-term investments) – (Long-term Debt + Debt in Current Liabilities + Preferred Stock)	
Z-Score: Altman's (1968) financial distress measure	$3.3 \times \frac{\text{Income before taxes and interest}}{\text{Total Assets}} + 0.999 \times \frac{\text{Sales}}{\text{Total Assets}} + 0.6 \times \frac{\text{Market Value of Equity}}{\text{Total Liabilities}} + 1.2 \times \frac{\text{Working Capital}}{\text{Total Assets}} + 0.4 \times \frac{\text{Retained Earnings}}{\text{Total Assets}}$	V25
Times Interest Earned	Income before taxes and interest / Interest Expense	V26
Return on Equity	Net Income / Total Common Equity	V27
Return on Average Prior Assets	Net Income / Average Total Assets$_{t-1}$	V28
One-Year Holding Period Return	$\frac{\text{Price at Balance Sheet Date − Price One Year Ago}}{\text{Price One Year Ago}}$ Following prior research by Beneish (1999b), if the "Price at Balance Sheet Date" is not available because the company is delisted the value of V29a is set to -100%.	V29a
Were analyst EPS forecasts achieved or exceeded?	IF (Actual EPS >= Estimated EPS) THEN "yes" ELSE "no" where EPS = Earnings Per Share	V30

[90] Although there are a variety of ways to estimate accruals (Lee et al. 1999), this research follows previous research (Beneish 1997; Lee et al. 1999; Perols 2011) by calculating Total Accruals as Income before extraordinary items - Net cash flow from operating activities (NCF_{OA}). As further support this calculation method (Beneish 1997) found similar results in a financial statement fraud modelling situation when total accruals were calculated using a different, longer method similar to (Dechow et al. 1996).

Specific Term	**Definition**	**Variable ID**
Was new common stock issued?	IF (Shares Outstanding > prior year's Shares Outstanding) THEN "yes" ELSE "no"	V31a
Was new long-term debt issued?	IF (New long-term debt issued > 0) THEN "yes" ELSE "no"	V31b
Proportion of common stock that is newly issued	$\frac{\text{Shares Outstanding} - \text{prior year's Shares Outstanding}}{\text{Shares Outstanding}}$	V32
Demand for financing (ex ante)	$\frac{\text{NCF}_{\text{OA}} - (\text{Average of prior 3 year's Capital Expenditure})}{\text{prior year's Current Assets}}$	V33a
Was there demand for financing (ex ante)?	IF (Demand for financing (ex ante) < -0.5) THEN "yes" ELSE "no"	V33b
Cash generated from operating and investment activities	NCF_{OA} – Cash Dividends – Capital Expenditure	V33c
Were there operating leases?	IF (Future operating lease obligations > 0) THEN "yes" ELSE "no"	V34
Was the auditor a Big Six firm?	IF (Auditor = Arthur Andersen, PricewaterhouseCoopers, Deloitte & Touche, Ernst & Young, KPMG or Coopers Lybrand) THEN "yes" ELSE "no"	V38
Number of changes of audit firm in the most recent four financial statements	IF($\text{Auditor}_t \neq \text{Auditor}_{t-1}$) THEN 1 ELSE 0 + IF($\text{Auditor}_{t-1} \neq \text{Auditor}_{t-2}$) THEN 1 ELSE 0 + IF($\text{Auditor}_{t-2} \neq \text{Auditor}_{t-3}$) THEN 1 ELSE 0[91]	V39
CEO Tenure	Balance Sheet Date – CEO Start Date (in days)	V40a
Number of CEO changes in the last three years	The number of new CEOs that started between time_t and time_{t-3} + IF(an existing CEO leaves without replacement during time_t and time_{t-3}) THEN 1 ELSE 0	V40b
CFO changed in the last three years?	IF(a new CFO started or an existing CFO left without replacement between time_t and time_{t-3}) THEN "yes" ELSE "no"	V41
Percentage of Directors who are also Executives[92]	$\frac{\text{Number of directors who are also executives at time}_t}{\text{Number of directors at time}_t}$	V42b
Percentage of Director shares owned by those who are also Executives[92]	$\frac{\text{Shares owned by directors who are also executives at time}_t}{\text{Shares owned by directors at time}_t}$	V42c

[91] Both auditor values had to be non-blank for it to be considered a change in auditor.
[92] Those people without any date information are not counted.

4.6 Data Sources

The primary data source is the Capital IQ database using original filings at the balance sheet date and not restated filings, because restated filings would not be available at the balance sheet date when the fraud detection model is designed to be used. The data field is from Capital IQ unless otherwise specified, while the corresponding Compustat field (where known) has also been included for convenience for those who have access to that database.

Data	**Data field**	**Corresponding Field (if known)**
Accounts receivable	IQ_AR	2
Total assets	IQ_TOTAL_ASSETS	6
AFDA	IQ_ALLOW_DOUBT_ACCT	67
Sales	IQ_REV	12
Inventory	IQ_INVENTORY	3
Inventory method	IQ_INV_METHOD	59
Cost of goods sold	IQ_COGS	41
Income before extraordinary items	IQ_NI_AVAIL_EXCL	18
Net cash flow from operating activities ($\mathrm{NCF_{OA}}$)	IQ_CASH_OPER	308
Property, Plant and Equipment (Gross)	IQ_GPPE	7
Property, Plant and Equipment (Net)	IQ_NPPE	8
Current Assets	IQ_TOTAL_CA	4
Cash & Short-term Investments	IQ_CASH_ST_INVEST	1
Current Liabilities	IQ_TOTAL_CL	5
Debt in Current Liabilities	IQ_CURRENT_PORT_DEBT	34
Long-term Investments	IQ_LT_INVEST	32
Total Liabilities	IQ_TOTAL_LIAB	181
Long-term Debt	IQ_LT_DEBT	9
Short-term Investments	IQ_ST_INVEST	193
Preferred Stock	IQ_PREF_EQUITY	130
Income before taxes and interest	IQ_NI_AVAIL_EXCL + IQ_INC_TAX + IQ_INTEREST_EXP	18 + 16 + 15

Data	**Data field**	**Corresponding Field (if known)**
Interest Expense	IQ_INTEREST_EXP	15
Market Value of Equity	Order of preference[93]: 1. HISTORICAL_MARKET_CAP (Bloomberg) 2. IQ_CLOSEPRICE × IQ_SHARESOUTSTANDING 3. PX_LAST (Bloomberg) × IQ_SHARESOUTSTANDING	25×199
Retained Earnings	IQ_RE	36
Working Capital	IQ_WORKING_CAP	179
New Long-term Debt Issued	IQ_LT_DEBT_ISSUED	111
Capital Expenditure	IQ_CAPEX	128
Cash Dividends	IQ_TOTAL_DIV_PAID_CF	
Shares Outstanding	IQ_SHARESOUTSTANDING	25
Net Income	IQ_NI	172
Price	PX_LAST (Bloomberg)	199
Estimated EPS	IQ_EPS_EST	I/B/E/S Database (Perols, 2011)
Actual EPS	IQ_EST_ACT_EPS	I/B/E/S Database (Perols, 2011)
Mergers & Acquisitions Transactions	"M&A Transactions" in IQ_TRANSACTION_LIST_MA	249
Future operating lease obligations	IQ_OL_COMM_NEXT_FIVE + IQ_OL_COMM_NEXT_FIVE	
Number of Employees	IQ_TOTAL_EMPLOYEES	29
Auditor (Audit Firm)	IQ_AUDITOR_NAME	149
CEO & CFO changes	Analysed from MGMT screen in Bloomberg	CompactD Database (Perols,2011)

[93] Bloomberg used in preference as it was much more complete for price and market capitalisation information.

Data	**Data field**	**Corresponding Field (if known)**
CEO Start Date	Analysed from MGMT screen in Bloomberg	CRSP, Compustat or Execucomp (Ndofor et al., 2013)
Balance Sheet Date	IQ_PERIODDATE_BS	
Number of Directors	Analysed from MGMT screen in Bloomberg	CompactD Database (Perols,2011)
Number of Directors who are also Executives	Analysed from MGMT screen in Bloomberg	CompactD Database (Perols,2011)
Shares owned by the two groups above	IQ_INSIDER_SHARES for CEO (using Macros in Excel)	CompactD Database (Perols,2011)
Shares owned by CEO	IQ_INSIDER_SHARES for CEO (using Macros in Excel)	CRSP,Compustat or Execucomp (Ndofor et al., 2013)
Year of Foundation	IQ_YEAR_FOUNDED[94]	
SIC code	IQ_PRIMARY_SIC_CODE	
Listing Exchange	IQ_EXCHANGE	
Percentage change in annual real GDP	IQ_REAL_GDP_GROWTH	
Percentage change in annual retail sales	IQ_RETAIL_SALES_YOY_PCT	
Unemployment rate	IQ_UNEMPLOY_RATE	
G-Index	Investor Responsibility Research Center (IRRC) - freely available information	
E-Index	Investor Responsibility Research Center (IRRC) - freely available information	
Operational and accounting complexity of the industry	As per 2-digit SIC mapping table in Seavey (2011)	

[94] A small number of blanks were filled in from multiple sources through Internet searches.

Chapter 5 Modelling and Results

This chapter presents the results from comparing multiple modelling techniques to detect financial statement fraud, after first describing the data and then the methodology being used. The specific research questions being addressed are

- RQ1 "Which supervised-learning modelling techniques are the most accurate at detecting financial statement fraud under varying assumptions about prior fraud probabilities and ratios of misclassification error costs?" and
- RQ2 "Which supervised-learning modelling technique is the best overall at detecting financial statement fraud, considering the entire range of assumptions investigated in RQ1?"

The motivation and benefits for this type of research are summarised in Section 2.3.5 of Chapter 2. The literature review in Chapter 3 highlights the need for further research into models for detection of financial statement fraud. The literature review also guides the selection of modelling techniques and other methodological decisions to utilise findings from prior research and make new contributions to the field. The research presented in this chapter uses the variables selected in the previous chapter using the variable selection process (described in Section 4.4) that is based on the newly developed schema (Section 4.1) and theoretical Fraud Detection Triangle framework (Section 4.2). An analysis of the importance of each explanatory variable follows in the next chapter.

5.1 Data

This study focuses on public companies listed on US stock exchanges. The study was restricted to publicly listed companies to promote homogeneity in the data in order to focus on differences between fraudulent and legitimate financial statements. The Over-The-Counter (OTC) markets have been excluded because there is less data transparency and availability, and a lack of required standards for companies listing on them (SEC 2014). The data were also restricted to publicly listed companies in order to be consistent with the aim of the study, which is to produce research findings that will be widely applicable to investors, regulators, auditors and other stakeholders. However, it is important to note that the methodology used to develop and evaluate models (as described later in Section 5.2) could be used on data from different types of organisations that produce financial statements. Investigating fraud in both private companies and those traded in the OTC markets is also worthwhile research. Consequently, the methodology used in this research is portable to future research on such companies, as well as not-for-profit organisations.

The US region was chosen because data are more readily accessible from there. However, the methodology that is used in this study can easily be adapted to other countries where financial statement fraud is also a problem. The ACFE (2012) found few differences across regions of the globe, and consequently models developed in this study might be able to be used in different countries without adaptation. There is a precedent for this in the field of business failure prediction where Altman’s (1968) Z-score model estimated on US data has been widely used in many other countries.

The selection of the fraudulent and legitimate cases used in this research is discussed next, followed by the selection of the final data set and univariate analysis of it. The explanatory variables that were defined and selected in the previous chapter are collected for each of these fraudulent and legitimate cases.

5.1.1 Determination of Fraud Observations

It is very important to establish benchmark data sets so that research into new techniques and explanatory variables can be compared with one another. Dechow et al. (2011) provide details on such a data set which is available through the Centre for Financial Reporting and Management at the Haas School of Business, UC Berkeley. The data set has been constructed from fraudulent misstatements in financial statements as alleged by the US

Securities and Exchange Commission (SEC) and disclosed in Accounting and Auditing Enforcement Releases (AAERs), and it has been periodically updated. This data set is used for this research and comprises data from AAERs issued from when they first started in 1982 until August 31, 2012. Even though these are only cases of alleged fraud, they will simply be referred to as cases of fraud in order to simplify the text. This is a procedure that has been followed in other studies of fraud.

AAERs are the most popular resource for empirical studies of financial statement fraud and most prior research has used it (Feroz et al. 1991; Persons 1995; Beasley 1996; Dechow et al. 1996; Beneish 1997; Green and Choi 1997; Fanning and Cogger 1998; Beneish 1999a, 1999b; Feroz and Kwon 1999; Lee et al. 1999; Feroz et al. 2000; Kaminski et al. 2004; Kiehl et al. 2005; Erickson et al. 2006; Bayley and Taylor 2007; Cotter and Young 2007; Hoogs et al. 2007; Ettredge et al. 2008; Skousen and Wright 2008; Brazel et al. 2009; Johnson et al. 2009; McKee 2009; Skousen et al. 2009; Beasley et al. 2010; Goel et al. 2010; Dechow et al. 2011; Perols 2011; Perols and Lougee 2011; Wang et al. 2011; Alden et al. 2012; Whiting et al. 2012; Cao et al. 2015). The benefits and drawbacks of using data from AAERs are discussed in the following subsection.

5.1.1.a An Analysis of the Use of Data from AAERs

SEC enforcement actions are an objective method for identifying companies with fraudulent financial statements, and as such AAERs provide an advantage over alternative measures of earnings management, such as restatements or accrual levels, in which accounting fraud may not exist (Cotter and Young 2007). AAERs are also publicly available, unlike data held by auditing firms that are rarely made available for research (Fanning and Cogger 1998). AAERs are one of the most comprehensive sources of detected cases of fraud and no better publicly available source exists (Beasley et al. 2010). Karpoff et al. (2014) analysed the most popular financial misconduct databases and found AAERs to be the best in terms of having a comprehensive scope, which is to say they excluded the smallest number of relevant cases. Data sets comprising fraud cases listed in AAERs are also free of such researcher classification bias that results from researchers themselves deciding which accounting errors are deemed fraudulent.

There is no perfect method of identifying firms that engage in accounting fraud (Erickson et al. 2006). There are limitations with an AAER data set, but the benefits outweigh them. One of the issues with fraud data sets is the reliability of the fraudulent classifications.

Karpoff et al. (2014) point out that AAERs include many cases that do not involve financial statement fraud and removing these extraneous cases is required before analysing the data. This is what has been done to produce the data set used in this research. Dechow et al. (2011) explain how only fraud cases were included in the data set by reading every AAER in the sample. The advantage of fraud cases from AAERs is that the reliability of the SEC allegations is very high, meaning that there is little chance that financial statements alleged to be fraudulent are in fact legitimate. The SEC has limited resources and is therefore most likely to pursue cases that have a greater assurance of fraud and less ambiguity (Dechow et al. 2010). It is important to note that there is a small chance that some companies do not dispute incorrect SEC allegations of fraud as they view it as their path of least cost (Fanning and Cogger 1998).

While there is little risk from false identification of fraudulent cases, there is no doubt that the SEC misses some frauds (Dyck et al. 2013; Karpoff et al. 2014) on account of their limited resources. One global fund manager wrote in a blog[95] on January 5, 2013, "I had reported many frauds to the SEC. Sometimes the SEC acted. Mostly it did not. When it acted it was often after the stock had gone to zero." This comment also highlights the long lag between the occurrence of fraud, and its detection and disclosure by the SEC in AAERs. This finding is consistent with Karpoff et al.'s (2014) analysis of financial misconduct databases. It also supports the need for models to speed up the detection of financial statement fraud.

Because some frauds are missed, there is a risk that when choosing legitimate observations a truly fraudulent observation may be chosen, even though there is no AAER issued against the company. Unfortunately, it is mostly an unavoidable problem with this type of research and it highlights the importance of not training overly-complex models that become too specific to the training data as they may contain some errors[96]. It is important to recall that techniques such as TreeNet and Random Forests (see Section 3.3.5 on page 56) that utilise random samples of the training data are much less likely to produce models overly-specific to the training data as they never use all of the training data together.

Despite firms of all sizes being vulnerable to fraud, there is some evidence that AAERs are biased towards larger companies (Dechow et al. 2011), presumably because their

[95] The blog is written by John Hempton and is currently available at http://brontecapital.blogspot.com.au/.

[96] Wang (2013) does present an alternative research framework that attempts to address the issue of undetected fraud being omitted.

fraud schemes could do more harm to more people, and also because there is more scrutiny from analysts and the press. This bias is also consistent with the SEC's duty to protect investors that probably means they more closely analyse repeat offenders and companies that are raising money (Dechow et al. 2010). The fact that the bias towards cases that could do more harm to more people will flow through into models developed using data from AAERs is arguably a benefit as the models will be trained to detect the costliest frauds. Additionally, a bias in detecting fraud that the SEC deems to be the most important is informative because an SEC investigation is a very important event for any company (Beasley et al. 2010).

5.1.2 Selection of Legitimate Companies

Supervised learning methods need to be trained on both fraudulent and legitimate cases, and therefore a matching set of legitimate companies has been chosen from those with no AAER issued against them. The matching process is important as it controls for external and unobservable factors (Fanning and Cogger 1998; Johnson et al. 2009). The matching was done one-to-one as a matched-pairs design, which selected one legitimate company for each fraudulent company. The vast majority of studies have used one-to-one matching to select legitimate observations. Although one-to-one matching is vastly different from real-world data sets that comprise all companies, it allows models to learn patterns without bias towards one particular class which may increase their discriminating ability (Alden et al. 2012). Consequently, matching is appropriate for classification problems such as financial statement fraud detection (McKee 2009) to improve the models' ability to effectively discriminate between fraudulent and legitimate statements (Persons 1995).

Matching has been performed in this study based on year of alleged fraud, industry and stock exchange, age of company and company size. Matching based on year, industry and size is common in the literature (Beasley 1996; Dechow et al. 1996; Green and Choi 1997; Fanning and Cogger 1998; Abbott et al. 2000; Abbott et al. 2004; Carcello and Nagy 2004; Kaminski et al. 2004; Uzun et al. 2004; Erickson et al. 2006; Bayley and Taylor 2007; Ettredge et al. 2008; Skousen and Wright 2008; Beasley et al. 2010; Wang et al. 2011; Alden et al. 2012). Matching using these parameters is intended to limit unwanted signals (noise) from seasonal earning patterns, unique industry effects in terms of business environment and reporting requirements, and company size. Matching was also performed on the age of the company as undertaken in previous research (Beneish 1999b; Desai et al. 2006; Perols and Lougee 2011) because the SEC, supported by prior academic research, more closely scrutinises younger companies as it perceives them to have a higher likelihood of financial statement fraud (Beneish 1999b). Matching has also been performed according to the stock exchange the fraudulent company is listed on, in order to control for changes in reporting requirements between exchanges. This is consistent with previous research (Beasley 1996; Abbott et al. 2000; Abbott et al. 2004; Beasley et al. 2010).

In summary, using the company screening tool in the Capital IQ database[97], matched legitimate companies are chosen based on their

- Existence in the specified year (Year),
- Being from the same industry according to two-digit SIC codes (Industry),
- Listing on the same stock exchange and reporting in US dollars (Exchange),
- Being in the same age range: over ten years old, five to ten years old and younger than five years (Age),
- Having the closest (non-zero and non-blank) size according to total assets (Size), and
- Not having financial statements alleged to be fraudulent in an AAER.

If a match could not be found using all of the criteria listed above, then the Age criterion was progressively relaxed until a match was found. For each company that produced fraudulent annual financial statements in consecutive years, a matching legitimate company is found for the first fraud year. The same legitimate company was used then as the matching company for the subsequent fraud years.

5.1.3 Selection of the Time Period to Study

The data set described in Section 5.1.1 contains information from all AAERs released until August 31, 2012. However, there is a long, multiple-year lag between the occurrence of the fraud and its disclosure in an AAER (Beneish 1997; Cotter and Young 2007; Beasley et al. 2010; Karpoff et al. 2014). Beasley et al. (2010) studied AAERs issued between 1998 and 2007 inclusive, but were not confident of drawing conclusions relating to frauds that occurred four years earlier than 2007 because of this time lag. Consequently, the end date of this study

[97] To complete this process, the details about the fraud companies listed in the AAER data set had to be obtained from the Capital IQ database. An automated script was run to match the company names in the AAER data set with those in the Capital IQ database, and in the case of any uncertainty manual conversion was done using additional information such as the nature of the business and the location of the headquarters to confirm the correct company was selected.

was chosen to be the end of 2007, which is slightly more than four years prior to August, 2012. This decision reduces the likelihood that AAERs issued after August 31, 2012 refer to fraud in the chosen time period and thereby influence the results of this study. There is further evidence to support choosing 2007 as the endpoint, because there is major concern that the years that follow are still incomplete with reference to fraud alleged in AAERs. This is because 2008, 2009 and 2010 are the only years in the data set that have more than a 50% decline in the number of frauds from the previous year. In future when data from 2008 onwards are more complete, it would be interesting to include the Global Financial Crisis (GFC) that began in the second half of 2007 and continued for several years. Testing models developed in this current research on data from during the GFC would give valuable insight into whether their accuracy is retained during a major financial crisis.

Longer time periods are preferred as they provide larger samples, but shorter time periods are preferable to exclude older fraud cases that are less relevant for current detection. To balance these competing preferences, a ten-year time period from 1998 to 2007 was chosen[98]. The ten-year period of this current study includes 464 fraudulent financial statements as alleged in AAERs, which increases the validity of results compared with many other studies that used smaller data sets. Using a larger number of fraud cases is important as it assists techniques such as logistic regression to reduce bias in estimates, and helps the more complicated techniques such as neural networks and decision trees to learn more effectively.

5.1.4 The Final Data Set

Within the entire data set created from AAERs in 1982 through to August, 2012, there were 789 businesses allegedly involved in producing 1838 fraudulent annual financial statements. From this, the final sample of 138 fraudulent businesses that produced 464 fraudulent financial statements was selected after the exclusions listed in the dot points below[99]. This process is also summarised in Table 5-1 below on page 140.

[98] The study also included both fraud schemes that started in the study time period and extended beyond it, and fraud schemes that occurred in the study time period but started earlier. This involved 44 company-years of fraud occurring outside the 1998-2007 range.

[99] Some studies also excluded banking, insurance and finance companies because they are subject to different reporting requirements, but these have been included in this study because explanatory variables measuring differences between industries have been included. These variables will allow models to cater for the differences between these industry sectors if they are important for distinguishing fraudulent from legitimate cases.

- 410 businesses were excluded because they were outside the selected time period of 1998-2007;
- A further 42 businesses were excluded because there were no data available, which was defined as no data on total assets, net income and equity;
- A further 190 businesses were excluded because they were not public companies listed on a stock exchange;
- A further 4 companies that did not report in US dollars were excluded, consistent with previous research (such as by Perols [2011]), because they are foreign companies that substantially differ from the rest of the sample. Given their small number, their exclusion reduces unwanted noisy data that would dilute the learning opportunity from the other data;
- A further 3 companies were excluded because a matched legitimate company could not be found, even when the Age and Exchange matching criteria (as defined in Section 5.1.2) were removed;
- Finally, 24 financial statements were excluded because they related to fraudulently understating (as opposed to overstating) financial performance or financial position. This was consistent with previous research such as by Dechow et al. (2011). The expected direction of association of the explanatory variables in understated cases is likely to be opposite to that for the more common overstated cases (which were presented in the previous chapter). As there is only a relatively small number of these cases, they have been excluded to allow models to concentrate on detecting more homogeneous fraud cases involving overstatement of financials. These 24 financial statements were from 10 different businesses, but there is only a drop by two in Table 5-1 as the other 8 committed fraud using overstatement in other years.

Unlike in the work of Perols (2011) who removed companies less than four years old (as some of their explanatory variables required four years of data), such companies have not been excluded in this study so not to limit and bias the results towards older companies. This will result in some missing values, but statistical techniques will be used to handle these and will be explained later.

The number of fraudulent businesses is smaller than the number of fraudulent financial statements because some businesses in the data set have produced more than one fraudulent financial statement. Figure 5-1 below reveals that the largest number of fraudulent businesses in the final data set produced only one fraudulent set of annual financial statements. However, many businesses also produced fraudulent annual financial statements in two, three, four or more years; this was done in consecutive years for all but three businesses. The median number of fraudulent financial statements produced by a fraudulent business is three. This is not surprising given that an earlier study by Beasley et al. (2010) found that financial statement fraud schemes lasted a median of two years and the trend at that time was that they were becoming longer. Overall, this means that models trained on this data will detect fraudulent statements from companies committing fraud once in isolation or from those who are repeat offenders.

Table 5-1. Summary of process to select the final fraud data set.

	Total number of alleged fraudulent	
	Businesses	**Financial Statements produced by those businesses**
Annual fraud in the entire data set created from AAERs in 1982 through to August, 2012	789	1838
Less: Not within the time period (1998-2007)	-410	-810
Less: Data not available	-42	-105
Less: Not publicly listed on a stock exchange	-190	-428
Less: Not reporting in US dollars	-4	-4
Less: No matched company found	-3	-3
Less: Fraud by understatement	-2	-24
Final Fraud Data set	**138**	**464**

Figure 5-1. The relative number of fraudulent businesses from the final data set that have produced fraudulent annual financial statements in one, two, three, four or more years.

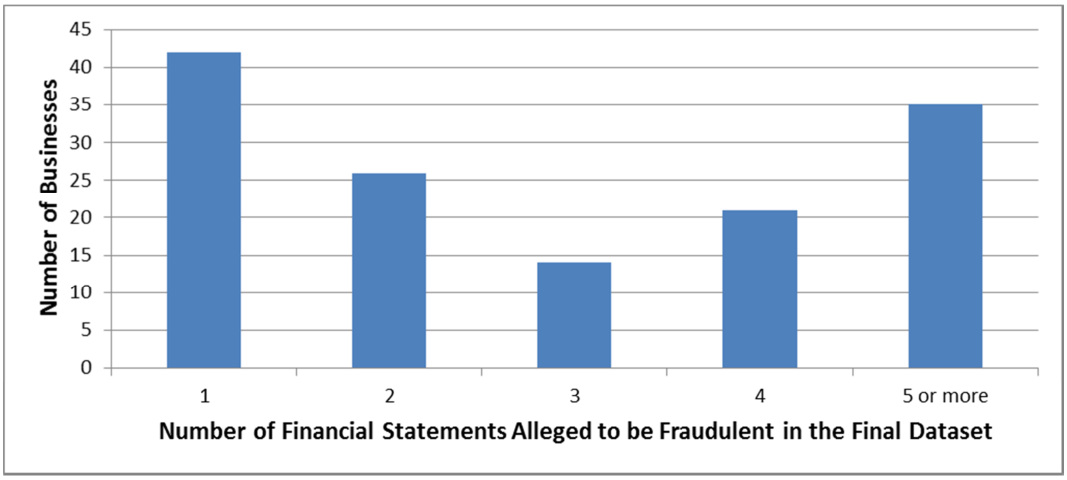


Table 5-2. Industry breakdown of fraudulent businesses in this study as compared to another recent study as well as the entire population of businesses.

Industry Category	Percentage of Fraudulent Businesses in		Percentage of Businesses in Overall Population (from Dechow et al. (2011))
	This Study	**Dechow et al. (2011)**	
Agriculture	0%	0%	0%
Mining and Construction	1%	3%	3%
Refining and Extractive	2%	1%	5%
Food and Tobacco	4%	3%	2%
Textile and Apparel	2%	3%	2%
Lumber, Furniture and Printing	1%	2%	3%
Chemicals	3%	2%	2%
Pharmaceuticals	4%	3%	3%
Durable Manufacturers	19%	19%	19%
Computers	**30%**	**20%**	**11%**
Transportation	4%	5%	6%
Utilities	1%	2%	3%
Retail	9%	13%	10%
Banks and Insurance	14%	12%	21%
Services	8%	13%	10%
Total	100%	100%	100%

Table 5-2 above reveals the proportion of fraudulent businesses according to industry categories using the breakdown specified by Frankel et al. (2002)[100], which has also been used by Dechow et al. (2011). The proportions in this study are highly correlated (with a coefficient of correlation of 0.93) with those in Dechow et al. (2011). The latter found the industry categories of Computer, Retail and Services contained notably more financial statement fraud relative to their proportion in the overall population of businesses. However,

[100] Industries are based on the following Standard Industrial Classification (SIC) codes: Agriculture 0100-0999; Mining and Construction 1000-1299,1400-1999; Refining and Extractive 1300-1399,2900-2999; Food and Tobacco 2000-2141; Textiles and Apparel 2200-2399; Lumber, Furniture and Printing 2400-2796; Chemicals 2800-2824,2840-2899; Pharmaceuticals 2830-2836,3829-3851; Durable Manufacturers 3000-3569, 3580-3669, 3680-3828,3852-3999; Computers 3570-3579,3670-3679,7370-7379; Transportation 4000-4899; Utilities 4900-4999; Retail 5000-5999; Banks and Insurance 6000-6999, Services 7000-7369,7380-9999.

in the data set used in this research, only the Computers industry category was overrepresented for fraud. The reason for this fact is likely that businesses in the Computers industry frequently overstate their sales by recognising premature and fictitious revenue (Dechow et al. 2011). This fact also justifies the use of an explanatory variable that measures industry representation primarily from technology and computer companies, such as V46 from the previous chapter. Overall, Table 5-2 reveals that the data set to be used in this study contains fraud that occurs across a broad range of industries. This is consistent with research findings that fraud does in fact take place across different industries (Beasley et al. 2010).

Table 5-3 below also shows that the fraudulent financial statements are fairly uniformly distributed across the range of business sizes according to their total assets. The smallest 50% of businesses produced 52% of the fraudulent financial statements. Similar to those of Dechow et al. (2011), these results were obtained by ranking businesses according to their size from when they first produced fraudulent financial statements. An even more uniform distribution was obtained when using market capitalisation (as was used by Dechow et al. [2011]) instead of total assets. The fact that fraud occurs across the range of business sizes during the time period selected for this study is consistent with other research mentioned in Chapter 2 (Beasley et al. 2010).

Table 5-3. The frequency of fraudulent financial statements for varying business sizes.

Size of the Fraudulent Businesses By total assets in the first year fraud was committed	**Number of Fraudulent Financial Statements in the Final Sample**		
	Frequency	**Percentage**	**Cumulative Percentage**
Smallest 10%: 0-10%	34	7%	7%
Next Largest 10%: 10-20%	55	12%	19%
Next Largest 10%: 20-30%	69	15%	34%
Next Largest 10%: 30-40%	41	9%	43%
Next Largest 10%: 40-50%	44	9%	52%
Next Largest 10%: 50-60%	51	11%	63%
Next Largest 10%: 60-70%	49	11%	74%
Next Largest 10%: 70-80%	40	9%	83%
Next Largest 10%: 80-90%	44	9%	92%
Largest 10%: 90-100%	37	8%	100%
Total	464	100%	

Together Tables 5-2 and 5-3 show that financial statement fraud occurs across a range of industries and company sizes in the final data set used in this research, which suggests that models developed in this current research can be applied to a broad range of companies in terms of their size and industry representation.

5.1.5 Partitions of the Data for Training and Testing

The final data set of fraudulent financial statements along with their matched legitimate statements was then partitioned in two. As stated in Chapter 3, chronologically partitioning the data set is desirable because it results in more realistic accuracy estimates as models are tested on data further into the future. Consequently, one partition comprising data from 1997 until 2002 is used for training models and the other partition from 2003 to 2007 is used as holdout data for testing models. Table 5-4 summarises the size of the data sets for training and testing that respectively contain 295 (63.6%) and 169 (36.4%) fraudulent financial statement cases, which is more than many other studies that used fewer than 100 cases. This partitioning allocates approximately one-third of the data for testing, which is common in modelling studies.

Table 5-4. Size of the data sets used in this research.

	Number of Financial Statements		
Data set	**Fraudulent**	**Legitimate**	**Total**
Training (≤ 2002)	295	295	590 (63.6%)
Holdout for Testing (≥ 2003)	169	169	338 (36.4%)
All	464	464	928 (100%)

5.1.6 Univariate Analysis

Univariate analysis provides further information about the final data, as well as an indication of the statistical significance of each explanatory variable on its own to differentiate fraudulent and legitimate financial statements. Tests on the entire final data set (not just the training data set) have been conducted to see if there is a statistically significant difference between means of individual variables for fraudulent and legitimate financial statements. For example, the mean of variable V1a for fraudulent statements will be compared with the mean of V1a for legitimate statements. The null hypothesis for each variable is that there is no significant difference in the means and the alternative hypothesis is that they are significantly different. P-values closer to zero indicate more support for the alternative hypothesis.

The first test that has been conducted is a standard t-statistic test for comparing means. Cases with missing values have been excluded, and consequently the number of data points for each test varies according to the number of missing values. As an alternative, the missing values have been replaced using a neutral value procedure outlined in the following subsection, and then an F-statistic one-way analysis-of-variance test has been used to compare means without any data loss from missing values. According to the Central Limit Theorem, violation of the normality assumption of these tests is not a practical concern because of the size of the data set.

5.1.6.a Replacement of Missing Values

Based on the concept used by Beneish (1999a), the neutral value of variables was used as the default replacement for missing values[101].

- The neutral value for variables involving percentage or absolute change calculations is zero, which represents that no change has occurred.
- The neutral value for variables that involve specific values of ratios, such as V28a *Return on Average Assets*, is one (as used by Beneish [1999a]), which represents the numerator being equal to the denominator.
- Similarly, the neutral value for variables calculated as the difference between an industry average and the company's figure is zero, which indicates the company has the same value as the industry average.

For the remaining variables, the mean is used as the default replacement for a missing value, except where specific neutral values have been identified from prior research. These exceptions (for V25, V33a and V50), along with all the replacements for all missing values, are described in Table 5-7 on page 150.

There are no missing values for binary variables such as V9 *Was Last-In, First-Out inventory valuation used?*, because their value is "no" unless there is evidence for a "yes" answer. In addition, variables V39c *Percentage of Director shares owned by those who are also Executives*, V40 *Percentage of total shares owned by the CEO* and V49b *Corporate*

[101] Missing values have occurred when the variable calculation involved trying to divide a non-zero number by zero, as well as when there are no data available.

Governance E-Index have been excluded from the missing value replacement process because more than 80% of their data are missing. Consequently, only t-statistic tests that exclude missing values have been performed on these variables, except for variable V49b which is removed altogether as it only has 29 (3%) non-missing values.

5.1.6.b Results from Univariate Tests

The results of all univariate analyses are presented in Table 5-7 (at the end of this section, on page 150) and a discussion of these results follows. It is clear that the results from the t-statistic that excludes missing values are almost identical to the F-statistic test that replaces missing values, because of the similarity in their p-values for each variable[102]. Therefore, for simplicity, the following analysis will refer only to the p-values corresponding to the t-test.

The explanatory variables that are statistically significant at a 10% level of significance are discussed in the following dot points according to their category in the overall schema presented in Section 4.1 (page 71). It is important to note that for variables with an expected direction of association with fraud (as presented in Section 4.5 starting on page 88) one-sided p-values are used as only one side is relevant. These one-sided p-values are calculated by dividing the two- p-values (that are presented in Table 5-7) by two.

Financial: Specific Account

- **Accounts Receivable:** Only V1a *Accounts Receivable* is statistically significant with a one-sided p-value for an expected positive association with fraud of 0.000. As expected, higher levels of accounts receivable have occurred in fraudulent financial statements.
- **Allowance for doubtful accounts:** No variables from this category are statistically significant.
- **Inventory:** Only V8a *Inventory to Sales* is statistically significant with a one-sided p-value for an expected positive association with fraud of 0.000. As expected, higher levels of inventory relative to sales have occurred in fraudulent financial statements.
- **Sales:** Only V14 *Sales from Acquisitions* is statistically significant in the expected direction with a one-sided p-value for a positive association with fraud of 0.000. As expected, acquisitions have more often been a contributor to sales in fraudulent financial

[102] In addition to comparing p-values, if comparing the t-statistic to the F-statistic then note that the relationship is $F = t^2$.

statements. In addition, V10b and V10d measuring *Sales Growth* and V11a *Sales to Total Assets* are statistically significant with two-sided p-values ranging from 0.003 to 0.067. While their expected directions of association with fraud were uncertain, these variables showed higher sales growth, but lower sales relative to total assets occurred in fraudulent financial statements.

Financial: General

- **Asset Composition:** Only V16 *Net Property Plant & Equipment (PP&E) to Total Assets* and V17 *Soft Assets to Total Assets* are statistically significant in their expected directions of association with fraud, having one-sided p-values of 0.006 and 0.044 respectively. As expected, lower values of PP&E and higher levels of soft assets, relative to total assets, have occurred in fraudulent financial statements.
- **General Accrual Measures**: Only V20 *Positive Accruals* is statistically significant in the expected direction with a one-sided p-value for a positive association with fraud of 0.000. In addition, V19 *Total Accruals to Total Assets* with an uncertain expected direction of association with fraud is statistically significant with a two-sided p-value of 0.057. Fraudulent financial statements had higher levels of total accruals and more often were produced by companies with positive accruals in both the specified year and the year prior.
- **Level of Debt and Financial Distress:** Only V26 *Times Interest Earned* is statistically significant in the expected direction with a one-sided p-value for a positive association with fraud of 0.004. As expected, more fraudulent financial statements were issued by companies with more difficulty in handling changes in interest rates in terms of servicing debt. In addition, V23 *Total Debt to Total Assets* was found to be statistically significant with a two-sided p-value of 0.021. However, surprisingly, the direction of association that was found indicated that fraudulent financial statements had lower levels of debt. Although this was not as expected, one possible reason for it is that the levels of debt have been fraudulently lowered to improve the financial position of the company.
- **Performance and Profitability**: All variables from this category have uncertain expected directions of association with fraud, but only V27a and V27b measuring *Return on Equity* (ROE) are statistically significant with two-sided p-values of 0.038 and 0.050 respectively. These variables showed that fraudulent financial statements had lower ROE values and lower ROE relative to the industry average.

- **Financing:** All variables from this category are statistically significant in the expected direction of association with fraud with one-sided p-values ranging from 0.000 to 0.097. As expected, issuing or needing to issue new securities and the use of operating leases have been associated with more fraudulent financial statements.

Non-Financial

- Only V39b *Percentage of Directors who are also Executives* is statistically significant with a one-sided p-value for a positive association with fraud of 0.038. As expected, there was a greater percentage of directors who were also executives in companies that produced fraudulent financial statements. In addition, V36 *Changes of Audit Firm* and V37b *Changes in CEO* with uncertain expected directions of associations with fraud are statistically significant with two-sided p-values of 0.079 to 0.046 respectively. These variables showed companies that produced fraudulent financial statements had fewer recent changes of audit firm, but more recent changes in CEO.

Comparison: Financial and Non-Financial

- No variables from this category are statistically significant.

Control

- No variables from this category are statistically significant, except for V45 *Company Size* that has an uncertain expected direction of association with fraud and is statistically significant with a two-sided p-value of 0.001. This variable suggests that financial statement fraud occurred more in larger companies, but it may be the result of an SEC bias towards investigating larger companies. The fact that there was a statistically significant control variable also indicates that the matching process was not exact.

New

- No variables from this category are statistically significant, but many of the variables were only expected to be statistically significant in multivariate models and also those that include variable interactions.

The Sales and Financing Financial categories and the Non-Financial category have the largest number of statistically significant variables. Furthermore, all subcategories of the General Financial category contain statistically significant variables. Table 5-5 reveals that relatively more variables in the General Financial and Non-Financial categories are statistically significant when compared to their contribution to the set of all chosen explanatory variables. Table 5-6 on the next page also shows that the statistically significant variables represent the four factors of the new Fraud Detection Triangle framework (Exploitable Opportunity, Incentive/Pressure, Integrity/Attitude/Rationalisation and Suspicious Information) in approximately the same proportions as the set of all chosen explanatory variables.

Table 5-5. The proportion of all chosen explanatory variables and those that are statistically significant from univariate analysis by variable category.

	Proportion of[103]	
Category of Explanatory Variable	**All Chosen Variables** from Table 4-1	**Significant Variables** from Univariate Analysis
Financial: Specific Account	28%	25%
Financial: General	40%	55%
Non-Financial	6%	15%
Comparison: Financial and Non-Financial	12%	0%
Control	8%	5%
New	6%	0%
Total	100%	100%

Table 5-6. The proportion of all chosen explanatory variables and those that are statistically significant from univariate analysis by each factor in the new Fraud Detection Triangle framework.

	Proportion of [103,104]	
Fraud Detection Triangle Factor	**All Chosen Variables** from Table 4-3	**Significant Variables** from Univariate Analysis
Exploitable Opportunity (O)	49%	52%
Incentive/Pressure (I)	38%	37%
Integrity/Attitude/Rationalisation (R)	5%	7%
Suspicious Information (S)	8%	4%

[103] Consistent with Chapter 4, these calculations exclude sub-types. For example, there are multiple ways of measuring variable V1 (V1a, V1b and V1c), but this only counts as one variable in this list.

[104] Variables that are associated with multiple factors are counted in all of them.

The ten most statistically significant variables according to two-sided p-values in both univariate tests are, in order from most significant (lowest p-value) to least significant (highest p-value):

1. V31b New Common Stock or Long-term Debt Issued?
2. <u>V20 Were the specified and the prior year's Total Accruals > Zero?</u>
3. V31a New Common Stock Issued?
4. V8a Inventory to Sales
5. <u>V33a Demand for financing</u>
6. V14 Were any Sales from Acquisitions?
7. V1a Accounts Receivable
8. V45 Company Size
9. V11a Sales to Total Assets
10. <u>V26 Four-period average of Times Interest Earned</u>

Nine of them are financial variables comprising three from the Financing and two from the Sales categories, which is consistent with the finding above that these categories had more statistically significant variables. Despite there being many statistically significant non-financial variables, none of them have made the ten most statistically significant listed above. Interestingly, only the three underlined variables contain temporal information despite more than 50% of all chosen explanatory variables containing temporal information (as shown previously in Table 4-2). This result is surprising given the usefulness of temporal information as previously discussed in Section 4.1.1, but again this result may change with the use of multivariate analysis.

Table 5-7. Results from univariate tests (t-statistic and F-statistic) comparing means of individual variables for fraudulent and legitimate financial statements. The null hypothesis for each variable is that the means are equal and the alternative hypothesis is that they are different, and so p-values closer to zero indicate more support for the alternative hypothesis. The t-test excludes cases with missing values, and the number of missing values for each explanatory variable in the final data set is shown in this table. The F-test uses all data whereby missing values are replaced with the replacement values shown in this table. Finally, the actual and expected direction of association with fraud is presented, such that a positive value indicates that higher values of the specified variable are associated with fraudulent financial statements, and negative values indicate the opposite.

Variable ID	**Variable Name**	**Missing Values**		**Univariate Test Results** Statistic (p-value)		**Direction of Association with Fraud**	
		Number	**Replacement**	**T-statistic**	**F-statistic**	**Actual**	**Expected**
Specific Account - Accounts Receivable							
V1	Accounts Receivable						
V1a	Value for the specified year	0		-3.333 (0.001)	11.106 (0.001)	Positive	Positive
V1b	Percentage change	8 (1%)	0	0.886 (0.376)	0.799 (0.372)	Negative	Positive
V1c	Was Percentage change > 10%?	0		0.131 (0.896)	0.017 (0.896)	Negative	Positive
V2	Percentage change in Accounts Receivable to Sales	10 (1%)	0	0.694 (0.488)	0.486 (0.486)	Negative	Positive
V3	Percentage change in Accounts Receivable to Total Assets	8 (1%)	0	1.036 (0.301)	1.078 (0.299)	Negative	Positive
Specific Account - Allowance for doubtful accounts (AFDA)							
V4	Percentage change in AFDA	36 (4%)	0	-1.291 (0.197)	1.731 (0.189)	Positive	Negative
V5	Percentage change in AFDA to Accounts Receivable	36 (4%)	0	-0.982 (0.326)	0.987 (0.321)	Positive	Negative
V6	Percentage change in AFDA to Sales	38 (4%)	0	-1.169 (0.243)	1.403 (0.236)	Positive	Negative
Specific Account - Inventory							
V7	Change in Inventory to average Total Assets	0		-0.979 (0.328)	0.958 (0.328)	Positive	Positive
V8	Inventory to Sales						
V8a	Value for the specified year	0		-4.035 (0.000)	16.283 (0.000)	Positive	Positive
V8b	Change	10 (1%)	0	-0.535 (0.593)	0.283 (0.595)	Positive	Positive
V9	Was Last-In, First-Out (LIFO) inventory valuation used?	0		-0.311 (0.756)	0.096 (0.756)	Positive	Negative

Variable ID	Variable Name	Missing Values		Univariate Test Results Statistic (p-value)		Direction of Association with Fraud	
		Number	**Replacement**	**T-statistic**	**F-statistic**	**Actual**	**Expected**
Specific Account - Sales							
V10	Sales Growth						
V10a	Percentage change	1 (0%)	0	-1.561 (0.119)	2.398 (0.122)	Positive	Uncertain
V10b	V10a minus the Industry Average	9 (1%)	0	-1.874 (0.061)	3.372 (0.067)	Positive	Uncertain
V10c	Previous year's Percentage change	11 (1%)	0	-1.674 (0.094)	2.689 (0.101)	Positive	Uncertain
V10d	Four-year growth rate	31 (3%)	0	-2.03 (0.043)	3.812 (0.051)	Positive	Uncertain
V10e	Previous year's percentage change in total assets	18 (2%)	0	-0.352 (0.725)	0.114 (0.735)	Positive	Uncertain
V11	Sales to Total Assets						
V11a	Value for the specified year	0		3.021 (0.003)	9.126 (0.003)	Negative	Uncertain
V11b	Percentage change	6 (1%)	0	0.733 (0.464)	0.72 (0.397)	Negative	Uncertain
V12	Gross Margin to Sales						
V12a	Percentage change	3 (0%)	0	1.065 (0.287)	1.136 (0.287)	Negative	Uncertain
V12b	Was percentage change > 10%?	0		0.891 (0.373)	0.794 (0.373)	Negative	Uncertain
V13	Cash Sales						
V13a	Percentage change	1 (0%)	0	-1.170 (0.242)	1.370 (0.242)	Positive	Uncertain
V13b	Was change < 0?	0		-0.564 (0.573)	0.318 (0.573)	Positive	Uncertain
V14	Were any sales from acquisitions?	0		-3.375 (0.001)	11.389 (0.001)	Positive	Positive
General Financial - Asset Composition							
V15	Current Assets to Total Assets	0		-1.010 (0.313)	1.019 (0.313)	Positive	Positive
V16	Net PP&E to Total Assets	0		2.540 (0.011)	6.452 (0.011)	Negative	Negative
V17	Soft Assets to Total Assets	0		-1.707 (0.088)	2.914 (0.088)	Positive	Positive
V18	Percentage Change in Assets other than Current Assets and Net PP&E to Total Assets	6 (1%)	0	1.387 (0.166)	1.945 (0.164)	Negative	Positive
General Financial - General Accrual Measures							
V19	Total Accruals to Total Assets	0		-1.909 (0.057)	3.644 (0.057)	Positive	Uncertain
V20	Were the specified and the prior year's Total Accruals > 0?	0		-4.393 (0.000)	19.295 (0.000)	Positive	Positive

Variable ID	Variable Name	Missing Values		Univariate Test Results Statistic (p-value)		Direction of Association with Fraud	
		Number	**Replacement**	**T-statistic**	**F-statistic**	**Actual**	**Expected**
V21	Total Discretionary Accruals	23 (2%)	Mean	0.221 (0.825)	0.049 (0.825)	Negative	Positive
V22	RSST (unadjusted) Accruals	0		-0.542 (0.588)	0.294 (0.588)	Positive	Positive
General Financial - Level of Debt and Financial Distress							
V23	Debt to Total Assets	0		2.319 (0.021)	5.379 (0.021)	Negative	Positive
V24	Debt to Equity	0		1.244 (0.214)	1.546 (0.214)	Negative	Positive
V25	Altman's (1968) financial distress measure (Z-score)	95 (10%)	2.675[105]	-0.430 (0.667)	0.318 (0.573)	Positive	Positive
V26	Four-period average of Times Interest Earned	119 (13%)	Mean	-2.649 (0.008)	6.999 (0.008)	Positive	Positive
General Financial - Performance and Profitability							
V27	Return on Equity						
V27a	Value for the specified year	0	0	2.078 (0.038)	4.317 (0.038)	Negative	Uncertain
V27b	Industry Average minus Specific Company	8 (1%)	0	1.980 (0.048)	3.839 (0.050)	Positive	Uncertain
V28	Return on Average Prior Assets						
V28a	Value for the specified year	6 (1%)	1	-0.875 (0.382)	0.955 (0.329)	Positive	Uncertain
V28b	Previous year	22 (2%)	1	0.540 (0.589)	0.243 (0.622)	Negative	Uncertain
V28c	Change	23 (2%)	0	-0.723 (0.470)	0.518 (0.472)	Positive	Uncertain
V29	Holding Period Return						
V29a	One-year	94 (10%)	0[106]	-0.040 (0.968)	0.061 (0.805)	Positive	Uncertain
V29b	Previous One-year	128 (14%)	0[106]	-0.176 (0.860)	0.211 (0.646)	Positive	Uncertain
V30	Were analyst Earnings Per Share forecasts achieved or exceeded?	0		-0.526 (0.599)	0.277 (0.599)	Positive	Uncertain

[105] The value of 2.675 was chosen as the neutral value because it was the cut-off value chosen by Altman (1968) as the best for his model to distinguish between predictions of company failure and non-failure.

[106] Holding Period Return is essentially a percentage change calculation and so zero is its missing replacement value.

Variable ID	Variable Name	Missing Values		Univariate Test Results Statistic (p-value)		Direction of Association with Fraud	
		Number	Replacement	T-statistic	F-statistic	Actual	Expected
General Financial - Financing							
V31	Were New Securities issued?						
V31a	Common Stock?	0		-4.064 (0.000)	16.520 (0.000)	Positive	Positive
V31b	Common Stock or Long-term Debt?	0		-5.068 (0.000)	25.687 (0.000)	Positive	Positive
V32	Proportion of common stock that is newly issued	0		-2.490 (0.013)	6.203 (0.013)	Positive	Positive
V33	Demand for financing						
V33a	Specific Value (ex ante)	6 (1%)	-0.5[107]	3.605 (0.000)	13.514 (0.000)	Negative	Negative
V33b	Was there demand (ex ante)?	0		-1.301 (0.193)	1.693 (0.193)	Positive	Positive
V33c	Cash from operating and investment activities	0		-1.897 (0.058)	3.600 (0.058)	Positive	Positive
V34	Were there operating leases?	0		-1.613 (0.107)	2.601 (0.107)	Positive	Positive
Non-financial - Key Roles and Positions							
V35	Was the auditor a Big Six firm?	0		-0.744 (0.457)	0.553 (0.457)	Positive	Negative
V36	Number of changes of audit firm in the most recent four financial statements	0		1.761 (0.079)	3.102 (0.079)	Negative	Uncertain
V37	CEO						
V37a	Tenure	218 (23%)	Mean	-0.524 (0.601)	0.274 (0.601)	Positive	Uncertain
V37b	Number of changes in the last three years	40 (4%)	0[108]	-2.009 (0.045)	3.974 (0.046)	Positive	Uncertain

[107] Using -0.5 as the neutral value for this variable is appropriate because -0.5 is used as the cut-off value to determine whether there is demand for financing in V33b and prior research (Dechow et al. 1996; Erickson et al. 2006).

[108] Zero is used in preference to the mean, as it indicates no changes in CEO, which is an appropriate neutral value over a three year period given the average CEO tenure in the data set is over 7 years.

Variable ID	Variable Name	Missing Values		Univariate Test Results Statistic (p-value)		Direction of Association with Fraud	
		Number	Replacement	T-statistic	F-statistic	Actual	Expected
V38	Has the CFO changed in the last three years?	0		-0.404 (0.686)	0.163 (0.686)	Positive	Uncertain
V39	Composition/Holdings of the Board						
V39a	Number of Directors	0		-1.212 (0.226)	1.469 (0.226)	Positive	Positive
V39b	Percentage of Directors who are also Executives	120 (13%)	Mean[109]	-1.782 (0.075)	3.174 (0.075)	Positive	Positive
V39c	Percentage of Director shares owned by those who are also Executives	797 (86%)	Exclude	0.258 (0.797)	Excluded	Negative	Positive
V40	Percentage of total shares owned by the CEO	781 (84%)	Exclude	0.919 (0.359)	Excluded	Negative	Uncertain
Comparing Financial and Non-financial							
V41	Percentage change in the number of Employees minus percentage change in Total Assets	97 (10%)	0	-0.915 (0.360)	0.767 (0.381)	Positive	Negative
V42	Percentage change in Sales minus percentage change in the number of Employees	97 (10%)	0	0.363 (0.717)	0.107 (0.744)	Negative	Positive
V43	Percentage Change in Sales to Employees: Specific Company minus Industry Average	118 (13%)	0	0.597 (0.551)	0.323 (0.57)	Negative	Positive

[109] The mean was used in preference to zero, because zero has an important specific meaning of no directors also being executives.

Variable ID	Variable Name	Missing Values		Univariate Test Results Statistic (p-value)		Direction of Association with Fraud	
		Number	Replacement	T-statistic	F-statistic	Actual	Expected
Control variables							
V44	Company Age: Number of years since foundation	0		1.250 (0.212)	1.563 (0.212)	Negative	Not expected to be significant given the matching procedure
V45	Company Size: natural log of Total Assets	0		-3.234 (0.001)	10.461 (0.001)	Positive	
V46	Industry: SIC code starts with a 3?	0		-0.069 (0.945)	0.005 (0.945)	Positive	
V47	Stock Exchange listed on						
V47a	NASDAQ?	0		0.066 (0.948)	0.004 (0.948)	Negative	
V47b	NYSE?	0		-0.066 (0.948)	0.004 (0.948)	Positive	
New variables							
V48	Macroeconomic indicators			No univariate results as the means are exactly the same because of the matching procedure. These variables will only be considered in multivariate models with variable interactions.			
V48a	Previous year's percentage change in annual real GDP	0					
V48b	Previous year's percentage change in annual retail sales	0					
V48c	Previous year's unemployment rate inverted	0					
V49	Corporate governance indices						
V49a	G-Index	501 (54%)	Mean	1.340 (0.181)	1.785 (0.182)	Negative	Positive
V49b	E-Index	889 (97%)	Remove	Excluded	Excluded		Positive
V50	Accounting complexity of the industry	136 (15%)	32.7[110]	-0.563 (0.574)	0.687 (0.408)	Positive	Positive

[110] 32.7 is the mean value of the accounting complexity measure from Seavey (2011) who introduced the measure, and so is an appropriate neutral value to be used as a missing value.

5.2 Modelling Methodology

This section describes the methodology to develop and assess a variety of models to detect financial statement fraud. It includes detailing the model-building techniques and the metric used to evaluate their performance. The evaluation of the models is conducted over a range of error cost ratios and prior probabilities (as specified later in this section) and is consistent with recent advances in methodology used by Perols (2011).

5.2.1 Performance Metric

Consistent with standard practice in previous cost-sensitive research (Persons 1995; Beneish 1997, 1999a; Feroz et al. 2000; Lin et al. 2003; Bayley and Taylor 2007; Perols 2011), the following Weighted Error Cost (WEC) metric is used to evaluate accuracy.

$$WEC = p_{IF} \times C_{IF} \times p(F) + p_{IL} \times C_{IL} \times p(L) \quad \text{where,}$$

- p_{IF} = the proportion (or percentage) of fraudulent statements incorrectly classified $= \frac{number\ of\ incorrectly\ classified\ fraudulent\ statements}{number\ of\ fraudulent\ statements}$, such that $0 \leq p_{IF} \leq 1$;
- p_{IL} = the proportion (or percentage) of legitimate statements incorrectly classified $= \frac{number\ of\ incorrectly\ classified\ legitimate\ statements}{number\ of\ legitimate\ statements}$, such that $0 \leq p_{IL} \leq 1$;
- C_{IF} / C_{IL} = the cost of incorrectly classifying a truly fraudulent/legitimate statement. The ratio of these two costs (C_{IF}: C_{IL}) is the relevant information and consequently C_{IL} is set to 1 and C_{IF} is varied to represent different ratios, which is consistent with past research (Persons 1995; Beneish 1997, 1999a; Feroz et al. 2000; Lin et al. 2003; Bayley and Taylor 2007; Perols 2011);
- $p(F)$ = the proportion of fraudulent statements in the real world population (not the sample data set), where $0 < p(F) < 1$. $p(F)$ is also known as the prior probability or *a priori* probability of fraud in financial statements; and,
- $p(L)$ = the proportion of legitimate statements in the real world population, where $0 < p(L) < 1$ and $p(L) = 100\% - p(F)$.

Simpler performance metrics would be either the overall percentage accuracy or the percentage accuracy for fraudulent statements $(1 - p_{IF})$ and for legitimate statements $(1 - p_{IL})$. However, WEC is preferred as it also considers differing costs of misclassification errors C_{IF} and C_{IL} and the proportion of fraudulent statements in the real world $[p(F)]$. These considerations are vital for two reasons. The first is that $C_{IF} > C_{IL}$ in the real world. In addition, the proportion of fraudulent financial statements in the real world [$p(F)$] is less than the 50% proportion in the one-to-one matched sample used in this research.

WEC is also preferred to using the area under the receiver operating characteristic (ROC) curve which averages performance over the range of possible ratios of misclassification error costs. The area under the ROC curve is not being used in this study because it includes ratios that might not be relevant to financial statement fraud detection (Perols 2011). It is important to note there has been mathematical demonstration (Hand 2009a; Hand 2009b) that it is inappropriate to use the area under the ROC curve to compare multiple models, because this assumes the costs (C_{IF} and C_{IL}) depend on the choice of model (rather than the underlying problem), which is absurd according to Hand.

The procedure for optimising the cut-off value for different ratios of error costs (C_{IF}:C_{IL}) and prior probabilities of fraud [$p(F)$] is discussed in the next subsection. The values of p_{IL} and p_{IF} depend on the accuracy of the models, while the values of C_{IF}, C_{IL} and $p(F)$ are discussed in the subsequent subsections.

5.2.1.a Optimisation of Cut-off Values

Each model generates output values, commonly a probability of being fraudulent, that when compared to a cut-off value determines whether each financial statement is classified as legitimate or fraudulent. For example, values of less than 0.5 could be classified as legitimate, while values greater than or equal to 0.5 are classified as fraudulent. Instead of using a default cut-off value such as 0.5, this current research uses the methodological improvement employed by Beneish (1997, 1999a) that has also been used by other researchers (Persons 1995; Bayley and Taylor 2007; Perols 2011). This improved methodology involves empirically determining the optimal cut-off values for each ratio of error costs (C_{IF}:C_{IL}) and prior probability of fraud [$p(F)$] as the value which minimises the weighted error cost (WEC) on the training data set. As cut-off values are optimised based on the same training data used to develop models, the presence of any sample selection bias would be common to both processes and consequently not troublesome (Skogsvik 2005). Because of this fact and in order to be consistent with prior research in the field, cut-off values have been empirically optimised in this study. The corresponding optimal cut-off value will then be used for making classifications on the holdout data set that has not been used during this optimisation process.

5.2.1.b Selection of the Ratios of Error Costs (C_{IF}:C_{IL})

While it is broadly agreed that missing fraud is the more costly error, the quantifiable difference in misclassification error costs varies depending on the particular conditions and stakeholders. Consequently, as suggested in Chapter 3, the analysis that follows will evaluate models using ratios of the cost of missing fraud relative to the cost of falsely alleging fraud (C_{IF}:C_{IL}) that range from 1:1 to 100:1.

Specifically, the values of C_{IF}:C_{IL} are 1:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1 and 100:1.

This decision is based on the summary in Table 3-6 (from the review in Chapter 3) that shows the ratio of error costs range between 1:1 and 1:100 in studies that are relevant to this research (Persons 1995; Beneish 1997, 1999a; Feroz et al. 2000; Lin et al. 2003; Bayley and Taylor 2007; McKee 2009; Perols 2011). Table 3-6 further reveals that most of the studies do not consider ratios greater than 50:1, and consequently no ratios in-between 60:1 and 100:1 are considered, which is consistent with research conducted by Beneish (1997, 1999a).

5.2.1.c Determination of the Prior Probabilities of Fraud [$p(F)$]

Based on the percentage of audits that involved fraud, 0.6% of financial statements were estimated to be fraudulent (Loebbecke et al. 1989; Bell and Carcello 2000), which is also a reminder of the inexperience of auditors in detecting fraud. Other estimates of the proportion of financial statements that are fraudulent based on fraud alleged in AAERs have included 2% (Persons 1995), 1% (Fanning and Cogger 1998; Feroz et al. 2000; Lin et al. 2003; McKee 2009), and most recently approximately 0.4% using AAERs between 1979 and 2002 (Dechow et al. 2011). A similar proportion of 0.69% was estimated using AAERs between 1982 and 1988 and an analysis of news articles (Beneish 1997). It is interesting to note that auditors themselves provide the lowest proportion of fraud within financial statements, estimated at 0.43% (Bernardi 1994), which could be part of the reason for auditors' poor ability to detect fraud (as was discussed in Chapter 2).

Wang (2013) noted that estimates such as those above exclude fraud that goes undetected. Dyck et al. (2013) go further and estimate the proportion of statements that are fraudulent, including those that often go undetected, at a substantially larger value of 14.5%, with a conservative lower bound of 5.6%. These figures were derived by estimating that the total proportion of fraudulent statements is nearly four times that of detected fraudulent statements, which explains their substantially larger estimate. The multiple of four was calculated using the natural experiment of the failure of Arthur Andersen (a former top accounting firm) that forced all its clients to appoint new auditors. Dyck et al. (2013) argue that the new auditors of these enforced auditor turnover cases "cleaned house" and conducted very thorough audits, detecting fraud that usually went undetected. They found approximately four times more fraudulent financial statements than were usually uncovered, and consequently used the multiple of four for estimating total fraud. It is questionable whether companies audited by the failed Arthur Andersen were representative of all companies, which would influence the validity of these estimates, but Dyck et al. also conducted other tests to further validate their estimate of 14.5% the lower bound.

There is no standardised way of handling these varying estimates. For example, Perols (2011) handled the varying estimates of the proportion of statements that are fraudulent [$p(F)$] by using values of 0.3%, 0.6% and 1.2%, while Bayley and Taylor (2007) considered values of $p(F)$ of 1%, 2%, 3% and 4%. This concept of considering multiple values of $p(F)$ is extended in this current analysis to include the substantially larger estimates by Dyck et al. (2013) that have not been considered by previous financial fraud detection modelling studies. From this discussion, the prior probability of fraud [$p(F)$] estimates used in this current analysis were decided to be:

- 0.4%, as the lowest estimate with credible evidence;
- 1%, as the most-used estimate in prior research; and
- 5.5% (= 5.6% rounded to the nearest 0.5%) and 14.5%, as the new largest estimates.

5.2.1.d Ratio of Error Costs Adjusted for Prior Probability of Fraud

In addition to optimising the cut-off value, some modelling techniques, such as decision trees and ensembles of them, can incorporate different ratios of error costs (C_{IF}:C_{IL}) and prior probabilities of fraud [$p(F)$] into the development of the model itself. Instead of varying both of these values, Perols (2011) showed that one of them can be varied to represent changes in both. While Perols chose to vary only the prior probability of fraud using undersampling of legitimate statements, this current study manipulates only the relative cost of missing fraud (adjusted for prior probabilities) during model development to retain the advantages of the one-to-one matched data set (as mentioned in Section 5.1.2), and because costs are more easily interpreted than prior probabilities. The cost of missing fraud relative to falsely alleging fraud that is adjusted for the prior probability of fraud is calculated according to the following equation.

$$Prior - adjusted\ Relative\ Cost\ of\ Missing\ Fraud\ (PaRC_{IF}) = \frac{C_{IF}}{C_{IL}} \times \frac{p(F)}{p(L)} = \frac{C_{IF}}{C_{IL}} \times \frac{p(F)}{100\% - p(F)}$$

For example, a ratio of error costs of 40:1 indicates that missing a fraud is 40 times more costly than a false allegation of fraud. However, if only 5.5% of financial statements are fraudulent then missing fraud is $\frac{40}{1} \times \frac{5.5\%}{100\% - 5.5\%} = 2$ times worse than falsely alleging fraud. The value of two also represents a ratio of error costs of 10:1 with 14.5% of financial statements being fraudulent as shown in Table 5-8, which presents all the prior-adjusted relative costs of missing fraud. In total there are 18 different values to be used ranging from 0.004 to 17, where larger numbers indicate higher costs of missing fraud (relative to falsely alleging fraud) or a larger prior probability of fraud, or both. The 18 different values are 0.004, 0.01, 0.04, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 1, 2, 3, 5, 6, 7, 8, 10 and 17.

Table 5-8. The corresponding prior-adjusted relative cost of missing fraud ($PaRC_{IF}$) for each combination of ratio of error cost and prior probability of fraud. Values appearing more than once have been shaded in the same colour.

Ratio of Error Costs (C_{IF}:C_{IL}) (missing fraud: falsely alleging fraud)	**Prior probability of fraud [$p(F)$]**			
	0.40%	**1.00%**	**5.50%**	**14.50%**
1:1	0.004	0.01	0.1	0.2
10:1	0.04	0.1	0.6	2
20:1	0.1	0.2	1	3
30:1	0.1	0.3	2	5
40:1	0.2	0.4	2	7
50:1	0.2	0.5	3	8
60:1	0.2	0.6	3	10
100:1	0.4	1	6	17

The prior-adjusted relative costs of missing fraud ($PaRC_{IF}$) are also used for presenting results in future sections because this enables the use of simpler graphs that facilitate improved readability and interpretability. As a consequence of cost and prior information being incorporated into $PaRC_{IF}$, the results can be presented in two-dimensional graphs, with $PaRC_{IF}$ on the horizontal axis and WEC on the vertical axis. Table 5-8 can then be used to find the appropriate prior-adjusted relative cost of missing fraud ($PaRC_{IF}$) for a desired prior probability of fraud [$p(F)$] and ratio of error costs (C_{IF}:C_{IL}). The way $PaRC_{IF}$ is incorporated into the WEC performance measure is explained below.

As stated above,

$$WEC = \ p_{IF} \times C_{IF} \times p(F) + p_{IL} \times C_{IL} \times p(L)\ and$$

$$PaRC_{IF} \ = \frac{C_{IF}}{C_{IL}} \times \frac{p(F)}{p(L)}$$

which reveals that the weight of missing fraud relative to falsely alleging fraud

$$\frac{p_{IF}}{p_{IL}} = \frac{C_{IF} \times p(F)}{C_{IL} \times p(L)}$$

$$= \frac{PaRC_{IF}}{1}\ .$$

Consequently, WEC can be reduced to $p_{IF} \times PaRC_{IF} + p_{IL}$*, or more conveniently*

$$WEC = \begin{cases} p_{IF} \times PaRC_{IF} \ + \ p_{IL} & PaRC_{IF} \geq 1 \\ p_{IF} \ + \ p_{IL} \times \frac{1}{PaRC_{IF}} & PaRC_{IF} < 1 \end{cases}\ .$$

The definition of WEC is split in this manner so that the weight of the least costly error (p_{IL} when $PaRC_{IF} \geq 1$ and p_{IF} when $PaRC_{IF} < 1$) is one, which results in greater standardisation across different values of $PaRC_{IF}$. This is also consistent with methods used in modern classification software such as CART decision trees. An example is a naïve model that classifies all financial statements as

- Fraudulent when $PaRC_{IF} \geq 1$ and
- Legitimate when $PaRC_{IF} < 1$.

This naïve model will always have a WEC value of one, because

$$WEC = \begin{cases} 0\% \times PaRC_{IF} + 100\% \ = \ 1 & PaRC_{IF} \geq 1 \\ 100\% + \ 0\% \times \frac{1}{PaRC_{IF}} \ = \ 1 & PaRC_{IF} < 1 \end{cases}\ .$$

This facilitates more intuitive assessment of performance, because WEC values less than one indicate performance better than a naïve model and WEC values greater than one indicate inferior performance.

5.2.2 Benchmark Models: M-score and F-score

Two benchmark models, selected from the review in Chapter 3, are used as estimated in previous studies. For both models, larger values indicate more evidence of the financial statement being fraudulent.

The first is the widely known and cited M-score model produced by Beneish (1999a) using probit analysis, which is defined below. Larger values of the M-score indicate a higher likelihood of financial statement fraud being present.

$$\text{M-score} = -4.84 + 0.92\text{DSR} + 0.528\text{GMI} + 0.404\text{AQI} + 0.892\text{SGI} + 0.115\text{DEPI} - 0.172\text{SGAI} - 0.327\text{LEVI} + 4.679\text{TATA} \ \textit{where}$$

- DSR = $\frac{Accounts\ Receivable_t/Sales_t}{Accounts\ Receivable_{t-1}/Sales_{t-1}} = V2 + 1$ in this current research;
- GMI = $\frac{Gross\ Margin_{t-1}/Sales_{t-1}}{Gross\ Margin_t/Sales_t} = \frac{1}{1+V12a}$ in this current research;
- AQI = $\frac{(1-Current\ Assets_t-PP\&E_t)/Total\ Assets_t}{(1-Current\ Assets_{t-1}-PP\&E_{t-1})/Total\ Assets_{t-1}} = V18 + 1$ in this current research;
- SGI = $\frac{Sales_t}{Sales_{t-1}} = V10a + 1$ in this current research;
- DEPI = $\frac{Depreciation^{111}{}_{t-1}/(Depreciation_{t-1}+PP\&E_{t-1})}{Depreciation_t/(Depreciation_t+PP\&E_t)}$ [112];
- SGAI = $\frac{Sales,General\ and\ Administrative\ Expense^{113}{}_t/Sales_t}{Sales,general\ and\ administrative\ expense_{t-1}/Sales_{t-1}}$ [112];
- LEVI = $\frac{(Long-term\ Debt_t+Current\ Liabilities_t)/Total\ Assets_t}{(Long-term\ Debt_{t-1}+Current\ Liabilities_{t-1})/Total\ Assets_{t-1}}$ [112];
- TATA = $\frac{Total\ Accruals_t}{Total\ Assets_t} = \text{V19}$ in this current research.

The second model is the more recent F-score model produced by Dechow et al. (2011) using logistic regression, which was chosen because the authors have been widely cited in the field and the study uses a relatively large sample. The authors actually developed three F-score models, but their Model 2 is chosen as representative in this current study because all three models reportedly produced similar results[114]. The F-score for Model 2 is calculated as shown below, where larger values indicate a higher likelihood of financial statement fraud being present.

---

[111] The depreciation value was obtained through Capital IQ as field IQ_DA minus field IQ_GW_INTAN_AMORT, while in Computat it is field 14 minus field 65.

[112] This variable is not used in this current research as it was not statistically significant in Beneish's study.

[113] The sales, general and administrative expense value was obtained through Capital IQ as field IQ_SGA, while in Computat it is field 189.

[114] Model 2 was chosen for two reasons. First, it includes non-financial variables. In addition, it excludes two variables requiring an estimate of the "annual buy-and-hold value-weighted market return", because there is some ambiguity to its calculation, which might introduce calculation error that will bias the F-score's performance.

$$F\text{-}score = \frac{e^{Ffirst}/\left(1+e^{Ffirst}\right)}{p(F)}$$

where $p(F)$ is the prior probability of fraud as defined above and

$$\begin{aligned}\text{Ffirst} = {} & -8.252 + 0.665\text{RSST} + 2.457\text{CREC} + 1.393\text{CINV} + 2.011\text{SOFT} + \\ & 0.159\text{CCSS} - 1.029\text{CROA} + 0.983\text{ISSU} - 0.15\text{ABEE} + \\ & 0.419\text{OLEAS} \;\; where\end{aligned}$$

- RSST = RSST Accruals = $V22$ in this current research;
- CREC = $\frac{Change\ in\ Accounts\ Receivable}{Average\ Total\ Assets}$, similar to $V3$ in this current research;
- CINV = Change in Inventory = $V7$ in this current research;
- SOFT = Percentage of Soft Assets = $V17$ in this current research;
- CCSS = Percentage Change in Cash Sales = $V13a$ in this current research;
- CROA = Change in Return on Assets = $V28c$ in this current research;
- ISSU = Existence of New Equity or Long-term Debt = $V31b$ in this current research;
- ABEE = Abnormal change in employees = $V41$ in this current research;
- OLEAS = Existence of Operating Leases = $V34$ in this current research;

The cut-off values that determine the classification (fraudulent or legitimate) for the M-score were empirically optimised by Beneish (1999a) as mentioned in Section 5.2.1.a above. Although Dechow et al. (2011) did not optimise the cut-off values for the F-score, cut-off values are optimised for both the M-score and F-score in this research to allow fair comparisons[115]. For both models, missing values are handled in the same way that was used in the univariate analysis presented in Section 5.1.6 and Table 5-7 (which was in part based on the way Beneish handled missing values for his M-score model). However, the difference is that for variables being replaced by the mean, the replacement value was calculated as the mean of data only from the training set to retain the testing set as a true holdout sample.

5.2.3 Model Building Techniques and Determining Parameters

The details and parameters of each of the model building techniques used in this study are presented in the following sub-sections. Only the training data are used for model development and setting parameters. The holdout data set is exclusively used to evaluate

[115] Optimising the cut-off values improved results on the holdout sample for both models.

model performance, which results in more realistic accuracy estimates that are not upwardly biased. Furthermore, all models exclude variables V39c *Percentage of Director shares owned by those who are also Executives*, V40 *Percentage of total shares owned by the CEO* and V49b *Corporate Governance E-Index* as they contain too many missing values. This is consistent with the methods used in Section 5.1.6 Univariate Analysis. With these exclusions, the total number of explanatory variables is 71, representing 49 of the main variable types (V1-V50) except V40.

5.2.3.a Standard Regression-based Models

Discriminant analysis and logistic regression models do not automatically handle missing values, and so missing values were replaced in the same way as they were for the benchmark models (as specified above in Section 5.2.2). This is preferable to removing financial statements that have at least one missing value, because this procedure (known as list-wise deletion) radically reduces the number of data points available to construct the model (Myrtveit et al. 2001).

Standard regression-based models are more influenced by outliers than other techniques such as decision trees. Therefore, some prior studies (Beneish 1999a; Bayley and Taylor 2007; Dechow et al. 2011) have winsorized explanatory variables at 1% and 99% to mitigate outliers and extreme values that could be created by small denominators in ratio variables[116]. Winsorizing at 1% and 99% involves replacing all data below the 1st percentile with the 1st percentile, and all data above the 99th percentile with the 99th percentile. The use of winsorizing in this study was empirically evaluated using standard discriminant analysis and logistic regression on the training data. Winsorizing the data at 1% and 99% resulted in similar, although reduced, ability to classify both fraudulent and legitimate statements for cases in the training data and also when cross-validation was used. Consequently, winsorizing has not been implemented in this study.

Both discriminant analysis and logistic regression models were produced using IBM's SPSS Statistics package. The probability of fraud is the output used for all models. Five thousand potential cut-off values were evaluated for each of the 18 values of $PaRC_{IF}$ for

[116] Removing outliers and mean-adjusting variables have also been evaluated in prior research (Skousen and Wright 2008; Skousen et al. 2009), but it did not yield improved results and so has not been used in this current study.

every model and the one that produced the lowest WEC on the training data was chosen. The five thousand values were equally distributed between zero and one.

Table 5-9 summarises the different models that have been developed. These models are based on prior studies that were reviewed in Chapter 3. Each model differs with the selection of explanatory variables, and models that use the following variables have been developed:

- All variables, as undertaken by multiple prior studies (Lee et al. 1999; Kaminski et al. 2004; Bayley and Taylor 2007);
- Only variables that are statistically significant at a 15% level in univariate tests, as were undertaken by Skousen and Wright (2008) and Skousen et al. (2009)[117], neither of whom specified a reason for using a 15% level instead of the more common 5% or 10% level. These univariate tests only use the training data, but are otherwise the same as those discussed above in Section 5.1.6;
- Only variables that are statistically significant at the 15% level in both univariate tests and logistic regression (as undertaken by Skousen and Wright (2008) and Skousen et al. (2009)[117]);
- Only one variable from each main variable (V1-V50), selected as the most statistically significant sub-type from univariate tests. For example, there are multiple ways of measuring variable V1 (V1a, V1b and V1c), but only the most statistically significant sub-type is included in this model. The exception is V37a and V37b that are both binary variables required to describe the exchange on which the company is listed.
- Variables selected by stepwise techniques as used in prior research (Persons 1995; Bell and Carcello 2000; Dechow et al. 2011). The maximum level of significance for entry was set at 5% and the minimum level of significance for removal was set at 10%, as was shown to be useful in business failure prediction and insurance fraud detection (Gepp and Kumar 2008; Gepp et al. 2012; Gepp and Kumar 2014).

[117] Skousen et al. used these variable selection techniques to produce a discriminant analysis model, but this research has extended their use to logistic regression models to test their usefulness as well.

Table 5-9. Standard Regression-based Models developed in this research.

Model Code	**Model Details**
Discriminant Analysis (DA)	
DA_All	Inclusion of all variables (except the missing value exclusions)
DA_U15%	Only variables statistically significant at the 15% level in univariate tests
DA_U15%LR	Only variables statistically significant at the 15% level in the LR_U15% model
DA_U1	Only the most statistically significant variable from each main variable (V1-V50)
DA_Step	Stepwise variable selection
Logistic Regression (LR)	
LR_All	Inclusion of all variables (except the missing value exclusions)
LR_U15%	Only variables statistically significant at the 15% level in univariate tests
LR_U15%LR	Only variables statistically significant at the 15% level in the LR_U15% model
LR_U1	Only the most statistically significant variable from each main variable (V1-V50)
LR_Step	Stepwise variable selection
New Multi-stage Logistic Regression (LR_MS)	
LR_MS_S	Multi-stage model using the overall schema of explanatory variables
LR_MS_F	Multi-stage model using the new Fraud Detection Triangle framework

Based on the multi-stage logistic regression used by Summers and Sweeney (1998), two new multi-stage logistic regression models have also been developed in this research.

1. The first model (LR_MS_S) utilises the newly constructed overall schema for explanatory variables (presented in Section 4.1), which comprises Financial, Non-Financial, Comparison, Control and New categories. In the first stage, five logistic regression models are estimated, each using explanatory variables from a different category. In the second stage, the LR_MS_S logistic regression model is estimated from the probability outputs of the first stage models. That is, the LR_MS_S model uses $\Pr(\mathrm{F})_{Financial}$, $\Pr(\mathrm{F})_{Non-Financial}$, $\Pr(\mathrm{F})_{Comparison}$, $\Pr(\mathrm{F})_{Control}$ and $\Pr(\mathrm{F})_{New}$ as explanatory variables, where $\Pr(\mathrm{F})_x$ is the probability of fraud estimated using a logistic regression with explanatory variables from the x category.
2. The second model (LR_MS_F) utilises the new Fraud Detection Triangle framework (presented in Section 4.2), which now includes Exploitable Opportunity (O), Incentive/Pressure (I), Integrity/Attitude/Rationalisation (R) and Suspicious Information (S) factors. As is the case with the LR_MS_S model, the first stage involves four logistic

regression models being estimated, each one using explanatory variables from a different factor[118]. The LR_MS_F model is then estimated using logistic regression with $\mathrm{Pr(F)}_O$, $\mathrm{Pr(F)}_I$, $\mathrm{Pr(F)}_R$ and $\mathrm{Pr(F)}_S$ as explanatory variables.

5.2.3.b Artificial Neural Network Models

As mentioned in the literature review in Chapter 3, artificial neural networks are better than standard regression-based models at handling outliers and at selecting important explanatory variables from a large set of variables. Consequently, artificial neural network models are built using all explanatory variables with missing values being replaced as they were for other models above. However for neural network models, data from each explanatory variable is normalised, so that all data values range between 0 and 1. This data transformation is performed to assist the learning efficiency of the neural networks (Feroz et al. 2000) and is consistent with past research (Feroz et al. 2000; Green and Choi 1997; Perols 2011). Normalising is performed independently for each explanatory variable according to the following formula:

$$\frac{data\ point - minimum\ value\ of\ associated\ explanatory\ variable}{range\ (maximum - minimum)\ value\ of\ associated\ explanatory\ variable}.$$

Normalising has no influence on binary variables, because their minimum value is zero and range is one and so the formula becomes $\frac{data\ point - 0}{1} = data\ point$.

Table 5-10. Artificial Neural Network Models developed in this research.

Model Code	Model Details
NN_BK	Backpropagation neural network with one hidden layer containing four neurons
NN_GA_1	Backpropagation neural network optimised by a genetic algorithm with the minimum learning rate set to 0.1
NN_GA_5	Backpropagation neural network optimised by a genetic algorithm with the minimum learning rate set to 0.5

Table 5-10 above summarises the artificial neural network models that are used in this research, which are explained in more detail in the remainder of this subsection. The pseudo probability of fraud is the model output used. This output is similar to a probability in that it normally lies between zero and one, and larger values indicate a higher likelihood of fraud.

[118] Variables that are associated with more than one factor are included in all associated factors.

However, it cannot be interpreted as a probability and is consequently termed pseudo because the values can be marginally outside the range of zero to one. Five thousand potential cut-off values between the minimum and maximum pseudo probability were evaluated for each of the 18 values of $PaRC_{IF}$, which is consistent with procedure for the standard regression-based models.

As stated in the literature review in Chapter 3, backpropagation artificial neural networks with one hidden layer have had encouraging empirical results in prior studies and consequently are used as another benchmark technique against which to compare newer models developed in this research. The backpropagation neural network (NN_BK) developed in this research used IBM's SPSS Statistics package's "Multilayer Perceptron" model, which is a feed-forward backpropagation network. The parameters of the neural network are described in the following dot-points.

- The commonly used sigmoid function is used for both the activation and transfer function, because it is generally the most useful (Shih 1994) and was empirically found to be the most useful by Feroz et al. (2000), as well as also being consistent with Green and Choi (1997).
- The number of hidden layers was one, consistent with prior research (Green and Choi 1997; Feroz et al. 2000).
- The weights are updated after each case as was done by Green and Choi (1997).
- Based on the methodology in prior research (Green and Choi 1997; Feroz et al. 2000; Perols 2011), three values (0.1, 0.3 and 0.5) were evaluated for the learning rate and momentum parameters, while four values (4, 8, 12 and 16) were evaluated for the number of neurons in the hidden layer. This totals $3 \times 3 \times 4 = 36$ alternative configurations that were evaluated on a randomly chosen 30% sample of the training data that is partitioned off during model development. The configuration that was chosen had the best classification accuracy percentage. This configuration had a
    - Learning rate of 0.5,
    - Momentum of 0.1, and
    - Four neurons in the hidden layer.

A model that optimised the parameters of a neural network with a genetic algorithm (NN_GA) has also been developed, because using a genetic algorithm is likely to be the best way to optimise neural networks (Bhattacharya et al. 2011b). Encouraging results have also been found using a similar approach in a prior study (Fanning and Cogger 1998). Furthermore, the use of a genetic algorithm removes the need to evaluate additional configurations of the standard backpropagation algorithm because, by the principle of natural selection, only the most optimal configuration is expected to survive the genetic algorithm optimisation (Bhattacharya et al. 2011b). The software program Neuralyst (version 1.4) that executes within Microsoft Excel was used to develop the genetically optimised neural network. The underlying neural network still uses backpropagation and sigmoid functions. However, the genetic algorithm selects which explanatory variables to include, and optimises the parameters of the neural network. These parameters include the number of hidden layers, the number of neurons in each hidden layer, the learning rate and the momentum rate. The maximum momentum rate parameter for the genetic algorithm was set to the default value of 1.0, because this is already greater than the maximum momentum value used in prior research (Perols 2011). The default value for the minimum learning rate is 0.5, which is greater than the lowest value of 0.1 used in prior research (Green and Choi 1997; Perols 2011). Consequently, two models were developed, one optimised with a minimum learning rate of 0.1 (NN_GA_1) and the other 0.5 (NN_GA_5). Both optimised neural networks had two hidden layers, the first with 30 neurons and the second with 10 neurons. The remainder of the genetic algorithm settings were set to Neuralyst's default values because Bhattacharya et al. (2011b) used them and stated that tweaking the parameters is unlikely to have much of an effect on the classification accuracy of the evolved neural network. These default genetic algorithm parameters include 10 as the generation count, 1 as the number of crossover points, 0.1 as the mutation rate and the error on the training data as the fitness criterion. Further details of the genetic algorithm used by Neuralyst are provided in the User's Guide (Shih 1994).

5.2.3.c Explanatory Variables for Decision Trees and Ensembles of Them

Decision trees treat numerical variables as ranked or ordinal data, because they use splitting rules that partition the data based on an explanatory variable being $\leq$ or $>$ a particular value. As a result, some of the variable sub-types are redundant and have been removed when developing the decision tree models; these are:

- V1c *Was Percentage change in Accounts Receivable > 10%* because it is equivalent to V1b *Percentage change in Accounts Receivable* and a split point of 10%.
- V12b *Was Percentage change in Gross Margin to Sales > 10%* because it is equivalent to V12a *Percentage change in Gross Margin to Sales* and a split point of 10%.

- V13b *Was change in Cash Sales <0* because it is equivalent to V13a *Percentage change in Cash Sales* and a split point of 0%.
- V33b *Was there demand for financing (ex ante)* because it is equivalent to V33a *Demand for financing* and a split point of -0.5.

As decision trees can also easily model categorical variables, a new categorical variable (V47c) that can take on three values (NASDAQ, NYSE and AMEX) was used in place of the two existing binary variables (V47a and V47b) that indicate the exchange on which each company is listed. Overall, this reduced by five the number of explanatory variables from 71 to 66.

5.2.3.d Individual Decision Tree Models

As mentioned in Chapter 3, there is a lack of research that uses decision tree techniques, particularly CART decision trees that have performed well on a Chinese data set (Bai et al. 2008) and have performed better than other decision trees in other classification problems in business. Consequently, CART decision tree models will be developed in this study using version 7 of Salford Predictive Modeler (SPM) by Salford Systems. CART decision trees are grown to their full extent and then they are pruned to result in the tree with the lowest cost (WEC) as determined by cross-validation on the training data set. Cut-off values between 0 and 1 at intervals of 0.01 were trialled and the optimal cut-off value for the probability of fraud output was determined as the one that produced the lowest cross-validated WEC[119].

The best parameter settings for CART decision trees were empirically determined by thousands of cross-validated trials on the training data set. The following CART parameters were varied.

- MINCHILD: the minimum number of data points allowed in a child node, which is any node other than the root node. Values above one limit the size of the tree, which can sometimes improve classification accuracy on holdout samples because nodes with a small number of cases can be unreliable when classifying new data. The following eight values were trialled for the MINCHILD parameter: 1, 2, 5, 10, 25, 50, 100 and 200.

[119] Intervals of 0.01 were the smallest available in the Salford Systems software used.

- ATOM: the minimum number of data points allowed in a parent (non-leaf) node. Similar to the effect of MINCHILD, values of ATOM above 2 limit the size of the tree, which sometimes improves classification accuracy. Because the minimum limits of MINCHILD and ATOM only affect small nodes that are likely to be in the lower levels of the tree, they generally only have small effects on classification accuracy (Steinberg 2014). The following eight values were trialled for the ATOM parameter: 2, 5, 10, 25, 50, 100, 200 and 500.
- CV: the number of folds in the cross-validation on the training data set. Cross-validation with 5, 20 and 50 folds were trialled to determine whether any of them improved classification accuracy as compared with using 10 folds.
- RULES: the impurity function used to determine the best splitting rule at each node. CART trees are grown by searching for the best splitting at each node from all explanatory variables and all possible split points in the learning data, and there are multiple ways to determine the best splitting rule. Specifically, the best is chosen as the one that has the lower impurity according to an impurity function. The Gini, Entropy and Twoing functions are standard impurity functions, but Twoing is equivalent to Gini for two-class (fraud and legitimate) classification problems such as in this study. Consequently, only the Gini and Entropy functions are evaluated[120]. Unlike the Gini function, Entropy places emphasis on splitting rules that produce subsets of similar sizes (in terms of the number of data points or financial statements). Although the choice of splitting rule does not make much difference for two-class classification problems (Breiman et al. 1984; Breiman 1996b), empirical investigation was conducted to determine whether there was a superior function when working with financial statement fraud data. More detailed theoretical and empirical analysis of different splitting rules is provided by Breiman et al. (1984) and Breiman (1996b).

The best settings were then chosen for each value of the Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$) as that which minimised the weighted error cost (WEC) using cross-validation on the training data set. These chosen settings are summarised in Table 5-11. The MINCHILD parameter was predominantly set to one, which does not influence the

---

[120] $Impurity\ (Gini) = 2 \times p_{node}(F) \times p_{node}(L)$ and
$Impurity\ (Entropy) = -p_{node}(F) \times \log p_{node}(F) - p_{node}(L) \times \log p_{node}(F)$,
where $p_{node}(F/L)$ is the proportion of fraud/legitimate cases in the node.

development of the tree models. The ATOM parameter was mostly set to 10, which is the CART's default value, while the CV parameter varies between 5, 10 and 20 folds. The RULES are mostly Gini, while the Entropy impurity function is best when $PaRC_{IF}$ is 0.4, 0.5 or 0.6.

Table 5-11. The CART parameters that result in the decision tree with the smallest cross-validated WEC on the training data. MINCHILD = the minimum number of data points allowed in a non-root node; ATOM = the minimum number of data points allowed in a non-leaf node; CV = the number of folds in cross-validation; RULES = the impurity function used to determine the splitting rules (G = Gini and E = Entropy). The shaded regions are those for which the best tree on the training data set classifies all statements the same way.

	$\boldsymbol{Prior - adjusted\ Relative\ Cost\ of\ Missing\ Fraud\ (PaRC_{IF})}$																	
Parameter	.004	.01	.04	.1	.2	.3	.4	.5	.6	1	2	3	5	6	7	8	10	17
MINCHILD					1	1	1	1	1	1	5	25	1	1	1	1		
ATOM					500	10	10	10	10	2	10	50	10	10	10	10		
CV					10	20	20	20	20	10	10	10	5	5	5	5		
RULES					G	G	E	E	E	G	G	G	G	G	G	G		

While it is desirable to have univariate splitting rules for ease of interpretation and analysis, using linear combinations of variables in splitting rules was investigated to determine whether it would improve classification accuracy and lower the WEC. Trials for all of the parameter settings above were repeated for trees allowing linear splitting rules, but no decrease in WEC on the training data set was found for any value of $PaRC_{IF}$. Hence, only models with exclusively univariate splits were used. CART's Missing Value Indicator (MVI) parameter was also modified to investigate the usefulness of penalising the inclusion of variables with missing values. The use of the MVI parameter also did not improve any models and so has not been incorporated into any final model.

Two final models using CART decision trees have been developed as shown in Table 5-12. The first, DT_Mult, is a suite of cost-sensitive CART decision trees, one for each value of $PaRC_{IF}$. Each decision tree uses the optimal parameters for the corresponding specific value of $PaRC_{IF}$ as per Table 5-11, as well as incorporates that cost ($PaRC_{IF}$) into the model development process. However, it is not certain that differences in these parameters result in substantially improved performance. Therefore, in order to investigate the value of developing a decision tree for each value of $PaRC_{IF}$, a second model, DT_One, was developed. DT_One is a single cost-insensitive CART decision tree used for all values of $PaRC_{IF}$. This decision tree is the tree from DT_Mult where $PaRC_{IF} = 1$.

Table 5-12. Individual Decision Tree Models developed in this research.

Model Code	**Model Details**
DT_Suite	Suite of cost-sensitive CART decision trees, one for each value of $PaRC_{IF}$
DT_One	One cost-insensitive CART decision tree used for all values of $PaRC_{IF}$

Missing Values for Individual CART Decision Trees

Missing values were treated differently in CART models, compared to all other models. The reason is that CART provides a sophisticated way of handling missing values using surrogates. When CART encounters a missing value at a particular node in the tree, it replaces the splitting rule at that node with a back-up *surrogate* splitting rule that is the most similar to the original splitting rule. The surrogate splitting rule is the most similar in action to the chosen rule, which is not necessarily the second best rule at reducing node impurity. That is, a surrogate splitting rule is that which most closely mimics the original rule in terms of partitioning the data into left and right child nodes. This process is performed at every node of the tree, which means that the surrogate splitter is akin to imputing each missing value on a case-by-case basis. This allows cases (financial statements) with a missing value for the same variable to be handled differently depending on the values of the other explanatory variables. This arguably results in better characterization of the data (Steinberg 2012) that may assist in improving accuracy and so this method for handling missing values has been used for all CART models[121].

5.2.3.e Ensembles of Decision Trees: Random Forests and TreeNet

As mentioned in Chapter 3, Random Forests and stochastic gradient boosting have shown promise in detecting financial statement fraud. However, in prior studies neither has been tested using a separate holdout sample or over a range of varying ratios of error costs, as has been done in this research. The models developed in this research are summarised in Table 5-13.

[121] The CART setting for the number of surrogates to calculate was set to five.

Version 7 of SPM, the software also used for CART decision trees, has been used to develop these models, where TreeNet is the program used to perform stochastic gradient boosting. The sophisticated method for replacing missing values in CART is not available in Random Forests or TreeNet, presumably because it would be too computationally difficult given that a large number of trees need to be developed for these ensemble techniques.

Table 5-13. Ensemble Decision Tree Models developed in this research.

Model Code	Model Details
Random Forests (RF)	
RF_8	Random Forests with 1000 trees and variable subset size of 8
RF_66	Random Forests with 1000 trees and variable subset size of 66
TreeNet (TN)	
TN	TreeNet with 0.01 learning rate and maximum number of nodes per tree of 12

Random Forests Parameters

Random Forests uses bootstrap samples such that each individual tree ignores 37% of the data, which is referred to as out-of-bag data (Salford Systems 2012a). Similar to cross-validation, results on out-of-bag training data can be used to empirically determine the best parameters. There are two main parameters required for Random Forests models, which were set as described below.

1. One thousand decision trees were grown, the same as grown by Whiting et al. (2012). Growing additional trees was investigated, but there was no noticeable improvement in performance on the training data set and so the parameter was left at 1000.
2. The size of the variable subset available at each node was set to eight, which is the square root of the total number of explanatory variables ($\sqrt{66} \approx 8$). This is suggested by prior research (Bhattacharya et al. 2011a) and the same as performed by Whiting et al. (2012). Empirical investigations were also performed to investigate whether other values were preferable. Values of 1, 4, 8, 16, 24, 32, 40, 48, 56 and 66 were trialled, and surprisingly the value of 66 produced the best results on out-of-bag training data. This is unexpected as 66 represents 100% of the explanatory variables, which removes the randomness associated with only having a limited subset of variables available at each node (the

random subspace method). Consequently, two models were developed, one with the size of the variable subset to 8 (RF_8) and the other to 66 (RF_66)[122].

Additionally, the minimum number of cases in a parent node was set to 2 to grow the trees to their maximum size according to the theoretical discussion of Random Forests in Section 3.3.5[123].

As with other models, the probability of fraud is the output used. As is the case with the CART models, cut-off values between 0 and 1 at intervals of 0.01 were trialled and the optimal cut-off value for the probability of fraud output was determined as the one that produced the lowest out-of-bag WEC[124].

TreeNet Parameters

Ten-fold cross-validation was performed to develop the final model and determine the best parameters. The choice of 10-folds was because it has been used by the only prior research in the field (Whiting et al. 2012), it was found to be the best for the individual decision tree DT_One model and because the use of 10-folds is common in the modelling literature. The subsampling rate was set to the default value of 0.5, because the value used in prior research in the field was not published, and very small values can reduce accuracy (Salford Systems 2012b). This means that a sample of only 50% of the data is used by each individual tree. The three parameters that were empirically determined were the learning rate, the number of trees and the size of the individual trees. The different values trialled for these parameters are explained in the following bullet points.

- A learning rate of 0.001 was used by Whiting et al. (2012) in the only financial statement fraud detection research that uses stochastic gradient boosting. Because 0.001 represents a very slow learning rate, values of 0.005 and 0.01 were also trialled.
- One thousand decision trees were grown by Whiting et al. (2012), which appears to be insufficient given the extremely low learn rate of 0.001. The following trials empirically confirmed that growing more trees is desired. The maximum number of trees (20,000) was grown for every trial in TreeNet, and then the number of trees that resulted in the

[122] Creating individual models for each value of $PaRC_{IF}$ by varying the "class weights" parameter was briefly investigated, but it did not result in improved performance and so was not implemented.

[123] Random Forests also requires other minor parameters to be set. This was done at their default values. In addition to this, the Gini impurity function was used to develop individual decision trees, as this was the best in the majority of the individual CART decision tree trials.

[124] Intervals of 0.01 were the smallest available in the Salford Systems software used.

smallest cross-validated training WEC was chosen. For the trials with a learning rate of 0.001, the number of trees with the best results ranged between 12,368 and 19,984 trees. The maximum number of nodes per tree with the best results also varied from 12 (with 12,368 trees) to 2 (with 19,984 trees). The number of trees required decreases as the number of nodes in each tree increases, because each individual tree becomes more complex. Although fewer trees are also required for higher learning rates, the maximum number of trees is grown because the only downside is the computation time.

- TreeNet and stochastic gradient boosting uses smaller trees. Whiting et al. (2012) limited the size of trees to three-way interactions between explanatory variables, which visually corresponds to limiting trees to having 4 levels (including the level comprising the root node). TreeNet controls tree size differently, by specifying the maximum number of terminal nodes. Three-way interactions can occur in trees with between four and eight terminal nodes. However, the choice was made to investigate a wider range of nodes between two and twelve, where two terminal nodes represents zero interactions between explanatory variables and higher numbers indicate that more complex interactions are possible.

These trials clearly indicated that interactions between the explanatory variables exist, because the cross-validated WEC was substantially lower for the trials with larger trees. The models with a maximum of 3-node trees, allowing a small amount of interaction, resulted in an average decrease in WEC of 12% compared to 2-node tree models. Furthermore, the maximum of 12-node trees that allowed much more interaction decreased the WEC by a further 18.5% on average. Overall, the chosen parameters that minimised the cross-validated WEC on the training data were a learn rate of 0.01, a maximum of 12 nodes per tree and 1,184 trees. A learn rate of 0.05 and either a maximum of 9 or 12 nodes per tree yielded the same minimum WEC, but these parameters were not preferred as they used more complicated trees, in accordance with the principle of parsimony[125].

[125] For completeness, TreeNet models with a learn rate of 0.05 and a maximum of 9 or 12 nodes were tested on the hold out data, but performed worse than the chosen TreeNet model. This is empirical support for the use of the principle of parsimony when choosing between alternative model configurations.

Varying the value of $PaRC_{IF}$ did not change the model output, and so one TreeNet model has been used for every value of $PaRC_{IF}$[126]. The probability of fraud output was used and the optimal cut-off values were chosen in the same way as for the CART models.

5.2.3.f Ensembles involving multiple modelling techniques

In addition to ensembles of decision trees, ensemble models involving multiple modelling techniques have been developed and are discussed below and summarised in Table 5-14.

Table 5-14. Summary of ensemble models involving multiple modelling techniques developed in this research.

Model Code	**Model Details**
NN_LR	Classifications from NN_BK included as an additional explanatory variable in the LR_All model
DT_LR	Classifications from DT_One included as an additional explanatory variable in the LR_All model
DTnode_LR_Step	Terminal node assignments from DT_One included as additional explanatory variables in the LR_Step model
NN-DTnode_LR_Step	Classifications from NN_BK included as an additional explanatory variable in the DTnode_LR_Step model
Vote5	Majority Vote between DT_One, TN, RF_8, NN_BK and LR_All
Vote3_RF_TN_DT	Majority Vote between TN, RF_8 and DT_One
Vote3_RF_TN_DA	Majority Vote between TN, RF_8 and DA_All
Vote3_RF_TN_NN	Majority Vote between TN, RF_8 and NN_BK
AV5_NoNN	Average between DT_One, TN, RF_8, LR_All and DA_All
AV2_RF_TN	Average between TN and RF_8
AV3_RF_TN_DT	Average between TN, RF_8 and DT_One
DT_One_DA	Discriminant analysis (based on DA_All) performed as a second stage on the riskier classifications from DT_One

[126] Creating individual models for each value of $PaRC_{IF}$ by varying the "class weights" rather than the "costs" parameter was briefly investigated, but it did not result in improved cross-validated performance on the training data set and so was not implemented.

Models Proposed by McKee

Two ensemble models have been included based on the models developed by McKee (2009).

1. NN_LR: A logistic regression model the same as LR_All, but with an additional explanatory variable, the classification from NN_BK when $PaRC_{IF} = 1$. A fraud classification from the NN_BK model was found to have a statistically significant positive effect on the probability of fraud in the resulting NN_LR model at the 1% significance level.
2. An individual decision tree model the same as DT_One, but with an additional explanatory variable, the classification from NN_LR when $PaRC_{IF} = 1$. This resulted in a simple tree that only used the NN_LR output, which is equivalent to NN_LR when $PaRC_{IF} = 1$. However, this new decision tree ensemble model is less able to adapt to different values of $PaRC_{IF}$ and so has not been included in the results section. The reason that it is less able to adapt is that it only has two terminal nodes that produce only two distinct probability of fraud outputs. This reasoning is further explained in the next subsection.

New Ensembles Developed from Analysing McKee's Models

The order of combining models was selected by McKee without a theoretical rationale being presented. Consequently, new ensembles with a different ordering of underlying techniques have been developed to utilise the theoretical strengths of each modelling technique.

A strength of logistic regression is that it produces a probability of fraud output that is continuous. A small change in an explanatory variable results in a small change in probability in a logistic regression model, unlike a decision tree that produces a discontinuous probability of fraud output. Individual decision trees assign the same probability to all financial statements allocated to the same terminal node and consequently cannot differentiate between them. For example, a tree with only six terminal nodes can only produce six different probabilities to the entire data set. In addition to this, the NN_BK neural network model only produces a pseudo-probability of fraud that cannot be interpreted as a probability. As a result, logistic regression will be used in the final stage of the new ensemble models that are proposed below because of its advantageous output.

DT_LR is a logistic regression model the same as LR_All, but with an additional explanatory variable, the classification from DT_One. The decision tree classification variable was found to have a statistically significant positive effect on the probability of fraud in the resulting DT_LR model at the 1% significance level. Adding the NN_BK classification as an explanatory variable to the DT_LR model was trialled, but the additional variable was not statistically significant even at the 15% level and there was no improvement in the classification accuracy on the training data set.

There is also an opportunity to use more information from a decision tree than simply the classification of either fraud or legitimate. Decision trees partition data into a collection of terminal nodes, each with a homogeneous subset of the original data. This information allows a logistic regression model to include specific effects from the homogeneous subset in each terminal node. Additionally, the logistic regression model would also have access to the original explanatory variables, allowing it to detect and incorporate patterns that are common across all nodes in the decision tree, which CART cannot easily do given its tree structure. Consequently, a new model DTnode_LR_Step has been developed where the terminal node assigned by DT_One is fed into a logistic regression. The terminal node information is provided by a set of binary dummy variables, one for each terminal node, that indicates whether each financial statement was assigned to a specific terminal node. This drastically increases the number of explanatory variables, and consequently stepwise logistic regression (LR_Step) was used.

As McKee also proposed integrating a neural network, adding the NN_BK classification as an additional explanatory variable to the DTnode_LR_Step was also trialled. The additional variable was found to be statistically significant even at the 1% level. The resulting model, termed NN-DTnode_LR_Step, also has slightly improved classification accuracy on the training data set.

Ensembles using Majority Vote

Combining techniques using a Majority Vote has shown promise on Greek and Chinese data (as per the review in Chapter 3). In a Majority Vote, the final classification is determined as the classification that has been assigned the most times (the mode) by a set of underlying models. For example, if two models classified a statement as fraudulent and three classified it as legitimate, then legitimate would be the final classification because three is bigger than two. The Greek study (Kotsiantis et al. 2007) actually found that combining the individual model results using a decision tree (with linear splitting rules) was superior to

Majority Vote. This was trialled in this study, but it was found to be inappropriate. The reason is that TreeNet and Random Forests models have a 100% accurate classification on the training set. This means that a decision tree (even with linear splitting rules) will revert to solely using either of those two models and consequently no ensemble is created. Hence, majority voting is used with either three or five underlying models. An odd number of underlying models is used to remove the possibility for both fraudulent and legitimate classifications to receive the same number of votes.

Vote5 is a Majority Vote model with the following five underlying models:

1. A decision tree, DT_One, which was chosen in preference to DT_Suite as it resulted in a voting ensemble that is more accurate on the training data;
2. The TreeNet model, TN;
3. A Random Forests model, RF_8, which was chosen in preference to RF_66 because it had slightly better performance on the training data at extreme values of $PaRC_{IF}$;
4. The standard neural network, NN_BK; and,
5. A standard regression-based model, either a discriminant analysis or logistic regression model, but not both, to limit the ensemble to five models. The reason that either discriminant analysis or logistic regression is omitted is that their outputs have the highest coefficient of correlation. As stated in Chapter 3, ensemble models work well when the underlying models are diverse and so only one standard regression model has been included. The logistic regression model with the best training performance is LR_All, while DA_All is the best discriminant analysis model. LR_All was chosen in the end because it resulted in a voting ensemble that is slightly more accurate on the training data[127].

Vote3 are Majority Vote models with three underlying models. Many different combinations of the models from the Vote5 model were trialled as the three underlying models. DA_All and LR_All were not included in the same Vote3 model because of their similarity. The chosen models all incorporate both TN and RF_8, because performance on the training data was noticeably improved if they were included[128]. The three Vote3 models are:

[127] The choice of DA_All or LR_All is not very important as they both result in very similar voting ensemble performance on both the training and holdout data sets.

[128] For completeness, models that did not include both TN and RF_8 were assessed on the hold out data too, and once again their performance was inferior to Vote3 models that included both TN and RF_8.

- Vote3_RF_TN_DT comprising TN, RF_8 and DT_One, which resulted in slightly better training performance than using DT_Suite;
- Vote3_RF_TN_DA comprising TN, RF_8 and DA_All, which resulted in slightly better training performance than using LR_All; and,
- Vote3_RF_TN_NN comprising TN, RF_8 and NN_BK.

Ensembles using Averaging

Ensembles are also developed by averaging the (pseudo) probability of fraud outputs from a set of underlying models. The resulting average probability is then compared to the averaged cut-off values to obtain a classification. This technique for developing ensembles had encouraging results on Greek data (Kotsiantis et al. 2007).

Unlike Majority Vote, there is no reason to have an odd number of underlying models. Additionally, the process of averaging is not negatively influenced by including both DA_All and LR_All with highly correlated outputs. Initially, using all six of the models mentioned for majority voting was trialled. Using DT_One instead of DT_Suite again resulted in an ensemble with better training performance. The training performance of the ensemble model improved by removing the weakest individual model, NN_BK. The resulting average of five models is the AV5_NoNN model that is based on the DT_One, TN, RF_8, DA_All and LR_All models. Two other ensembles were developed by averaging the probability outputs of underlying models:

1. AV2_RF_TN that includes the TreeNet model and a Random Forests (RF_8) model, which are the two best performers on the training data; and,
2. AV3_RF_TN_DT that includes a single decision tree (DT_One) in addition to the TN and RF_8 models.

Other Ensembles

Nagadevara (2010) studied the classification problem of whether or not customers were going to change their telecommunications provider. He found encouraging empirical results from using a decision tree to classify the data and then using discriminant analysis to classify the customers predicted to change by the decision tree. Based on this concept, the DT_One_DA model has been developed in this research. This involves using discriminant analysis (as per DA_All) to further classify the riskier classifications from the DT_One model. The riskiest classifications are those that could result in the more costly error; specifically, the riskier classifications are

- Legitimate when $PaRC_{IF} > 1$, because it could result in the more costly error of missing fraud, and
- Fraud when $PaRC_{IF} < 1$, because it could result in the more costly error of falsely alleging fraud.

When $PaRC_{IF} = 1$, the DT_One_DA model is equivalent to the DT_One model. As the classifications of DT_One change with $PaRC_{IF}$, a separate discriminant analysis model was developed for each value of $PaRC_{IF}$.

A summary of all the models developed is provided in Table 5-15 on the next page.

Table 5-15. Summary of the models developed in this research.

Model Code	Model Details
Benchmark Models as estimated in prior research	
M-score	Probit analysis model developed by Beneish (1997, 1999a)
F-score	Stepwise logistic regression model developed by Dechow et al. (2011)
Discriminant Analysis (DA) and Logistic Regression (LR)	
DA_All	Inclusion of all variables (except the missing value exclusions)
DA_U15%	Only variables statistically significant at the 15% level in univariate tests
DA_U15%LR	Only variables statistically significant at the 15% level in the LR_U15% model
DA_U1	Only the most statistically significant variable from each main variable
DA_Step	Stepwise variable selection
LR_All	Inclusion of all variables (except the missing value exclusions)
LR_U15%	Only variables statistically significant at the 15% level in univariate tests
LR_U15%LR	Only variables statistically significant at the 15% level in the LR_U15% model
LR_U1	Only the most statistically significant variable from each main variable
LR_Step	Stepwise variable selection
LR_MS_S	Multi-stage LR model using the overall schema of explanatory variables
LR_MS_F	Multi-stage LR model using the new Fraud Detection Triangle framework
Artificial Neural Network (NN)	
NN_BK	Backpropagation neural network with 1 hidden layer containing 4 neurons
NN_GA_1	Backpropagation neural network optimised by a genetic algorithm with a minimum learning rate of 0.1 (NN_GA_1) or 0.5 (NN_GA_5)
NN_GA_5	
Decision Trees (DTs) and Ensembles of them: Random Forests (RF) and TreeNet (TN)	
DT_Suite	Suite of cost-sensitive CART decision trees, one for each value of $PaRC_{IF}$
DT_One	One cost-insensitive CART decision tree used for all values of $PaRC_{IF}$
RF_8	Random Forests with 1000 trees and variable subset size of 8
RF_66	Random Forests with 1000 trees and variable subset size of 66
TN	TreeNet with 0.01 learning rate and maximum number of nodes per tree of 12
Ensembles involving multiple modelling techniques	
NN_LR	NN_BK classifications included as an additional explanatory variable in LR_All
DT_LR	DT_One classifications included as an additional explanatory variable in LR_All
DTnode_LR_Step	Terminal node assignments from DT_One included as additional explanatory variables in the LR_Step model
NN-DTnode_LR_Step	NN_BK classifications included as an additional explanatory variable in the DTnode_LR_Step model
Vote5	Majority Vote between DT_One, TN, RF_8, NN_BK and LR_All
Vote3_RF_TN_DT	Majority Vote between TN, RF_8 and DT_One
Vote3_RF_TN_DA	Majority Vote between TN, RF_8 and DA_All
Vote3_RF_TN_NN	Majority Vote between TN, RF_8 and NN_BK
AV5_NoNN	Average between DT_One, TN, RF_8, LR_All and DA_All
AV2_RF_TN	Average between TN and RF_8
AV3_RF_TN_DT	Average between TN, RF_8 and DT_One
DT_One_DA	Discriminant analysis (based on DA_All) performed as a second stage on the riskier classifications from DT_One

5.3 Results

The results are presented in the order of some general findings, an analysis by modelling technique and then an overall comparative analysis. In addition, appendices A and B tabulate the WEC values for every model on the holdout and training data respectively, and appendices C and D tabulate the percentage accuracy of classifying fraudulent, legitimate and all financial statements on the holdout and training data respectively. In the analysis, lower Weighted Error Cost (WEC) values are preferred because they indicate better performance from a reduced cost of classification errors.

5.3.1 General Findings

5.3.1.a A Comparison of Training, Cross-validated and Holdout Performance

The summary Table 5-15 above contains 34 models that have each been evaluated for all 18 different values of $PaRC_{IF}$ on both the training and holdout data sets. Out of the total of $34 \times 18 = 612$ evaluations, the weighted error cost (WEC) was lower on the training data set in 606 cases. The six exceptions were the DA_Step and LR_Step models for the three lowest values of $PaRC_{IF}$. A t-test for comparing the means of a matched pairs sample indicated that the WEC is lower on the training data set (compared to the holdout data set), even at a statistical significance level of 0.001%. The associated t-statistic with 611 degrees of freedom for holdout WEC minus training WEC was $t = \frac{-1.21}{4.95/\sqrt{612}} = -6.05$.

Superior performance on the training data, as indicated by lower values of WEC, was expected because these models have been designed on the training data and the cut-off values have been optimised for the training data. The substantially different performance between the training and holdout data reinforces the value in using holdout data sets that provide more realistic estimates of model performance. This difference is illustrated in Table 5-16. This table also shows that although NN_GA_5 has substantially superior performance on the training data, it is inferior on the holdout data. This illustrates the fact that the best model on the training data is not necessarily the best model on the holdout data. This fact further substantiates the importance of testing models on a holdout data set.

Table 5-16. Illustration of the potential difference between training and testing performance.

	$PaRC_{IF} = 1$	
Model	**Training WEC**	**Holdout WEC**
LR_All	0.48	0.76
NN_GA_5	0.05	0.85

The decision tree, Random Forests and TreeNet models were all developed based on cross-validation. This provided an opportunity to compare cross-validated performance with performance on the holdout data set. Although the cross-validated WEC is higher than training WEC, a t-test for comparing the means of a matched pairs sample indicated that the cross-validated WEC is lower than the holdout WEC, even at a statistical significance level of 0.001%. The associated t-statistic with 89 degrees of freedom for holdout WEC minus cross-validated WEC is $t = \frac{-0.23}{0.28/\sqrt{90}} = -7.80$. Although many prior studies (discussed in Chapter 3) used cross-validation to estimate real-world accuracy, this result highlights that using holdout data sets provides more accurate estimates, particularly those that occur chronologically after the training data as happens in real-world scenarios. The use of such a holdout data set in this research provides a valuable contribution to the evaluation of previously used modelling techniques, as well as new techniques.

5.3.1.b A Comparison of Performance in This Research with That in Prior Research

Comparing the performance of models that have been obtained from different data sets can be misleading because of differences in the data sets. Models and techniques from previous studies have been included in this research, so that comparisons of their performance can be made using the same data. This also allows model performance in this research to be compared with that found in prior research. These comparisons, which are discussed below, indicate that models have found that fraud is more difficult to detect in the data used in this research. The statistics presented in this subsection are percentages instead of WEC values for easier interpretability and comparability because many prior studies did not use a WEC performance measure, mostly because they did conduct cost-sensitive research.

The M-score (Beneish 1997, 1999a) and F-score (Dechow et al. 2011) models have been used as they were presented in the original research, including using the same variable coefficients. Furthermore, both models were originally evaluated on holdout data sets that occurred chronologically after the training data, which is consistent with this research. This enabled useful comparisons of the data sets to be made. The accuracy of the M-score model decreases by almost 20% on average on the current data, across a range of different ratios of error costs ($C_{IF}: C_{IL}$) as demonstrated in Table 5-17. The M-score model was originally evaluated on only 24 cases of fraud compared with 169 in this study. The larger number of fraud cases in this study might include a more diverse range of fraud cases than the limited original study. Increased diversity usually results in more difficulty in classification, and so this could be a contributor to the M-score's reduced accuracy on the data used in this study. The difference in accuracy might also indicate that the detection of fraudulent financial statements has become more difficult over time, as this study was published 16 years prior to this dissertation. This could be as a result of fraudsters becoming more sophisticated and better at concealing fraud.

Table 5-17. The accuracy of the M-score model in this research compared to its original study.

Scenario[129]	**Study**	**Percentage Accuracy**		
		Fraud	**Legitimate**	**Overall**
$C_{IF}: C_{IL} = 10: 1, p(F) = 2.84\%$ $\therefore PaRC_{IF} = 0.3$	Original	38%	94%	67%
	This research	0%	99%	50%
$C_{IF}: C_{IL} = 20: 1, p(F) = 2.84\%$ $\therefore PaRC_{IF} = 0.6$	Original	50%	93%	71%
	This research	1%	98%	49%
$C_{IF}: C_{IL} = 40: 1, p(F) = 2.84\%$ $\therefore PaRC_{IF} = 1$	Original	54%	91%	73%
	This research	56%	50%	53%

Similar to the M-score, the accuracy of the F-score model reduced substantially on the data set used in this research as shown in Table 5-18. On this data set, the F-score model has reduced accuracy for detecting both fraudulent and legitimate statements when using the same cut-off value as the original study. Furthermore, if the cut-off is varied so that the accuracy of detecting fraud is the same, then the accuracy of detecting legitimate statements is nearly half of the accuracy in the original study. The F-score does however retain its relatively superior ability to detect fraudulent financial statements, compared with legitimate. Unlike the M-score, it was evaluated on many more fraud cases (107). This means that the

[129] The corresponding $PaRC_{IF}$ was calculated using the same equation as above, $PaRC_{IF} = \frac{C_{IF}}{C_{IL}} \times \frac{p(F)}{p(L)}$.

size of the data set is less likely to explain the difference in performance between the F-score in the original study and in this research. However, the F-score was originally tested on holdout data from 1999 to 2002, which is still earlier than the current holdout data set from 2003 to 2007. Consequently, the reduced performance of the F-score might also be associated with fraud becoming more difficult to detect over time.

Table 5-18. The accuracy of the F-score model in this research compared to its original study.

	Percentage Accuracy		
Study	**Fraud**	**Legitimate**	**Overall**
Original	74%	62%	68%
This research: same cut-off as original study	59%	47%	53%
This research: cut-off varied so fraud accuracy is 74%	74%	33%	53%

There is another potential reason for the reduced performance of the M-score and F-score models in this current research. If the nature of fraud and consequently the indicators of fraud have changed, then these models developed on older data might be searching for outdated patterns that no longer relate to fraud. Consequently, their performance would be reduced on the more recent data in this study. This highlights the importance of regularly updating models and re-assessing their accuracy. This was also empirically shown on a related problem of financial distress prediction, where newer updated models clearly outperformed the most famous and widely-used, older model (Gepp and Kumar 2014).

The performance of models as recorded in their original study has been compared with the performance of models developed using the same modelling technique on the training data in this study. Unlike the M-score and F-score models, these comparisons are not subject to the influence of potentially searching for outdated patterns. Nevertheless, similar results have been found, as shown in Table 5-19. Accuracy on the data used in this study is consistently worse than that presented in the corresponding original study. Potential reasons for this are that:

- The original studies, with the exception of Whiting et al. (2012), tested their models on far fewer fraud cases than the 169 used in this research. As stated above, this might make detection in the current study more difficult. However, it is also more similar to the real world and so results in more accurate estimates of performance in this study;

- Some of the original studies have not used holdout data that occur chronologically after the training data to obtain more realistic real-world performance assessments. This would have caused their performance levels to be inflated;
- Fraud may be becoming more difficult to detect over time, which would cause reduced accuracy in this study because it uses more recent data than these prior studies. This is actually an advantage of this study, because the results are more applicable to today even if the reported accuracy is lower.

Table 5-19. A comparison of the performance of modelling techniques in their original study with the performance in this research (indicated by shading).

Study	Test Data	**Accuracy: F**raud, **L**egitimate, **O**verall		
		F	**L**	**O**
Standard regression-based				
From review in Chapter 3	Varied	Overall: 67% to 73%		
This research ($PaRC_{IF} = 1$)	Chronological holdout	Overall: 57% to 62%		
Backpropagation Neural Network				
Green and Choi (1997)	Random holdout	74%	68%	71%
McKee (2009)	Unclear	48%	96%	71%
NN_BK ($PaRC_{IF} = 1$)	Chronological holdout	57%	56%	59%
Optimised Neural Network				
Fanning and Cogger (1998)	Chronological holdout	66%	59%	63%
NN_GA_5 ($PaRC_{IF} = 1$)	Chronological holdout	65%	50%	58%
Ensemble: Neural network and logistic regression				
McKee (2009)	Unclear	71%	83%	77%
NN_LR ($PaRC_{IF} = 1$)	Chronological holdout	51%	70%	61%
Ensemble: Random Forests ($variable\ subset\ size\ = \sqrt{number\ of\ variables}$)				
Whiting et al. (2012)	Cross-validation	82%	83%	82%
RF_8 ($PaRC_{IF} = 1$)	Chronological holdout	67%	64%	66%
Ensemble: TreeNet / Stochastic Gradient Boosting				
Whiting et al. (2012)	Cross-validation	74%	84%	79%
TN ($PaRC_{IF} = 1$)	Chronological holdout	55%	80%	67%

Overall, it is clear that models find it more difficult to detect fraud in the data used to evaluate model performance in this study than the data used in prior studies. Directly comparing results from this research with that in prior research is therefore likely to be misleading. To increase the validity of comparisons, models developed using techniques from previous studies have been included in this research so that comparisons that use the same data set can be made.

5.3.2 Analysis by Modelling Technique

The accuracy of the models measured by WEC on the holdout data are discussed according to the modelling technique used. Lower values of WEC indicate better performance. As stated in Section 5.2.1.d, a naïve model that classifies all statements as fraudulent when $PaRC_{IF} \geq 1$ and all statements as legitimate when $PaRC_{IF} < 1$ always has a WEC of one. Consequently, WEC values lower than one indicate accuracy that is superior to a naïve model that could be practically useful. Performance on the holdout data is the most important result, but performance on the training data is also included to assist interpretation and explanation of the holdout performance.

5.3.2.a Benchmark Models

The performance of the M-score and F-score models are shown in Figure 5-2. Their WEC values are the same for high values of $PaRC_{IF}$, but the M-score model has much higher WEC for low values of $PaRC_{IF}$. The performance is relatively similar for mid-range values of $PaRC_{IF}$. It is not surprising that the two models respond differently to varying values of $PaRC_{IF}$, because they are based on different explanatory variables (as previously defined in Section 5.2.2). The F-score model responds to relatively high and low levels of $PaRC_{IF}$ with a stable WEC of one. More importantly, the only time the WEC is lower than one for either model is when $PaRC_{IF}$ is one. In this case, the WEC is 0.94 and 0.96 for the M-score and F-score model respectively. This means that neither model is practically useful when the $PaRC_{IF}$ is not equal to one, as a naïve model is superior.

Although the M-score is superior to the F-score when $PaRC_{IF}$ is 0.5, 0.6, 1 or 3, the F-score model more consistently has lower WEC. Hence, the F-score model is used in figures hereafter when only one benchmark model is presented in order to reduce the complexity of figures.

Figure 5-2. A comparison of the holdout performance of the benchmark models. The M-score model is not shown when $PaRC_{IF} = 0.004$ because it has a WEC of 2.48.

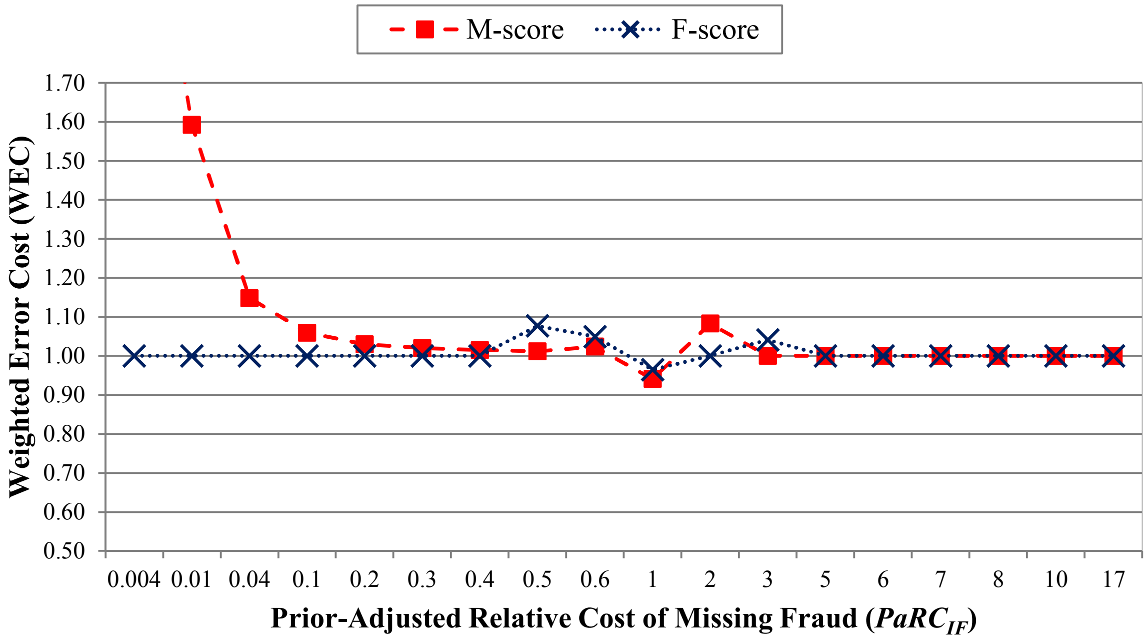
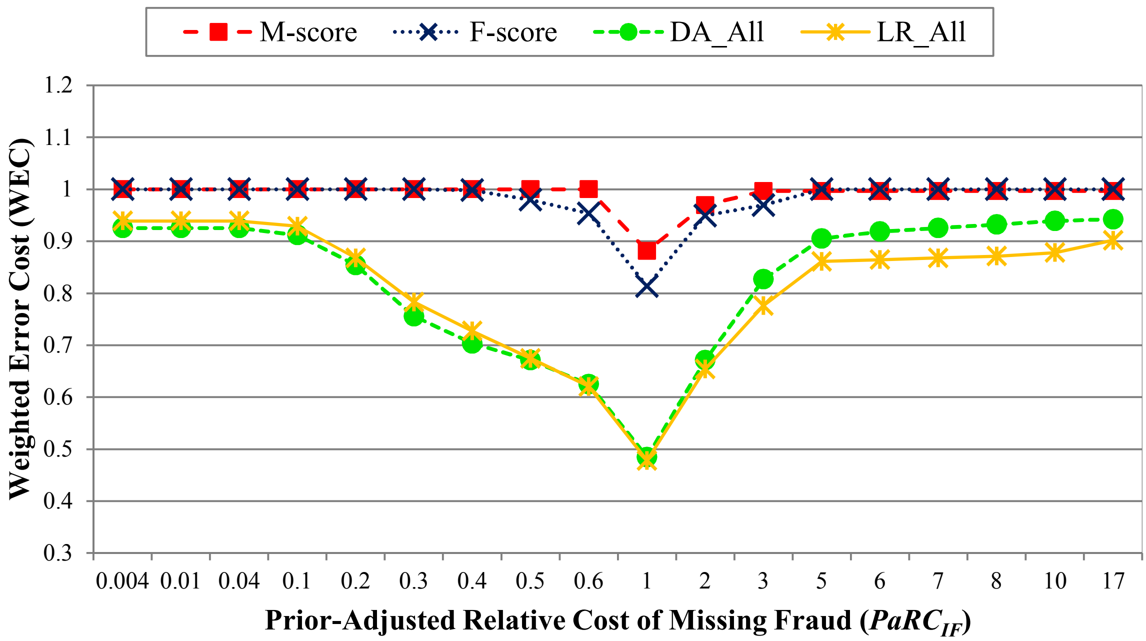


Figure 5-3. The training performance of the benchmark models compared to two standard regression-based models.



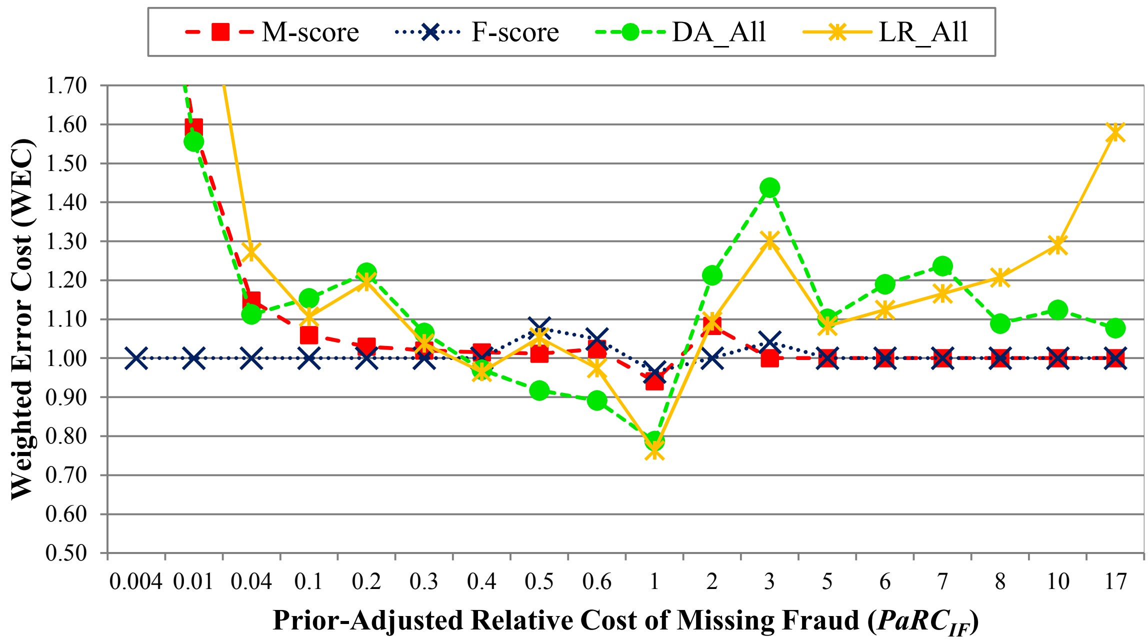
5.3.2.b Standard Regression-based Models

Benchmark, DA_All and LR_All Models

Figure 5-3 above illustrates that standard regression-based models, including the benchmark models, perform better on the training data when $PaRC_{IF}$ is closer to one. This is shown by smaller values of WEC as $PaRC_{IF}$ approaches one, which creates a v-shape. The performance of the discriminant analysis model DA_All and the logistic regression model LR_All are similar with the former being better for lower values of $PaRC_{IF}$ and the latter for higher values of $PaRC_{IF}$. The figure also shows that the accuracy of benchmark models on the training data is substantially inferior with much higher WECs. This is expected as the benchmark models, unlike all other models, have not been developed on the training data and so it is more akin to holdout data for the benchmark models. The fact that both data sets are new to the benchmark models is illustrated by their performance on the holdout data in Figure 5-2 being similar to their performance on the training data in Figure 5-3. The main difference is that the M-score does not have higher WECs at low levels of $PaRC_{IF}$ on the training data (Figure 5-3), because the cut-off values were optimised on that data.

The performance on the holdout data is illustrated in Figure 5-4. The performance of DA_All and LR_All is drastically different on the holdout data, which highlights the importance of its use. It is again demonstrated that models perform best when $PaRC_{IF}$ approaches one. However, on the training data, the benchmark models were inferior in all cases, but they are actually superior on the holdout data when $PaRC_{IF} \leq 0.3$ or $PaRC_{IF} \geq 2$. The LR_All model has the lowest WEC when $PaRC_{IF}$ equals 1 or 0.4. Nevertheless, the DA_All model is the best model overall when $0.4 \leq PaRC_{IF} \leq 1$.

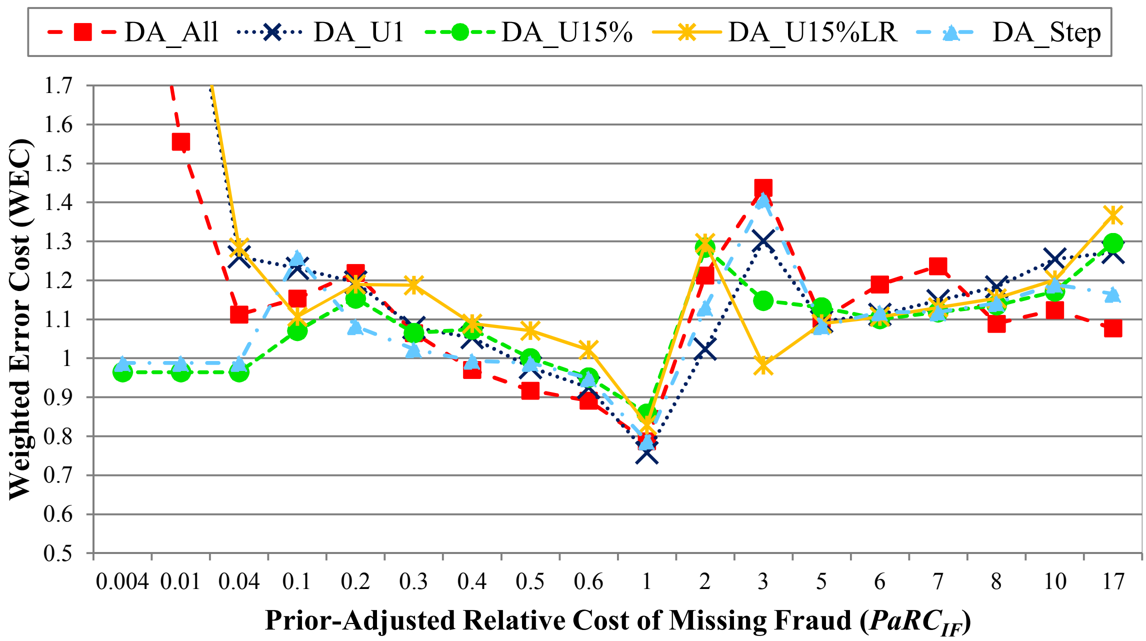
Figure 5-4. The holdout performance of the benchmark models compared to two standard regression-based models. The DA_All, LR_All and M-score models are not shown for some low values of $PaRC_{IF}$ as the WECs are greater than two.



Discriminant Analysis (DA) Models

The performance of all the discriminant analysis models on the holdout data is shown in Figure 5-5. Once again, the models perform best (lower WEC) when the value of $PaRC_{IF}$ approaches one. The performance is similar overall with no model standing out as the worst or the best. The best model varies depending on the situation in terms of $PaRC_{IF}$. Every model is the best in at least one situation, as well as the worst in at least one situation. DA_U15%LR is the only model to be best in only one situation ($PaRC_{IF} = 3$), but in that situation it is noticeably superior and the only model with a WEC lower than one. At the same time, this model is the worst in many other situations and so the choice of model can be important depending on the value of $PaRC_{IF}$. Overall, the WECs are above one for much of the graph, which means that in many cases a naïve model that classifies all statements the same way would be superior. At least one of the models has a WEC lower than one in 8 out of the 18 values of $PaRC_{IF}$, specifically when $PaRC_{IF} \leq 0.04$, $0.4 \leq PaRC_{IF} \leq 1$ and when $PaRC_{IF} = 3$.

Figure 5-5. A comparison of the holdout performance of the discriminant analysis models. When $PaRC_{IF} = 0.004$, the DA_All (WEC = 2.44), DA_U1 (WEC = 3.92) and DA_U15%LR (WEC = 3.95) models are not shown. When $PaRC_{IF} = 0.01$, the DA_U1 (WEC = 2.15) and DA_U15%LR (WEC = 2.17) models are also not shown. The models are ordered below from most number of explanatory variables (left) to least number (right).

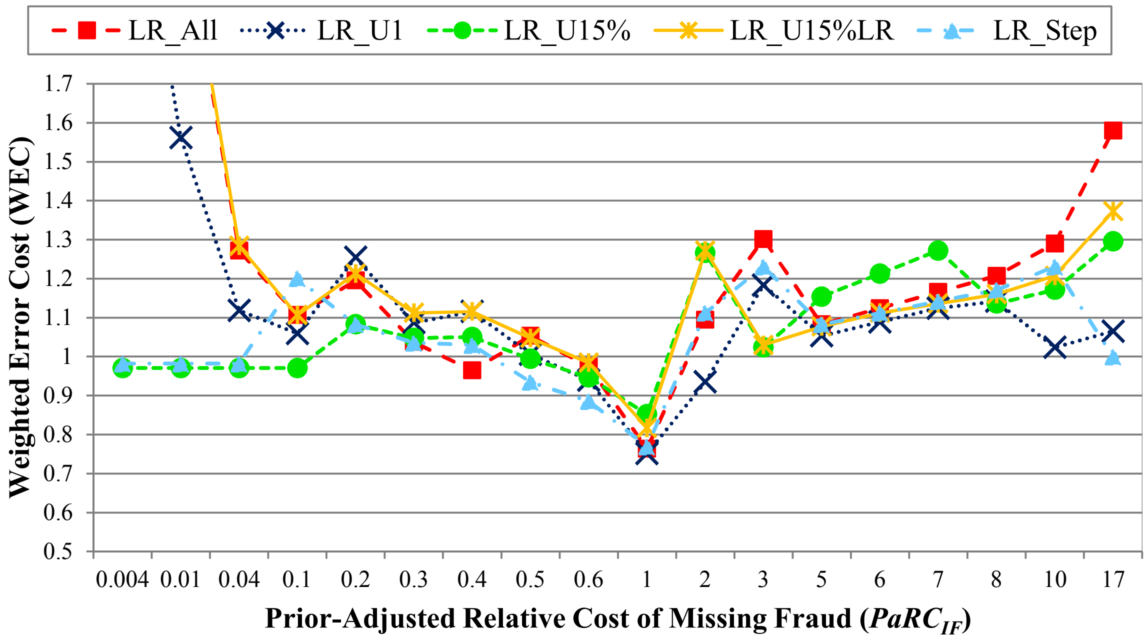


Logistic Regression (LR) Models

The performance of all the logistic regression models on the holdout data is shown in Figure 5-6. Again, the models perform best when $PaRC_{IF}$ approaches one and the results are similar overall without an obvious best model. There are also other findings that are consistent with those made from discriminant analysis models:

- The best model varies depending on the situation in terms of $PaRC_{IF}$;
- Every model is the worst and the best in at least one situation, except the LR_U15%LR model that has at best the second lowest WEC when $PaRC_{IF} = 3$. This is the same situation in which the corresponding DA model DA_U15%LR performed best; and,
- The Step and U15% models are the best for the lower values of $PaRC_{IF}$, and the U1 model is the best when $PaRC_{IF} = 1$.

Figure 5-6. A comparison of the holdout performance of the logistic regression models. When $PaRC_{IF} = 0.004$, the LR_All (WEC = 3.93), LR_U1 (WEC = 2.45) and LR_U15%LR (WEC = 3.95) models are not shown. When $PaRC_{IF} = 0.01$, the LR_All (WEC = 2.16) and LR_U15%LR (WEC = 2.17) models are also not shown. The models are ordered according to the number of explanatory variables they use, from most (left) to least (right).



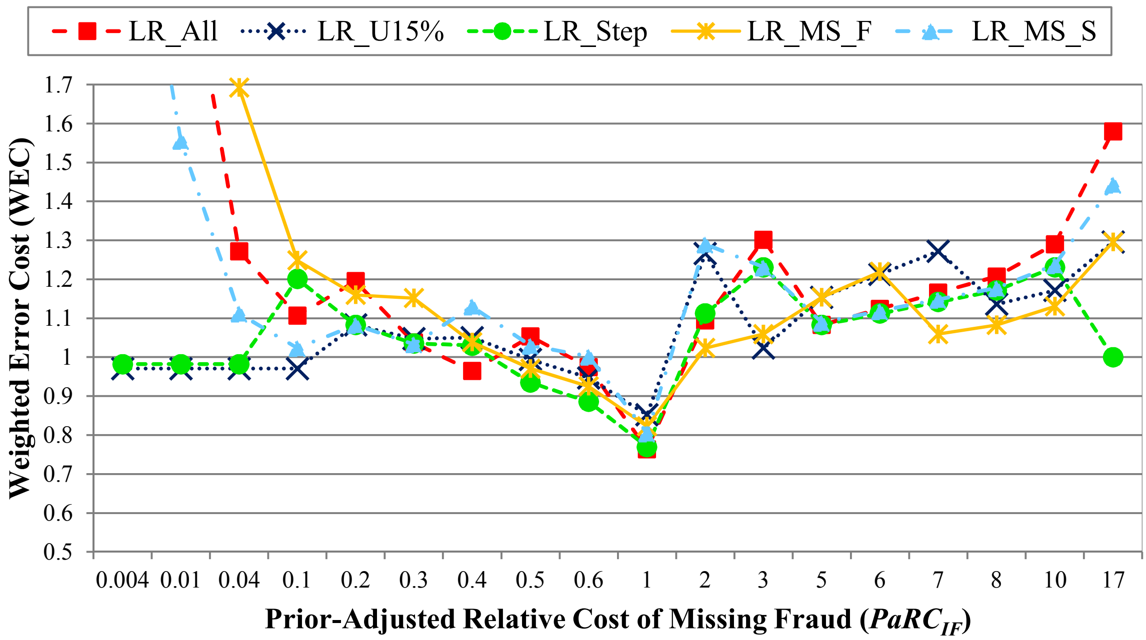
DA and LR models do not always produce relatively similar results when the same (number of) variables are used. For example, the All model is the best DA model for the largest values of $PaRC_{IF}$, but the All model is the worst LR model in the same situation. Overall, the WEC values are again above one for much of the graph. At least one of the models has a WEC lower than one in 9 out of the 18 values of $PaRC_{IF}$, specifically when

$PaRC_{IF} \leq 0.01$ and when $0.4 \leq PaRC_{IF} \leq 2$. Similar to the DA model results, some LR models find it easier to adjust to lower (rather than higher) values of $PaRC_{IF}$, while some models appear unable to adjust with extremely high WEC values that are too high to be shown in Figure 5-6.

Multi-stage Logistic Regression (LR_MS) Models

Although they are not consistently superior, the multi-stage logistic regressions have similar performance to other logistic regression models, as illustrated in Figure 5-7. The three comparison logistic regression models were chosen as the models using the largest, median and smallest number of explanatory variables. Similar to some of the LR and DA models, the LR_MS models have extremely high WEC at the lowest levels of $PaRC_{IF}$. When compared with all the logistic regression models, LR_MS_F has the lowest WEC when $PaRC_{IF}$ is 7 or 8 and LR_MS_S has the lowest WEC when $PaRC_{IF}$ is 0.3. Overall, the LR_MS_F model has a WEC lower than a naïve model that classifies all statements the same way (WEC = 1) when $PaRC_{IF}$ is 0.5, 0.6 or 1, while this only occurs for the LR_MS_S model when $PaRC_{IF}$ is 1. Once again, all the models have lower WECs as $PaRC_{IF}$ approaches one.

Figure 5-7. The holdout performance of the multi-stage logistic regression models compared with other logistic regression models. The LR_All, LR_MS_F and LR_MS_S models are not shown for values of $PaRC_{IF} \leq 0.01$ as their WECs are greater than two.



A Comparison of Discriminant Analysis (DA) and Logistic Regression (LR) Models

The performance of both discriminant analysis and logistic regression models on the holdout data is shown in Figure 5-8. Some of the models were not included because displaying all of them results in an overly complex graph that is very difficult to interpret and additional models were not needed to illustrate the comparison between DA and LR models. The Step and U15% models were included as they have more stable WECs in Figures 5-6 and 5-7. Furthermore, amongst all the DA models, the Step and U15% models were the best two according to the average of ranking the techniques according to:

- The number of values of $PaRC_{IF}$ for which the WEC is less than one (more is better), which measures how many times the model is superior to a naïve model classifying all statements the same way;
- The number of values of $PaRC_{IF}$ for which the WEC is the lowest model (more is better), which measures how many times the model is the best model; and,
- The average ranking between the models across all values of $PaRC_{IF}$, which measures performance across all values of $PaRC_{IF}$.

The same is true for the Step and U15% models amongst all the LR models, including the multi-stage LR models.

Figure 5-8. A comparison of the holdout performance of discriminant analysis and logistic regression models. The LR_All, LR_MS_F and LR_MS_S models are not shown for values of $PaRC_{IF} \leq 0.01$ as their WECs are greater than two.

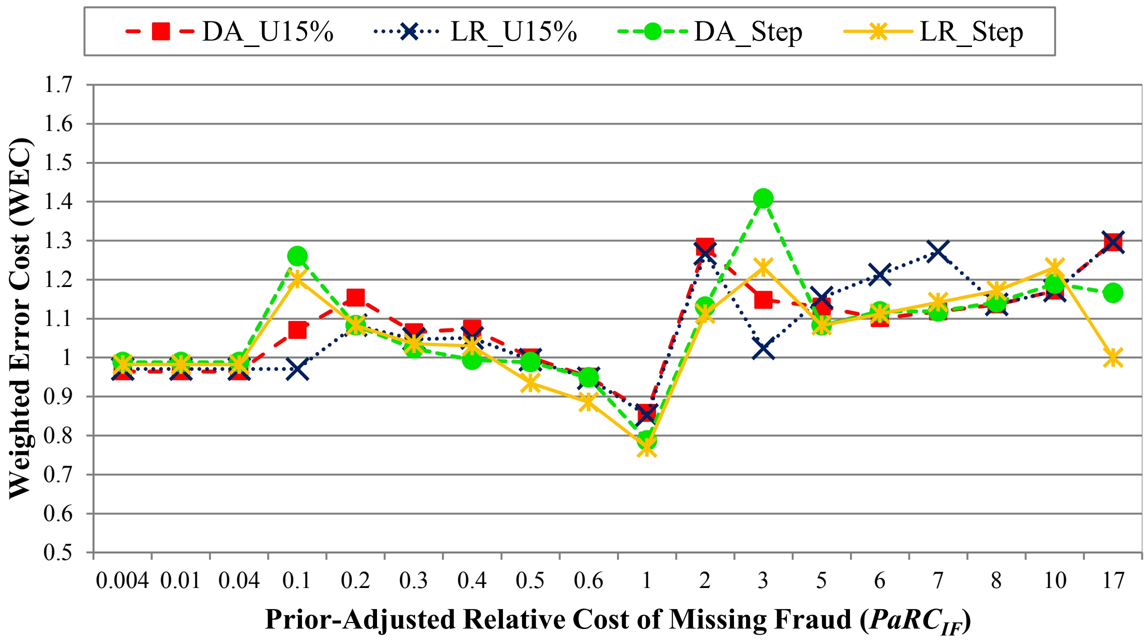


Figure 5-8 reveals that overall the logistic regression and discriminant analysis models have very similar performance in terms of WEC (as also occurred between the DA_All and LR_All models in Figure 5-4). All models have a V-shape near where $PaRC_{IF}$ is one, where the best performance occurs. Additionally, both logistic regression and discriminant analysis are better when $PaRC_{IF}$ is lower, rather than greater, than one. This is highlighted by no model having a WEC less than one when $PaRC_{IF} > 1$. The most notable exception to similar performance is when $PaRC_{IF}$ is equal to three.

Overall, the performance is similar between both discriminant analysis and logistic regression models. However, it is useful to select one model as the best. This model is to be used for comparisons with other models. LR_Step has been chosen as the best model out of the four presented in Figure 5-8 as it produces the most consistent relatively-low WEC. The LR_Step model has the best average ranking across three criteria; specifically it has:

3. A WEC less than one for six values of $PaRC_{IF}$, which is the second best;
4. The lowest WEC for five values of $PaRC_{IF}$, which is the second highest amount; and,
5. An average ranking of 2.3 across all values of $PaRC_{IF}$, which is the best.

It is noteworthy that all of the models compared in this subsection, including the representative LR_Step model, use a subset of the total number of explanatory variables. This suggests that variable reduction techniques, such as selection by statistical significance according to univariate analysis or stepwise selection, assist in developing discriminant analysis and logistic regression models for detecting financial statement fraud. Using univariate analysis to select variables for a discriminant analysis model has been trialled in previous research (Skousen et al. 2009). This research has shown that this variable selection technique also works well for a logistic regression model. On the other hand, stepwise variable selection has been trialled for logistic regression in multiple prior studies (Persons 1995; Bell and Carcello 2000; Dechow et al. 2011), and this research has shown that stepwise variable selection also works well for discriminant analysis models that detect financial statement fraud.

5.3.2.c Artificial Neural Network (NN) Models

The performance of the three neural network models on the training and holdout data is shown in Figure 5-9 and Figure 5-10 respectively. The results on the training data indicate the consistent result of superior performance (as indicated by lower WEC) when $PaRC_{IF} = 1$. It was expected that the genetically optimised neural network (NN_GA) models would have more accurate training results. This was true for the NN_GA_5 model that has a substantially lower WEC in all cases when compared with the standard backpropagation NN_BK model. This result is shown in Figure 5-9 that also reveals the NN_GA_1 model predominantly produced classifications with higher WEC on the training data. This poor performance also extended to the holdout data as shown in Figure 5-10. The holdout WEC for the NN_GA_1 model is never below one, which means that the model is not practically useful because a naïve model classifying all statements the same way is superior in all cases. Thus, changing the minimum learning rate of the NN_GA model from the default of 0.5 (as used in NN_GA_5) to 0.1 as used in NN_GA_1 is not recommended when developing models to detect financial statement fraud.

Figure 5-9. A comparison of the training performance of the artificial neural network models.

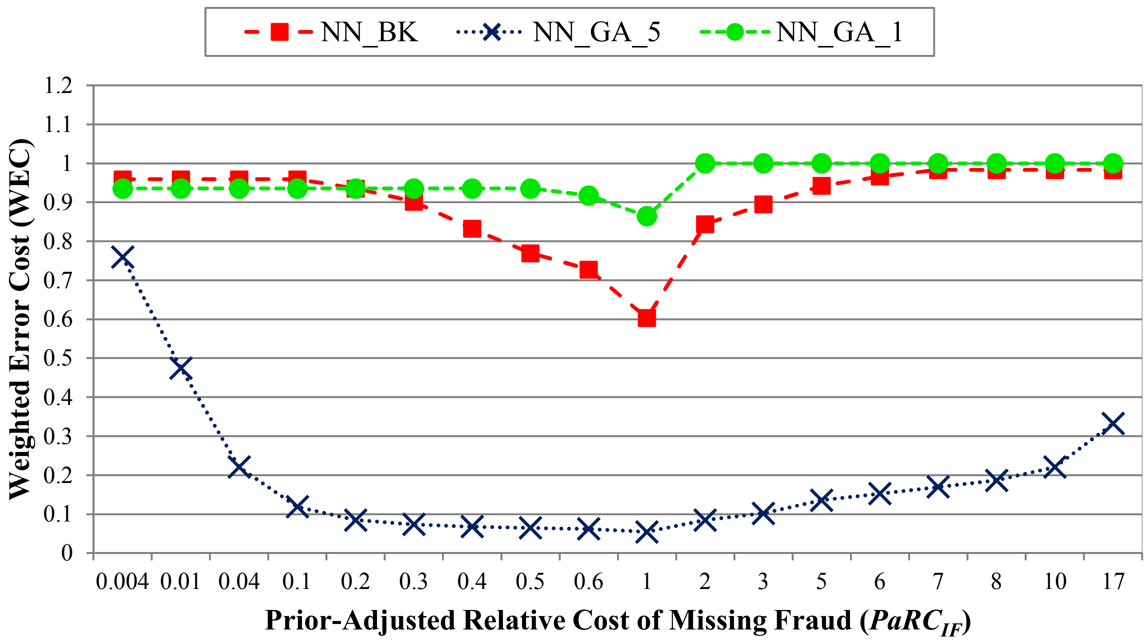
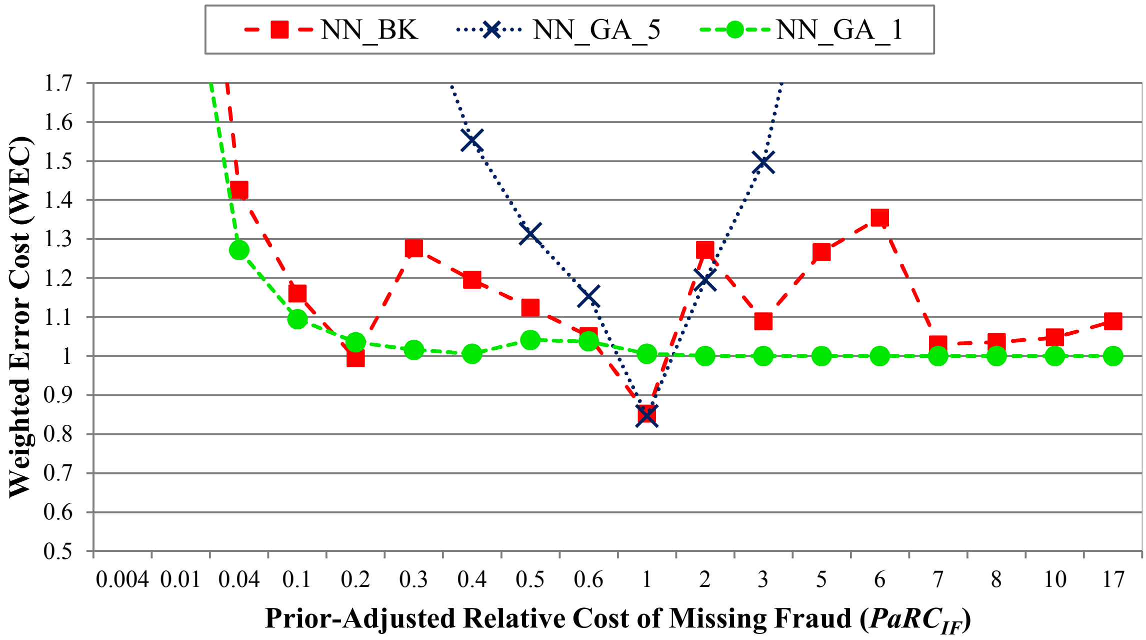


Figure 5-10. A comparison of the holdout performance of the artificial neural network models. The NN_BK and NN_GA1 models are not shown for the lowest two values of $PaRC_{IF}$ as their WEC ranges between 2.16 and 5.42. Most of the NN_GA5 model is not shown because its WEC is too high, ranging up to 43.62 when $PaRC_{IF}$ is 0.004.



The NN_BK and NN_GA_5 models only have a WEC less than one when $PaRC_{IF}$ is one. In that situation, the WEC of the genetically optimised NN_GA_5 model is 0.846, slightly lower than NN_BK's WEC of 0.852. However, the WEC of the NN_GA_5 model quickly becomes relatively high for $PaRC_{IF}$ values that differ from one. This is an important result that extends previous encouraging findings from optimised neural network models that did not consider differing values of $PaRC_{IF}$ (Fanning and Cogger 1998). The poor performance of the genetically optimised neural networks is unexpected. Although the NN_GA_5 model more accurately learnt patterns in the training data set, those patterns have not persisted into the future holdout data. This highlights the importance of choosing models based on their holdout performance and not their training performance alone.

Overall, the genetically optimised neural networks have not been useful in detecting financial statement fraud on the data set used in this research. The only advantage revealed was the reduced time and effort required to develop them, which was also found by Fanning and Cogger (1998). Although the NN_GA_5 model is slightly better when $PaRC_{IF} = 1$, the standard backpropagation NN_BK model is the best neural network model.

5.3.2.d Individual Decision Tree (DT) Models

The performance of the two decision tree models on both the training and holdout data is shown in Figure 5-11. Both models, particularly DT_One, have low WEC on the training data for the middle range of $PaRC_{IF}$ values. Both models also have a WEC equal to one for the highest and lowest values of $PaRC_{IF}$, both on the training data and holdout data. Although this only represents performance equivalent to a naïve model for extreme values of $PaRC_{IF}$, it is relatively good performance compared to most of the other models with WECs of greater than one. The WEC of the DT_One model is less than one when $PaRC_{IF}$ is 0.5, 0.6 or 1, while this is only true for the DT_Suite model when $PaRC_{IF}$ is 1 (when the two models are identical). The only WEC above one for the DT_One model is 1.05 when $PaRC_{IF}$ is 2, which makes it the most stable model analysed so far in terms of WEC. Overall, it is clear that DT_One is superior to DT_Suite because it has equal or lower WECs in every case on both the training and holdout data.

DT_Suite comprises a set of cost-sensitive decision tree models, a tree developed for each value of $PaRC_{IF}$, while DT_One utilises the tree developed for $PaRC_{IF} = 1$ in all scenarios. That is, integrating different misclassification costs into the model development process is a feature of CART that has been utilised by DT_Suite. However, the DT_Suite model has been outperformed by the DT_One model. It illustrates that for this financial statement fraud detection data set, changing values of $PaRC_{IF}$ is better modelled by manipulating the cut-off values on a single cost-insensitive tree. A possible reason for this can be found when analysing the complexity of the decision trees in the DT_Suite model, which is shown in Table 5-20 below. The most complex tree is used when $PaRC_{IF} = 1$, which means that the DT_One model always uses the most complex tree from DT_Suite. Therefore, less complex trees performed better on the training data when $PaRC_{IF} \neq 1$. However, the additional complexity of the DT_One resulted in a substantially smaller reduction in accuracy when moving to the holdout data. Thus, the more complex DT_One model performed better on the holdout data.

Figure 5-11. A comparison of the training and the holdout performance of the decision tree models.

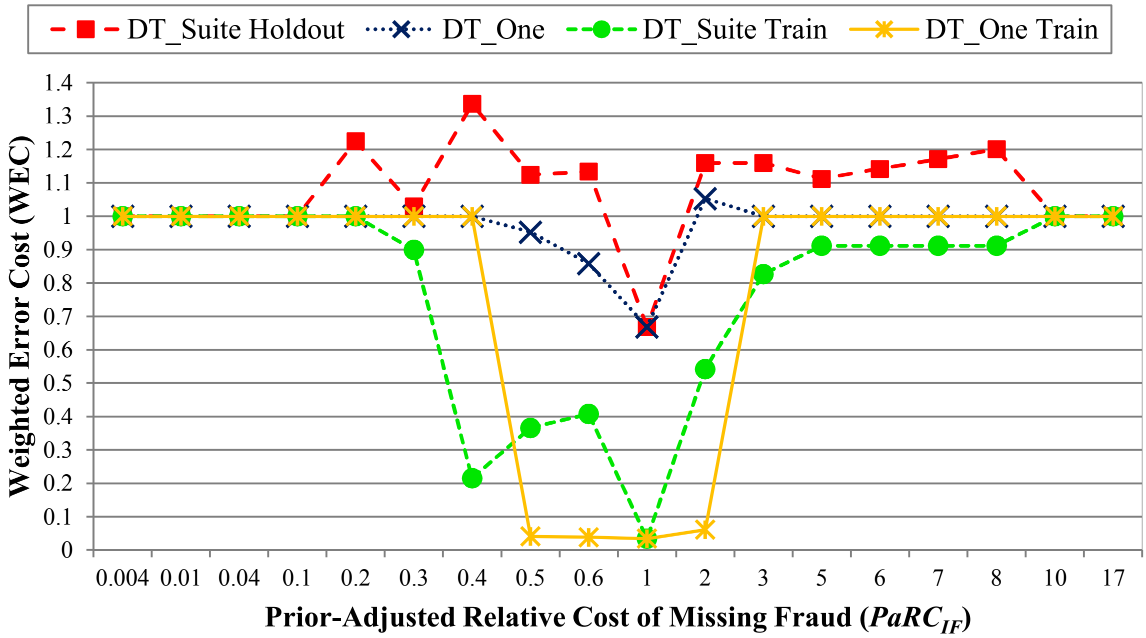


Table 5-20. The complexity of the decision trees in the DT_Suite model, as measured by the number of terminal nodes. The shaded regions indicate the situations in which the model did not develop a tree, but rather reverted to the naïve model of classifying all financial statements the same way.

	Prior – adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
	.004	.01	.04	.1	.2	.3	.4	.5	.6	1	2	3	5	6	7	8	10	17
Number of Terminal Nodes					4	5	46	27	21	70	16	7	3	3	3	3		

5.3.2.e Ensembles of Decision Trees: Random Forests (RF) and TreeNet (TN)

All three decision tree ensembles have very low WECs as shown in Figure 5-12. In fact, they have perfect accuracy (WEC = 0) for values of $PaRC_{IF}$ between 0.3 and 5. It is clear that the models have accurately learnt patterns in the training data. However, unlike genetically optimised neural networks, many of these patterns have persisted into the future holdout data as evidenced by many holdout WECs less than one as shown in Figure 5-13. These models, particularly the TN model, adapt relatively well to higher values of $PaRC_{IF}$. They are the first models with multiple WEC values below one when $PaRC_{IF} > 1$.

Figure 5-12. A comparison of the training performance of the decision tree ensembles.

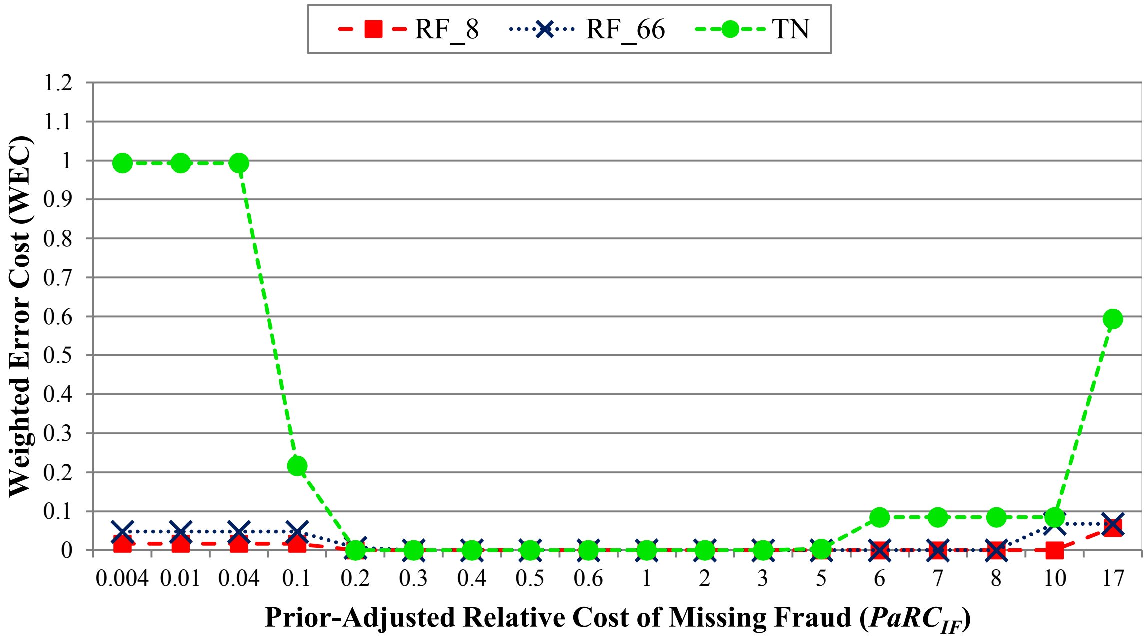
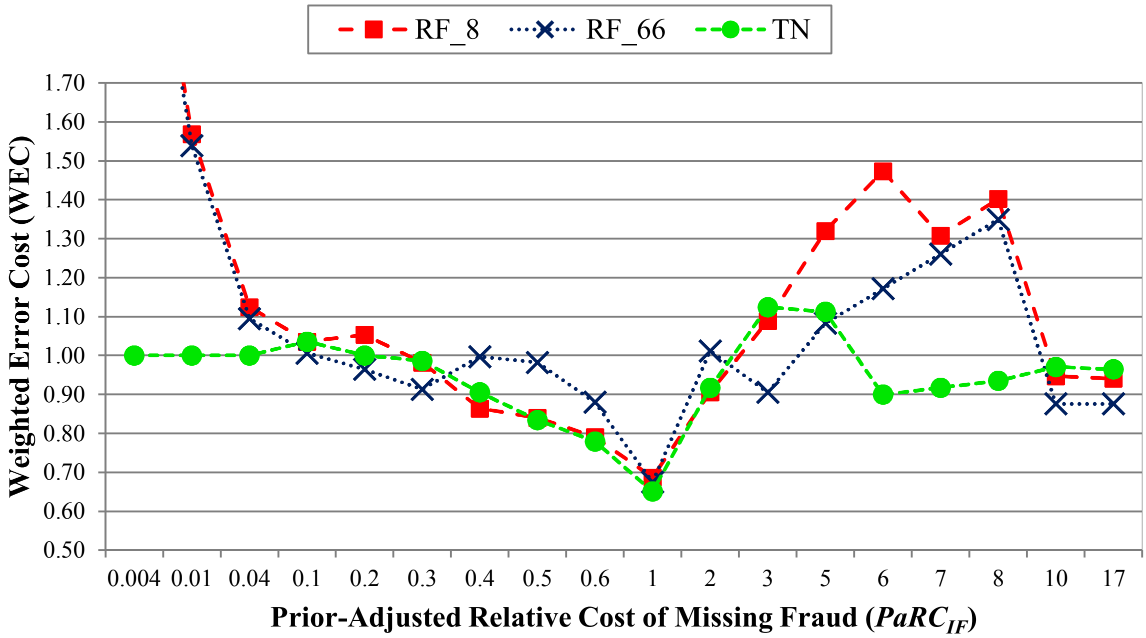


Figure 5-13. A comparison of the holdout performance of the decision tree ensembles. The RF_8 and RF_66 models are not shown when $PaRC_{IF} = 0.004$ as their WECs are 2.46 and 2.43 respectively.



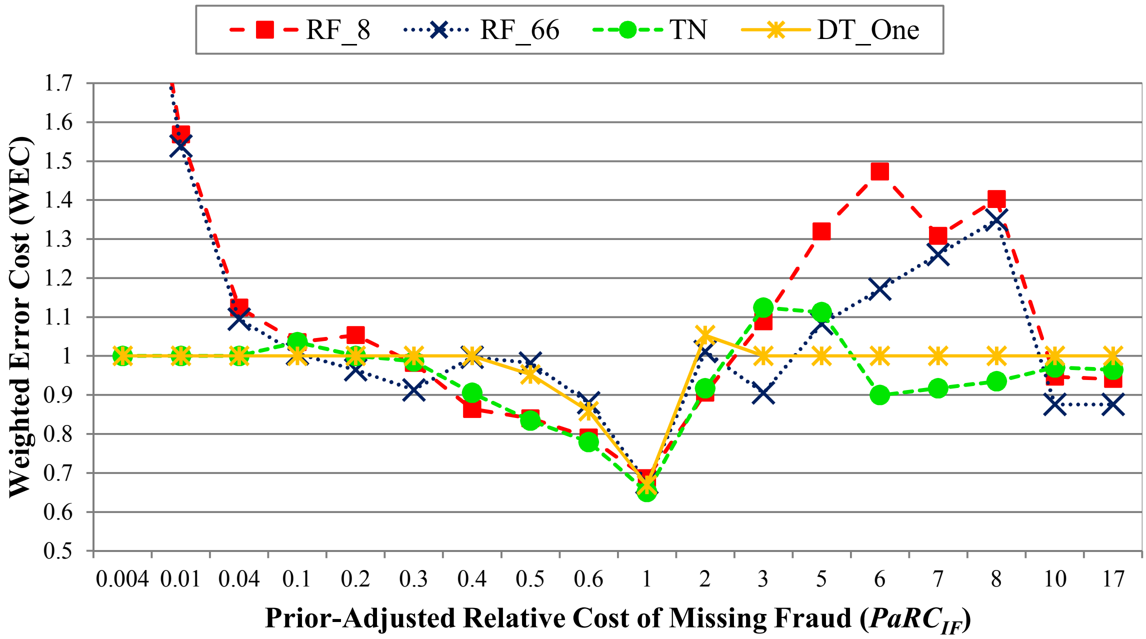
The importance of using holdout data for evaluation is again apparent. While the TN model is the worst on the training data, it is arguably the best overall on the holdout data. The TN model has a WEC less than one the largest number of times (11). The TN model also has the lowest WEC in 9 situations, which is more than either RF model. Furthermore, TN has more stable results over the different values of $PaRC_{IF}$. When TN does not have the lowest

WEC, it usually has a WEC comparable with the lowest WEC. While the TN model is the best overall, the RF_66 model is noticeably superior in some situations, the most notable being when $PaRC_{IF}$ is 3. In contrast, Whiting et al. (2012) previously found Random Forests to outperform stochastic gradient boosting (used in the TN model), but this was using cross-validated evaluation. The methodological improvements used to develop the TN model (described in Section 5.2.3.e) could also explain this different result.

The RF_66 model has a lower WEC than RF_8 for 14 out of the 18 values of $PaRC_{IF}$. However, the RF_66 is noticeably inferior in the other 4 cases that occur when $PaRC_{IF}$ is between 0.4 and 2. The superiority of the RF_66 model, albeit minor, is unexpected. RF_66 incorporates only one random element by random sampling (with replacement) the data. In contrast, the RF_8 model incorporates an additional random element as suggested in prior research (Bhattacharya et al. 2011a) by randomly selecting a subset of 8 variables available for consideration at every tree node. Overall, the results are not conclusive about the best size of the random variable subset to be used when developing Random Forests models. Because of its accuracy in this research, it is advised that future research include a Random Forests model that has all variables available at every tree node.

Figure 5-14 provides insight into the performance of ensembles of decision trees relative to an individual decision tree. While the single decision tree (DT_One) is rarely the best technique, it does have comparable performance for many values of $PaRC_{IF}$. DT_One has varied performance relative to the two RF models, but it is outperformed by TreeNet most of the time. DT_One only has a lower WEC than TreeNet for three values of $PaRC_{IF}$. However, it should be noted that DT_One is not practically useful when $PaRC_{IF}$ is outside of the range 0.5 to 2 as it reverts to classifying all financial statements the same way. The tree ensemble models on the other hand provide classification models that function for all values of $PaRC_{IF}$, and which have WEC values lower than one in many cases. Overall, TreeNet remains the best individual or ensemble decision tree model.

Figure 5-14. The holdout performance of ensembles of decision trees compared with a single decision tree (DT_One).



5.3.2.f Ensembles involving multiple modelling techniques

Ensembles using Majority Vote (Vote)

The WEC of the four ensembles that use Majority Vote on the holdout data are very similar, as shown in Figure 5-15 below. Visually, the Vote5 and Vote3_RF_TN_NN models stand out the most for having relatively high WEC values indicating comparatively worse performance.

A number of measures (the same as used in Section 5.3.2.b) were calculated to compare overall performance, and are presented in Table 5-21 below.

- The number of values of $PaRC_{IF}$ for which the WEC is less than one, measuring how many times the model is superior to a naïve model classifying all statements the same way;
- The number of values of $PaRC_{IF}$ for which the WEC is the lowest model, measuring how many times the model is the best model; and,
- The average of the ranking (where lower WEC is preferred) for each value of $PaRC_{IF}$, measuring performance across all values of $PaRC_{IF}$.

In a Majority Vote ensemble, increasing the number of underlying models from three in the other models to five in the Vote5 model offers little advantage. The four models have similar accuracy, but Table 5-21 illustrates that the Vote3_RF_TN_DT model has the best overall performance. The Vote3_RF_TN_DT model was the only Majority Vote model with underlying models that were exclusively decision-tree based (TN, RF_8 and DT_One).

Figure 5-15. A comparison of the holdout performance of the ensemble models using Majority Vote.

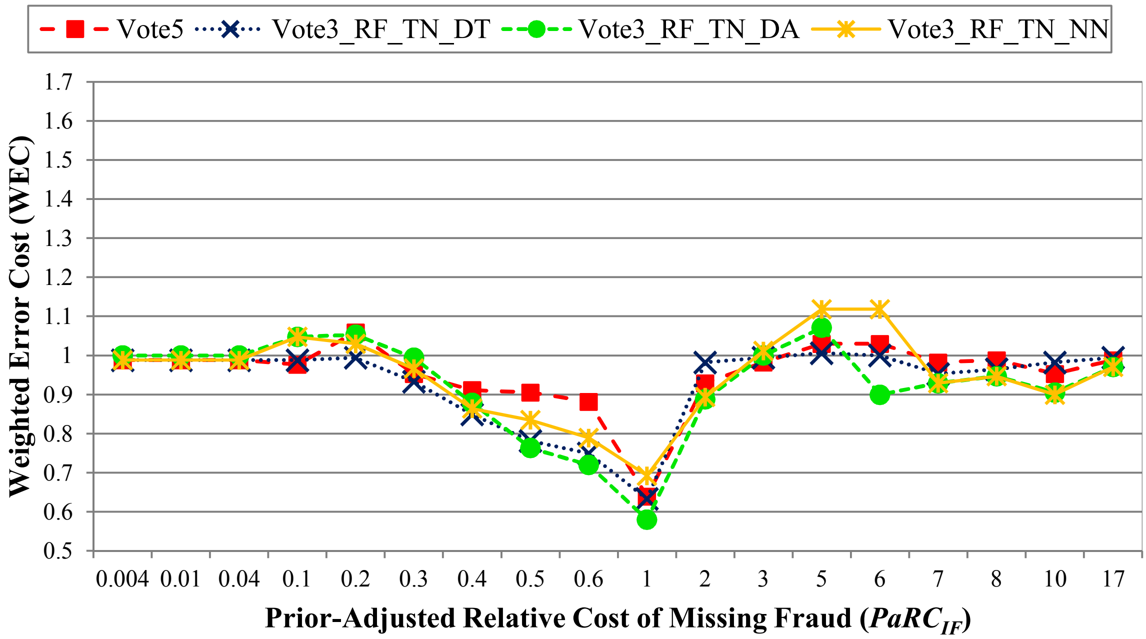


Table 5-21. A comparison of the holdout performance of the ensemble models using Majority Vote.

Model	WEC Measure across all values of $PaRC_{IF}$ (Ranking: 1 is best – 4 is worst)			Final Average Rank
	WEC < 1	Lowest WEC	Average Rank	
Vote3_RF_TN_DT	16 (rank 1)	7 (rank 1)	1.9 (rank 1)	1
Vote3_RF_TN_DA	11 (rank4)	7 (rank 1)	2.4 (rank 2)	2.33
Vote5	15 (rank 2)	5 (rank 3)	2.6 (rank 3)	2.67
Vote3_RF_TN_NN	12 (rank 3)	5 (rank 3)	2.6 (rank 4)	3.33

Ensembles using Averages (AV)

The holdout performance of the four ensembles that use averages are also very similar, as shown in Figure 5-16. The AV5_NoNN model is clearly the best for the lowest levels of $PaRC_{IF}$. Visually the AV5_NoNN model is also probably the best model overall, which is a finding supported by all the measures in Table 5-22. Unlike the Majority Vote ensembles, the ensembles averaging over multiple underlying models have improved as the number of underlying models has increased. The best model (AV5_NoNN) averages over the largest number (5) of underlying models: DA_All, LR_All, DT_One, RF_8 and TN models.

Figure 5-16. The holdout performance of the ensemble models using averages.

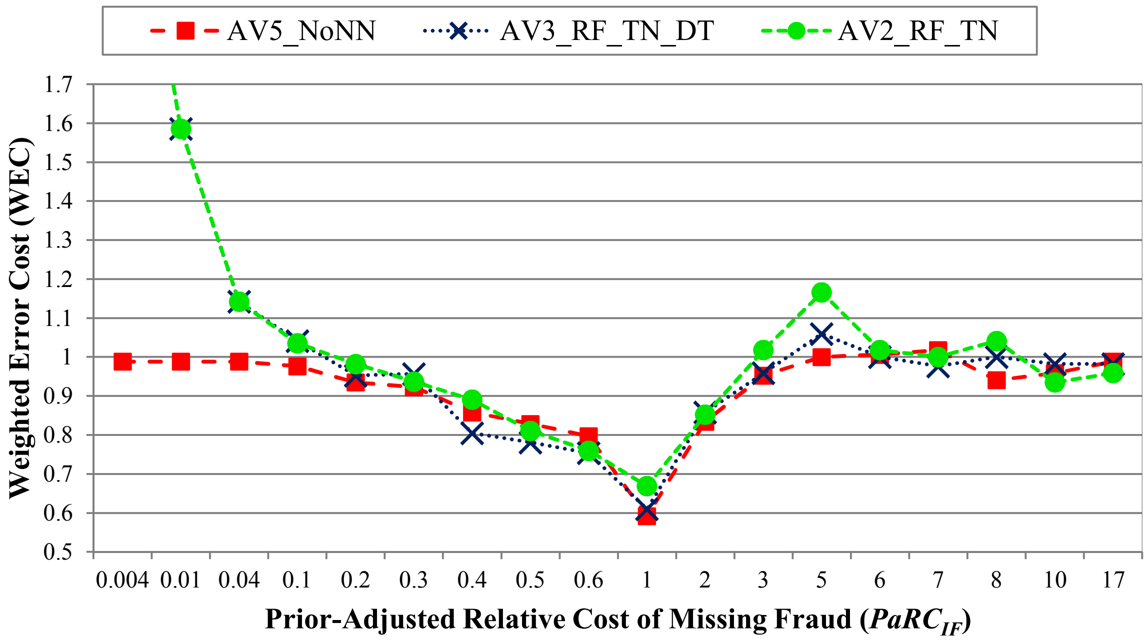


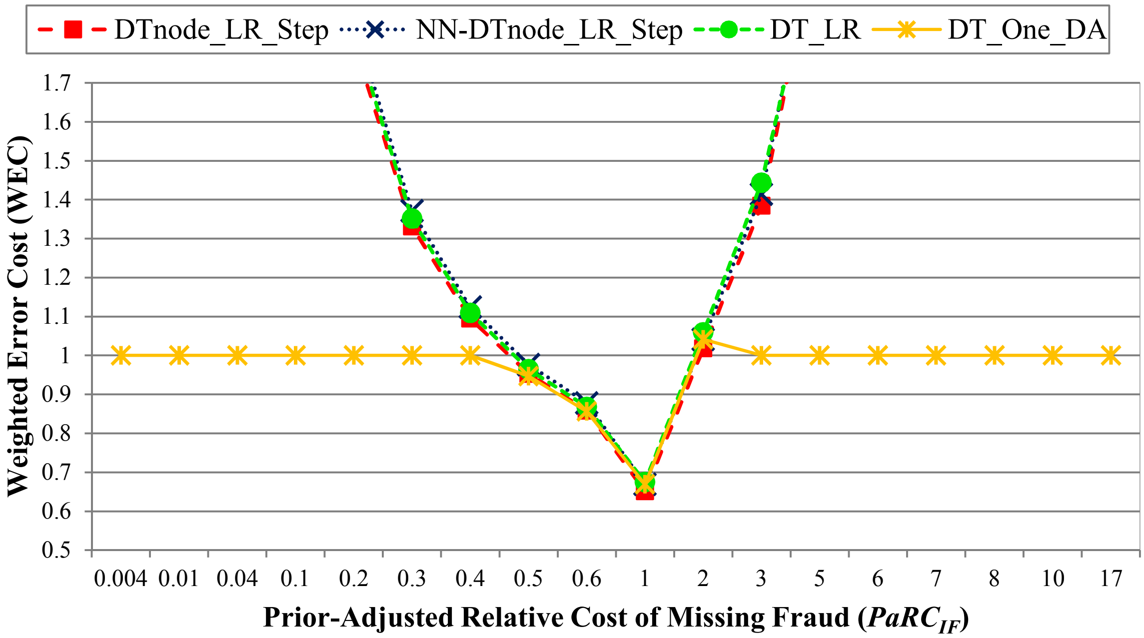
Table 5-22. A comparison of the holdout performance of the ensemble models using averages.

Model	WEC Measure across all values of $PaRC_{IF}$ (Ranking: 1 is best – 3 is worst)			Final Average Rank
	WEC < 1	Lowest WEC	Average Rank	
Av5_NoNN	15 (rank 1)	7 (rank 1)	1.8 (rank 1)	1
AV3_RF_TN_DT	11 (rank 2)	5 (rank 2)	1.9 (rank 2)	2
AV2_RF_TN	9 (rank 3)	2 (rank 3)	2.4 (rank 3)	3

Other Ensembles Models

The performance on the holdout data of the models proposed by McKee (2009) and new models developed from analysing his model are presented in Figure 5-17 below. The DTnode_LR_Step, NN-DTnode_LR_Step and DT_LR models have almost identical performance in Figure 5-17, although they do differ in the cases that have a WEC too high to be shown in the figure. It isn't noticeable visually, but the DTnode_LR_Step model has the lowest WEC of the three models in all cases shown on the graph and in 14 out of the 18 values of $PaRC_{IF}$. Although the difference is very minor, this empirically supports the theoretical analysis (from Section 5.2.3.f) to include the more rich information provided by terminal node classifications from a decision tree model rather than simply the overall classification used by McKee (2009). However, all three McKee-related models are outperformed by the DT_One_DA model, which uses discriminant analysis in a second stage after the DT_One model. The DT_One_DA model has extremely similar performance when $\mathrm{PaRC_{IF}}$ is between 0.5 and 2, but then has a stable WEC for more extreme values of $\mathrm{PaRC_{IF}}$ when the other models become ineffective with very high WECs.

Figure 5-17. A comparison of the holdout performance of other ensemble models. The majority of the first three models are not shown with high WECs that range up to 66.96 when $PaRC_{IF}$ is 0.004.



5.3.3 Overall Analysis

5.3.3.a Best Model(s) for Each Value of $\mathrm{PaRC_{IF}}$

The model with the lowest WEC depends on the value of $PaRC_{IF}$. This demonstrates that considering different values of $PaRC_{IF}$ is important, because no model has the lowest in every value of $PaRC_{IF}$.

Table 5-23 below lists the best model(s) (out of all 34) for each value of $PaRC_{IF}$, as well as any models that are within 2% of the lowest WEC. This table summarises the answer to research question RQ1 (defined on page 132) and can be used to select the most appropriate model with the lowest WEC when information about the value of $PaRC_{IF}$ is known.

Table 5-23. The best model(s) according to the lowest holdout Weighted Error Cost (WEC) for each Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$).

$PaRC_{IF}$	**Best Model(s) with the Lowest Holdout WEC** **F/L**: Accuracy at Detecting Fraud/Legitimate				**Model(s) within 2% of the Lowest WEC**
	WEC	**F**	**L**	**Model Code**	
0.004	0.964	4%	100%	DA_U15%	LR_U15% (within 1%) LR_Step Vote3_RF_TN_DA
0.01	0.964	4%	100%	DA_U15%	LR_U15% (within 1%) LR_Step Vote3_RF_TN_DA
0.04	0.964	4%	100%	DA_U15%	LR_U15% (within 1%) LR_Step Vote3_RF_TN_DA
0.1	0.970	3%	100%	LR_U15%	AV5_NoNN (within 1%) Vote5
0.2	0.935	9%	99%	AV5_NoNN	AV3_RF_TN_DT
0.3	0.913	28%	94%	RF_66	AV5_NoNN
0.4	0.805	28%	96%	AV3_RF_TN_DT	
0.5	0.763	50%	87%	Vote3_RF_TN_DT	
0.6	0.720	50%	87%	Vote3_RF_TN_DT	
1	0.580	66%	76%	Vote3_RF_TN_DT	
2	0.834	85%	46%	AV5_NoNN	
3	0.905	91%	36%	RF_66	
5	1.000	100% 100% 100% 98%	0% 0% 0% 12%	M-score / F-score NN_GA_1 DT_One / DT_One_DA AV5_NoNN	Vote5
6	0.899	98%	21%	TN Vote3_RF_TN_DT	
7	0.917	98%	21%	TN	Vote3_RF_TN_NN Vote3_RF_TN_DT
8	0.935	98%	21%	TN	AV5_NoNN (within 1%) Vote3_RF_TN_DT Vote3_RF_TN_NN NN_LR Vote5
10	0.876	100%	12%	RF_66	Vote3_RF_TN_DA
17	0.876	100%	12%	RF_66	

As stated earlier in Section 5.2.1.d, the value of $PaRC_{IF}$ can be estimated from the proportion of fraudulent statements in the real-world population [also known as the prior probability of fraud, $p(F)$] and the cost of missing fraud relative to the cost of falsely alleging fraud (C_{IF}:C_{IL}). The calculation is $PaRC_{IF} = \frac{C_{IF}}{C_{IL}} \times \frac{p(F)}{p(L)} = \frac{C_{IF}}{C_{IL}} \times \frac{p(F)}{1-p(F)}$. Alternatively, Table 5-24 be used to obtain the corresponding value of $PaRC_{IF}$ for given values of $p(F)$ and C_{IF}:C_{IL}. The estimates of $p(F)$ from prior research is discussed above (in Section 5.2.1.c), but the ratio C_{IF}:C_{IL} varies depending on the stakeholder. As an example, Beneish (1999a) estimates that C_{IF}:C_{IL} is likely to be in the range between 20:1 and 30:1 for investors.

Table 5-24. (Table 5-8 reproduced.) The corresponding prior-adjusted relative cost of missing fraud for each combination of ratio of error cost and prior probability of fraud. Values appearing more than once have been shaded in the same colour.

Ratio of Error Costs (C_{IF}:C_{IL}) (missing fraud: falsely alleging fraud)	**Prior probability of fraud [$p(F)$]**			
	0.40%	**1.00%**	**5.50%**	**14.50%**
1:1	0.004	0.01	0.1	0.2
10:1	0.04	0.1	0.6	2
20:1	0.1	0.2	1	3
30:1	0.1	0.3	2	5
40:1	0.2	0.4	2	7
50:1	0.2	0.5	3	8
60:1	0.2	0.6	3	10
100:1	0.4	1	6	17

Table 5-23 on the previous page reveals that the lowest WECs are less than one, except when $PaRC_{IF}$ is equal to five the lowest WEC is equal to one. That means that in every other case, a model developed in this research is superior to a naïve model (defined above in Section 5.2.1.d) that classifies all financial statements the same way. Consistent with all previous results, it is also evident in Table 5-23 that the WEC when $PaRC_{IF}$ is equal to one is noticeably lower than any other WEC. When $PaRC_{IF}$ increases above one, models compensate by correctly detecting more fraud at the expense of correctly detecting legitimate financial statements. The opposite occurs when values of $PaRC_{IF}$ are less than one. It is also noticeable that for the mid-range values of $PaRC_{IF}$, no models are within 2% of the best, as indicated by the last column being empty. However, for the more extreme values of $PaRC_{IF}$ (high and low) there are usually models that have a WEC within 2% of the best model.

The discriminant analysis and logistic regression models DA_U15% and LR_U15% have the lowest WEC for the lowest values of $PaRC_{IF}$. These two models only used variables that were statistically significant at the 15% level in univariate analysis. Ensemble models dominate the remainder of the values of $PaRC_{IF}$ and decision tree techniques, and ensembles of them, often have the lowest WEC. Ensembles that use Majority Vote (Vote) and averages (AV) also have the lowest WEC (or within 2% of the lowest) in many cases, particularly for values of $PaRC_{IF}$ between 0.4 and 1 when ensembles of TreeNet, Random Forests and a CART decision tree (RF_TN_DT) have the lowest WEC. In addition to Random Forests (RF_66) being the best model when $PaRC_{IF}$ is equal to 0.3 and 3, it along with TreeNet are the best for the highest values of $PaRC_{IF}$.

5.3.3.b Selection of the Best Model Overall

The best models from each modelling technique (from Section 5.3.2) are compared in the following subsections across all values of $PaRC_{IF}$ to determine the best model overall, which addresses research question RQ2 (defined on page 132).

The analysis presented in the following subsections shows that the AV5_NoNN is the best model when considering the whole range of $PaRC_{IF}$ values analysed in this study. Consequently, it would be an excellent model to choose when the value of $PaRC_{IF}$ is unknown. Table 5-25 below presents the percentage accuracy and Weighted Error Cost (WEC) of the AV5_NoNN model for each value of $PaRC_{IF}$. As the relative cost of missing fraud increases, indicated by larger values of $PaRC_{IF}$, the percentage of fraud detected increases at the expense of a decline in the percentage of legitimate detected. At the extreme values of $PaRC_{IF}$, the AV5_NoNN is able to detect either fraud or legitimacy with perfect accuracy while still maintaining a small (non-zero) percentage accuracy of detecting the other. The overall percentage accuracy declines as the value of $PaRC_{IF}$ moves further away from one, either above or below. This further illustrates that fraud is easiest to detect when $PaRC_{IF}$ is equal to one.

Although AV5_NoNN is the best model overall, it is not the best model for every specific value of $PaRC_{IF}$. Table 5-25 also reveals in the last column how the AV5_NoNN model compares to lowest WEC (of any model) for each value of $PaRC_{IF}$. The AV5_NoNN model is comparable in most situations and has the lowest WEC in three, when $PaRC_{IF}$ is equal to 0.2, 2 and 5. However, its WEC is more than 10% higher than the lowest for four values of $PaRC_{IF}$: 0.6, 6, 7 and 17. Consequently, if the value of $PaRC_{IF}$ is known then it is better to use the best model for that particular value of $PaRC_{IF}$, as presented in the previous section in Table 5-23.

Table 5-25. The holdout performance of the best model overall, AV5_NoNN, and a comparison to the lowest WEC (of any model) for each value of $PaRC_{IF}$.

$PaRC_{IF}$	WEC	Percentage Accuracy			Comparison with WEC of the Best Model for each $PaRC_{IF}$
		Fraud	Legitimate	Overall	
0.004	0.988	1%	100%	51%	2% higher WEC
0.01	0.988	1%	100%	51%	2% higher WEC
0.04	0.988	1%	100%	51%	2% higher WEC
0.1	0.976	2%	100%	51%	1% higher WEC
0.2	0.935	9%	99%	54%	Equal WEC
0.3	0.923	14%	98%	56%	1% higher WEC
0.4	0.858	20%	98%	59%	7% higher WEC
0.5	0.828	30%	93%	62%	9% higher WEC
0.6	0.797	33%	92%	63%	11% higher WEC
1	0.592	75%	66%	70%	2% higher WEC
2	0.834	85%	46%	66%	Equal WEC
3	0.953	89%	37%	63%	5% higher WEC
5	1.000	98%	12%	55%	Equal WEC
6	1.006	99%	7%	53%	12% higher WEC
7	1.018	99%	7%	53%	11% higher WEC
8	0.941	100%	6%	53%	1% higher WEC
10	0.959	100%	4%	52%	9% higher WEC
17	0.988	100%	1%	51%	13% higher WEC

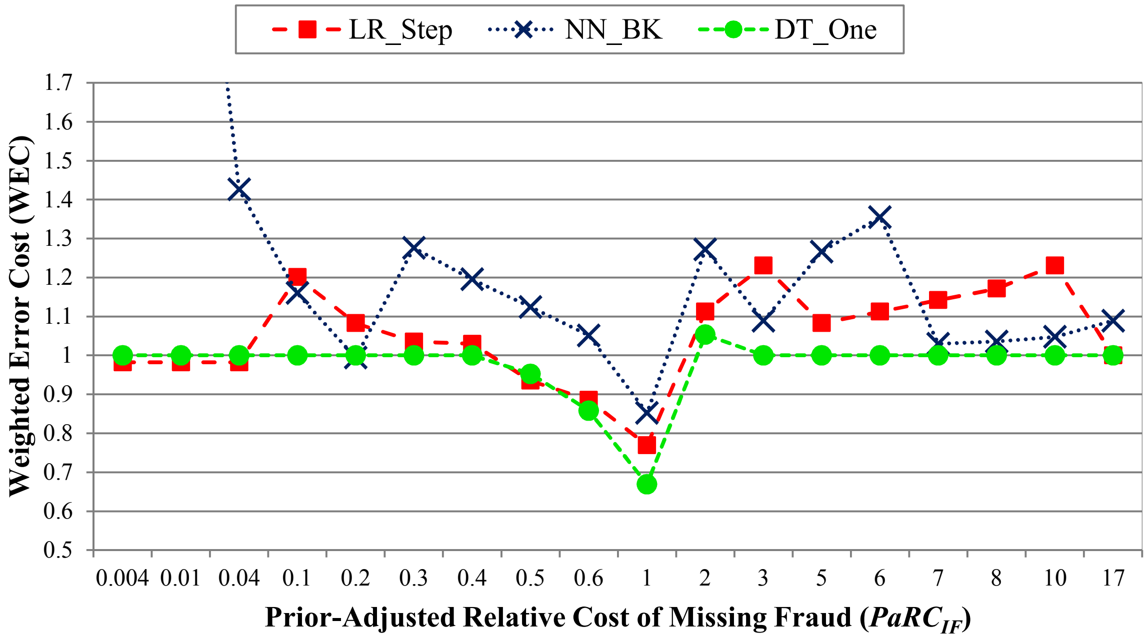
The analysis to determine that the AV5_NoNN is the best model overall now follows.

A Comparison of Individual (non-ensemble) Models

The three individual (non-ensemble) models that performed the best in the previous section were a logistic regression model (LR_Step), a neural network (NN_BK) and a decision tree (DT_One). Figure 5-18 reveals that the decision tree DT_One clearly has the lowest holdout WEC overall, although LR_Step has a slightly lower WEC for four values of $PaRC_{IF}$ (0.004, 0.01, 0.04 and 0.5). A neural network had previously outperformed ID3 decision trees on Greek data (Kirkos et al. 2007), but this result was clearly reversed using CART decision trees in this more comprehensive comparison.

This research also provides more clarity on the relative accuracy of logistic regression and neural networks for financial statement fraud detection. Although Feroz et al. (2000) found a logistic regression and neural network model comparable, Perols (2011) found that logistic regression was superior using cross-validated results. Liou (2008) also found a similar result on Taiwanese data without considering differing error costs. The superior results of logistic regression compared to a neural network have been substantiated in this study with increased reliability because a holdout data set from the US has been used. As illustrated in Figure 5-18, the LR_Step model is superior to NN_BK for 12 out of the 18 values of $PaRC_{IF}$, including when $PaRC_{IF}$ is one (which is equivalent to ignoring the effects of error costs and priors).

Figure 5-18. A comparison of the holdout performance of different individual (non-ensemble) models.



The Best Individual Model Is Outperformed by the Ensemble Model DT_One_DA

The WEC of the best individual model, DT_One, can be very slightly reduced by using discriminant analysis as a second stage classifier, as shown in Table 5-26. The DT_One_DA model uses discriminant analysis as a second stage classification motivated by the work by Nagadevara (2010). However, unlike the substantial improvements found by Nagadevara on a different classification problem, the improvements over a standard decision tree (DT_One) are extremely small, as shown in Table 5-26. DT_One_DA's improvements in WECs are so small that they could be considered to be outweighed by the negative of the model's additional complexity.

Table 5-26. A comparison of the holdout WEC for the DT_One and DT_One_DA models.

Model	$\boldsymbol{Prior - adjusted\ Relative\ Cost\ of\ Missing\ Fraud}\ (\boldsymbol{PaRC_{IF}})$					
	≤ 0.4	0.5	0.6	1	2	≥ 3
DT_One	1	0.953	0.858	0.669	1.053	1
DT_One_DA	1	0.947	0.856	0.669	1.041	1

A Comparison of Ensemble Models: DT_One_DA

Consistent with previous research (presented in Chapter 3), the best models in this study are ensembles of multiple models. The best model so far, DT_One_DA, is compared in Figure 5-19 with the remaining ensemble models that performed best in the previous section. It can be seen that these models have WECs less than one far more often than in previous graphs. However, there are too many models in Figure 5-19 to be able to select the best model, so the comparison measures in Table 5-27 are used.

Figure 5-19. A comparison of the holdout performance of different ensemble models.

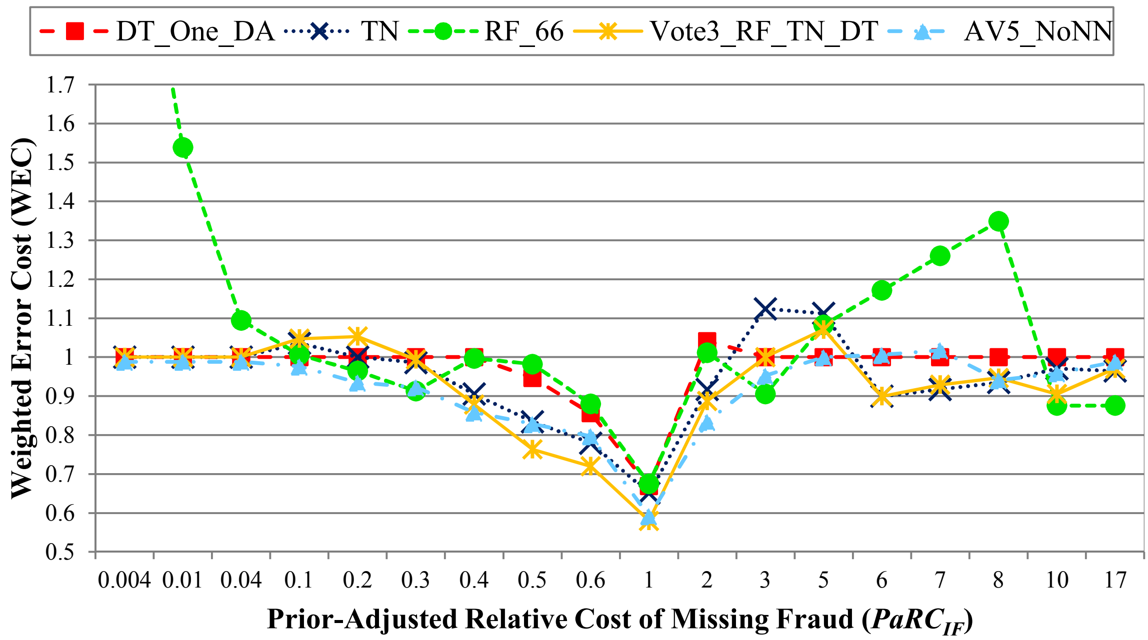
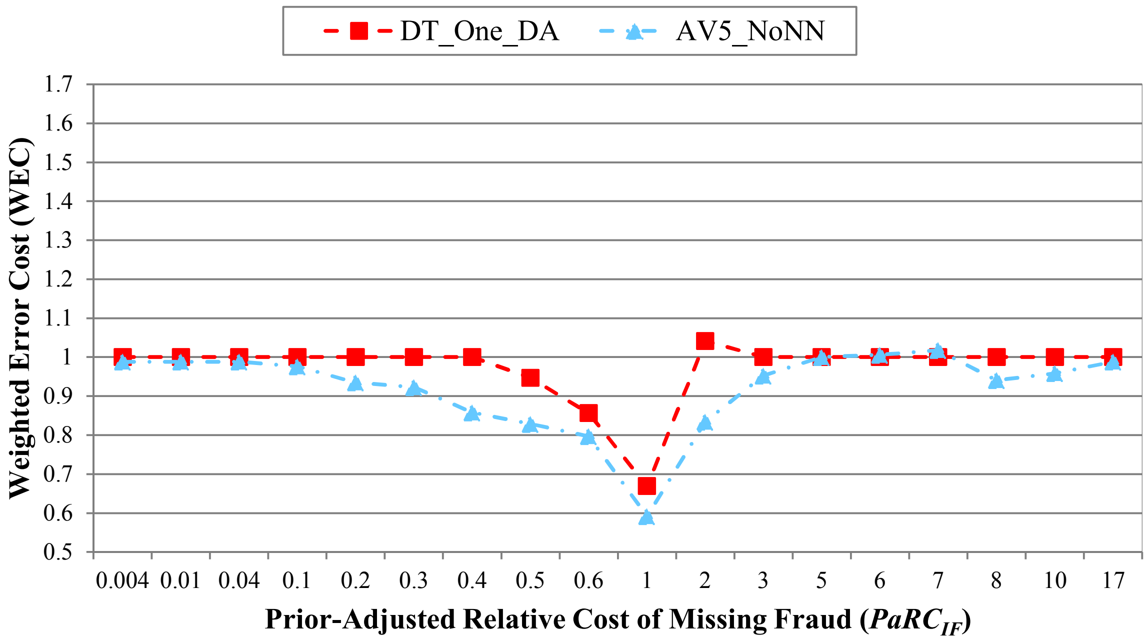


Table 5-27. A comparison of the holdout performance of different ensemble models.

Model	WEC Measure across all values of $PaRC_{IF}$ (Ranking: 1 is best – 5 is worst)			Final Average Rank
	WEC < 1	Lowest WEC	Average Rank	
AV5_NoNN	15 (rank 1)	8 (rank 1)	2.0 (rank 1)	1
Vote3_RF_TN_DT	11 (rank 2)	4 (rank 2)	2.5 (rank 2)	2
TN	11 (rank 2)	3 (rank 4)	2.7 (rank 3)	3
RF_66	9 (rank 4)	4 (rank 2)	3.7 (rank 5)	3.67
DT_One_DA	3 (rank 5)	1 (rank 5)	3.4 (rank 4)	4.67

Although DT_One_DA rarely has the highest WEC, there are models with a lower WEC than DT_One_DA in all situations in Figure 5-19 except for when $PaRC_{IF} = 5$. Furthermore, the comparison with AV5_NoNN in Figure 5-20 clearly shows that DT_One_DA has a higher WEC overall. The DT_One_DA model is also the worst model (lowest rank) according to all the measures in Table 5-27. As is the case with the standard decision tree DT_One, DT_One_DA it is only better than a naïve model (WEC < 1) three times and so it only has practical use in these situations.

Figure 5-20. The holdout performance of AV5_NoNN compared with DT_One_DA.

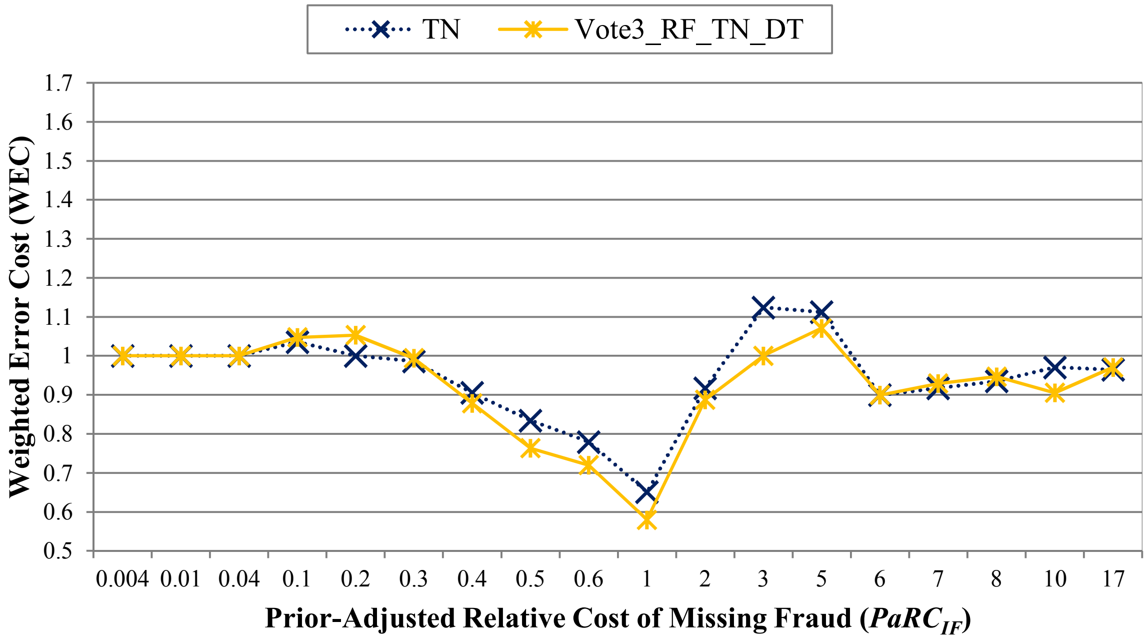


A Comparison of Ensemble Models: Decision Tree Ensembles

There have been mixed results when comparing Random Forests (or bagging using decision trees) and logistic regression using cross-validation. Perols (2011) found logistic regression to be superior, while Whiting et al. (2012) found Random Forests to be the better technique. Additional clarity on this comparison is provided in this study by using a holdout sample, instead of cross-validation. The results in Table 5-27 above show that Random Forests (RF_66) has outperformed DT_One_DA in terms of WEC, which in turn outperformed the non-ensemble models including logistic regression. Thus, the results support the finding of Whiting that overall Random Forests are superior to logistic regression at detecting financial statement fraud.

It is clear from Figure 5-19 above that while Random Forests (RF_66) performs really well with low WEC in some cases, it is the more volatile in terms of the WEC varying greatly across different values of $PaRC_{IF}$. TreeNet, the other decision tree ensemble, is, however, more consistent with its relatively good performance and WECs less than one. Thus, as mentioned above in Section 5.3.2.e, TreeNet is likely to be the better choice over the whole range of values of $PaRC_{IF}$. However, the comparison between TN and Vote3_RF_TN_DT in Figure 5-21 below reveals that the TreeNet model is not the best as it has equal or higher WEC for most values of $PaRC_{IF}$. Nevertheless, when excluding the ensemble models that use averages and Majority Vote (which incorporate both TreeNet and Random Forests), ensembles of decision trees are the best techniques.

Figure 5-21. The holdout performance of Vote3_RF_TN_DT compared with TN.



A Comparison of Ensemble Models: Ensembles that use Majority Vote or Averages

Table 5-27 (on page 213 above) clearly shows that the best model overall is the ensemble AV5_NoNN. This ensemble model averages the outputs of five other models: Random Forests, TreeNet, a CART decision tree, discriminant analysis and logistic regression. The second best model is the ensemble Vote3_RF_TN_DT that takes the Majority Vote between Random Forests, TreeNet and a CART decision tree. It is interesting to note that the ensemble formed using averages (AV5_NoNN) uses a larger, broader selection of models, while the Majority Vote ensemble (Vote3_RF_TN_DT) utilises only decision trees and ensembles of them. Both the AV5_NoNN and Vote3_RF_TN_DT models, particularly the former, predominantly outperform the benchmark F-score model, as shown in Figure 5-22 below. This figure also shows graphically that the AV5_NoNN model is the best overall with a more consistently lower WEC than the Vote3_RF_TN_DT model. As shown in Table 5-27 above, the AV5_NoNN model's WEC is below one for 15 out of the 18 values of $PaRC_{IF}$. In the remaining three situations, it has a WEC equal to one once and slightly greater than one twice, as shown above in Table 5-25.

Figure 5-23 below illustrates that the best model overall, AV5_NoNN, has a WEC lower than both benchmark models in every situation except when $PaRC_{IF}$ is equal to 6 or 7, when it is slightly higher. This figure also illustrates that the AV5_NoNN model has the lowest WEC when $PaRC_{IF}$ approaches the value of one, which is the same for every model.

Thus, models find it easier to detect fraud when the prior-adjusted relative cost of missing fraud is roughly equal to the prior-adjusted relative cost of falsely alleging fraud, as indicated by a $PaRC_{IF}$ value of one.

Figure 5-22. The holdout performance of the best two models compared to the best benchmark model (F-Score).

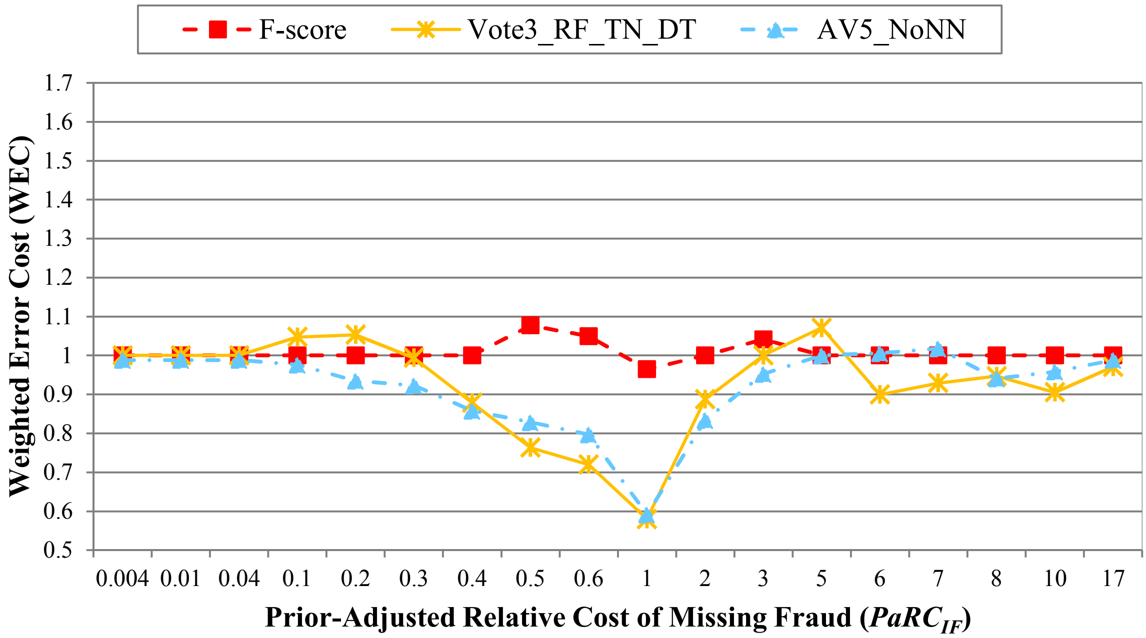
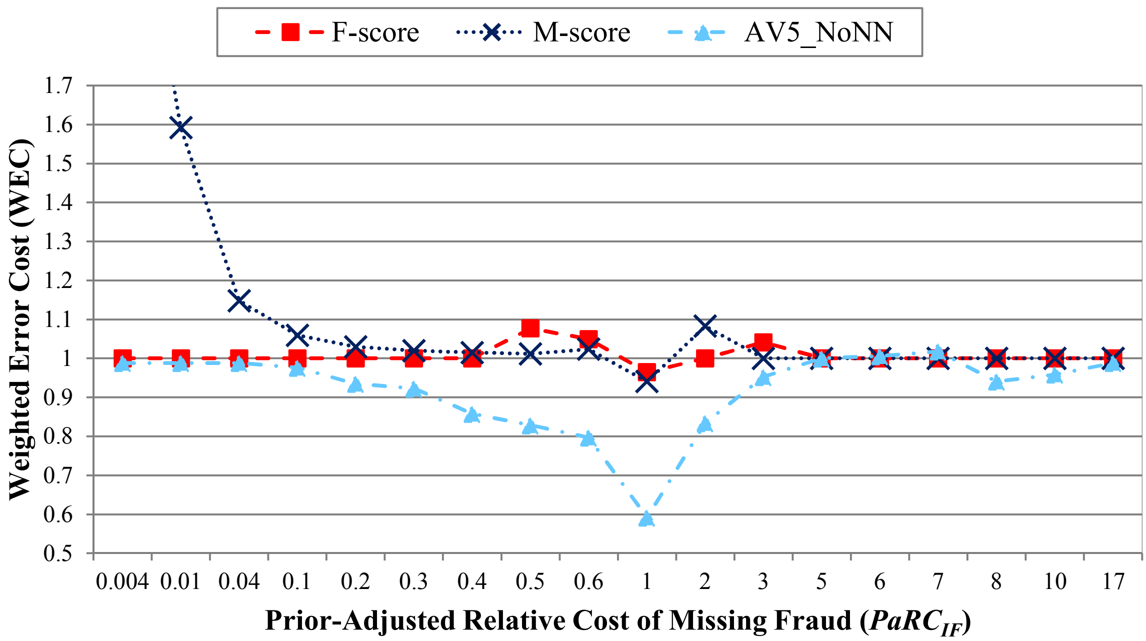


Figure 5-23. The holdout performance of the best model compared to the benchmark models.



5.3.3.c A Comparison of Models When $PaRC_{IF}$ is Equal to One

The analysis so far has considered different values of $PaRC_{IF}$ that represent varying error costs and prior probabilities. This is more applicable to the real world than many prior studies that present results ignoring these considerations. To assist a complete interpretation of results, an analysis of results when $PaRC_{IF}$ is equal to one follows, because it is equivalent to ignoring these considerations. An advantage of this analysis is that results can be presented in terms of percentage accuracy, which is more intuitive than the WEC measure. However, the following analysis is limited to those occasions when $PaRC_{IF}$ is estimated to be close to or equal to one.

Table 5-28 below lists the percentage of both fraudulent statements and legitimate statements that each model classifies correctly from the holdout data set, as well as the average overall accuracy. The results in the table are ordered from highest overall accuracy to the lowest and reveal that:

- Ensemble models, particularly those involving multiple modelling techniques, are the most accurate. In addition, CART decision trees are the most accurate non-ensemble technique. This is the same result as found in the overall analysis of different $PaRC_{IF}$ values presented in Section 5.3.3.b;
- The model with the highest overall accuracy is the ensemble model Vote3_RF_TN_DT that averages the probability outputs of a Random Forest model, a TreeNet model and a CART decision tree;
- There are two models with equal second highest overall accuracy, which are AV5_NoNN and AV3_RF_TN_DT. The latter model is also the most accurate at detecting fraud with an accuracy of 78%. These two models were also the best models overall as presented in Section 5.3.3.b above;
- AV5_NoNN is one of only five (out of 34) models that correctly detected more fraudulent than legitimate statements. This result is consistent with many prior studies (discussed in the review in Chapter 3);
- The genetically optimised neural network NN_GA_1 classifies the most legitimate statements correctly, but at the expense of having the lowest rate of correctly detecting fraud. It is the worst model with the lowest overall classification accuracy of 50%; and,
- The benchmark models, F-score and M-score, had the next lowest overall accuracy, which reveals that all models developed in this research except NN_GA_1 are more accurate than the benchmark models when $PaRC_{IF}$ is equal to one.

Table 5-28. The accuracy of models on the holdout data set when $PaRC_{IF} = 1$. The order is from highest to lowest overall accuracy, and shading indicates the maximum in that column.

Model	**Accuracy: Fraud, Legitimate, Overall**		
	F	**L**	**O**
Vote3_RF_TN_DT	66%	76%	71%
AV5_NoNN	75%	66%	70%
AV3_RF_TN_DT	78%	62%	70%
Vote5	60%	76%	68%
TN	55%	80%	67%
AV2_RF_TN	57%	76%	67%
Vote3_RF_TN_DA	60%	75%	67%
DTnode_LR_Step	63%	72%	67%
DT_Suite	62%	72%	67%
DT_One	62%	72%	67%
DT_One_DA	62%	72%	67%
NN-DTnode_LR_Step	63%	70%	67%
DT_LR	62%	71%	66%
RF_66	63%	69%	66%
RF_8	67%	64%	66%
Vote3_RF_TN_NN	60%	70%	65%
LR_Step	47%	76%	62%
LR_U1	56%	69%	62%
LR_All	55%	69%	62%
DA_U1	59%	65%	62%
DA_All	50%	72%	61%
DA_Step	50%	71%	61%
NN_LR	51%	70%	61%
LR_MS_S	49%	71%	60%
DA_U15%LR	49%	69%	59%
LR_MS_F	51%	66%	59%
LR_U15%LR	54%	64%	59%
NN_GA_5	65%	50%	58%
DA_U15%	49%	66%	57%
LR_U15%	49%	66%	57%
NN_BK	56%	59%	57%
M-score	56%	50%	53%
F-score	31%	72%	52%
NN_GA_1	15%	85%	50%

5.3.4 Summary

This chapter has presented analyses of models to detect fraud in the financial statements of publicly listed companies in the US. The analyses have used a data set with substantially more fraud cases than most previous studies and an extensive set of explanatory variables. Furthermore, the data used can act as a benchmark data set for future research. Unlike many prior studies, a holdout data set that occurred chronologically after the training data has also been used to provide results that are more applicable to the real world. Compared to such a holdout data set, analyses in this chapter also revealed that using cross-validation did not provide similar results, which is an important finding for future research.

The analyses in this chapter has been conducted over the widest range of estimates for the prior probability of fraud, compared with earlier studies. This range includes a substantially higher estimate of 14.5% from recent research (Dyck et al. 2013) that has not been incorporated into previous financial statement fraud detection modelling studies. As the ratio of error costs depends on the stakeholder, ratios have been considered ranging from 1:1 to 1:100 for the cost of missing fraud relative to the cost of falsely alleging fraud. Consistent with best practice in previous cost-sensitive research, models have been compared using a weighted error cost measure after their cut-off values have been empirically optimised. Overall, models were most accurate when the cost of missing fraud was equal to the cost of falsely alleging fraud after adjusting for the prior probability of fraud.

Finally, a total of 32 models have been developed after analysing many more in order to choose the parameters for each model. Some of the models were based on those in prior studies such as discriminant analysis, logistic regression, artificial neural network and ensemble models. New models have also been developed. For example, variable selection techniques, previously trialled exclusively for either discriminant analysis or logistic regression models, were used for both modelling techniques in this research. Two new multi-stage logistic regression models were also developed in addition to new models based on CART decision trees, Random Forests and TreeNet with more refined parameters. New ensembles were also developed including two-stage models incorporating CART decision trees and logistic regression, as well as ensembles that used Majority Vote and averaging individual model outputs. All 32 models have been developed and tested on the same data so that valid comparisons could be made. An additional two benchmark models (F-score and M-score), exactly as developed in their original studies, were also included for comparison.

Discriminant analysis and logistic regression models had very similar accuracy, and both were superior to artificial neural network models. Despite the fact that decision trees were not evaluated in prior research using US data, a CART decision tree was the most accurate non-ensemble modelling technique overall. However, none of the non-ensemble models adjusted well to differing values of prior probabilities and error costs.

Ensemble models predominantly outperformed non-ensemble models. The ensemble models Random Forests and TreeNet outperformed a single CART decision tree. All three of these models used cross-validation on the training data, which is designed to prevent models from learning patterns specific to the training data that do not persist in holdout data. The comparatively good performance of these models on the holdout data suggests the use of cross-validation in model development has worked well. Ensembles that combined multiple modelling techniques with Majority Vote or averaging of probability outputs were the most accurate at detecting financial statement fraud on this data set. However, the best ensembles all included both a TreeNet and Random Forests model, and many also included a CART decision tree as underlying models.

Many models had a lower weighted error cost than the benchmark F-score and M-score models, which indicates that improved performance over existing models is possible. Overall, the best model is the ensemble AV5_NoNN, which outperforms both benchmark models. This ensemble model averages the probability outputs of five other models: Random Forests, TreeNet, a CART decision tree, discriminant analysis and logistic regression. It is a good choice of model when the prior probability and ratio of error costs is unknown. However, as the best model varies with these values, if they are known, Table 5-23 can be used to select the model best suited to the circumstances, which might not be the AV5_NoNN model. The fact that no model is superior in all contexts illustrates the importance for research to consider a variety of prior probabilities and ratio of error costs.

Although ensemble models were the most accurate, they are more complex and consequently more difficult to interpret and analyse than non-ensembles such as decision trees, logistic regression and discriminant analysis. Furthermore, variable reduction techniques, such as stepwise selection or using univariate analysis, benefited discriminant analysis and logistic regression models in their ability to detect financial statement fraud. Whether a relatively less complex model using a reduced set of variables can be comparable to the weighted error cost of the best ensemble models presented above is investigated in the next chapter.

Chapter 6 Analysis of Variables and Development of Simpler Models

The previous chapter addressed the first two research questions (RQ1 and RQ2) by investigating which modelling techniques are the most accurate at detecting financial statement fraud. It is also of interest to investigate the explanatory variables that were used in these models. This chapter addresses research question RQ3, which is "Which explanatory variables are the most useful in models that detect financial statement fraud?" The results from this analysis form the basis for developing less complex models, and subsequently for investigating research question RQ4, which is "How do simpler models compare with the those developed for the first two research questions (as presented in Chapter 5) in their ability to detect financial statement fraud?" This is an important question because simpler models are sometimes preferred to more complex ones, in accordance with the principle of parsimony that states that the simpler of two competing theories is preferred.

6.1 Analysis of Explanatory Variables

This section addresses research question RQ3. It analyses explanatory variables in terms of their number in each model, followed by their statistical significance and direction of association with financial statement fraud. The statistical significance of variables is then evaluated relative to the overall schema of explanatory variables and then the new Fraud

Detection Triangle framework[130]. The relative importance of variables in models that allow interactions is then analysed.

This research has used a more comprehensive set of explanatory variables than prior studies. Further to this, the number of fraud cases in the training sample in this research is larger, often more than double that used in prior studies. Additionally, both standard regression-based models and models that allow complex interactions between explanatory variables are analysed. Consequently, the results presented in this section contribute greatly to better understanding the variables that are important in models that detect financial statement fraud and how they associate with fraud.

6.1.1 Number of Variables in Each Model

Decision trees, ensembles of decision trees and genetically optimised neural networks all have in-built processes to select the most appropriate variables to use. As these processes differ, the number of variables used in these models varies. More details on the way these models rank the importance of variables are provided later in Section 6.1.5.a. Standard regression-based models that use differing numbers of explanatory variables were also developed in Chapter 5, which enables further investigation into whether that number influences model performance.

Table 6-1 shows the number of variables used by each model from the analysis in Chapter 5. The following models were excluded from the table:

- Benchmark models because they were not developed using the same data and have already been discussed in Section 5.2.2;
- Ensembles of multiple modelling techniques as they combine two or more models from this table; and,
- DT_Suite because the model was never superior to the other decision tree model (DT_One). Furthermore, it should be noted that the number of variables for DT_Suite varies with the prior-adjusted relative cost of missing fraud ($PaRC_{IF}$).

The total number of explanatory variables available, including variable sub-types[131], is 71, except for decision trees and ensembles of them that had 66 variables available. Some

[130] The new Fraud Detection Triangle framework is based on Cressey's (1953) Fraud Triangle, as detailed in Section 4.2, starting on page 77.

variables were omitted for decision-tree based models as they are equivalent to each other from a decision tree's perspective (refer to Section 5.2.3.c for more details).

Table 6-1. The number of explanatory variables used by each model.

Model	**Number of Variables**[131]
Standard Regression-based Models using variables selected by	
All variables (DA_All, LR_All and multi-stage models LR_MS_S and LR_MS_F)	71
One variable from each main type (DA_U1 and LR_U1)	50
Univariate Analysis (DA_U15% and LR_U15%)	25
Univariate Analysis and Logistic Regression (DA_U15%LR and LR_U15%LR)	14
Stepwise Techniques (DA_Step and LR_Step)	13 and 14
Artificial Neural Networks	
Standard backpropagation network (NN_BK)	71
Genetically-optimised backpropagation network (NN_GA_1 and NN_GA_5)	35 and 45
One CART Decision Tree (DT_One)	56
Ensembles of Decision Trees (RF and TN models)	66

Some standard regression-based models used all explanatory variables that were available and others used the smallest number. The models identified as the better overall performers in Chapter 5, the Step and U15% models, used a reduced number of variables ranging between 13 and 25. Meanwhile, the CART decision tree that performed relatively well in Chapter 5 used a large number of variables (56), because its model building algorithm found them to be useful in discriminating between fraudulent and legitimate statements. Furthermore, the superior performance of the ensembles of decision trees in Chapter 5 is likely the result of better classifications close to region boundaries as a consequence of using more variables (all 66 available) and a large number of trees. This is consistent with the theoretical analysis presented above in Section 3.3.5.

Overall, there are inconsistent results in terms of the number of explanatory variables that were used. The standard regression-based models benefited from a reduced set of variables. In contrast, a decision tree and particularly ensembles of trees were able to gain benefit from using more explanatory variables. Meanwhile, all the neural network models

[131] These calculations include sub-types. For example, there are multiple ways of measuring variable V1, V1a, V1b and V1c, and each of these sub-types counts as a separate variable.

performed relatively poorly regardless of the number of variables used. This suggests that the in-built feature of the decision trees (and ensembles of them) to select the most appropriate variable for each stage of the tree performs relatively well on this financial statement fraud data set. However, this does not guarantee that the ensembles of decision trees would have drastically reduced performance if they used a reduced set of variables. This will be investigated later in this chapter.

As pointed out in the review in Section 3.3.1, multicollinearity is an issue for all the standard regression-based models. The use of stepwise variable selection substantially reduces the likelihood of multicollinearity occurring, which may explain the good results (in Chapter 5) of the Step models relative to the other standard regression-based models. In contrast, decision tree techniques do not suffer from problems of multicollinearity, which helps to explain why they are better able to handle a larger number of variables.

6.1.2 Analysis of Statistical Significance and Direction Using Standard Regression-based Models

This section extends the univariate analysis of variables (in Section 5.1.6) to multivariate analysis that takes other variables into account when investigating the influence of a particular variable.

Standard regression-based models are used because they are easier to interpret than the neural networks and decision tree ensembles that contain complex interactions amongst explanatory variables and for which the internal logic is often more hidden. For standard regression-based models, statistical tests are also available to evaluate the contribution of an individual variable to a specified level of statistical significance. Although easy interpretability is an advantage of relatively small decision trees, the DT_One model has 70 terminal nodes, which makes it far too large to interpret easily.

Logistic regression models are used instead of discriminant analysis models because they have fewer underlying assumptions that are likely violated (as stated in Section 3.3.1). Variable coefficients in a logistic regression model are also relatively straightforward to interpret as they influence the odds of fraud occurring, which is a concept explained in the next subsection before its use in those that follow.

6.1.2.a Odds of Fraud

Changes in the odds of fraud are relatively easy to interpret because if the odds increase (or decrease), the probability also increases (or decreases). Specifically, the

$$odds\ of\ fraud\ = \frac{probability\ of\ fraud\ occurring}{probability\ of\ fraud\ not\ occurring} = \frac{probability\ of\ fraud}{1 - probability\ of\ fraud},$$

which can be re-arranged as

$$probability\ of\ fraud = \frac{odds\ of\ fraud}{1 + odds\ of\ fraud}.$$

As an example, when the probability of fraud is 75% the odds of fraud occurring equate to $\frac{80\%}{1-80\%} = 4$ (to 1). The odds of 1 (to 1) equate to a $\frac{1}{1+1} = 50\%$ probability, and higher (or lower) odds result in a higher (or lower) probability.

Although odds and probability are positively correlated, the percentage change in odds is always greater than the percentage change in probability. Moreover, the difference between the percentage changes is greater when larger values are involved. For example, if the odds increase by 20% from:

- 1 to 1.1, then the probability increases by 9.09% from 0.5 to 0.55;
- 4 to 4.8 (larger values), then the probability increases by 3.45% (bigger difference from 20%) from 0.8 to 0.83.

6.1.2.b Results

Table 6-2 presents details on the variables in the stepwise logistic regression model (LR_Step). LR_Step was chosen because it was selected as the best overall standard regression-based model in Chapter 5 (Section 5.3.2.b) and it is not likely to suffer from multicollinearity issues because of the stepwise variable selection. This is important because multicollinearity reduces the reliability of interpretations of variable coefficients.

Table 6-3 presents results from all the logistic regression models, as well as stepwise discriminant analysis[132]. Only statistically significant variables are included, which is defined as those that are statistically significant at a 10% level in at least one model. Variables that are exclusively statistically significant in the LR_All model are also excluded, because of the

[132] The stepwise discriminant analysis model is included because it is unlikely to have multicollinearity problems and it serves as an alternative technique to logistic regression.

increased likelihood of multicollinearity (from using all variables) that reduces the reliability of the results. Overall, this table shows that the direction of association that variables have with fraud is often consistent across these models. Further analysis of these variables according to the categories in the overall schema of explanatory variables is presented on page 234 following the tables below.

Details of the variables in the LR_All model are presented in Table 6-4 for completeness, because LR_All provides information on every explanatory variable. However, findings from Table 6-4 should not be heavily relied upon because of the high likelihood of problems associated with multicollinearity.

Table 6-2. An analysis of the stepwise logistic regression model (LR_Step), where β values are the estimated variable coefficients. For each increase by one in a variable, the odds of a financial statement being fraudulent change by $e^{\beta} - 1$, holding constant all other explanatory variables. For example, for each increase in V33a by one, the odds of a financial statement being fraudulent decrease by 44.5%. Wald is the test statistic associated with the null hypothesis that a variable's coefficient (β) is zero and the alternative hypothesis that β is not zero, which indicates that the variable contributes to the model. Wald is calculated as β / SE_{β}, where SE is the standard error. The Chi-square distribution (with one degree of freedom) is then used to calculate the p-value for this test statistic, where values closer to zero indicate more support for the alternative hypothesis. Finally, the direction of association with fraud is summarised, along with the expected direction according to the theoretical analysis in Chapter 4. A positive value indicates that higher values of the specified variable are associated with a higher probability of fraudulent financial statements, and negative values indicate the opposite. Shading indicates that a variable is statistically significant at the 10% level. **Bold** indicates unexpected statistically-significant results.

Variable ID	Variable Name	Logistic Regression (LR_Step)				Direction of Association with Fraud	
		β	$e^{\beta} - 1$	Wald	P-value	LR_Step	Expected
V8a	Inventory to Sales	2.785	15.202	7.928	0.005	Positive	Positive
V11a	Sales to Total Assets	-0.412	-0.338	10.537	0.001	Negative	Uncertain
V19	Total Accruals to Total Assets	1.266	2.547	3.845	0.050	Positive	Uncertain
V20	Were the specified and the prior year's Total Accruals > 0?	1.564	3.778	14.990	0.000	Positive	Positive
V24	Debt to Equity	-0.083	-0.079	14.913	0.000	**Negative**	Positive
V27a	Return on Equity	-0.772	-0.538	13.789	0.000	Negative	Uncertain
V28a	Return on Average Prior Assets	0.886	1.426	4.132	0.042	Positive	Uncertain
V31b	Was New Common Stock or Long-term Debt Issued?	1.383	2.986	12.139	0.000	Positive	Positive
V33a	Demand for financing (ex ante)	-0.589	-0.445	8.053	0.005	Negative	Negative
V34	Were there operating leases?	0.737	1.090	7.470	0.006	Positive	Positive
V39b	Percentage of Directors who are also Executives	1.394	3.029	4.038	0.044	Positive	Positive
V44	Company Age: Number of years since foundation	-0.010	-0.010	13.605	0.000	**Negative**	Not expected to be significant given the matching procedure
V45	Company Size: natural log of Total Assets	0.389	0.476	41.632	0.000	**Positive**	
V46	Industry: SIC code starts with a three?	-0.486	-0.385	4.336	0.037	**Negative**	
Constant		-3.966	36.419	-0.981	0.000		

Table 6-3. The direction of association between fraud and the logistic regression models and stepwise discriminant analysis. Shading, * and ** indicate a variable is statistically significant at a 10%, 5% and 1% level respectively. **Bold** indicates unexpected results.

Variable ID	Variable Name	Direction of Association with Fraud						
		DA_Step	LR_Step	LR_U15%LR	LR_U15%	LR_U1	LR_All	Expected
V8a	Inventory to Sales	Positive**	Positive**	Positive*	Positive*	Positive*	Positive*	Positive
V10c	Previous year's Percentage change in Sales			Positive	Positive		Positive	Uncertain
V11a	Sales to Total Assets	Negative**	Negative**	Negative*	Negative*	Negative**	Negative*	Uncertain
V19	Total Accruals to Total Assets		Positive*		Positive	Positive*	Positive	Uncertain
V20	Were the specified and the prior year's Total Accruals > 0?	Positive**	Positive**	Positive**	Positive**	Positive**	Positive**	Positive
V24	Debt to Equity	**Negative****	**Negative****			**Negative****	**Negative****	Positive
V26	Four-period average of Times Interest Earned			Positive	Positive	Positive	Positive	Positive
V27a	Return on Equity	Negative**	Negative**			Negative**	Negative*	Uncertain
V28a	Return on Average Prior Assets	Positive*	Positive*			Positive	Positive	Uncertain
V31b	Was New Common Stock or Long-term Debt Issued?	Positive**	Positive**	Positive**	Positive*	Positive**	Positive*	Positive
V32	Proportion of common stock that is newly issued	Positive*			Positive	Positive	Positive	Positive
V33a	Demand for financing (ex ante)	Negative**	Negative**	Negative**	Negative*	Negative*	Negative*	Negative
V34	Were there operating leases?		Positive**	Positive*	Positive	Positive*	Positive**	Positive
V36	Number of changes of audit firm in the most recent four financial statements			Negative	Negative	Negative	Negative	Uncertain
V39b	Percentage of Directors who are also Executives	Positive*	Positive*	Positive*	Positive*	Positive	Positive	Positive
V44	Company Age: Number of years since foundation	**Negative****	**Negative****	**Negative****	**Negative****	**Negative****	**Negative****	Not expected to be significant given the matching procedure
V45	Company Size: natural log of Total Assets	**Positive****	**Positive****	**Positive****	**Positive***	**Positive****	**Positive****	
V46	Industry: SIC code starts with a three?		**Negative***			Negative	**Negative***	
V48a	Previous year's percentage change in annual real GDP					**Negative***	Negative	
V49a	Corporate Governance G-Index	Positive**		Positive**	Positive**	Positive	Positive	Positive

Table 6-4. An analysis of the logistic regression model using all variables (LR_All). The column definitions are the same as for the previous table. Shading indicates a variable is statistically significant at the 10% level. **Bold** indicates unexpected statistically-significant results.

Variable ID	Variable Name	Logistic Regression (LR_All)				Direction of Association with Fraud	
		β	$e^{\beta}-1$	Wald	P-value	LR_All	Expected
Specific Account - Accounts Receivable							
V1	Accounts Receivable						
V1a	Value for the specified year	0.000	0.000	0.328	0.567	Positive	Positive
V1b	Percentage change	-1.006	-0.634	4.326	0.038	**Negative**	Positive
V1c	Was Percentage change > 10%?	-0.690	-0.498	5.912	0.015	**Negative**	Positive
V2	Percentage change in Accounts Receivable to Sales	0.463	0.588	0.590	0.442	Positive	Positive
V3	Percentage change in Accounts Receivable to Total Assets	1.110	2.035	3.275	0.070	Positive	Positive
Specific Account - Allowance for doubtful accounts (AFDA)							
V4	Percentage change in AFDA	0.112	0.119	0.257	0.612	Positive	Negative
V5	Percentage change in AFDA to Accounts Receivable	0.011	0.011	0.003	0.958	Positive	Negative
V6	Percentage change in AFDA to Sales	-0.214	-0.193	0.420	0.517	Negative	Negative
Specific Account - Inventory							
V7	Change in Inventory to average Total Assets	1.294	2.646	0.310	0.578	Positive	Positive
V8	Inventory to Sales						
V8a	Value for the specified year	2.759	14.782	5.460	0.019	Positive	Positive
V8b	Change	-0.189	-0.172	0.433	0.511	Negative	Positive
V9	Was Last-In, First-Out (LIFO) inventory valuation used?	0.347	0.415	0.888	0.346	Positive	Negative
Specific Account - Sales							
V10	Sales Growth						
V10a	Percentage change	1.951	6.034	3.294	0.070	Positive	Uncertain
V10b	V10a minus the Industry Average	-0.054	-0.053	0.004	0.947	Negative	Uncertain
V10c	Previous year's Percentage change	0.240	0.271	1.995	0.158	Positive	Uncertain
V10d	Four-year growth rate	-0.778	-0.541	0.926	0.336	Negative	Uncertain
V10e	Previous year's percentage change in total assets	-0.040	-0.039	0.233	0.629	Negative	Uncertain

Variable ID	Variable Name	Logistic Regression (LR_All)				Direction of Association with Fraud	
		β	$e^{\beta}-1$	Wald	P-value	LR_All	Expected
V11	Sales to Total Assets						
V11a	Value for the specified year	-0.455	-0.366	4.532	0.033	Negative	Uncertain
V11b	Percentage change	-1.637	-0.805	8.628	0.003	Negative	Uncertain
V12	Gross Margin to Sales						
V12a	Percentage change	-0.036	-0.035	0.562	0.453	Negative	Uncertain
V12b	Was percentage change > 10%?	-0.400	-0.330	1.676	0.195	Negative	Uncertain
V13	Cash Sales						
V13a	Percentage change	-0.189	-0.172	0.321	0.571	Negative	Uncertain
V13b	Was change < 0?	0.215	0.239	0.516	0.473	Positive	Uncertain
V14	Were any sales from acquisitions?	0.382	0.465	2.368	0.124	Positive	Positive
General Financial - Asset Composition							
V15	Current Assets to Total Assets	-0.353	-0.298	0.124	0.724	Negative	Positive
V16	Net Property, Plant & Equipment (PP&E) to Total Assets	-0.528	-0.410	0.139	0.709	Negative	Negative
V17	Soft Assets to Total Assets	0.369	0.446	0.107	0.743	Positive	Positive
V18	Percentage Change in Assets other than Current Assets and Net PP&E to Total Assets	-0.008	-0.007	0.118	0.731	Negative	Positive
General Financial - General Accrual Measures							
V19	Total Accruals to Total Assets	1.714	4.554	3.115	0.078	Positive	Uncertain
V20	Were the specified and the prior year's Total Accruals > 0?	1.832	5.247	15.089	0.000	Positive	Positive
V21	Total Discretionary Accruals	-0.051	-0.050	0.344	0.558	Negative	Positive
V22	RSST (unadjusted) Accruals	-1.087	-0.663	1.679	0.195	Negative	Positive
General Financial - Level of Debt and Financial Distress							
V23	Debt to Total Assets	-1.394	-0.752	4.362	0.037	Negative	Positive
V24	Debt to Equity	-0.082	-0.078	11.363	0.001	Negative	Positive
V25	Altman's (1968) financial distress measure (Z-score)	0.001	0.001	0.064	0.801	Positive	Positive
V26	Four-period average of Times Interest Earned	0.001	0.001	2.619	0.106	Positive	Positive

Variable ID	Variable Name	Logistic Regression (LR_All)				Direction of Association with Fraud	
		β	$e^{\beta}-1$	Wald	P-value	LR_All	Expected
General Financial - Performance and Profitability							
V27	Return on Equity						
V27a	Value for the specified year	-1.436	-0.762	4.883	0.027	Negative	Uncertain
V27b	Industry Average minus Specific Company	-0.657	-0.482	1.139	0.286	Negative	Uncertain
V28	Return on Average Prior Assets						
V28a	Value for the specified year	0.462	0.587	0.293	0.588	Positive	Uncertain
V28b	Previous year	0.320	0.377	0.178	0.673	Positive	Uncertain
V28c	Change	0.337	0.400	0.186	0.666	Positive	Uncertain
V29	Holding Period Return						
V29a	One-year	0.045	0.046	0.263	0.608	Positive	Uncertain
V29b	Previous One-year	-0.002	-0.002	0.000	0.986	Negative	Uncertain
V30	Were analyst Earnings Per Share forecasts achieved or exceeded?	-0.120	-0.113	0.205	0.651	Negative	Uncertain
General Financial - Financing							
V31	Were New Securities issued?						
V31a	Common Stock?	0.228	0.256	0.679	0.410	Positive	Positive
V31b	Common Stock or Long-term Debt?	1.186	2.274	6.021	0.014	Positive	Positive
V32	Proportion of common stock that is newly issued	0.794	1.212	1.261	0.261	Positive	Positive
V33	Demand for financing						
V33a	Specific Value (ex ante)	-0.771	-0.537	5.714	0.017	Negative	Negative
V33b	Was there demand (ex ante)?	-2.663	-0.930	5.440	0.020	Negative	Positive
V33c	Cash from operating and investment activities	0.000	0.000	1.141	0.285	Positive	Positive
V34	Were there operating leases?	0.863	1.371	6.916	0.009	Positive	Positive

Variable ID	Variable Name	Logistic Regression (LR_All)				Direction of Association with Fraud	
		β	$e^{\beta}-1$	Wald	P-value	LR_All	Expected
Non-financial - Key Roles and Positions							
V35	Was the auditor a Big Six firm?	0.067	0.069	0.025	0.874	Positive	Negative
V36	Number of changes of audit firm in the most recent four financial statements	-0.662	-0.484	3.838	0.050	Negative	Uncertain
V37	CEO						
V37a	Tenure (days)	0.000	0.000	1.061	0.303	Positive	Uncertain
V37b	Number of changes in the last three years	0.174	0.190	0.425	0.514	Positive	Uncertain
V38	Has the CFO changed in the last three years?	0.191	0.211	0.637	0.425	Positive	Uncertain
V39	Composition/Holdings of the Board						
V39a	Number of Directors	-0.040	-0.040	3.089	0.079	Negative	Positive
V39b	Percentage of Directors who are also Executives	1.529	3.612	3.175	0.075	Positive	Positive
V39c	Percentage of Director shares owned by those who are also Executives	Excluded because of insufficient data					Positive
V40	Percentage of total shares owned by the CEO	Excluded because of insufficient data					Uncertain
Comparing Financial and Non-financial							
V41	Percentage change in the number of Employees minus percentage change in Total Assets	-0.094	-0.090	0.236	0.627	Negative	Negative
V42	Percentage change in Sales minus percentage change in the number of Employees	-0.164	-0.151	0.398	0.528	Negative	Positive
V43	Percentage Change in Sales to Employees: Specific Company minus Industry Average	-0.015	-0.015	0.041	0.839	Negative	Positive

Variable ID	Variable Name	Logistic Regression (LR_All)				Direction of Association with Fraud	
		β	$e^{\beta}-1$	Wald	P-value	LR_All	Expected
Control variables							
V44	Company Age: Number of years since foundation	-0.010	-0.010	8.079	0.004	Negative	Not expected to be significant given the matching procedure
V45	Company Size: natural log of Total Assets	0.349	0.417	9.718	0.002	Positive	
V46	Industry: SIC code starts with a three?	-0.688	-0.498	5.223	0.022	Negative	
V47	Stock Exchange listed on						
V47a	NASDAQ?	0.022	0.022	0.000	0.994	Positive	
V47b	NYSE?	-0.256	-0.226	0.007	0.934	Negative	
New variables							
V48	Macroeconomic indicators						Not expected to be significant given the matching procedure
V48a	Previous year's percentage change in annual real GDP	-7.102	-0.999	0.163	0.686	Negative	
V48b	Previous year's percentage change in annual retail sales	-0.791	-0.546	0.027	0.870	Negative	
V48c	Previous year's unemployment rate inverted	-0.042	-0.041	0.392	0.531	Negative	
V49	Corporate governance indices						
V49a	G-Index	0.082	0.085	3.755	0.053	Positive	Positive
V49b	E-Index	Excluded because of insufficient data					Positive
V50	Accounting complexity of the industry	-0.003	-0.003	0.023	0.881	Negative	Positive
Constant		-0.111	-0.105	0.001	0.977		

6.1.2.c Analysis of Financial Variables That Measure Specific Accounts

Inventory

The inventory to sales ratio (V8a) is statistically significant at a 1% level in the stepwise logistic regression model (LR_Step) with a p-value of 0.5% (see Table 6-2), as well as being statistically significant at a 5% level in all the other models in Table 6-3[133]. Larger values of V8a increased the probability of financial statement fraud, which is consistent with increased opportunity to fraudulently overvalue inventory by including obsolete inventory (see Section 4.5.4). According to the LR_Step model presented in Table 6-2, for each increase by one in the inventory to sales ratio, the odds that financial statement fraud has occurred increases by 1520%, holding constant all other explanatory variables in the model. This appears extremely high, but the inventory to sales ratio is extremely unlikely to increase by one; for example, from 15% to 115%. A more relevant interpretation is for each increase by 0.01 in the inventory to sales ratio, the odds of financial statement fraud having occurred increases by 2.82%, holding constant all other explanatory variables in the model. This 2.82% value is calculated as $(1 + 1520\%)^{1\%} - 1$ or $e^{2.785*0.01} - 1$, where 2.785 is the coefficient of V8a in the LR_Step model.

Sales

Sales growth in the previous year (V10c) is statistically significant at a 10% level only in one model (LR_U15%) in Table 6-3. It had a positive association with the probability of financial statement fraud in this model. This offers limited support for the theory that high sales growth in the prior period both increases pressure on managers to achieve that growth rate again and increases opportunities to commit fraud through reduced controls and monitoring (see Section 4.5.5).

The sales to total assets ratio (V11a) is statistically significant at a 1% level in the stepwise logistic regression model (LR_Step) with a p-value of 0.1%, as well as being statistically significant at a 5% level in all the other models shown in Table 6-3. However, in contrast to V10c, V11a has a negative association with fraud, such that larger values decreased the probability of financial statement fraud. Larger values of V11a indicate greater efficiency and ability to generate sales from assets, which decreases the pressure to fraudulently improve performance (see Section 4.5.5). According to the LR_Step model

[133] No other inventory variables were listed in Table 6-3.

presented in Table 6-2, for each increase by 0.01 in the sales to assets ratio, the odds of financial statement fraud having occurred decrease by 0.41%, holding constant all other explanatory variables in the model; the relevant calculation is $(1 - 33.8\%)^{1\%} - 1 = -0.41\%$.

The expected direction of association of sales variables was largely uncertain because of mixed results and theories in past research (see Section 4.5.5). This current research adds additional empirical evidence to better understand the influence of the sales variables. Overall, sales growth from the prior year increased the probability of fraud, possibly through increased pressure and opportunity to commit fraud. However, sales from the specified year that occurred with greater efficiency relative to total assets reduced the probability of fraud having occurred, likely through decreased pressure to inflate sales.

Accounts Receivable

The percentage change in both accounts receivable (V1b and V1c) and accounts receivable to sales (V3) are statistically significant variables at a 10% level in the LR_All model (see Table 6-4 above). V3 is positively associated with fraud as expected, but V1b and V1c are both unexpectedly negatively related to the probability of fraud. The conflicting result could indicate the theory underlying them (see Section 4.5.2) needs adjustment, but it is more likely the result of multicollinearity problems in LR_All and should not be relied upon.

Although accounts receivable variables were statistically significant in the LR_All model, no accounts receivable variables are listed in Table 6-3. This indicates that none of them are statistically significant at a 10% level in other models, despite their use in many prior studies (as presented in Section 4.5.2). The main reason for their inclusion was the increased opportunity to commit fraud because accounts receivable is an account relatively difficult to audit. This research is more comprehensive than previous research in terms of the number of explanatory variables used. Consequently, these variables might not be statistically significant because other variables that measure the opportunity to commit fraud are preferred, such as inventory variables V8a and accrual variables V19 and V20. This would explain why V1a measuring accounts receivable is statistically significant at a 1% level with the expected positive direction of association in univariate analysis (see Section 5.1.6), but it was not the case in this multivariate analysis.

Allowance for Doubtful Accounts (AFDA)

None of the AFDA variables are statistically significant at a 10% level in Tables 6-2, 6-3 and 6-4. These variables had previously been used by two studies that used artificial neural networks (Green and Choi 1997; Lin et al. 2003), but no tests were conducted to assess the individual contribution of these variables. While the earlier study cited the variables' use in auditing as a reason for inclusion, it was acknowledged that there was no supporting theory. Furthermore, this research empirically finds that the AFDA variables do not contribute to models to detect financial statement fraud at a 10% level of statistical significance.

6.1.2.d Analysis of General Financial Variables

Asset Composition

While one variable (V16) is statistically significant at a 10% level in univariate analysis (in Section 5.1.6), no asset composition variables are statistically significant at a 10% level in multivariate analysis presented in Tables 6-2, 6-3 and 6-4. This indicates that the useful information contained in these variables is better captured by variables that measure specific asset accounts and other variables that measure opportunity for fraud to occur.

General Accrual Measures

The total accruals to total assets ratio (V19) is statistically significant at a 5% level in both the LR_Step and LR_U1 models, as well as at a 10% level of significance in the LR_All model (see Table 6-3). Larger values of this ratio increased the probability of financial statement fraud, supporting the theory that this results in an increased opportunity to manipulate the non-cash accruals (see Section 4.5.7). According to the LR_Step model presented in Table 6-2, for each increase by 0.01 in the total accruals to total assets ratio, the odds of financial statement fraud having occurred increased by $(1 + 255\%)^{1\%} - 1 = 1.27\%$, holding constant all other explanatory variables in the model.

Variable V20 measures whether there were positive accruals in the specified year as well as the prior year. It is statistically significant at a 1% level in the LR_Step model with a p-value of 0.01%, as well as being statistically significant at a 1% level in all the other models in Table 6-3. Prior accruals reduce legitimate ways to manage earnings and consequently increase the pressure to turn to fraud (see Section 4.5.7). Consistent with this theoretical expectation and the findings for V19 (above), V20 increased the probability of financial statement fraud. Specifically, according to the LR_Step model presented in Table 6-2, the odds of the financial statement fraud having occurred increase by 378% when there are positive accruals in both the specified year and the year prior, holding constant all other explanatory variables in the model.

There have been mixed results in prior studies in terms of which measures of accruals are best for financial statement detection models. Consistent over a range of standard regression-based models, this study shows that measures of total accruals were significant, while variables using either discretionary accruals or RSST unadjusted accruals were not statistically significant at a 10% level of significance.

Level of Debt and Financial Distress

The debt to equity ratio (V24) is statistically significant at a 1% level in both stepwise models (DA_Step and LR_Step) as well as LR_U1 and LR_All. Based on prior research, increased debt levels were expected to be associated with more pressure to commit fraud (see Section 4.5.8). However, higher debt levels (relative to equity) decreased the probability of financial statement fraud having occurred, although only by a relatively small amount. According to the LR_Step model presented in Table 6-2, for each increase by 0.01 in the debt to equity ratio, the odds of financial statement fraud having occurred decreased by 0.08%, holding constant all other explanatory variables in the model; the relevant calculation is $(1-7.9\%)^{1\%}-1=-0.08\%$.

Although the finding of a negative association between debt and fraud was contrary to previous fraud detection research, it was not completely unexpected. As discussed in Section 4.5.8, higher debt levels are likely to increase monitoring and scrutiny from creditors (those who are owed money), which is consistent with the well-established agency theory (Jensen and Meckling 1976). This increased scrutiny then reduces the opportunity to commit and conceal fraud. This new empirical result, which is supported by well-established theory, might have been revealed in this research as a consequence of using a more comprehensive set of explanatory variables. Furthermore, the number of fraud cases in the training sample in this research is larger, usually more than 100% larger, than any of the prior studies that found the contrary result.

In addition to the size of debt (relative to equity), financial distress as measured by the difficulty that changes in interest rates would cause in terms of making debt repayments (V26) was found to be statistically significant at a 10% level, but only in the LR_U1 model.

As expected, more financial distress was associated with more fraud, which is consistent with increased pressure and incentive to commit fraud.

Performance and Profitability

Accounting measures of profitability are statistically significant at a 10% level (see Table 6-3), in contrast to the market-based measures of performance and the comparison to an industry average. This indicates that accounting measures of profitability contributed more to standard regression-based models when detecting financial statement fraud.

Return on equity (V27a) is statistically significant at a 1% level in both stepwise models (DA_Step and LR_Step) and LR_U1, as well as at a 5% level of significance in the LR_All model. More return on equity decreased the probability of financial statement fraud. According to the LR_Step model presented in Table 6-2, for each increase by 0.01 in return on equity, the odds of financial statement fraud having occurred decreased by 0.77%, holding constant all other explanatory variables in the model; the relevant calculation is $(1 - 53.8\%)^{1\%} - 1 = -0.77\%$.

Return on average prior assets (V28a) is also statistically significant, at a 5% level in both stepwise models (DA_Step and LR_Step) and at a 10% level in the LR_U1 model. However, in contrast to V27a, more return on average prior assets increased the probability of financial statement fraud. According to the LR_Step model presented in Table 6-2, for each increase by 0.01 in return on average prior assets, the odds of financial statement fraud having occurred increased by $(1 + 143\%)^{1\%} - 1 = 0.89\%$, holding constant all other explanatory variables in the model.

The probability of fraud decreasing with larger returns on equity is consistent with reduced pressure to fraudulently improve financial statements during times of good profitability (see Section 4.5.9). In contrast, higher profits relative to prior assets increased the probability of fraud. To understand these differing results, it is important to note that the only difference between them is their denominators: equity in the specified year for V27a and average prior assets for V28a. The denominator of V28a being a measurement of prior years could be a key consideration. These findings are consistent with higher current profitability (V27a) reducing fraud (from reduced pressure), but higher profitability relative to past resources (V28a) is suspicious and more likely to be fraudulent. Thus, V28a might be related to the new Suspicious Information (S) factor of the Fraud Detection Triangle, unlike V27a that is related to the Pressure/Incentive (I) factor.

Financing

The financing category has the largest number of statistically significant variables at a 10% level and they all have the expected direction of association with fraud, which is consistent with the univariate analysis (presented in Section 5.1.6).

Variable V31 measures whether new common stock or new long-term debt was issued and is statistically significant at a 1% level in the LR_Step model with a p-value of 0.05%, as well as being statistically significant at least at a 5% level in all the other models shown in Table 6-3. V31b increased the probability of financial statement fraud. According to the LR_Step model presented in Table 6-2, the odds of the financial statement fraud having occurred increase by 299% when there are positive accruals in both the specified year and the year prior, holding constant all other explanatory variables in the model. In addition, the proportion of common stock that is newly issued (V32) is also statistically significant at a 5% level in the stepwise discriminant analysis model DA_Step. Consistent with V31b, a higher proportion of newly issued equity indicated a higher probability of fraud.

The demand for financing (ex ante) (V33a) is also statistically significant, at a 1% level in both stepwise models (DA_Step and LR_Step) and LR_U15%LR, and at a 5% level in the other models in Table 6-3. The probability of fraud increased when the demand for financing was higher, as indicated by lower values of V33a. According to the LR_Step model presented in Table 6-2, the odds of the financial statement fraud having occurred increases by 44.5% for each decrease by one in V33a (which indicates an increased demand for financing), holding constant all other explanatory variables in the model.

Variable V34 measures whether operating leases have been used and is statistically significant at a 1% level in the LR_Step and LR_All models, as well as being statistically significant at least at a 10% level in the other logistic regression models in Table 6-3. The probability of financial statement fraud increased if operating leases are being used. Specifically, according to the LR_Step model presented in Table 6-2, the odds of the financial statement fraud having occurred increase by 109% when operating leases have been used.

The findings for the V31b, V32 and V33a variables support the view that there is an increased incentive to fraudulently improve financials when new securities are being issued, or are likely to be needed to be issued in the near future. Additionally, the results for V34 support the view that the use of operating leases is associated with managers who are more likely to commit fraud as they find it easier to rationalise fraud by focusing on the short-term (see Section 4.5.10).

6.1.2.e Analysis of Non-Financial Variables

The number of changes of audit firm in the most recent four financial statements (V36) is statistically significant at a 10% level in all models in Table 6-3 except for the stepwise models. More changes decreased the probability of financial statement fraud. This offers some support for the theory that changes of audit firm help to avoid the complacency and the overconfidence that lead to lower quality audits. This is contrary to the argument that the quality of audit improves as the auditor tenure increases, and that changes of audit firm can indicate management more prone to rationalise fraud as they are more likely to be engaged in auditor “opinion-shopping” (see Section 4.5.11). The expected direction of association of this variable was uncertain because of competing theories. Consequently, this finding is additional empirical evidence to better understand the influence of changes of audit firm.

The percentage of directors who are also executives (V39b) is statistically significant at a 5% level in the stepwise logistic regression model with a p-value of 4.4% (see Table 6-2), as well as being at least statistically significant at a 10% level in all the other models in Table 6-3. A larger percentage increased the probability of financial statement fraud. This supports the theory that directors who are also executives are less independent and more of them result in a reduced monitoring from the board of directors. This reduction in monitoring creates more opportunities for management to commit and conceal financial statement fraud (as per Section 4.5.11). According to the LR_Step model presented in Table 6-2, for each increase by 0.01 in the percentage of directors who are also executives, the odds of financial statement fraud having occurred increase by $(1 + 303\%)^{1\%} - 1 = 1.4\%$, holding constant other explanatory variables in the model.

Overall, the occurrence of financial statement fraud decreased with more changes of audit firm, and fewer directors who are also executives.

6.1.2.f Analysis of Variables That Compare Financial and Non-Financial Information

No variables comparing financial and non-financial information are statistically significant at a 10% level in multivariate analysis (see Tables 6-2, 6-3 and 6-4) or univariate analysis (see Section 5.1.6). The motivation for including these variables was that increases in sales or assets (financial) without corresponding increases in the number of employees (non-financial) raises the suspicion of fraud, because sales or assets are much easier to fraudulently manipulate than the number of employees. Nevertheless, these variables were not a good signal of fraud in these models. Further, these variables are not important in models that allow interactions with other explanatory variables, as presented later in Section 6.1.5. For future research, it is recommended that the financial and non-financial information being compared should be different from sales (or assets) and the number of employees.

6.1.2.g Analysis of Control Variables

Company age (V44) is statistically significant at a 1% level in all of the models in Table 6-3. Company size (V45) is also statistically significant at a 1% level in all the models except LR_U15%, in which it is statistically significant at a 5% level. Industry membership (V46) is also statistically significant at a 5% level in LR_Step and LR_All. The findings are revealing, because these control variables were not expected to be significant because of the matching procedure. First, this means that these variables are important when detecting financial statement fraud. Secondly, it reveals that the matching procedure was not exact, which was acknowledged previously in Section 4.5.13 and was part of the reason for including these variables in the first place. An example of the matching being inexact is that companies were only matched based on being more than ten years old, five to ten years old and younger than five years. Furthermore, this criterion was relaxed in some cases to enable a matching company to be found (see Section 5.1.2).

Younger companies were associated with more fraud, which is consistent with the theory that there are more incentives to commit fraud around the time of the initial public offering (as per Section 4.5.13). Larger companies were also associated with more fraud. This finding is contrary to the thought that smaller firms have weaker internal controls and so there is more opportunity to commit fraud within them. Consequently, this could indicate that larger firms have weaker controls or it could possibly indicate that the US Securities Exchange Commission (SEC) has a bias to prosecute larger firms (as suggested in Section 5.1.1.a). Finally, firms with Standard Industrial Classification (SIC) codes starting with a 3, which are primarily computer companies, committed less fraud in this study. This is the opposite finding of previous research, and indicates this result might be purely an empirical result specific to the time period being studied.

6.1.2.h Analysis of New Variables

The new accounting complexity variable V50 is not significant in any model in Table 6-3, including LR_All. This does not provide support for the use of the new variable V50 in fraud detection models. Further to this, V50 is not important in artificial neural network and decision-tree based models that incorporate interactions between explanatory variables, as shown later in Section 6.1.5.

A macroeconomic variable measuring the change in annual real GDP (V48a) is statistically significant at a 5% level in the LR_U1 model. This indicates that in this model V48a was able to assist in distinguishing fraudulent statements from legitimate statements that are not their matched counterparts[134]. This is very limited support for the use of macroeconomics variables. Similarly limited backing for the use of macroeconomic variables is also provided by neural network models that incorporate interactions (see Section 6.1.5).

Weaker corporate governance, as indicated by larger values of the G-index (V49a), were associated with more financial statement fraud. This variable is statistically significant at a 1% level in DA_Step, LR_U15% and LR_U15%LR, as well as at a 10% level in LR_U1 and LR_All, as shown in Table 6-3. This result is in contrast to prior studies that did not find the G-index to be statistically significant using a much smaller number of fraud cases (as per Section 4.5.15). Overall, using a larger number of fraud cases, this research has found empirical support for the hypothesis that financial statement fraud is related to the aggregation of many corporate governance initiatives, not just the presence or absence of a particular one.

6.1.2.i Summary of Variable Analysis

This section summarises the variables most statistically significant in the multivariate analysis presented above. This is defined as those variables that are statistically significant at a 10% level in at least two models in Table 6-3, providing that at least one is a stepwise model because they are the least affected by the problem of multicollinearity.

Overall, the analysis above reveals that the probability a financial statement is fraudulent is increased by:

- Higher levels of inventory relative to sales (V8a);
- Lower levels of sales relative to total assets (V11a);
- Larger total accruals relative to total assets (V19) and positive total accruals in the specified year and the one prior (V20);

[134] These variables could not distinguish between a fraudulent statement and its matched counterpart because they are from the same year and consequently have the same annual macroeconomic information.

- Lower levels of debt relative to equity (V24), contrary to results in prior studies;
- Smaller profits relative to equity (V27a) and larger profits relative to the average asset base in the prior two years (V28a);
- New securities (common stock or long-term debt) being issued (V31b) and likely to be needed to be issued in the near future (V33a);
- The use of operating leases (V34);
- A larger proportion of directors also being executives (V39b);
- Younger and larger companies (V44 and V45); and
- Weaker corporate governance as measured by more provisions that restrict shareholder rights (V49a).

6.1.3 Analysis According to the Overall Schema of Explanatory Variables

The variables that are most statistically significant, as listed above in the previous section, include measures of specific financial accounts (inventory and sales), general accrual measures, debt, accounting profitability (in preference to market or industry based profitability), financing needs and activities, composition of the board of directors (non-financial), company characteristics (age and size) and overall corporate governance. These variables represent a variety of categories within the overall schema of explanatory variables as shown in Table 6-5 below. This table also compares the most statistically significant variables with the set of all variables in terms of their proportions in each category. The general financial variables represent the largest proportion in both cases, but that proportion is larger for the most statistically significant variables. There is also a greater proportion of control variables, and a lower proportion of financial variables measuring specific accounts, amongst the most statistically significant variables. Despite having many missing values, the corporate governance G-index is the statistically significant variable in the New category. This provides a strong case for its future use in financial statement fraud detection models.

The comparison category is the only one not represented in the most statistically significant set of variables. However, these variables might be important contributors when they interact with other explanatory variables, which is investigated later in Section 6.1.5. In addition, other comparison variables could be investigated in future research (as suggested in Section 6.1.2.f).

Table 6-5. The proportion of all explanatory variables and those that were the most statistically significant from multivariate analysis by category within the overall schema.

Category within the Overall Schema of Explanatory variables	Proportion of explanatory variables[135]	
	All from Table 4-1	**Most statistically significant** from Section 6.1.2.i
Financial: Specific Account	28%	14%
Financial: General	40%	57%
Non-Financial	6%	7%
Comparison: Financial and Non-Financial	12%	0%
Control	8%	14%
New	6%	7%
Total	100%	100%

6.1.3.a Analysis of the Multi-stage Logistic Regression Model That Utilises the Overall Schema of Explanatory Variables

The multi-stage logistic regression model LR_MS_S utilises the categories within the overall schema of explanatory variables. Consequently, an analysis of it provides insight into the contributions that each category (as a whole) makes towards detecting financial statement fraud.

In the first stage of the LR_MS_S model, five logistic regression models are estimated, each using explanatory variables from a different category: Financial (Finc), Non-Financial (Non-Finc), Comparison (Comp), Control (Ctrl) and New. In the second stage, the LR_MS_S logistic regression model is estimated from the probability outputs of the first stage models, as shown below where $\Pr(\mathrm{F})_x$ is the probability of fraud estimated using a logistic regression with explanatory variables from the x category. The estimated coefficients and results from testing their statistical significance are then shown in Table 6-6.

$$\Pr(\mathrm{F})_{LR\_MS\_S} = \frac{e^{\mathrm{Log\_odds}_{LR\_MS\_S}}}{1 + e^{\mathrm{Log\_odds}_{LR\_MS\_S}}}, where$$

$$\mathrm{Log\_odds}_{LR\_MS\_S} = \beta_0 + \beta_1 \Pr(\mathrm{F})_{Finc} + \beta_2 \Pr(\mathrm{F})_{Non-Finc} + \beta_3 \Pr(\mathrm{F})_{Comp} + \beta_4 \Pr(\mathrm{F})_{Ctrl} + \beta_5 \Pr(\mathrm{F})_{New}.$$

[135] Different sub-types of the same main variable, such as V1a and V1b are only counted as one variable.

Table 6-6. An analysis of the multi-stage logistic regression model based on the overall schema (LR_MS_S). The column definitions are the same as for Table 6-2 with the following exception. For each increase by 0.01 in a variable, the odds of the financial statement being fraudulent change by $e^{0.01\beta} - 1$, holding constant all other explanatory variable. For example, for each increase by 0.01 in the probability of fraud according to the first stage Financial model ($Pr(F)_{Financial}$), the odds of a financial statement being fraudulent increase by 5%.

	LR_MS_S			
Variable	$\boldsymbol{\beta}$	$\boldsymbol{e^{0.01\beta} - 1}$	**Wald**	**P-value**
$\mathrm{Pr(F)}_{Financial}$	4.848 (β_1)	0.050	96.934	0.000
$\mathrm{Pr(F)}_{Non-Financial}$	3.384 (β_2)	0.034	5.510	0.019
$\mathrm{Pr(F)}_{Comparison}$	5.734 (β_3)	0.059	0.575	0.448
$\mathrm{Pr(F)}_{Control}$	2.724 (β_4)	0.028	7.974	0.005
$\mathrm{Pr(F)}_{New}$	3.707 (β_5)	0.038	2.697	0.101
Constant	-10.207 (β_0)	-0.097	5.970	0.015

As expected, all the variable coefficients are positive, indicating that a higher estimated probability of fraud in each first stage model translates to a higher probability of fraud in the second stage model. The outputs from the first stage models based on the Control and Financial variables were statistically significant at a 1% level and also had the greatest magnitude of influence by having the largest $e^{0.01\beta} - 1$ values in Table 6-6. The output from the Non-Financial variables model was also statistically significant at a 5% level. Consistent with the results presented above, the output from the model using Comparison variables was not statistically significant at a 10% level with a p-value of 44.8%. Furthermore, no variable in the first stage model using Comparison variables was statistically significant at a 10% level as all of their p-values were greater than 45%. Finally, the output from the model using the New variables contributes to the second stage model at a statistical significance level of 15% (but not 10%) with a p-value of 10.1%. Consistent with prior results the G-index variable is the only one in the first stage model using New variables that is statistically significant at a 10% level.

Overall, the LR_MS_S model provides additional support for the use of the overall schema of explanatory variables, except for the comparison category.

6.1.4 Analysis According to the New Fraud Detection Triangle Framework

Table 6-7 shows the proportion of the most statistically significant variables (as listed above in Section 6.1.2.i) that are associated with each factor of the new Fraud Detection

Triangle framework. The proportions are similar to those for all the explanatory variables, which are also shown in Table 6-7. The most statistically significant variables are primarily associated with the Exploitable Opportunity (O) and Incentive/Pressure (I) factors, and a small proportion is associated with the Integrity/Attitude/Rationalisation (R) factor. This supports the use of explanatory variables that in total measure the O, I and R factors. Beyond this, there is an opportunity for further research into variables that measure the R factor. As mentioned in Section 4.2.3, "the relatively few variables associated with the R factor are not an indication that it is unimportant, but rather an indication that less focus has been placed on it in prior research, probably because it is the most difficult factor to measure (Skousen et al. 2009)".

Table 6-7. The proportion of all explanatory variables and those that were the most statistically significant from multivariate analysis by each factor in the new Fraud Detection Triangle framework. The final column represents an adjustment to the most statistically significant column based on the findings from Section 6.1.2.

	Proportion of explanatory variables[135]		
Fraud Detection Triangle Factor	**All** from Table 4-3	**Most statistically significant** from Section 6.1.2.i	**With Adjustments**
Exploitable Opportunity (O)	49%	47%	47%
Incentive/Pressure (I)	38%	47%	40%
Integrity/Attitude/Rationalisation (R)	5%	6%	7%
Suspicious Information (S)	8%	0%	7%

The major difference between the proportions for all variables and the most statistically significant variables is that none of the latter set are associated with the new Suspicious Information (S) factor, because it is a new factor. However, the findings from analysing each variable's direction of association with fraud (in Section 6.1.2 above) indicate that there were some changes to the factors that each variable is associated with; these changes are:

- More debt (V24) was found to be associated with reduced opportunity to commit fraud, rather than with increased pressure to commit fraud to satisfy creditors. Consequently, variable V24 was adjusted from being associated with both the O and I factors to being solely associated with the O factor;
- Larger values of V11a indicate greater ability to generate sales from assets. This decreased the pressure to fraudulently improve performance, rather than increased the

opportunity to commit fraud from higher levels of sales, which is a relatively difficult account to audit. Consequently, variable V24 was adjusted from being associated with both the O and I factors to being solely associated with the I factor. Thus, this change offsets the change for V24 above from a proportion perspective in Table 6-7;

- While higher current profitability (V27a) indicated reduced pressure to commit fraud, higher profitability relative to past resources (V28a) was found to be suspicious information and associated with more fraud. Consequently, variable V28a was adjusted from being associated with the I factor to being associated with the S factor.

When adjustments are made for these new findings, the most statistically significant variables do include a measurement of the S factor, as shown by the final column of Table 6-7. Furthermore, the proportions of all categories very closely approximate the proportions in the set of all explanatory variables. The relatively few variables associated with the S factor do not necessarily indicate that this factor is relatively unimportant, but is a likely consequence of relatively few variables having been tested, given that it is a new factor proposed in this research. Overall, there is support for the use of variables that measure the S factor and further research into such variables would be beneficial.

Overall, the adjusted results in Table 6-7 show that all factors of the newly proposed Fraud Detection Triangle framework have contributed to models to detect financial statement fraud. This is empirical evidence to support the use of this framework during the variable selection process for such models.

6.1.4.a Analysis of the Multi-stage Logistic Regression Model That Utilises the Framework

The multi-stage logistic regression model LR_MS_F utilises the new Fraud Detection Triangle framework. Consequently, an analysis of it provides insight into the contributions that each framework factor (as a whole) makes towards detecting financial statement fraud.

In the first stage of the LR_MS_F model, four logistic regression models are estimated, each using explanatory variables from a different factor: Exploitable Opportunity (O), Incentive/Pressure (I), Integrity/Attitude/Rationalisation (R) and Suspicious Information (S). In the second stage, the LR_MS_F logistic regression model is estimated from the probability outputs of the first stage models, as shown below where $\Pr(\mathrm{F})_y$ is the probability of fraud estimated using a logistic regression with explanatory variables from the y factor.

The estimated coefficients and results from testing their statistical significance are then shown in Table 6-8.

$$\Pr(\mathrm{F})_{LR\_MS\_F} = \frac{e^{\mathrm{Log\_odds}_{LR\_MS\_F}}}{1 + e^{\mathrm{Log\_odds}_{LR\_MS\_F}}}, where$$

$$\mathrm{Log\_odds}_{LR\_MS\_F} = \beta_0 + \beta_1 \Pr(\mathrm{F})_O + \beta_2 \Pr(\mathrm{F})_I + \beta_3 \Pr(\mathrm{F})_R + \beta_4 \Pr(\mathrm{F})_S$$

Table 6-8. An analysis of the multi-stage logistic regression model based on the new Fraud Detection Triangle (LR_MS_F). The column definitions are the same as for Table 6-2 with the following exception. For each increase by 0.01 in a variable, the odds of the financial statement being fraudulent change by $e^{0.01\beta} - 1$, holding constant all other explanatory variables. For example, for each increase by 0.01 in the probability of fraud according to the O model, the odds of a financial statement being fraudulent increase by 3%.

	LR_MS_F			
Variable	$\boldsymbol{\beta}$	$\boldsymbol{e^{0.01\beta} - 1}$	**Wald**	**P-value**
$\Pr(\mathrm{F})_{Exploitable\ Opportunity\ (O)}$	2.972 (β_1)	0.030	27.538	0.000
$\Pr(\mathrm{F})_{Incentive/Pressure\ (I)}$	3.653 (β_2)	0.037	42.689	0.000
$\Pr(\mathrm{F})_{Integrity/Attitude/Rationalisation\ (R)}$	3.680 (β_3)	0.037	5.313	0.021
$\Pr(\mathrm{F})_{Suspicious\ Information\ (S)}$	1.692 (β_4)	0.017	0.283	0.595
Constant	-5.985 (β_0)	-0.058	11.438	0.001

As expected, all the variable coefficients are positive, indicating that higher estimated probability of fraud in each first stage model translates to higher probability of fraud in the second stage model. The output from the first stage models based on the O and I factors were statistically significant at a 1% level, while that based on the R factor was statistically significant at a 5% level. The I and R factors also had the largest magnitude of influence as shown by having the largest $e^{0.01\beta} - 1$ values in Table 6-8.

The output from the S-factor based model was not statistically significant with a p-value of almost 60%. No variable in the S-factor based model was statistically significant at a 10% level as all their p-values were greater than 35%. It is noteworthy that the comparison variables that were found not to be statistically significant (in Section 6.1.2.f) are all associated with the S factor. However, the results in Table 6-8 do not include the adjustments made in the previous section that associates V28a with the S factor.

Overall, the LR_MS_S model provides additional support for the use of the new Fraud Detection Triangle Framework, except for the S factor. However, if the LR_MS_F model is re-estimated using the entire data set, not just the training data, the S-factor based model becomes statistically significant at a 15% level. This provides preliminary empirical support for the newly proposed S factor being useful to financial statement fraud detection models. It is also possible that the S factor is becoming more useful as time progresses, because it was more statistically significant when using the entire data set that contains more recent data than the training data set.

6.1.5 Analysis of Models That Incorporate Interactions between Variables

The fact that the TreeNet (TN) model comprised trees with twelve terminal nodes in preference to fewer indicates that there are useful interactions amongst the explanatory variables because more nodes allow for more interactions. The artificial neural network and decision-tree based models used in this research automatically incorporate interactions between explanatory variables. Consequently, an analysis of the variables in these models provides insight into what variables are most important when interactions between them are possible. This analysis is presented after an explanation of the format of the results and then the results themselves in Table 6-9.

6.1.5.a Evaluating Variable Importance in Models Incorporating Interactions

Unlike the standard regression-based techniques, there aren't standardised statistical tests for determining the significance of variables. The results for the contribution of individual variables to the artificial neural network and decision-tree based models are presented in Table 6-9 below and varies between the different modelling techniques:

- The backpropagation neural network model (NN_BK) was developed using the statistics software package SPSS. Sensitivity analysis for each explanatory variable is conducted in SPSS to determine the magnitude of its influence on the output, which is then used to compute an importance score. Each importance score is then divided by the largest score to obtain a relative importance score that ranges between 0 (unimportant) and 100% (most important).
- The backpropagation neural network optimised by a genetic algorithm (NN_GA_1 and NN_GA_5) was developed using the software package Neuralyst. This package provides the list of variables that were included in the final model. It uses a process akin to the evolutionary process of natural selection whereby the variables that survive through the generations are those having the best "genetic fitness", that is, they contribute the maximum amount of information in classifying financial statements as fraudulent or

legitimate. No further information is provided on the relative importance of the explanatory variables, which is a disadvantage of genetically optimised neural networks developed in Neuralyst.

- The decision-tree based models (DT_One, TN, RF_8 and RF_66) were developed using Salford Systems' SPM software package. As was done with the NN_BK model, raw importance scores from the decision-tree based models are divided by the largest score to obtain a relative importance score that ranges between 0 and 100%. However, these raw importance scores were calculated based on how many splitting rules use each variable, what fraction of the data passes through those splitting rules and how well those splitting rules perform (Salford Systems 2012a). For the individual CART decision tree model (DT_One) there is only one tree to examine[136], but for TreeNet (TN) and Random Forests (RF_8 and RF_66) the importance scores are summed across many trees.

In Table 6-9, the relative importance scores are colour-coded for each model; A higher ranking in terms of variable importance is indicated by a colour that is more green (than yellow) followed by more yellow (than red); the greater the difference in colour, the greater the difference in the ranking.

[136] Instead of only considering variables in the tree, a feature of CART is its use of surrogates to assess the contribution of variables in the model as well as the contributions that variables would have had if they were included (as discussed in Section 3.3.4).

Table 6-9. The relative importance of each variable in the artificial neural network and decision-tree based models. The values range from 0 (unimportant) to 100% (most important), except for genetically optimised neural networks (NN_GA_5 and NN_GA_1) that only indicate whether each variable was included in the model. For each model, the most important is coloured green, the median is coloured yellow and the least important is coloured red; other values are mixtures of these colours depending on their ranking in terms of importance. Colours can be compared between models such that a higher ranking in terms of variable importance is indicated by a colour that is more green (than yellow) followed by more yellow (than red). Blank cells indicate that the variable was not used in the model.

Variable ID	**Variable Name**	**Measure of importance in the following models**						
		NN_BK	**NN_GA_5**	**NN_GA_1**	**DT_One**	**TN**	**RF_8**	**RF_66**
	Specific Account - Accounts Receivable							
V1	Accounts Receivable							
V1a	Value for the specified year	37%	included	included	66%	96%	84%	73%
V1b	Percentage change	4%	included	included	75%	86%	47%	29%
V1c	Was Percentage change > 10%?	44%	included	included				
V2	Percentage change in Accounts Receivable to Sales	2%			48%	65%	21%	11%
V3	Percentage change in Accounts Receivable to Total Assets	20%	included		50%	74%	41%	28%
	Specific Account - Allowance for doubtful accounts (AFDA)							
V4	Percentage change in AFDA	23%	included	included	54%	57%	21%	8%
V5	Percentage change in AFDA to Accounts Receivable	11%	included	included	29%	52%	19%	8%
V6	Percentage change in AFDA to Sales	8%			59%	50%	12%	5%
	Specific Account - Inventory							
V7	Change in Inventory to average Total Assets	12%			77%	57%	27%	11%
V8	Inventory to Sales							
V8a	Value for the specified year	46%	included	included	56%	93%	38%	28%
V8b	Change	12%	included		71%	67%	25%	16%
V9	Was Last-In, First-Out (LIFO) inventory valuation used?	10%			1%	6%	1%	0%
	Specific Account - Sales							
V10	Sales Growth							
V10a	Percentage change	9%	included		64%	62%	36%	12%
V10b	V10a minus the Industry Average	13%	included	included	74%	67%	38%	14%
V10c	Previous year's Percentage change	6%	included	included	43%	95%	70%	56%

Variable ID	**Variable Name**	**Measure of importance in the following models**						
		NN_BK	**NN_GA_5**	**NN_GA_1**	**DT_One**	**TN**	**RF_8**	**RF_66**
V10d	Four-year growth rate	7%	included		65%	70%	57%	28%
V10e	Previous year's percentage change in total assets	8%	included	included	62%	63%	31%	10%
V11	Sales to Total Assets							
V11a	Value for the specified year	80%	included	included	78%	97%	87%	52%
V11b	Percentage change	19%	included	included	39%	67%	24%	10%
V12	Gross Margin to Sales							
V12a	Percentage change	12%	included	included	25%	74%	35%	14%
V12b	Was percentage change > 10%?	24%						
V13	Cash Sales							
V13a	Percentage change	17%	included	included	29%	60%	38%	8%
V13b	Was change < 0?	13%	included	included				
V14	Were any sales from acquisitions?	33%	included	included		21%	4%	3%
	General Financial - Asset Composition							
V15	Current Assets to Total Assets	13%	included	included	62%	87%	30%	17%
V16	Net PP&E to Total Assets	32%	included		56%	95%	88%	82%
V17	Soft Assets to Total Assets	31%	included	included	85%	83%	50%	28%
V18	Percentage Change in Assets other than Current Assets and Net PP&E to Total Assets	13%	included	included	21%	70%	28%	17%
	General Financial - General Accrual Measures							
V19	Total Accruals to Total Assets	31%	included	included	40%	88%	100%	91%
V20	Were the specified and the prior year's Total Accruals > 0?	98%	included	included	8%	51%	25%	18%
V21	Total Discretionary Accruals	25%	included	included	34%	80%	33%	13%
V22	RSST (unadjusted) Accruals	2%	included	included	18%	66%	20%	6%
	General Financial - Level of Debt and Financial Distress							
V23	Debt to Total Assets	37%			20%	64%	34%	16%
V24	Debt to Equity	25%			19%	55%	29%	9%
V25	Altman's (1968) financial distress measure (Z-score)	4%			10%	81%	44%	30%
V26	Four-period average of Times Interest Earned	32%			37%	82%	30%	14%

Variable ID	**Variable Name**	**Measure of importance in the following models**						
		NN_BK	**NN_GA_5**	**NN_GA_1**	**DT_One**	**TN**	**RF_8**	**RF_66**
	General Financial - Performance and Profitability							
V27	Return on Equity							
V27a	Value for the specified year	37%			62%	69%	31%	17%
V27b	Industry Average minus Specific Company	2%			34%	73%	41%	22%
V28	Return on Average Prior Assets							
V28a	Value for the specified year	2%			49%	79%	34%	19%
V28b	Previous year	35%			49%	82%	51%	26%
V28c	Change	21%			33%	68%	31%	17%
V29	Holding Period Return							
V29a	One-year	3%			5%	62%	15%	9%
V29b	Previous One-year	2%			10%	73%	21%	10%
V30	Were analyst Earnings Per Share forecasts achieved or exceeded?	13%				16%	1%	0%
	General Financial - Financing							
V31	Were New Securities issued?							
V31a	Common Stock?	40%	included	included		15%	2%	0%
V31b	Common Stock or Long-term Debt?	100%	included	included	13%	40%	1%	1%
V32	Proportion of common stock that is newly issued	34%	included	included	6%	86%	91%	100%
V33	Demand for financing							
V33a	Specific Value (ex ante)	62%	included	included	17%	80%	73%	40%
V33b	Was there demand (ex ante)?	19%						
V33c	Cash from operating and investment activities	22%	included	included	44%	67%	33%	12%
V34	Were there operating leases?	29%	included	included		29%	2%	1%
	Non-financial - Key Roles and Positions							
V35	Was the auditor a Big Six firm?	27%	included	included		19%	1%	0%
V36	Number of changes of audit firm in the most recent 4 financial statements	48%	included	included	4%	20%	2%	1%
V37	CEO							
V37a	Tenure	21%	included	included	71%	88%	32%	27%
V37b	Number of changes in the last three years	3%	included		5%	15%	1%	0%

Variable ID	Variable Name	Measure of importance in the following models						
		NN_BK	NN_GA_5	NN_GA_1	DT_One	TN	RF_8	RF_66
V38	Has the CFO changed in the last three years?	20%	included			23%	1%	0%
V39	Composition/Holdings of the Board							
V39a	Number of Directors	22%			16%	73%	17%	10%
V39b	Percentage of Directors who are also Executives	84%			23%	86%	46%	38%
V39c	Percentage of Director shares owned by those who are also Executives	Excluded because of insufficient data						
V40	Percentage of total shares owned by the CEO	Excluded because of insufficient data						
	Comparing Financial and Non-financial							
V41	Percentage change in the number of Employees minus percentage change in Total Assets	3%			12%	76%	38%	19%
V42	Percentage change in Sales minus percentage change in the number of Employees	10%			18%	71%	31%	15%
V43	Percentage Change in Sales to Employees: Specific Company minus Industry Average	14%			10%	56%	21%	6%
	Control variables							
V44	Company Age: Number of years since foundation	59%			14%	88%	47%	25%
V45	Company Size: natural log of Total Assets	59%			77%	100%	78%	53%
V46	Industry: SIC code starts with a 3?	27%				22%	1%	0%
V47	Stock Exchange listed on							
V47a	NASDAQ? (One categorical variable for DT_One, TN, RF_8 and RF_66)	16%	included	included		27%	2%	1%
V47b	NYSE? (One categorical variable for DT_One, TN, RF_8 and RF_66)	8%	included	included				
	New variables							
V48	Macroeconomic indicators							
V48a	Previous year's percentage change in annual real GDP	4%	included			26%	3%	1%
V48b	Previous year's percentage change in annual retail sales	13%	included	included		33%	3%	0%
V48c	Previous year's unemployment rate inverted	12%	included		8%	28%	2%	1%
V49	Corporate governance indices							
V49a	G-Index	32%	included		100%	68%	36%	31%
V49b	E-Index	Excluded because of insufficient data						
V50	Accounting complexity of the industry	56%	included	included	18%	53%	8%	4%

6.1.5.b Analysis of Results

The results shown in Table 6-9 indicate that the variables found to be important (indicated by green shading) by the neural network models are largely different from those found by decision-tree based models. The findings are more similar between the decision-tree based models, although the TreeNet (TN) and Random Forests (RF_8 and RF_66) models are most similar in terms of their relative importance scores. The following subsections analyse the importance of the individual variables in Table 6-9 relative to their statistical significance in standard-regression based models without interactions (see Section 6.1.2).

Analysis of the most statistically significant variables from models without interactions (see Section 6.1.2)

The most statistically significant variables from Section 6.1.2 are evaluated in the following dot points in terms of their importance to models that allow interactions. All of these variables were at least moderately important (as indicated by some green in the colour coding in Table 6-9) in one of the models, except for the ratio of debt to equity (V24).

- V8a *Inventory to Sales* is particularly important in the TN model. It was also moderately important in the other models, having been included in the NN_GA models and having mostly green (with some yellow) shading for the other models;
- V11a *Sales to Total Assets* is important in all models as indicated by the green shading;
- V19 *Total Accruals to Total Assets* is very important in the ensembles of decision trees, including being the most important variable in the TN model. This variable was less important in DT_One and NN_BK, but it was also included in the NN_GA models;
- V20 *Were the specified and the prior year's Total Accruals > 0?* is very important in the neural network models, but was not in the decision-tree based models with yellow, orange and red shading;
- V24 *Debt to Equity* is relatively unimportant in all the models as indicated by red shading allowing interactions in contrast to the standard regression-based models;
- V27a *Return on Equity* is moderately important with yellow shading and some green, but it is not included in the NN_GA models;
- V28a *Return on Average Prior Assets* is not important in the neural networks, but was moderately important in the decision-tree based models with yellow and green shading. Interestingly, V28b (the value of V28a one year prior) was equally or more important in every model;

- V31b *Was New Common Stock or Long-term Debt issued?* is the most important to the neural networks, but it was relatively unimportant in the decision-tree based models;
- V33a *Demand for financing (ex ante)* is reasonably important in all models, except DT_One;
- V34 *Were there operating leases?* is not included or not important to the decision-tree based models, but was somewhat important to the neural networks;
- V39b *Percentage of Directors who are also Executives* is important in the TN and NN_BK models, moderately important in the RF models, but relatively unimportant in DT_One and not used by the NN_GA models;
- V44 *Company Age* is important in the TN and NN_BK models, reasonably important in the RF models and relatively unimportant in DT_One;
- V45 *Company Size* is very important in all models, including being the most important in the TN model;
- V49a *Corporate Governance G-Index* is the most important variable in the DT_One model, but only slightly important in the other models.

Analysis of variables that were not the most statistically significant in models without interactions (see Section 6.1.2), but are important in models with interactions

Some variables are important in models that incorporate interactions, even though they were not the most statistically significant in Section 6.1.2. The following dot points list these variables. Amongst the techniques that incorporate interactions, the decision-tree based models performed relatively well at detecting financial statement fraud (in Chapter 5), while the neural network models did not. Consequently, only variables that are important in decision-tree based models (DT_One, TN and RF models), as indicated by consistently green shading in Table 6-9, are listed. Results from ensembles of decision trees (TN and RF models) were given additional weight when determining important variables (as indicated by green shading) from Table 6-9. The reason for this is that each variable ranking in these models is aggregated over many opportunities to be used (as a result of many trees).

- V1a *Accounts receivable* is important in all models, and was most important variable measuring accounts receivable. This in contrast to the multivariate analysis without interactions (see Section 6.1.2.c) in which no accounts receivable variable is statistically significant at a 10% level;

- V10c *Previous year's Percentage change in Sales* is important in the ensembles of decision trees, in addition to its being statistically significant at a 10% level in the LR_U15% model (see Section 6.1.2.c);
- V16 *Net PP&E to Total Assets* is consistently important, particularly in the ensembles of decision trees. Other variables from the asset composition category are also relatively important, but V16 was the most important according to the ensembles of decision trees. This is in contrast to no asset composition variables being statistically significant at a 10% level in the multivariate analysis without interactions (see Section 6.1.2.d);
- V32 *Proportion of common stock that is newly issued* is very important in the ensembles of decision trees including being the most important in the RF_66 model, although it is not so important in other models. This variable was also statistically significant at a 5% level in the DA_Step model (see Section 6.1.2.d);
- V37a *CEO Tenure* is important in the decision tree (DT_One) and TreeNet models, although not so in the Random Forests models;
- V41 is the most important variable that compares financial and non-financial information, but it was only moderately important in the TreeNet model and relatively unimportant in the other models.

Other findings in models with interactions that are supported by those in models without interactions (see Section 6.1.2)

The results in Table 6-9 from models that incorporate interactions indicate the following variables are relatively unimportant to financial statement fraud detection models.

- Variables (V4-V6) measuring the allowance for doubtful debts (AFDA), although they are somewhat important in the decision tree model DT_One;
- Newly proposed variables measuring macroeconomic conditions (V48) and accounting complexity (V50), although they are somewhat important in neural network models;
- Discretionary accruals or RSST unadjusted accruals, which are less important than measures of total accruals;
- Market-based measures of performance, which are less important than accounting-based measures.

6.2 Simpler Models

Financial statement fraud detection models are widely applicable to investors, regulators, auditors and other stakeholders (as discussed in Section 2.3.5). Considering this wide range of stakeholders, it is likely that some will prefer simpler models, of course conditionally upon their being relatively accurate. The term simple applies to both models that use fewer variables and less complex modelling techniques that are easier to interpret.

Fanning and Cogger (1998) indicated that simpler models would be useful to auditors. Simpler models could also facilitate their wider use, akin to the widespread application of Altman's Z-score model for predicting financial distress and bankruptcy in companies. Altman's Z-score model is a relatively simple discriminant analysis model that uses five explanatory variables. Models with fewer variables might also be preferred by users to make the data collection and calculation easier and faster. It is also possible that models with fewer variables could have better classification accuracy. For example, the standard regression-based models already developed were more accurate when they used a reduced number of variables (see Section 6.1.1). However, it is not that simple because fewer variables could also result in poorer classification accuracy. Given that the analyses in Chapter 5 investigated accuracy without regard for complexity, it is clearly useful to develop and evaluate simpler models, which addresses research question RQ4 (defined on page 221).

6.2.1 The Role for the More Complex Models Already Developed

All the explanatory variables used in this research were selected because of their public availability and ease of access; variables too difficult to obtain are unlikely to be used in a practical context (Perols 2011). Furthermore, all models in this research could be implemented as computer software that automatically evaluated new financial statements and periodically updated the fraud detection model being used. Once developed, such automated systems could handle more complex models just as easily as simpler models. Consequently, the accuracy of a model is likely to be its key factor regardless of its complexity.

It should also be noted that some people might be more willing to rely on models with more variables because they can be perceived to be more sophisticated, as was found in an experiment with senior auditors (Boatsman et al. 1997). For such people, the more complex models developed in Chapter 5 are more appropriate than the simpler models presented in the next section.

6.2.2 Development of Simpler Models

Three modelling techniques are considered for developing simpler models: TreeNet, logistic regression and an individual CART decision tree.

Ensembles involving multiple modelling techniques were the best models overall at detecting financial statement fraud (see Section 5.3.3.b). However, these models are relatively difficult to automate because they are developed using a number of different software packages. Therefore, these models are too complex to be used for developing simpler ones. In contrast, decision tree-based models such as TreeNet are relatively easy to develop into automated systems, particular when using software produced by Salford Systems. Further to this, the TreeNet model TN was the best model overall aside from the ensembles of multiple modelling techniques (see Section 5.3.3.b). Hence, using a reduced number of variables in a TreeNet model is investigated in the next subsection.

Even with a reduced number of variables, a TreeNet model is difficult to interpret because it is still an ensemble model. Consequently, using fewer variables in less complex non-ensemble models is also investigated. Logistic regression models are also trialled because they have interpretation advantages (see Section 6.1.2) and have shown their classification accuracy can improve as a result of reducing the number of variables (see Section 6.1.1). A less complex CART decision tree is also evaluated, because a CART decision tree was the best non-ensemble model in Section 5.3.3.b.

6.2.2.a TreeNet Models with Fewer Explanatory Variables

The TreeNet model developed above (in Chapter 5) used all 66 available variables. The TN_90%_6 model was instead developed using the six variables that had a minimum importance score of 90% in the TN model. Alternative thresholds of 85% and 80% were also trialled, resulting in the use of 13 and 18 variables respectively. However, the additional (7 or 12) variables did not result in a consistently lower WEC and so the model with the fewest variables (TN_90%_6) was chosen according to the principle of parsimony.

The six variables in the TN_90%_6 model are all important as shown in Table 6-10, but they only measure the O and I factors of the new Fraud Detection Triangle framework. Adding the most important variable (according to the TN model) from each of the R and S factors was also trialled. Once again, it did not result in a consistently lower Weighted Error Cost (WEC) and so the additional variables were not included in future analysis.

Table 6-10. The relative importance of the explanatory variables included in the TN_90%_6 model.

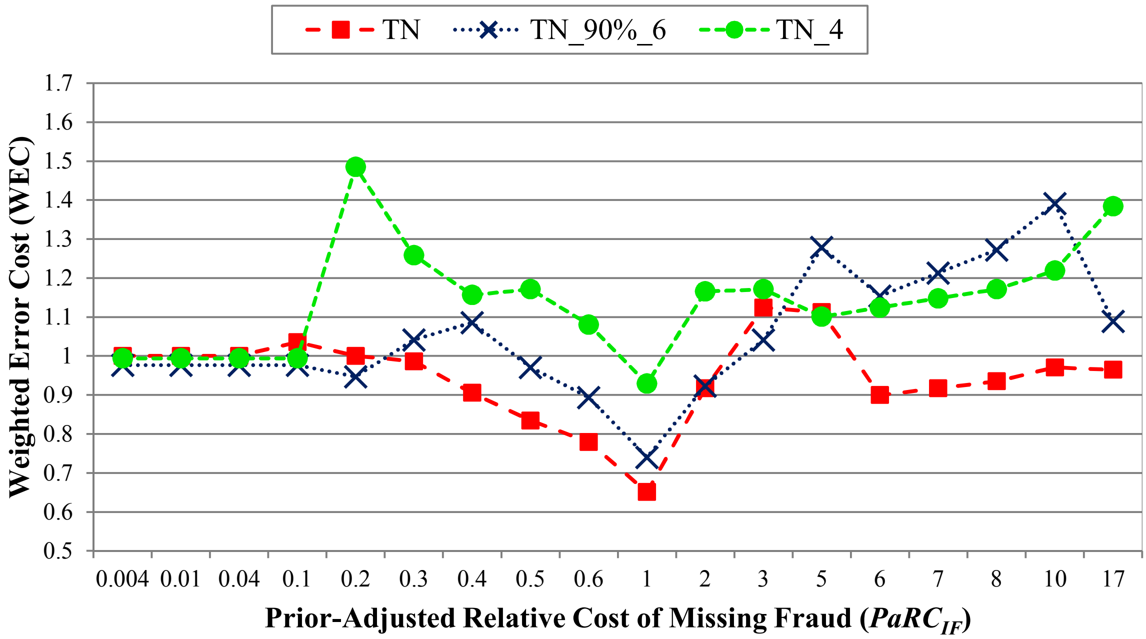
Variable ID	**Variable Name**	**Relative Importance**
V1a	Accounts Receivable	95%
V8a	Inventory to Sales	89%
V10c	Previous year's Percentage change in Sales	92%
V11a	Sales to Total Assets	97%
V16	Net PP&E to Total Assets	100%
V45	Company Size: natural log of Total Assets	92%

Another model TN_4 was also developed and evaluated. It used only four variables; the most important variable (according to the TN model) from each of the four factors (O, I, R and S) of the new Fraud Detection Triangle framework. However, Figure 6-1 below reveals that TN_90%_6 has higher classification accuracy on the holdout data as indicated by lower WEC values, except when the Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$) is greater than five[137]. Both models have WEC values greater than one when $PaRC_{IF}$ is greater than five. This indicates they are both worse than a naïve model (WEC = 1) that classifies all financial statements the same way. Consequently, as a result of greater accuracy, TN_90%_6 is chosen as the best TreeNet model that uses relatively few explanatory variables.

Compared to the TN_90%_6 model that uses six variables, the model that uses all 66 variables (TN) has a noticeably lower WEC in many cases, particularly for high values of $PaRC_{IF}$. This indicates that overall there has been a loss in accuracy from reducing the number of variables. However, the model with fewer variables has similar accuracy for some values of $PaRC_{IF}$ and better accuracy (lower WEC) when $PaRC_{IF}$ is less than 0.3 or equal to three.

[137] In addition to the results presented in the rest of this section, appendices A and B tabulate the WEC values for every model on the holdout and training data respectively, and appendices C and D tabulate the percentage accuracy of classifying fraudulent, legitimate and all financial statements on the holdout and training data respectively.

Figure 6-1. The holdout performance of the TreeNet model using all available explanatory variables (TN) with those using a reduced number of explanatory variables (TN_90%_6 and TN_4).



6.2.2.b Logistic Regression Models with Fewer Explanatory Variables

LR_Step uses only 14 variables and was the best standard regression-based model overall (see Section 5.3.2.b). Three logistic regression models were developed using even fewer variables.

LR_Step_11 Model

LR_Step_11 uses 11 of the 14 variables found in the LR_Step model. The remaining 11 variables still measure all four factors of the new Fraud Detection Triangle framework. The three variables that have been removed are[138]:

- The accrual measure V19 because it is less statistically significant (lower p-value in Table 6-2) than the other accrual measure (V20) in LR_Step;
- The demand for financing measure V33a because it is less significant (lower p-value in Table 6-2) than another financing variable that is also associated with the I factor of the Fraud Detection Triangle framework. This other variable measures whether new securities have been issued (V31b). Another financing variable (V34) that measures whether operating leases have been used is retained because it measures the R factor, rather than the I factor;

[138] Removing more variables was investigated, but the resulting classification accuracy was poor.

- The industry membership measure V46, because it was the only variable that was not statistically significant at a 5% level when V19 and V33a were removed.

LR_TN_6 Model

LR_TN_6 only uses variables with a minimum of 90% as their importance score in the TN model, as is the case for the TN_90%_6 model. TreeNet was chosen to select the variables for two reasons. First, it averages each variable's importance score over many trees (see Section 6.1.5.a). Secondly, from extensive consulting experience, the CEO of Salford Systems said[139] this often achieves results as good as, if not better than, a model performing its own variable selection.

LR_Int_11 Model

The use of interaction terms in logistic regression was investigated to develop this model, because there is evidence (see Section 6.1.5.b) to support the presence of useful interactions between explanatory variables.

The set of the 14 most statistically significant variables as listed in Section 6.1.2.i on page 242 was used as a starting point to investigate potential interaction terms. Section 6.1.5.b on page 255 identifies that some of these 14 variables are also important in models that incorporated interactions, particularly V11a, V19 and V45 (which are defined again in the next paragraph). Some variables were also identified as being important in models incorporating interactions even though they were not in the set of 14; specifically, V1a, V10c, V16, V28b, V32 and V37a (defined again below). These variables that are important in models that incorporate interactions might contribute to a logistic regression model in terms of an interaction term. Consequently, the following interactions between explanatory variables were investigated.

- CEO tenure (V37a) is thought to be associated with increased opportunity and capability to commit fraud (see Section 4.5.11). The extent of this association may increase with:
    - Increased pressure to fraudulently improve financials because of high prior sales growth (V10c);
    - Increased incentive to commit fraud from issuing new securities (V32);

[139] Personal communication during his visit to Bond University in September 2012 to present a seminar.

- Increased opportunity to commit fraud as measured by higher levels of accounts receivables (V1a) or total accruals (V19);
- Increased opportunity to commit fraud from weaker internal controls in smaller companies (V45[140]) or
- Suspiciously high returns relative to asset levels in prior periods (V28b).

- The amount that the probability of fraud increases with the incentive from a need for financing (V33a[140]) might increase with an additional pressure from low operating efficiency (V11a[140]) or increased opportunity to commit fraud (V16[140]).

These interactions were empirically tested by adding them one at a time to the set of 14 most statistically significant individual variables (see 6.1.2.i). Specifically, this was done by adding an interaction term, which is the multiplication of the two interacting variables, as well as adding the individual variables themselves if they are not already one of the original 14. None of the interaction terms were statistically significant at a 10% level, apart from the interaction between company size (V45) and CEO tenure (V37a). This model was then reduced to 11 variables using the same process as for the LR_Step_11 model that involved removing (one variable at a time) the less significant of two similar variables and removing those not statistically significant at a 5% level. A 5% level was chosen instead of 10% to promote a simpler model with fewer variables.

Comparison of Simpler Logistic Regression Models

Table 6-11 below reveals that the coefficients of variables are fairly stable across the simpler logistic regression models. The coefficients are also consistent with the results and findings from the analysis of the statistical significance of each variable (presented in Section 6.1.2), which improves the reliability of those findings. The models shown in Table 6-11 also use similar variables except for the LR_TN_6 model. This is not surprising given that the variables are selected by a different modelling technique, TreeNet.

---

[140] This statement is indicated by a lower value in this variable.

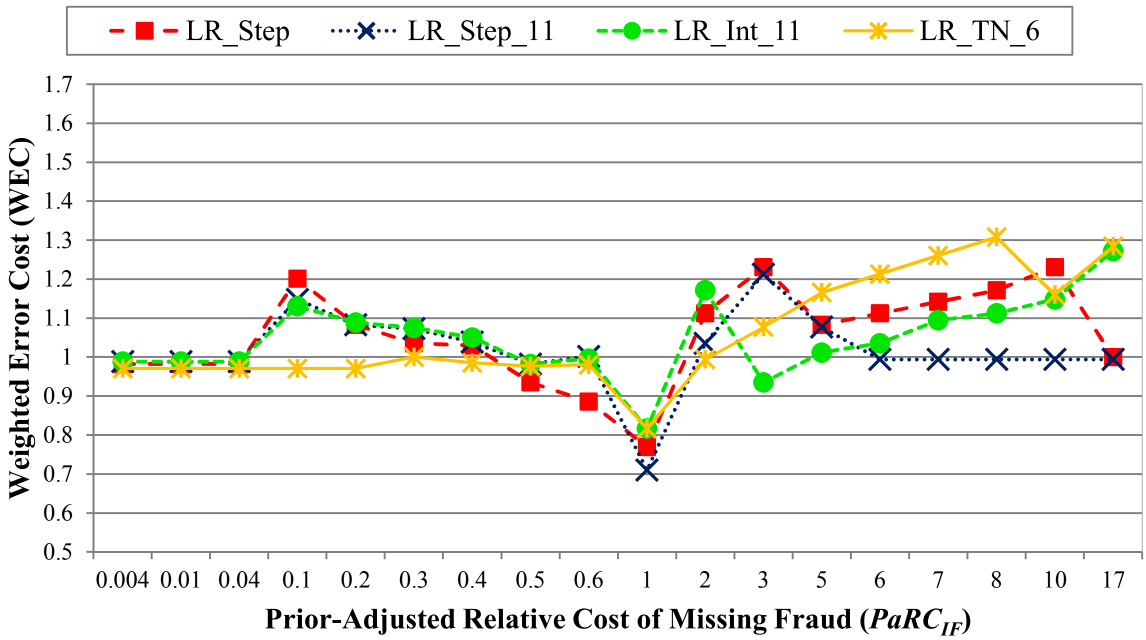
Table 6-11. The coefficients of variables in the simpler logistic regression models. All variables are statistically significant at a 10% level; * indicates at a 5% level and ** indicates at a 1% level.

Variable ID	Variance Name	Coefficients of variables in the following models			
		LR_Step	**LR_Step_11**	**LR_Int_11**	**LR_TN_6**
V1a	Accounts Receivable				0.000
V8a	Inventory to Sales	2.785**	2.421**	2.409*	2.579**
V10c	Previous year's Percentage change in Sales				0.232
V11a	Sales to Total Assets	-0.412**	-0.333**	-0.256**	-0.224*
V16	Net PP&E to Total Assets				-0.985
V19	Total Accruals to Total Assets	1.266*			
V20	Were the specified and the prior year's Total Accruals > 0?	1.564**	1.888**	1.968**	
V24	Debt to Equity	-0.083**	-0.062**	-0.054**	
V27a	Return on Equity	-0.772**	-0.551**	-0.402**	
V28a	Return on Average Prior Assets	0.886*	0.9*		
V31b	Was New Common Stock or Long-term Debt Issued?	1.383**	1.349**	1.346**	
V33a	Demand for financing (ex ante)	-0.589**			
V34	Were there operating leases?	0.737**	0.714**	0.701**	
V37a	CEO Tenure (years)			0.243**	
V39b	Percentage of Directors who are also Executives	1.394*	1.762**		
V44	Company Age: Number of years since foundation	-0.010**	-0.008**	-0.009**	
V45	Company Size: natural log of Total Assets	0.389**	0.353**	0.484**	0.083
V46	Industry: SIC code starts with a 3?	-0.486*			
V37a × V45 Interaction term				-0.030**	
Constant		-3.966	-4.383	-5.190	-0.664

All models reveal that more fraud occurs in larger firms (V45), which is probably a result of the SEC bias towards larger companies (see Section 5.1.1.a). Interestingly, the V37a × V45 interaction term in LR_Int_11 is also statistically significant at a 1% level. Although CEO tenure (V37a) was not statistically significant without this interaction term, the LR_Int_11 model indicates that longer serving CEOs are associated with increased levels of fraud. Furthermore, this effect is magnified for smaller companies as indicated by the negative coefficient of the V37a×V45 interaction term. This could mean that the CEO is more capable of exploiting the opportunity to commit fraud as a result of weaker internal controls, which would be consistent with the Exploitable Opportunity (O) factor of the new Fraud Detection Triangle framework.

The holdout performance of the LR_Step, LR_Step_11, LR_Int_11 and LR_TN_6 models is shown in Figure 6-2. Most models perform comparably apart from LR_TN_6 being the best (lowest WEC) for lower values of $PaRC_{IF}$, LR_Int_11 being the best when $PaRC_{IF}$ is equal to 3 or 5 and LR_Step 11 being the best for the highest values of $PaRC_{IF}$. The comparable performance of the LR_Int_11 model provides additional support for the use of an interaction term between variables measuring company size (V45) and CEO tenure (V37a).

Figure 6-2. The holdout performance of the logistic regression models that use relatively few explanatory variables.



A naïve model that classifies all statements the same way has a constant WEC of one (see Section 5.3.2). Consequently, models are only practically useful if they have WEC values lower than one. While LR_TN_6 is preferred in terms of the fewest variables, Figure 6-2 shows that the only time its WEC is substantially less than 1 is when $PaRC_{IF}$ is equal to one. At this point, LR_TN_6 and LR_Int_11 have the equal highest WEC, while the LR_Step model's WEC is 13% lower and the LR_Step_11 model's WEC is 8% lower again. As LR_Step_11 is simply LR_Step with three variables removed, this is further evidence that standard regression-based models can benefit from using fewer explanatory variables. Overall, the WECs of the LR_Step model and the LR_Step_11 model are comparable for most values. Therefore, in accordance with the principle of parsimony, the LR_Step_11 is chosen as the best logistic regression model that uses relatively few explanatory variables.

6.2.2.c A CART Decision Tree with Reduced Complexity

The individual CART decision tree model DT_One included 56 variables and 70 terminal nodes. A smaller tree was developed by further pruning the DT_One tree until it had only nine terminal nodes. The resulting tree used seven explanatory variables. However, the only time its holdout WEC was substantially lower than one was when $PaRC_{IF}$ is equal to one. Furthermore, at this point its WEC is 0.78, which is 10% higher than LR_Step_11 (0.71). Thus, the decision tree did not respond well to using fewer variables, and consequently the LR_Step_11 model is preferred.

6.2.3 Analysis of the Simpler Models

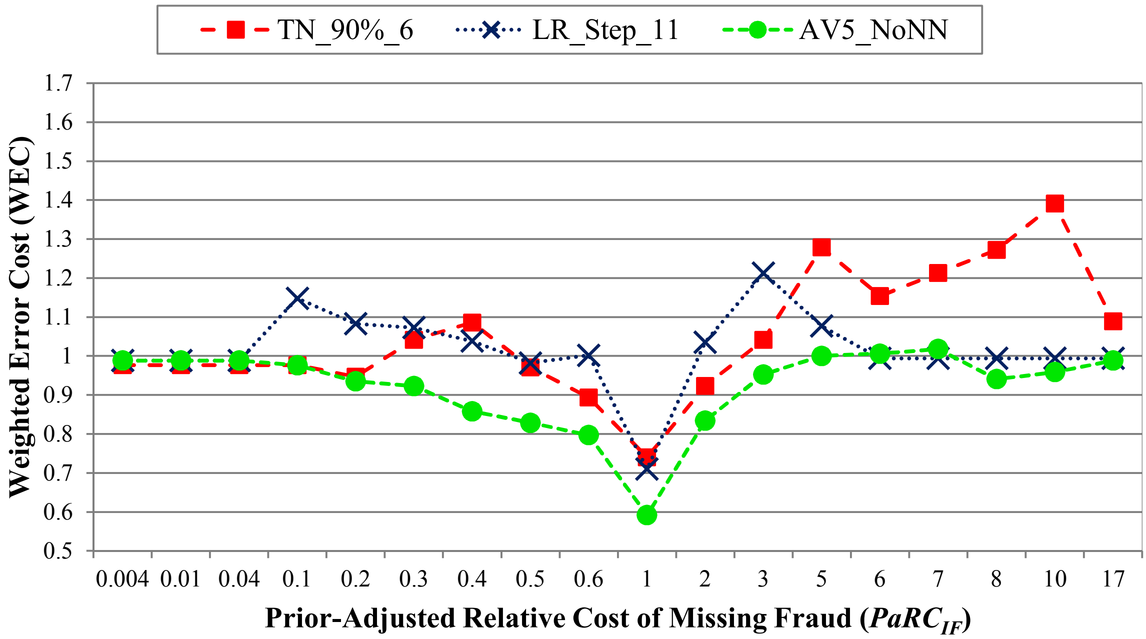
The TN_90%_6 and LR_Step_11 models were chosen in the previous section as the best simpler models. In this section, they are analysed in terms of their performance on the holdout dataset, followed by an analysis of the explanatory variables they use.

6.2.3.a Evaluation of Holdout Performance

Table 6-3 below compares the holdout performance of these simpler models, TN_90%_6 and LR_Step_11, with AV5_NoNN, the best model overall in Chapter 5 (see Section 5.3.3.b). It is clear that the more complex ensemble of multiple modelling techniques, AV5_NoNN, has better accuracy as indicated by lower WEC. However, as mentioned the two simpler models have advantages in terms of easier implementation and interpretability.

When comparing TN_90%_6 and LR_Step_11 it is clear that the former has lower WECs for lower values of $PaRC_{IF}$, while the latter has substantially lower WECs for the highest values of $PaRC_{IF}$. However, as presented in Chapter 5, it is again clear that models are better (have lower WEC) at detecting financial statement fraud when $PaRC_{IF}$ is close to one.

Figure 6-3. A comparison of holdout performance of the best simpler models (LR_Step_11 and TN_90%_6) and the best model overall (AV5_NoNN).



The WEC of TN_90%_6 is noticeably lower than one when $PaRC_{IF}$ is equal to 0.6, 1 or 2. This only occurs when $PaRC_{IF}$ is equal to 1 for LR_Step_11, but in that case it has a lower WEC than TN_90%_6. Because models are not practically useful when WEC is equal to or greater than one (see Section 5.3.2),

- The TN_90%_6 model is best only used when $PaRC_{IF}$ is between 0.6 and 2; while,
- The LR_Step_11 model is best only used when $PaRC_{IF}$ is equal to or close to one.

Thus, the TreeNet model that only uses six variables (TN_90%_6) is useful for a wider range of values of $PaRC_{IF}$ than LR_Step_11. Importantly, both models are substantially more accurate than the benchmark models, F-score and M-score as shown in Table 6-12 below. These models were only compared to the benchmark models when $PaRC_{IF}$ is equal to one, because that was the only case in which the benchmark models were useful with a WEC lower than one.

Table 6-12. The accuracy of the chosen simpler models (LR_Step_11 and TN_90%_6) on the holdout data set when $PaRC_{IF} = 1$. The best model overall from Chapter 5 (AV5_NoNN) and the benchmark models (M-score and F-score) are provided as a comparison.

Model	**Percentage Accuracy when $PaRC_{IF} = 1$**
AV5_NoNN	70%
LR_Step_11	64%
TN_90%_6	63%
M-score	53%
F-score	52%

TreeNet models are more complex and cannot be interpreted as easily because they are ensemble models. Consequently, TN_90%_6 might be well-suited for a context where the use of a small number of variables is strongly preferred. However, if interpretability is more important then the LR_Step_11 model is the best suited. In either case, the limitations on their use as listed in the dot points above should be acknowledged. If, however, classification accuracy is a more important consideration than complexity, then the AV5_NoNN model remains the best overall. Alternatively, if the value of $PaRC_{IF}$ can be reasonably estimated then the best model for that value of $PaRC_{IF}$ should be used, as per Table 5-23 on page 207.

Models are designed to be used for classifying or ranking financial statements

Both the TreeNet and logistic regression models have the probability of fraud as an output. However, interpreting this figure as a standard probability estimate is unreliable. The reason is that models have been evaluated and chosen based on their classifications, specifically those that result in the lowest WEC. Furthermore, it has been shown that accurate classification does not necessarily imply unbiased models with precise probability estimates (Friedman 1997; Hand 2009a). It has also been claimed that one-to-one matched pairs studies, as used in this research, result in biased probability estimates (Skogsvik 2005). However, this bias does not negatively influence the ranking of financial statements according to their probability of fraud. Consequently, these, and all other models presented in this research, are best suited for classifying fraudulent statements or ranking them according to their probability of being fraudulent.

6.2.3.b Analysis of Explanatory Variables Used in the Simpler Models

TN_90%_6 uses the same variables as the LR_TN_6 model. Hence, Table 6-11 above can also be used to compare the variables used in the LR_Step_11 and TN_90%_6 models. These two models use very different variables. The TN_90%_6 model is primarily made up of financial variables measuring specific accounts, while the LR_Step_11 model uses mostly general financial variables. LR_Step_11 also includes a non-financial measure unlike TN_90%_6. While the TN_90%_6 model only measures the O and I factors of the new Fraud Detection Triangle framework, the LR_Step_11 model includes variables that also measure the other factors, R and S.

There is a similarity between these models in terms of exogenous and temporal variables that were suggested as potentially useful (in Sections 4.1.1 and 4.1.2). 17-18% of the variables used in these models incorporated information from different time periods, which supports the inclusion of variables incorporating the additional information provided by measuring multiple time periods. However, neither model used any variable with exogenous information.

Further Analysis of the LR_Step_11 Model

One advantage of the LR_Step_11 model, which is defined below, is the simpler interpretation of the model compared with TN_90%_6.

$$\text{Log\_odds} = -4.38 + 2.42\text{V8a} - 0.33\text{V11a} + 1.89\text{V20} - 0.06\text{V24} - 0.55\text{V27a} + 0.9\text{V28a} + 1.35\text{V31b} + 0.71\text{V34} + 1.76\text{V39b} - 0.01\text{V44} + 0.35\text{V45}$$

$$\text{and the Probability of Fraud} = \frac{e^{\text{Log\_odds}_{LR\_Step\_11}}}{1 + e^{\text{Log\_odds}_{LR\_Step\_11}}}$$

According to the LR_Step_11 model the probability of fraud increases with:

- Increased opportunities to commit fraud as measured by higher levels of inventory (V8a Inventory to Sales) and more directors also being executives (V39b);
- Increased opportunity and increased pressure from high levels of current and prior accruals (V20);
- Increased incentives facing companies that have just issued new securities (V31b) and facing younger firms around the period of the initial stock offering (lower values of V44 Company Age);
- Managers who focus on the short-term, which might make it easier to rationalise fraud (V34 Were operating leases used?);
- Suspiciously higher returns relative to asset levels in prior periods (V28a Return on Average Prior Assets);
- Larger firms (V45 Company Size), which probably reflects an SEC bias in prosecution;

And decreases with:

- Reduced pressure from increases in efficiency of generating sales from assets (V11a Sales to Assets) and increases in accounting profitability (V27a Return on Equity); and
- Reduced opportunity from increases in monitoring from creditors (V24 Debt to Equity).

Testing the Influence of Major Events That Occurred During the Study Period

Some noteworthy fraud-related events occurred during the time period of the study (1998-2007). These events could be thought to have substantially influenced the results in this research, but the following paragraphs show that this was not the case.

The first notable event was the introduction of the Sarbanes-Oxley Act of 2002, which enforced major changes to corporate governance and financial practice requirements in US organizations. One outcome from the Act was hoped to be a reduction in the amount of financial statement fraud. Some data do not support the effectiveness of Sarbanes-Oxley (Deloitte 2008; Hogan et al. 2008; Deloitte 2009). However, even if it has not been effective at reducing the quantity of financial statement fraud, it is possible that the nature (and consequently the indicators) of fraud might have changed. This proposition was tested in this research by including a dummy variable indicating before or after Sarbanes-Oxley. This test was performed by Wang et al. (2011), who did not find it statistically significant in a logistic regression model. Consistent with this finding, the Sarbanes-Oxley dummy variable is not statistically significant even at a 15% level with a p-value of 19.6% in the LR_Step_11 model using all the data. The full data set was used, because there are no data after 2002 in the training data set. This finding is supported by the TN_90%_6 model, because the Sarbanes-Oxley dummy variable has an importance score of less than 25%, while the other variables' scores are all greater than 90%.

The second notable event was the large number of frauds from 1998-2000 related to the internet company (dot-com) crash (Dechow et al. 2011). As a result, a dummy variable set to one only during 1998-2000 was used to test whether this period is fundamentally different and had unduly influenced results in this study. The dot-com dummy variable is not statistically significant even at a 15% level with a p-value of 40% in the LR_Step_11 model using all the data. The p-value was even higher at 73% if only the training data set was used. This finding is again supported by the TN_90%_6 model, because the dot-com dummy variable has an importance score of less than 27%, while the other variables' scores are all greater than 90%.

It is also noteworthy that all the variable coefficients in the LR_Step_11 model remained similar and statistically significant at a 10% level during all of these robustness tests. This finding increases the reliability of the interpretation of the model's variable coefficients discussed in this chapter.

6.3 Summary

The analysis of variables presented in this chapter used a more comprehensive set of explanatory variables than prior studies. The number of fraud cases analysed is often more than double that which is used in prior studies. In addition, both standard regression-based models and models that allow complex interactions between explanatory variables are analysed. Importantly, the main findings are also consistent across multiple models. Consequently, the findings greatly contribute to a better understanding of the important variables in models that detect financial statement fraud and also of the way these variables are associated with fraud.

There is empirical evidence to support the use of the new Fraud Detection Triangle framework to assist in the selection of variables for models to detect financial statement fraud. Variables from all factors of the framework have been useful to such models, particularly the Exploitable Opportunity (O) and Incentive/Pressure (I) factors. Additional research into variables that measure the Integrity/Attitude/Rationalisation (R) and Suspicious Information (S) factors would be beneficial as less focus has been placed on them in prior research.

Some notable findings relating to individual variables are listed below.

- Contrary to findings in prior studies, higher debt levels (relative to equity) reduce fraud. This is likely to be the result of a decrease in the opportunity to commit and conceal fraud because of increased monitoring and scrutiny from creditors. This new empirical finding is consistent with Jensen and Meckling's (1976) well-established agency theory.
- Weaker corporate governance as measured by a corporate governance index (G-index) is associated with more financial statement fraud. This is the first study to provide empirical support for the hypothesis that financial statement fraud is related to the aggregation of many corporate governance initiatives, not just the presence or absence of a particular one. This is also consistent with the outcome of a large fraud case at Tyco International

Limited, where improving overall corporate governance was made the number one priority after the fraud was uncovered (Farber 2005);

- Longer serving CEOs are associated with increased levels of fraud. Further to this, the effect is magnified for smaller companies as indicated by a statistically significant interaction with company size. This possibly indicates that the CEO is more capable of exploiting any opportunities to commit fraud that result from weaker internal controls in smaller companies, which would be consistent with the O factor of the new Fraud Detection Triangle framework. This is the first study to propose or test this new interaction variable;
- Larger firms are associated with increased levels of fraud, which empirically supports the theory that the US Securities Exchange Commission (SEC) has a bias to prosecute larger firms (as suggested in Section 5.1.1.a);
- While mixed results have occurred in prior studies, measures of total accruals were noticeably more important than measures of discretionary or unadjusted accruals; and
- Accounting-based measures of performance are more useful to fraud detection models than market-based measures.

Overall, variables from a variety of categories were found to be useful. This includes both financial and non-financial variables, but variables comparing financial and non-financial measures were not found to be useful to financial statement fraud detection models. Future research could investigate variables that compare financial information other than sales (or assets) with non-financial information other than the number of employees.

Additional findings that are useful when developing financial statement fraud detection models include:

- The variables that are most statistically significant in standard regression-based models are general financial measures, particularly those relating to financing. In contrast, the decision trees and ensembles of them that incorporate interactions find more financial measures of specific accounts (accounts receivable, sales and inventory) to be important;
- Standard regression-based techniques benefit from having fewer explanatory variables, but decision trees and ensembles of them benefit from having a larger number of variables. This is probably due to the fact that decision-tree based models do not suffer from multicollinearity problems; and

- The relative ranking of variable importance provided by TreeNet has shown promise for selecting the explanatory variables to be used by a different modelling technique.

Less complex fraud detection models were also developed and evaluated. The two best models were a TreeNet model that uses only six variables, TN_90%_6, and a logistic regression that uses 11 variables, LR_Step_11. TN_90%_6 might be well-suited for a situation where collected data on only a small number of variables is strongly preferred. However, TreeNet models are relatively difficult to interpret because they are ensemble models. Consequently, when interpretability is more important, the LR_Step_11 model is the better-suited. Both models only achieved useful levels of classification accuracy when the Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$) is close to one, particularly the LR_Step_11 model, and so their use should be limited to those circumstances.

Finally, if model complexity is not as important as accuracy then the AV5_NoNN model developed in Chapter 5 remains the best overall. Alternatively, if the value of $PaRC_{IF}$ can be reasonably estimated then the best model for that value of $PaRC_{IF}$ should be used, as per Table 5-23 on page 207.

Chapter 7 Overall Conclusions and Future Work

Fraud is ubiquitous (ACFE 2014) and financial statement fraud in particular continues to be extremely costly to society. Financial statement fraud is difficult to detect and better decision aids are needed to assist in making its detection more effective. Fraud detection models are an example of such an aid, and have been investigated in this research. Beneficiaries of statistically reliable fraud detection models include:

- Those that have a responsibility to detect or deter fraud including auditors, regulators, company management, the board of directors and its audit committee. Detection models can assist all of them in deciding whether to investigate a certain set of financial statements further.
- Victim entities including investors, financiers, employees, customers, suppliers, analysts and other stakeholders. Detection models can assist all of them in trying to avoid association with fraudulent companies.

The major conclusions and contributions of this research are presented next, followed by suggestions for future work.

7.1 Conclusions and Contributions of This Research

Consistent with the aims of this project, this research has advanced the field of detecting financial statement fraud in terms of a better understanding of the:

- Most appropriate modelling techniques and their best parameter settings;
- Suitability and performance of different detection models, both new and existing; and,

- Explanatory variables that are important for use in detection models and how these variables are associated with fraud.

New theory and corresponding supporting empirical evidence were also presented. This theory can assist in the selection of explanatory variables for future research. The contribution of this research is strengthened because:

- It is the first study to evaluate models both on separate holdout data for accurate evaluation of model performance and under varying assumptions about prior fraud probabilities and ratios of misclassification error costs;
- Models are evaluated on holdout data that occur chronologically after the training data used to develop them, which results in more realistic estimates of model performance;
- The number of fraud cases used in this research is substantially greater than in most prior studies, and is often more than double that used previously;
- The set of explanatory variables used in this research is more comprehensive than any prior studies;
- More than 35 different models have been compared, after initially analysing many more in order to choose the parameters for each model; and,
- All models have been developed and tested on the same data so that comparisons between them are valid. Additional benchmark models have also been included for comparison; these models are used exactly as they were developed in their original studies.

The following subsections discuss the main findings according to the four main research questions driving this research.

7.1.1 Research Question 1 (RQ1)

RQ1: *Which supervised-learning modelling techniques are the most accurate at detecting financial statement fraud under varying assumptions about the prior probability of fraud and ratios of misclassification error costs?*

The best model to detect financial statement fraud, as indicated by the lowest Weighted Error Cost (WEC), varies depending on the assumptions about prior fraud probabilities and ratios of misclassification error costs. This is illustrated in the number of different model codes in the last column of Figure 7-1. The fact that no model is superior under all assumptions illustrates the need for research to consider a variety of prior probabilities and ratio of error costs, as has been done in this research.

The cost of missing fraud relative to falsely alleging it differs depending on the circumstance and the stakeholder. Beneish (1999a) estimated that the loss associated with investing in a company that produces fraudulent financial statements is 20 to 30 times higher than the lost opportunity to invest in a company because it was falsely accused of being fraudulent. Using this estimate, the ratio of the cost of missing fraud to the cost of falsely alleging it (C_{IF}:C_{IL}) is between 20:1 and 30:1 for investors. As a hypothetical case, consider an investor who faces a ratio of error costs of 20:1 and who agrees with recent estimates that the prior probability of fraud [$p(F)$] is 14.5% (Dyck et al. 2013). As indicated in Figure 7-1, the smaller table would be used to calculate the corresponding value of the prior-adjusted relative cost of missing fraud ($\mathrm{PaRC_{IF}}$), which is equal to three. Alternatively, the formula $PaRC_{IF} = \frac{C_{IF}}{C_{IL}} \times \frac{p(F)}{100\% - p(F)}$ could be used. This value would then be used in the second table in Figure 7-1 to select the best model (as indicated by the lowest WEC) for this hypothetical investor to use, which is RF_66. The RF_66 model is a Random Forests model that uses 1000 trees and all of the available variables.

Figure 7-1. An illustration of how to select the model with the lowest Weighted Error Cost (WEC) for a given ratio of error costs (C_{IF}:C_{IL}) and prior probability of fraud [$p(F)$]. This figure is an amalgamation of Tables 5-23 and 5-24.

	Prior probability of fraud [$p(F)$]			
Ratio of Error Costs (C_{IF}:C_{IL}) (missing fraud: falsely accusing fraud)	**0.40%**	**1.00%**	**5.50%**	**14.50%**
1:1	0.004	0.01	0.1	0.2
10:1	0.04	0.1	0.6	2
20:1	0.1	0.2	1	3
30:1	0.1	0.3	2	5
40:1	0.2	0.4	2	7
50:1	0.2	0.5	3	8
60:1	0.2	0.6	3	10
100:1	0.4	1	6	17

	Best Model(s) with the Lowest WEC **F/L**: Accuracy at Detecting Fraud/Legitimate			
$PaRC_{IF}$	**WEC**	**F**	**L**	**Model Code**
0.004	0.964	4%	100%	DA_U15%
0.01	0.964	4%	100%	DA_U15%
0.04	0.964	4%	100%	DA_U15%
0.1	0.970	3%	100%	LR_U15%
0.2	0.935	9%	99%	AV5_NoNN
0.3	0.913	28%	94%	RF_66
0.4	0.805	28%	96%	AV3_RF_TN_DT
0.5	0.763	50%	87%	Vote3_RF_TN_DT
0.6	0.720	50%	87%	Vote3_RF_TN_DT
1	0.580	66%	76%	Vote3_RF_TN_DT
2	0.834	85%	46%	AV5_NoNN
3	0.905	91%	36%	RF_66
5	1.000	100% 100% 100% 98%	0% 0% 0% 12%	M-score / F-score NN_GA_1 DT_One / DT_One_DA AV5_NoNN
6	0.899	98%	21%	TN Vote3_RF_TN_DT
7	0.917	98%	21%	TN
8	0.935	98%	21%	TN
10	0.876	100%	12%	RF_66
17	0.876	100%	12%	RF_66

While their use in prior studies is limited, ensembles of decision trees were the best model on numerous occasions in this research, particularly for high values of $PaRC_{IF}$. This is shown in the larger table in Figure 7-1 by the occurrence of model codes starting with RF (Random Forests) or TN (TreeNet). As indicated by the occurrence of model codes beginning with Vote or AV, newly developed ensembles of multiple modelling techniques including both Random Forests and TreeNet are also the best models on numerous occasions. This provides strong empirical support for the use of ensembles of decision trees on their own, as well as in ensembles with other modelling techniques.

As illustrated in Table 7-1, using the best model for each value of $PaRC_{IF}$ is an improvement over using either of the benchmark models (M-score or F-score). The holdout WEC of the best model is lower than both benchmark models on every occasion with the exception of the WEC being equal when $PaRC_{IF}$ is equal to five.

Table 7-1. The holdout performance of the best model for each value of $PaRC_{IF}$ compared to the benchmark models.

$PaRC_{IF}$	Percentage improvement (decrease) in WEC of the best model compared to	
	M-score	**F-score**
0.004	61% lower WEC	4% lower WEC
0.01	39% lower WEC	4% lower WEC
0.04	16% lower WEC	4% lower WEC
0.1	8% lower WEC	3% lower WEC
0.2	9% lower WEC	7% lower WEC
0.3	10% lower WEC	9% lower WEC
0.4	21% lower WEC	20% lower WEC
0.5	25% lower WEC	29% lower WEC
0.6	30% lower WEC	31% lower WEC
1	38% lower WEC	40% lower WEC
2	23% lower WEC	17% lower WEC
3	9% lower WEC	13% lower WEC
5	Equal WEC	Equal WEC
6	10% lower WEC	10% lower WEC
7	8% lower WEC	8% lower WEC
8	7% lower WEC	7% lower WEC
10	12% lower WEC	12% lower WEC
17	12% lower WEC	12% lower WEC

7.1.2 Research Question 2 (RQ2)

RQ2: *Which supervised-learning modelling technique is the best overall at detecting financial statement fraud, considering the entire range of assumptions investigated in RQ1?*

When the ratio of error costs and prior probability of fraud cannot be reliably estimated, $PaRC_{IF}$ is unknown. In this case, the approach outlined above (under RQ1) does not work and the AV5_NoNN model should be used, because it is the best model considering all investigated values of $PaRC_{IF}$. This newly developed ensemble model averages the probability outputs of five other models: a Random Forests, TreeNet, CART decision tree, discriminant analysis and a logistic regression model. This is the first study of financial statement fraud to create an ensemble of multiple modelling techniques that incorporate TreeNet or Random Forests[141].

Overall, the AV5_NoNN model is a considerable improvement on the benchmark models (M-score and F-score) with lower WECs in almost all cases, as shown in Table 7-2. Nonetheless, when the value of $PaRC_{IF}$ can be reasonably estimated then the best model for that particular value (see Figure 7-1 above) should be used in preference to AV5_NoNN as it often results in even lower WECs.

Table 7-2. The holdout performance of the AV5_NoNN model compared to the benchmark models.

	Percentage improvement (decrease) in WEC of the best model compared to	
$\boldsymbol{PaRC_{IF}}$	**M-score**	**F-score**
0.004	60% lower WEC	1% lower WEC
0.01	38% lower WEC	1% lower WEC
0.04	14% lower WEC	1% lower WEC
0.1	8% lower WEC	2% lower WEC
0.2	9% lower WEC	7% lower WEC
0.3	9% lower WEC	8% lower WEC
0.4	15% lower WEC	14% lower WEC
0.5	18% lower WEC	23% lower WEC
0.6	22% lower WEC	24% lower WEC
1	37% lower WEC	39% lower WEC
2	23% lower WEC	17% lower WEC
3	5% lower WEC	9% lower WEC
5	Equal WEC	Equal WEC
6	1% higher WEC	1% higher WEC
7	2% higher WEC	2% higher WEC
8	6% lower WEC	6% lower WEC
10	4% lower WEC	4% lower WEC
17	1% lower WEC	1% lower WEC

[141] Perols (2011) created an ensemble that included a model that used bagging and See4.5 decision trees, which is similar to, but not the same as, Random Forests (see Section 3.3.7).

Research from RQ1 and RQ2 also yielded additional findings, the most notable being:

- Decision trees were the best individual modelling technique, when compared with discriminant analysis, logistic regression and artificial neural networks. This is in contrast to research by Perols (2011) who found logistic regression to be superior. However, unlike Perols, this study used a data set with more than 100 fraud cases and holdout data that occurred chronologically after the training data. This research was also the first to evaluate decision trees using data comprising more than 100 cases of fraud and the first to use CART decision trees with US data;
- The AV5_NoNN model benefited from excluding neural networks, both standard backpropagation and genetically optimised networks. This is consistent with the fact that neural networks were found to be the worst modelling technique in this research, which was the first to evaluate neural networks over numerous ratios of error costs using a relatively large holdout data set that is chronologically after the training data set. The neural networks did successfully identify patterns in the training data, but those patterns did not consistently persist into the holdout data;
- Ensemble models were better at detecting financial statement fraud than individual models. Furthermore, combining multiple modelling techniques through majority voting[142] or averaging model outputs resulted in the best models overall. The decision tree ensemble TreeNet (that uses stochastic gradient boosting) is the next best model. In contrast to its sole usage in previous research (Whiting et al. 2012), this modelling technique benefited from using more than 1000 trees and a faster learning rate (0.01 compared to 0.001). All of these ensembles were superior to models that included the results from one model as an explanatory variable in another model, such as those proposed by McKee (2009) who only tested them on data comprising 50 fraud cases;
- The relatively poor performance of the two benchmark models (M-score and F-score), developed using older data, supports the importance of regularly updating fraud detection models over time instead of using static models. This is expected because both business and fraudsters change over time.

[142] As an example of majority voting, if three models classified a financial statement as fraudulent and two classified it as legitimate then the final classification would be fraudulent because three is greater than two.

7.1.3 Research Question 3 (RQ3)

RQ3: *Which explanatory variables are the most useful in models that detect financial statement fraud?*

An aim of this research is to produce findings that are widely applicable to investors, regulators, auditors and other stakeholders. Consequently, all explanatory variables used in this research are publicly available, because variables too difficult to obtain are unlikely to be used in a practical context (Perols 2011). Overall, variables from a variety of categories were found to be useful to financial statement fraud detection models. This includes both financial and non-financial variables.

The selection of variables is crucial to developing fraud detection models, but their selection in prior financial statement fraud detection research is not standardised by an overall theory (Perols and Lougee 2011). Consequently, a new theoretical framework has been developed for this purpose, which is illustrated in Figure 7-2 and based on Cressey's (1953) famous Fraud Triangle. This new Fraud Detection Triangle framework indicates that the probability of financial statements being fraudulent increases with either (or both)

- the presence of any of the precursors to fraud: Exploitable Opportunity (O factor), Pressure/Incentive (I factor) or Integrity/Attitude/Rationalisation (R factor);
- the presence of Suspicious Information (S factor) that might have occurred as a consequence of fraud (even though it is not a precursor to fraud).

Variables from all factors of the new framework have been useful to such financial statement fraud detection models, particularly the O and I factors as shown in Table 7-3. In addition, the following variables were very important in terms of complex interactions with other variables: level of accounts receivable (O factor), net property, plant and equipment to total assets (O factor), previous year's percentage change in sales (O and I factor) and the proportion of common stock that is newly issued (I factor). Additional research into variables that measure the R and S factors would be beneficial as less focus has been placed on them in prior research.

Figure 7-2. (Figure 4-3 reproduced.) The new Fraud Detection Triangle framework[143].

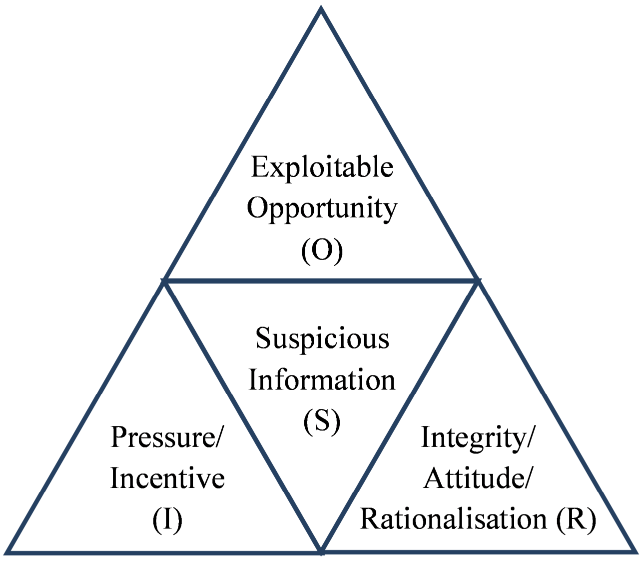


Table 7-3. The most statistically significant variables according to the factors of the new Fraud Detection Triangle framework. The probability of financial statements being fraudulent increases with company size, but this is likely the result of the US Securities Exchange Commission's prosecution bias towards larger companies. * indicates a variable that is associated with both the O and I factors. **Bold** indicates a variable with new empirical support.

	The probability of financial statements being fraudulent increases with	
	Higher levels of…	**Lower levels of…**
O Factor	$\frac{\text{Inventory}}{\text{Sales}}$ $\frac{\text{Total Accruals}}{\text{Total Assets}}$ Percentage of directors who are also executives Positive total accruals (current & prior)* **CEO tenure, at a larger rate for smaller companies**	$\frac{\text{Debt}}{\text{Equity}}$ **Overall corporate governance** (indicated by larger G-Index values)
I Factor	New equity or long-term debt having been issued Demand for financing (ex ante)	$\frac{\text{Net Income}}{\text{Total Common Equity}}$ $\frac{\text{Sales}}{\text{Total Assets}}$ Company age
R Factor	The use of operating leases	
S Factor	$\frac{\text{Net Income}}{\text{Average Prior Assets}}$	

[143] It is acknowledged that the layout of this diagram is very similar to the model presented by Kassem and Higson (2012).

Table 7-3 also illustrates the following main findings relating to individual variables:

- Contrary to findings in prior studies, higher debt levels (relative to equity) were associated with fewer frauds. This is likely to be the result of a decrease in the opportunity to commit and conceal fraud because of increased monitoring and scrutiny from creditors. This new empirical finding is consistent with Jensen and Meckling's (1976) well-established agency theory.
- Weaker corporate governance, as measured by larger values of the corporate governance G-index, is associated with more fraud. This is the first study to provide empirical support for the hypothesis that financial statement fraud is related to the aggregation of many corporate governance initiatives, not just the presence or absence of a particular one;
- Longer serving CEOs are associated with increased levels of fraud and this effect is magnified for smaller companies. This possibly indicates that the CEO is more capable of any exploiting opportunities to commit fraud that result from weaker internal controls in smaller companies. This is the first study to propose or test this new interaction variable;
- While mixed results have occurred in prior studies, measures of total accruals were noticeably more important than a measure of discretionary or unadjusted accruals. Future research could investigate this further by trialling other measures of accruals such as discretionary accruals that include industry averaged measures; and
- Accounting-based measures of performance are more useful to fraud detection models than market-based measures.

7.1.4 Research Question 4 (RQ4)

RQ4: *How do simpler models compare with those developed for the first two research questions in their ability to detect financial statement fraud?*

Considering the wide range of stakeholders, it is likely that some have a preference for simpler models provided they are still relatively accurate. For these stakeholders, simpler models have been developed that are easier to interpret and that use fewer explanatory variables. In terms of the fewest variables, a TreeNet model that uses six variables (TN_90%_6) was the best. In terms of the model easiest to interpret, a logistic regression model that uses 11 variables (LR_Step_11) was the best. The LR_Step_11 model was developed by starting with the stepwise logistic regression model with 14 variables and removing three variables that were theoretically similar according to the new Fraud Detection Triangle framework. Interestingly, reducing the number of variables using the new framework improved the performance of logistic regression, which is further empirical support for the use of it when selecting explanatory variables. Overall, standard regression-based techniques benefited from having fewer explanatory variables, but decision trees and ensembles of them benefit from having a larger number of variables. This is probably because decision-tree based models do not suffer from multicollinearity problems.

Table 7-4. (Table 6-12 reproduced.) A comparison of the holdout accuracy of the simpler models (LR_Step_11 and TN_90%_6), the best model overall (AV5_NoNN) and the benchmark models (M-score and F-score) when $PaRC_{IF} = 1$.

Model	**Percentage Accuracy when $PaRC_{IF} = 1$**
AV5_NoNN	70%
LR_Step_11	64%
TN_90%_6	63%
M-score	53%
F-score	52%

The results presented in Table 7-4 above show that the simpler models TN_90%_6 and LR_Step_11 are at least 10% more accurate than the benchmark models when $PaRC_{IF}$ is equal to one. However, this same table also reveals that they are less accurate than the AV5_NoNN model. Furthermore, to avoid low levels of classification accuracy and high WECs,

- The TN_90%_6 model is most appropriately used when $PaRC_{IF}$ is between 0.6 and 2; and,
- The LR_Step_11 model is most appropriately used when $PaRC_{IF}$ is equal to or close to one.

Thus, the simpler models are suitable for people who prefer simplicity with the trade-off of slightly reduced classification accuracy, so long as they can estimate their value of $PaRC_{IF}$ to be within the ranges shown above. However, they are not appropriate for situations where classification accuracy with the lowest WEC is preferred.

It is important to note that even the most complex ensemble models developed in this research can be automated. While not precisely recorded, all of the models with the lowest WEC presented above in Figure 7-1 took less than one hour to run using a computer with an Intel® Core™ i7 2.50GHz processor and 8MB of RAM, once the parameters were chosen. This is substantially faster than waiting for the SEC to formally allege and prosecute fraud, which has historically taken several years. Given that fraud schemes become more costly the longer they go undetected, models that speed up the process would assist in mitigating the costs of financial statement fraud. According to a blog post published by the Association of Certified Fraud Examiners, when a fraud occurs the best thing that can be done is to detect it as quickly as possible. (Warren 2012).

7.2 Future Work

The ultimate goal is to have models that are able to detect financial statement fraud with perfect accuracy. Since this goal is realistically unattainable there is always scope for additional research into better models to detect financial statement fraud. The results presented in this research can be used as benchmarks for future research.

Some specific suggestions follow for future research that is related to the research presented in this dissertation. In addition, suggestions have also been made in other sections of the dissertation.

7.2.1 Extending This Research with New Data

This research uses publicly available information that is relatively easy to obtain because its aim was to produce widely applicable findings. However, stakeholders who have access to additional information, such as regulators, might benefit from incorporating their private information into the models developed in this research[144].

This research is limited to publicly listed companies in the US, but the methodology that was used is portable to future research on other entities that produce financial statements, conditional upon the availability of data. For example, fraudulent financial reporting also affects companies outside of the US, as well as private companies, not-for-profit organisations and companies traded in the Over-The-Counter markets. This research could also be extended to incorporate quarterly filings, in addition to annual financial statements.

In future when fraud data from 2008 onwards are more complete, it would be interesting to include the Global Financial Crisis (GFC) that began in the second half of 2007 and continued for several years. Evaluating models developed in this current research on data

[144] Section 4.5.17 on page 125 provides a list of explanatory variables that were excluded from this research, largely because they were relatively difficult to obtain.

from during the GFC would give valuable insight into their accuracy during a major financial crisis. In addition, given that high-value financial frauds could very well be endogenous to a financial crisis, impartial classification models such as those developed in this research might be particularly beneficial additional to detection solely by human experts.

Obtaining information about the magnitude of the loss associated with each instance of fraud is a serious challenge because the SEC in the US does not consistently report the magnitude of fraud (Wang et al. 2011). However, if this challenge were to be overcome then it would be very useful to incorporate the magnitude of fraud into models so that only frauds above a certain threshold of magnitude (set by the stakeholder) would be considered. However, it is important to note that models developed in this research are already likely to be biased towards larger frauds, because they are developed from fraud alleged by the SEC which is commonly thought to be biased in that way.

7.2.2 Using This Research to Assist in Predicting Company Failure

The concepts of financial statement fraud and company failure are linked. Beasley et al. (2010) found that companies accused of committing financial statement fraud by the SEC were more than twice as likely to fail. Another study (Deloitte 2008) found that there is three times more fraud alleged by the SEC amongst failed companies compared with those that did not fail.

It would be interesting to investigate whether the output of a model to detect financial statement fraud (developed in this research) could be used to improve a model to predict company failure. Promising results have been found using a similar approach with neural network models (Yang 2008). This is particularly encouraging given that many of the models in this research outperformed neural network models. Finally, this type of research would also help address the unanswered question identified by Beaver et al. (2010) about whether accounting quality affects the utility of financial ratios to assist models to predict company failure.

7.2.3 Models as Decision Aids

Computerised models to detect financial statement fraud complement, not replace, human experts such as auditors and forensic investigators. It might be beneficial to develop an aid, such as a simple automated questionnaire, to assist people in determining their appropriate ratio of error costs (failing to detect fraud to falsely alleging fraud). This information can then be used to help them select the most appropriate financial statement fraud detection model.

Despite research showing decision aids mostly have a positive effect, decision makers are often reluctant to rely on them (Eining et al. 1997). Consequently, valuable research questions include how to best present and explain to stakeholders the models developed in this research, whether there is a preference for simpler models, and how the answers to these questions differ between stakeholder groups. That is, there is an opportunity for future research to investigate ways to improve the interaction between computerised models of financial statement fraud detection and human users.

Bibliography

Abbott, L. J., Y. Park, and S. Parker. 2000. The effects of audit committee activity and independence on corporate fraud. *Managerial Finance* 26 (11):55-68.

Abbott, L. J., S. Parker, and G. F. Peters. 2004. Audit Committee Characteristics and Restatements. *Auditing: A Journal of Practice & Theory* 23 (1):69-87.

ACFE. 2009. Occupational Fraud: A Study of the Impact of an Economic Recession. www.acfe.com: Association of Certified Fraud Examiners.

———. 2010. Report to the Nations on Occupational Fraud and Abuse. www.acfe.com: Association of Certified Fraud Examiners.

———. 2012. Report to the Nations on Occupational Fraud and Abuse. www.acfe.com: Association of Certified Fraud Examiners.

———. 2013a. Financial Statement Fraud Conference, 21-22 March, at Chicago, US.

———. 2013b. Course Materials. Financial Statement Fraud Conference, 21-22 March, at Chicago, US.

———. 2014. Report to the Nations on Occupational Fraud and Abuse. www.acfe.com: Association of Certified Fraud Examiners.

AICPA. 1997. Statement on Auditing Standards (SAS) No. 82: Consideration of Fraud in a Financial Statement Audit. New York, USA: American Institute of Certified Public Accountants.

———. 2002. Statement on Auditing Standards (SAS) No. 99: Consideration of Fraud in a Financial Statement Audit. New York, USA: American Institute of Certified Public Accountants.

———. 2011. Statement on Auditing Standards (SAS) Nos. 122–124: No. 122, Statements on Auditing Standards: Clarification and Recodification; No. 123, Omnibus Statement on Auditing Standards; No. 124, Financial Statements Prepared in Accordance With a Financial Reporting Framework Generally Accepted in Another Country. New York, USA: American Institute of Certified Public Accountants.

Albrecht, C., and W. S. Albrecht. 2001. Can auditors detect fraud: A review of the research evidence. *Journal of Forensic Accounting* 1:1-12.

Albrecht, W. S., K. R. Howe, and M. B. Romney. 1984. *Deterring fraud: the internal auditor's perspective*. Altamonte Springs, Florida, USA: Institute of Internal Auditors Research Foundation.

Alden, M. E., D. M. Bryan, B. J. Lessley, and A. Tripathy. 2012. Detection of Financial Statement Fraud Using Evolutionary Algorithms. *Journal of Emerging Technologies in Accounting* 9:71-94.

Aldhous, P. 2012. Specialist Knowledge is Useless and Unhelpful. *New Scientist* 216 (2893):28-29.

Altman, E. I. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23 (4):589-609.

Altman, E. I., R. G. Heldeman, and P. Marayanan. 1977. ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance* 1:29-54.

Amershi, A. H., and E. H. Feroz. 2000. The occurrence of fibonacci numbers in time series of financial accounting ratios: Anomalies or indicators of firm survival, bankruptcy and fraud? An exploratory study. *Managerial Finance* 26 (11):5-20.

Anthony, R. N., D. E. Hawkins, and K. A. Merchant. 2004. *Accounting: Text and Cases*. 11th ed. New York, USA: McGraw-Hill/Irwin.

Apparao, G., A. Singh, G. S. Rao, B. L. Bhavani, K. Eswar, and D. Rajani. 2009. Financial Statement Fraud Detection by Data Mining. *International . J. of Advanced Networking and Applications* 1 (3):159-163.

Argyrou, A., and A. Andreev. 2011. A semi-supervised tool for clustering accounting databases with applications to internal controls. *Expert Systems with Applications* 38 (9):11176-11181.

Bai, B., J. Yen, and X. Yang. 2008. False financial statements: characteristics of China's listed companies and CART detecting approach. *International Journal of Information Technology & Decision Making* 07 (02):339-359.

Bayley, L., and S. L. Taylor. 2007. Identifying Earnings Overstatements: A Practical Test. *Working paper*: Available at SSRN: http://ssrn.com/abstract=995957.

Beasley, M. S. 1996. An empirical analysis of the relation between the board of director composition and financial statement fraud. *The Accounting Review* 71 (4):443-465.

Beasley, M. S., J. V. Carcello, D. R. Hermanson, and P. D. Lapides. 2000. Fraudulent financial reporting: Consideration of industry traits and corporate governance mechanisms. *Accounting Horizons* 14 (4):441-454.

Beasley, M. S., J. V. Carcello, D. R. Hermanson, and T. L. Neal. 2010. Fraudulent Financial Reporting: 1998 - 2007. An Analysis of U.S. Public Companies. www.coso.org: The Committee of Sponsoring Organizations of the Treadway Commission (COSO).

Beaver, W. H., M. Correia, and M. F. McNichols. 2010. Financial Statement Analysis and the Prediction of Financial Distress. *Foundations and Trends in Accounting* 5 (2):99-173.

Bebchuk, L., A. Cohen, and A. Ferrell. 2009. What Matters in Corporate Governance? *Review of Financial Studies* 22 (2):783-827.

Becker, G. 1968. Crime and punishment: an economic approach. *Journal of Political Economy* 76:169-217.

Begley, J., J. Ming, and S. Watts. 1996. Bankruptcy Classification Errors in the 1980s: An Empirical Analysis of Altman's and Ohlson's Models. *Review of Accounting Studies* 1:267-284.

Belhadji, E., G. Dionne, and F. Tarkhani. 2000. A Model for the Detection of Insurance Fraud. *The Geneva Papers on Risk and Insurance* 25 (4):517-538.

Bell, T. B., and J. V. Carcello. 2000. A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing: A Journal of Practice & Theory* 19 (1):169-184.

Bell, T. B., S. Szykowny, and J. J. Willingham. 1991. Assessing the likelihood of fraudulent financial reporting: A cascaded logit approach. *Working paper, KPMG Forensic*.

Beneish, M. D. 1997. Detecting GAAP violation: implications for assessing earnings management among firms with extreme financial performance. *Journal of Accounting and Public Policy* 16 (3):271-309.

———. 1999a. The detection of earnings manipulation. *Financial Analysts Journal* 55 (5):24-36.

———. 1999b. Incentives and penalties related to earnings overstatements that violate GAAP. *The Accounting Review* 74 (4):425-457.

Bernardi, R. A. 1994. Fraud detection: The effect of client integrity and competence and auditor cognitive style. *Auditing: A Journal of Practice & Theory* 13:68-84.

Bhattacharya, S., S. Jha, K. Tharakunnel, and J. C. Westland. 2011a. Data mining for credit card fraud: A comparative study. *Decision Support Systems* 50 (3):602-613.

Bhattacharya, S., D. Xu, and K. Kumar. 2011b. An ANN-based auditor decision support system using Benford's law. *Decision Support Systems* 50 (3):576-584.

Blocher, E., and J. C. Cooper. 1988. A Study of Auditors' Analytical Review Performance. *Auditing: A Journal of Practice & Theory* 7 (2):1-28.

Block, S. B. 1999. A study of financial analysts: Practice and theory. *Financial Analysts Journal* 55:86-95.

Blumenstein, M., X. Y. Liu, and B. Verma. 2007. An investigation of the modified direction feature for cursive character recognition. *Pattern Recognition* 40 (2):376-388.

Boatsman, J. R., C. Moeckel, and B. K. W. Pei. 1997. The Effects of Decision Consequences on Auditors' Reliance on Decision Aids in Audit Planning. *Organizational Behavior and Human Decision Processes* 71 (2):211-247.

Bolton, R. J., and D. J. Hand. 2002. Statistical fraud detection: a review. *Statistical Science* 17:235-255.

Bonner, S. E., Z.-V. Palmrose, and S. M. Young. 1998. Fraud Type and Auditor Litigation: An Analysis of SEC Accounting and Auditing Enforcement Releases. *The Accounting Review* 73 (4):503-532.

Bose, I., S. Piramuthu, and M. J. Shaw. 2011. Quantitative methods for Detection of Financial Fraud. *Decision Support Systems* 50 (3):557-558.

Boulter, T., A. Mukherjee, and S. Bhattacharya. 2013. Motivation for Occupational Fraud: An Analysis of the "Fraud Triangle" Using Economic Logic. *The International Journal of Interdisciplinary Organizational Studies* 7 (1):47-57.

Brazel, J. F., K. L. Jones, and M. F. Zimbelman. 2009. Using Nonfinancial Measures to Assess Fraud Risk. *Journal of Accounting Research* 47 (5):1135-1166.

Breiman, L. 1996a. Bagging Predictors. *Machine Learning* 24 (2):123-140.

———. 1996b. Technical Note: Some Properties of Splitting Criteria. *Machine Learning* 24:41-47.

———. 2001a. Random Forests. *Machine Learning* 45 (1):5-32.

———. 2001b. Statistical Modeling: The Two Cultures. *Statistical Science* 16 (3):199-215.

Breiman, L., J. H. Friedman, R. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*: Wadsworth & Brooks.

Brennan, N. M., and M. McGrath. 2007. Financial Statement Fraud: Some Lessons from US and European Case Studies. *Australian Accounting Review* 17 (42):49-61.

Burns, N., and S. Kedia. 2006. The impact of performance-based compensation on misreporting. *Journal of Financial Economics* 79 (1):35-67.

Burton, F. G., T. J. Wilks, and M. Zimbelman. 2011. The impact of audit penalty distributions on the detection and frequency of fraudulent reporting. *Review of Accounting Studies* 16 (4):843-865.

Busta, B., and R. Weinberg. 1998. Using Benford's law and neural networks as a review procedure. *Managerial Auditing Journal* 13 (6):356-366.

Cao, Z., F. Leng, E. Feroz, and S. Davalos. 2015. Corporate governance and default risk of firms cited in the SEC's Accounting and Auditing Enforcement Releases. *Review of Quantitative Finance and Accounting* 44 (1):113-138.

Carcello, J. V., and A. L. Nagy. 2004. Audit Firm Tenure and Fraudulent Financial Reporting. *Auditing: A Journal of Practice & Theory* 23 (2):55-69.

Carson, T. L. 2003. Self-Interest and Business Ethics: Some Lessons of the Recent Corporate Scandals. *Journal of Business Ethics* 43 (4):389-394.

Cerullo, M. J., and V. Cerullo. 1999a. Using neural networks to predict financial reporting fraud: Part 1. *Computer Fraud & Security* 1999 (5):14-17.

———. 1999b. Using neural networks to predict financial reporting fraud: Part 2. *Computer Fraud & Security* 1999 (6):14-17.

Chai, W., B. K. Hoogs, and B. T. Verschueren. 2006. Fuzzy Ranking of Financial Statements for Fraud detection. In *IEEE International Conference on Fuzzy System*, 152-158.

Chen, H.-J., S.-Y. Huang, and C.-L. Kuo. 2009. Using the artificial neural network to predict fraud litigation: Some empirical evidence from emerging markets. *Expert Systems with Applications* 36 (2, Part 1):1478-1484.

Coglitore, F., and R. G. Berryman. 1988. Analytical Procedures: A Defensive Necessity. *Auditing: A Journal of Practice & Theory* 7 (2):150-163.

Comunale, C. L., R. L. Rosner, and T. R. Sexton. 2010. The auditor's assessment of fraud risk: A fuzzy logic approach. *Journal of Forensic and Investigative Accounting* 3 (1):149-194.

Cotter, J., and S. M. Young. 2007. Do Analysts Anticipate Accounting Fraud? *Working paper*: Available at SSRN: http://ssrn.com/abstract=981484.

Cox, D. R. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society B* 34:187-220.

Crapp, H. R., and M. Stevenson. 1987. Development of a Method to Assess the Relevant Variables and the Probability of Financial Distress. *Australian Journal of Management* 12 (2):221-236.

Cressey, D. 1953. *Other people's money: a study in the social psychology of embezzlement*. Glencoe, Illinois, USA: Free Press.

Debreceny, R. S., and G. L. Gray. 2010. Data mining journal entries for fraud detection: an exploratory study. *International Journal of Accounting Information Systems* 11 (3):157-181.

———. 2011. Data Mining of Electronic Mail and Auditing: A Research Agenda. *Journal of Information Systems* 25 (2):195-226.

Dechow, P., R. Sloan, and A. Sweeney. 1996. Causes and consequences of earnings manipulations: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research* 13 (1):1-36.

Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan. 2011. Predicting Material Accounting Misstatements. *Contemporary Accounting Research* 28 (1):17-82.

Dechow, P. M., W. Ge, and C. Schrand. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50 (2–3):344-401.

Dechow, P. M., R. G. Sloan, and A. P. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2):193-225.

Deloitte. 2008. Ten things about bankruptcy and fraud - A review of bankruptcy filings. www.deloitte.com/forensiccenter: Deloitte Forensic Centre.

———. 2009. Ten things about financial statement fraud - third edition. A review of SEC enforcement releases, 2000-2008. www.deloitte.com/forensiccenter: Deloitte Forensic Centre.

Derrig, R. A., and L. A. Francis. 2008. Distinguishing the Forest from the TREES: A Comparison of Tree-Based Data Mining Methods. *Variance* 2 (2):184-208.

Desai, H., C. E. Hogan, and M. S. Wilkins. 2006. The Reputational Penalty for Aggressive Accounting: Earnings Restatements and Management Turnover. *The Accounting Review* 81 (1):83-112.

Deshmukh, A., J. Romine, and P. H. Siegel. 1997. Measurement and Combination of Red Flags to Assess the Risk of Management Fraud: A Fuzzy Set Approach. *Managerial Finance* 23 (6):35-48.

Deshmukh, A., and L. Talluru. 1998. A rule-based fuzzy reasoning system for assessing the risk of management fraud. *Intelligent Systems in Accounting, Finance & Management* 7 (4):223-241.

Dikmen, B., and G. Küçükkocaoğlu. 2010. The detection of earnings manipulation: the three-phase cutting plane algorithm using mathematical programming. *Journal of Forecasting* 29 (5):442-466.

Dorminey, J., A. S. Fleming, M.-J. Kranacher, and R. A. Riley, Jr. 2012. The Evolution of Fraud Theory. *Issues in Accounting Education* 27 (2):555-579.

Durtschi, C., W. Hillison, and C. Pacini. 2004. The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting* 5:17-34.

Dyck, A., A. Morse, and L. Zingales. 2010. Who Blows the Whistle on Corporate Fraud? *The Journal of Finance* 65 (6):2213-2253.

Dyck, A., A. Morse, and L. Zingales. 2013. How Pervasive is Corporate Fraud? *Rotman School of Management Working Paper No. 2222608*: Available at SSRN: http://ssrn.com/abstract=2222608.

Efendi, J., A. Srivastava, and E. P. Swanson. 2007. Why do corporate managers misstate financial statements? The role of option compensation and other factors. *Journal of Financial Economics* 85 (3):667-708.

Eining, M. M., D. R. Jones, and J. K. Loebbecke. 1997. Reliance on decision aids: an examination of auditors' assessment of management fraud. *Auditing: A Journal of Practice & Theory* 16 (2):1-19.

Elliot, J. A., and D. B. Kennedy. 1988. Estimation and prediction of categorical models in accounting research. *Journal of Accounting Literature* 7:202-242.

Eppler, M. J., and J. Mengis. 2004. The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society* 20 (5):325-344.

Erickson, M., M. Hanlon, and E. L. Maydew. 2006. Is There a Link between Executive Equity Incentives and Accounting Fraud? *Journal of Accounting Research* 44 (1):113-143.

Ettredge, M. L., L. Sun, P. Lee, and A. A. Anandarajan. 2008. Is Earnings Fraud Associated with High Deferred Tax and/or Book Minus Tax Levels? *Auditing: A Journal of Practice & Theory* 27 (1):1-33.

Fama, E. F., and M. C. Jensen. 1983. Separation of ownership and control. *The journal of law & economics* 26 (2):301-325.

Fanning, K., K. Cogger, and R. Srivastava. 1995. Detection of Management Fraud: A Neural Network Approach. *International Journal of Intelligent Systems in Accounting Finance & Management* 4 (2):113-126.

Fanning, K. M., and K. O. Cogger. 1998. Neural Network Detection of Management Fraud Using Published Financial Data. *International Journal of Intelligent Systems in Accounting Finance & Management* 7 (1):21-41.

Farber, D. B. 2005. Restoring Trust after Fraud: Does Corporate Governance Matter? *The Accounting Review* 80 (2):539-561.

Feroz, E. H., and T. M. Kwon. 1999. Self Organizing Fuzzy and MLP Approaches to Detecting Fraudulent Financial Reporting. In *Soft Computing in Financial Engineering*, edited by R. A, Z. H and K. J. Heidelberg, Germany, 499-506.

Feroz, E. H., T. M. Kwon, V. S. Pastena, and K. Park. 2000. The efficacy of red flags in predicting the SEC's targets: an artificial neural networks approach. *Intelligent Systems in Accounting, Finance & Management* 9 (3):145-157.

Feroz, E. H., K. Park, and V. S. Pastena. 1991. The Financial and Market Effects of the SEC's Accounting and Auditing Enforcement Releases. *Journal of Accounting Research* 29:107-142.

Finkelstein, S. 1992. Power in Top Management Teams: Dimensions, Measurement, and Validation. *The Academy of Management Journal* 35 (3):505-538.

Fisher, R. A. 1938. The statistical utilization of multiple measurements. *Annals of Eugenics* 8:376-386.

Ford, V., A. Siraj, and W. Eberle. 2014. Smart grid energy fraud detection using artificial neural networks. In *IEEE Symposium on Computational Intelligence Applications in Smart Grid*, 1-6.

Francis, J. R. 2004. What do we know about audit quality? *The British Accounting Review* 36 (4):345-368.

Francis, J. R., and S. E. Seavey. 2012. Industry Complexity, Earnings Quality and Auditor Expertise. In *Presentation and Working Paper given at Bond University*.

Frankel, R. M., M. F. Johnson, and K. K. Nelson. 2002. The Relation between Auditors' Fees for Nonaudit Services and Earnings Management. *The Accounting Review* 77:71-105.

Friedman, J. H. 1997. On Bias, Variance, 0/1--Loss, and the Curse-of-Dimensionality. *Data Mining and Knowledge Discovery* 1 (1):55-77.

Friedman, J. H. 1999. Greedy function approximation: a gradient boosting machine. Technical Report, Department of Statistics, Stanford University.

Friedman, J. H. 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis* 38 (4):367-378.

Gao, L., and R. Srivastava. 2011. The decomposition of management fraud schemes: Analyses and implications. *Indian Accounting Review* 15 (1):1-23.

Gepp, A., and K. Kumar. 2008. The Role of Survival Analysis in Business Failure Prediction. *International Research Journal of Finance and Economics* 16:13-34.

———. 2014. Financial distress prediction: A cost-sensitive study using accounting ratios that compares multiple modelling techniques. In *26th Asian-Pacific Conference on International Accounting Issues*. Taipei, Taiwan.

Gepp, A., K. Kumar, and S. Bhattacharya. 2010. Business failure prediction using decision trees. *Journal of Forecasting* 29 (6):536-555.

Gepp, A., J. H. Wilson, K. Kumar, and S. Bhattacharya. 2012. A Comparative Analysis of Decision Trees Vis-à-vis Other Computational Data Mining Techniques in Automotive Insurance Fraud Detection. *Journal of Data Science* 10 (3):537-561.

Gillett, P. R., and N. Uddin. 2005. CFO Intentions of Fraudulent Financial Reporting. *Auditing: A Journal of Practice & Theory* 24 (1):55-75.

Glancy, F. H., and S. B. Yadav. 2011. A computational model for financial reporting fraud detection. *Decision Support Systems* 50 (3):595-601.

Goel, S., J. Gangolly, S. R. Faerman, and O. Uzuner. 2010. Can Linguistic Predictors Detect Fraudulent Financial Filings? *Journal of Emerging Technologies in Accounting* 7:25-46.

Gompers, P., J. Ishii, and A. Metrick. 2003. Corporate Governance and Equity Prices. *The Quarterly Journal of Economics* 118 (1):107-155.

Gray, G. L., and R. S. Debreceny. 2014. A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems* 15 (4):357-380.

Green, B. P., and J. H. Choi. 1997. Assessing the risk of management fraud through neural network technology. *Auditing: A Journal of Practice & Theory* 16 (1):14-28.

Gullkvist, B., and A. Jokipii. 2013. Perceived importance of red flags across fraud types. *Critical Perspectives on Accounting* 24 (1):44-61.

Gupta, R., and N. S. Gill. 2013. Financial Statement Fraud Detection using Text Mining. *International Journal of Advanced Computer Science and Applications* 3:189-191.

Hand, D. 2009a. Mismatched models, wrong results, and dreadful decisions: on choosing appropriate data mining tools. In *15th ACM SIGKDD International Conference on*

Knowledge Discovery and Data Mining. Paris: Association for Computing Machinery, 1-2.

Hand, D. J. 2009b. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning* 77 (1):103-123.

Hansen, J. V., J. B. McDonald, W. F. Messier, Jr., and T. B. Bell. 1996. A generalized qualitative-response model and the analysis of management fraud. *Management Science* 42 (7):1022-1032.

Healy, P. M. 1985. The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics* 7:85 - 107.

Hogan, C. E., Z. Rezaee, R. A. Riley, and U. K. Velury. 2008. Financial Statement Fraud: Insights from the Academic Literature. *Auditing: A Journal of Practice & Theory* 27 (2):231-252.

Hoogs, B., T. Kiehl, C. Lacomb, and D. Senturk. 2007. A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud. *Intelligent Systems in Accounting, Finance & Management* 15 (1-2):41-56.

Huang, S.-M., D. C. Yen, L.-W. Yang, and J.-S. Hua. 2008. An investigation of Zipf's Law for fraud detection. *Decision Support Systems* 46 (1):70-83.

Huang, S.-Y., R.-H. Tsaih, and W.-Y. Lin. 2014. Feature Extraction ff Fraudulent Financial Reporting through Unsupervised Neural Networks. *Neural Network World* 24 (5):539-560.

Huang, S. Y., R. H. Tsaih, and W. Y. Lin. 2012. Unsupervised neural networks approach for understanding fraudulent financial reporting. *Industrial Management & Data Systems* 112 (2):224-244.

Humpherys, S. L., K. C. Moffitt, M. B. Burns, J. K. Burgoon, and W. F. Felix. 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems* 50 (3):585-594.

Hurley, D. A., and D. Boyd. 2007. Sarbanes Oxley Act Section 404: Effective Internal Controls or Overriding Internal Controls? *Forensic Examiner* 16 (2):19-21.

Ittelson, T. 2009. *Financial Statements: A Step-By-Step Guide to Understanding and Creating Financial Reports*. New Jersey, USA: Career Press.

Jans, M., M. Alles, and M. Vasarhelyi. 2013. The case for process mining in auditing: Sources of value added and areas of application. *International Journal of Accounting Information Systems* 14 (1):1-20.

Jensen, M. C. 1993. The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems. *The Journal of Finance* 48 (3):831-880.

Jensen, M. C., and W. H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4):305-360.

Jiambalvo, J. 1996. Discussion of "Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC". *Contemporary Accounting Research* 13 (1):37-47.

Johnson, S. A., H. E. Ryan, and Y. S. Tian. 2009. Managerial Incentives and Corporate Fraud: The Sources of Incentives Matter. *Review of Finance* 13 (1):115-145.

Jones, J. J. 1991. Earnings Management During Import Relief Investigations. *Journal of Accounting Research* 29 (2):193-228.

Jones, M., and E. B. L. E. Library. 2011. *Creative accounting, fraud and international accounting scandals*. Chichester: Wiley.

Kaminski, K. a., T. S. Wetzel, and L. Guan. 2004. Can financial ratios detect fraudulent financial reporting? *Managerial Auditing Journal* 19 (1):15-28.

Karpoff, J. M., A. Koester, D. S. Lee, and G. S. Martin. 2014. Database Challenges in Financial Misconduct Research. *Georgetown McDonough School of Business Research Paper No. 2012-15*: Available at SSRN: http://ssrn.com/abstract=2112569.

Kassem, R., and A. Higson. 2012. The New Fraud Triangle Model. *Journal of Emerging Trends in Economics and Management Sciences* 3 (3):191-195.

Kedia, S., and T. Philippon. 2009. The Economics of Fraudulent Accounting. *The Review of* Financial Studies 22 (6):2169-2199.

Kent, P., J. Stewart, and J. Christensen. 2012. Alternative Markets for Non-audit Services in a Voluntary Setting. In 24th Asia-Pacific Conference on International Accounting Issues. Maui, Hawaii.

Khan, A. U. S., N. Akhtar, and M. N. Qureshi. 2014. Real-Time Credit-Card Fraud Detection using Artificial Neural Network Tuned by Simulated Annealing Algorithm. In *International Conference on Recent Trends in Information, Telecommunication and Computing*. Chandigarh, India, 113-121.

Kiehl, T., B. Hoogs, C. LaComb, and D. Senturk. 2005. Evolving Multi-Variate Time-Series Patterns for the Discrimination of Fraudulent Financial Filings. In *Genetic and Evolutionary Computation Conference (GECCO)*. Wahington, DC, USA.

Kirkos, E., C. Spathis, and Y. Manolopoulos. 2007. Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications* 32:995-1003.

Koskivaara, E. 2004. Artificial neural networks in analytical review procedures. *Managerial Auditing Journal* 19 (2):191-223.

Kotsiantis, S., E. Koumanakos, D. Tzelepis, and V. Tampakas. 2007. Forecasting fraudulent financial statements using data mining. *International Journal of Computational Intelligence* 3 (2):104-110.

KPMG Forensic. 2013. A survey of fraud, bribery and corruption in Australia & New Zealand 2012. Retrieved from http://www.kpmg.com/au/en/issuesandinsights/articlespublications/fraud-survey/pages/default.aspx.

Krambia-Kapardis, M., C. Christodoulou, and M. Agathocleous. 2010. Neural networks: the panacea in fraud detection? *Managerial Auditing Journal* 25 (7):659-678.

Kranacher, M.-J., R. Riley, and J. T. Wells. 2011. *Forensic accounting and fraud examination*. Hoboken, N.J: John Wiley.

Kwon, T. M., and E. H. Feroz. 1996. A multilayered perceptron approach to prediction of the SEC's investigation targets. *IEEE Transactions on Neural Networks* 7 (5):1286-1290.

Laitinen, T., and M. Kankaanpää. 1999. Comparative analysis of failure prediction methods: the Finnish case. *The European Accounting Review* 8 (1):67-92.

Land, J. K. 2010. CEO turnover around earnings restatements and fraud. *Pacific Accounting Review* 22 (3):180-198.

Lane, R., and B. T. O'Connell. 2009. The changing face of regulators' investigations into financial statement fraud. *Accounting Research Journal* 22 (2):118-143.

Lane, W. R., S. W. Looney, and J. W. Wansley. 1986. An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance* 10 (4):511-531.

Layton, A., T. Robinson, and I. B. Tucker. 2012. *Economics for Today*. Fourth Asia Pacific ed: Cengage Learning.

Lee, T. A., R. W. Ingram, and T. P. Howard. 1999. The difference between earnings and operating cash flow as an indicator of financial reporting fraud. *Contemporary Accounting Research* 16 (4):749-786.

Lenard, M. J., and P. Alam. 2004. The use of fuzzy logic and expert reasoning for knowledge management and discovery of financial reporting fraud. In *Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance*, edited by H. Nemati and C. Barko. Hershey, PA, USA: Idea Group Inc.

Lenard, M. J., A. L. Watkins, and P. Alam. 2007. Effective Use of Integrated Decision Making: An Advanced Technology Model for Evaluating Fraud in Service-Based Computer and Technology Firms. *Journal of Emerging Technologies in Accounting* 4:123-137.

Lennox, C. 2000. Do companies successfully engage in opinion-shopping? Evidence from the UK. *Journal of Accounting and Economics* 29 (3):321-337.

Lin, J., M. Hwang, and J. Becker. 2003. A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal* 18 (8):657-665.

Liou, F. M. 2008. Fraudulent financial reporting detection and business failure prediction models: a comparison. *Managerial Auditing Journal* 23 (7):650-662.

Loebbecke, J. K., M. M. Eining, and J. J. Willingham. 1989. Auditors' Experience with Material Irregularities: Frequency, Nature, and Detectability. *Auditing: A Journal of Practice & Theory* Fall:1-28.

Lou, Y.-I., and M.-L. Wang. 2009. Fraud risk factor of the fraud triangle assessing the likelihood of fraudulent financial reporting. *Journal of Business & Economics Research* 7 (2):61-78.

McKee, T. E. 2009. A Meta-Learning Approach to Predicting Financial Statement Fraud. *Journal of Emerging Technologies in Accounting* 6:5-26.

Morgan, J. N., and R. C. Messenger. 1973. THAID: a sequential search program for the analysis of nominal scale dependent variables. Retrieved from http://catalog.hathitrust.org/Record/000469743: Institute for Social Research, University of Michigan, USA.

Myers, J. N., L. A. Myers, and T. C. Omer. 2003. Exploring the Term of the Auditor-Client Relationship and the Quality of Earnings: A Case for Mandatory Auditor Rotation? *The Accounting Review* 78 (3):779-799.

Myrtveit, I., E. Stensrud, and U. H. Olsson. 2001. Analyzing data sets with missing data: an empirical evaluation of imputation methods and likelihood-based methods. *IEEE Transactions on Software Engineering* 27 (11):999-1013.

Nagadevara, V. 2010. Hybrid models and error weighting for predicting customer churn in telecom industry. *Review of Business Research* 10 (1):83-87.

National Commission on Fraudulent Financial Reporting. 1987. Report of the National Commission on Fraudulent Financial Reporting. www.coso.org.

Ndofor, H. A., C. Wesley, and R. L. Priem. 2013. Providing CEOs With Opportunities to Cheat: The Effects of Complexity-Based Information Asymmetries on Financial Reporting Fraud. *Journal of Management* forthcoming.

Negnevitsky, M. 2011. *Artificial intelligence: a guide to intelligent systems*. Harlow, England: Addison Wesley/Pearson.

Ngai, E. W. T., Y. Hu, Y. H. Wong, Y. Chen, and X. Sun. 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems* 50 (3):559-569.

Nieschwietz, R. J., J. J. Schultz, Jr., and M. F. Zimbelman. 2000. Empirical research on external auditors' detection of financial statement fraud. *Journal of Accounting Literature* 19:190-246.

O'Connor, J. J., R. L. Priem, J. E. Coombs, and K. M. Gilley. 2006. Do CEO Stock Options Prevent or Promote Fraudulent Financial Reporting? *The Academy of Management Journal* 49 (3):483-500.

Owusu-Ansah, S., G. D. Moyes, O. Peter Babangida, and D. Hay. 2002. An empirical analysis of the likelihood of detecting fraud in New Zealand. *Managerial Auditing Journal* 17 (4):192-204.

Pai, P.-F., M.-F. Hsu, and M.-C. Wang. 2011. A support vector machine-based model for detecting top management fraud. *Knowledge-Based Systems* 24 (2):314-321.

Palmrose, Z.-V., and S. Scholz. 2004. The Circumstances and Legal Consequences of Non-GAAP Reporting: Evidence from Restatements. *Contemporary Accounting Research* 21 (1):139-180.

Partnoy, F. 2010. Infectious Greed : How Deceit and Risk Corrupted the Financial Markets. London: Profile Books.

Pathak, J., N. Vidyarthi, and S. L. Summers. 2005. A fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims. *Managerial Auditing Journal* 20 (6):632-644.

Patterson, E., and J. Noel. 2003. Audit Strategies and Multiple Fraud Opportunities of Misreporting and Defalcation. *Contemporary Accounting Research* 20 (3):519-549.

Peasnell, K. V., P. F. Pope, and S. Young. 2000. Detecting earnings management using cross-sectional abnormal accruals models. *Accounting and Business Research* 30 (4):313-326.

Perols, J. 2011. Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. *Auditing: A Journal of Practice & Theory* 30 (2):19-50.

Perols, J. L. 2008. Detecting financial statement fraud: Three essays on fraud predictors, multi-classifier combination and fraud detection using data mining. Ph.D. Thesis, University of South Florida, Scholar Commons, Theses and Dissertations.

Perols, J. L., and B. A. Lougee. 2011. The relation between earnings management and financial statement fraud. *Advances in Accounting* 27 (1):39-53.

Persons, O. S. 1995. Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research* 11 (3):38-46.

Phua, C., V. Lee, K. Smith, and R. Gayler. 2010. A Comprehensive Survey of Data Mining-based Fraud Detection Research. *Computing Research Repository (CoRR). Available online at http://arxiv.org/abs/1009.6119*.

Quinlan, J. R. 1993. *C4.5: programs for machine learning*. California, US: Morgan Kaufmann Publishers.

Ragothaman, S., J. Carpenter, and T. Buttars. 1995. Using rule induction for knowledge acquisition: An expert systems approach to evaluating material errors and irregularities. *Expert Systems with Applications* 9 (4):483-490.

Ramamoorti, S. 2008. The Psychology and Sociology of Fraud: Integrating the Behavioral Sciences Component Into Fraud and Forensic Accounting Curricula. *Issues in Accounting Education* 23 (4):521-533.

Ravisankar, P., V. Ravi, G. R. Rao, and I. Bose. 2011. Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems* 50 (2):491-500.

Reed, R., and D. Pence. 2005. Detecting Fraud in Financial Statements: The Use of Digital Analysis as an Analytical Review Procedure. *Journal of Forensic Accounting* 6:135-146.

Rezaee, Z., and R. Riley. 2010. *Financial statement fraud: prevention and detection*. 2nd Edition ed. Hoboken, N.J: Wiley.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and İ. Tuna. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics* 39 (3):437-485.

Rokach, L., and O. Maimon. 2015. Introduction to Decision Trees. In *Data Mining with Decision Trees: Theory and Applications*. Singapore: World Scientific Publishing, 1-16.

Romney, M. B., P. J. Steinbart, J. M. Mula, R. P. McNamara, and T. Tonkin. 2013. Fraud, controls and audit in accounting information systems. In *Accounting information systems*. Frenchs Forest, NSW: Pearson Australia, 338-373.

Rosner, R. L. 2003. Earnings Manipulation in Failing Firms. *Contemporary Accounting Research* 20 (2):361-408.

Salford Systems. 2012a. *SPM Users Guide: Introduction to RandomForests*. Available at http://www.salford-systems.com/support/spm-user-guide.

———. 2012b. *SPM Users Guide: Introduction to TreeNet*. Available at http://www.salford-systems.com/support/spm-user-guide.

Sallehuddin, R., S. Ibrahim, and A. Hussein Elmi. 2014. Classification of SIM Box Fraud Detection Using Support Vector Machine and Artificial Neural Network. *International Journal of Innovative Computing* 4 (2):19-27.

Sathya, R., and A. Abraham. 2013. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence* 2 (2):34-38.

Schmitz, G. P. J., C. Aldrich, and F. S. Gouws. 1999. ANN-DT: an algorithm for extraction of decision trees from artificial neural networks. *IEEE Transactions on Neural Networks* 10:1392-1401.

Scott, I. 2012. Accounting and auditing enforcement releases: Cash flow evidence associated with revenue recognition fraud. Ph.D. Thesis, The University of Memphis, Ann Arbor.

Seavey, S. E. 2011. Industry Complexity and the Importance of Industry Leadership Versus Engagement Office Size on the Quality of Audit Outcomes. Ph.D. Thesis, Faculty of the Graduate School, University of Missouri.

SEC. 2014. *Listing and Delisting Requirements* [cited September 2014]. Available from http://www.sec.gov/answers/listing.htm.

Sharma, A., and P. K. Panigrahi. 2012. A Review of Financial Accounting Fraud Detection based on Data Mining Techniques. *International Journal of Computer Applications* 39 (1):37-47.

Shiguo, W. 2010. A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research. In *International Conference on Intelligent Computation Technology and Automation (ICICTA)*. Changsha, China, 50-53.

Shih, K.-H., C.-C. Cheng, and Y.-H. Wang. 2011. Financial information fraud risk warning for manufacturing industry - using logistic regression and neural network. *Romanian Journal of Economic Forecasting* 14 (1):54-71.

Shih, Y. 1994. *Neuralyst version 1.4: User's Guide*. Pasadena: Cheshire Enginerring Corporation.

Shumway, T. 2001. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business* 74 (1):101-124.

Silver, S. E., A. S. Fleming, and R. A. Riley, Jr. 2008. Preventing and Detecting Collusive Management Fraud. *The CPA Journal* 78 (10):46-48.

Skogsvik, K. 2005. On the Choice-Based Sample Bias in Probabilistic Business Failure Prediction. *SSE/EFI Working Paper Series in Business Administration*: Retrieved from http://EconPapers.repec.org/RePEc:hhb:hastba:2005_2013.

Skousen, C., and C. Wright. 2008. Contemporaneous risk factors and the prediction of financial statement fraud. *Journal of Forensic Accounting* IX:37-62.

Skousen, C. J., K. R. Smith, and C. J. Wright. 2009. Detecting and predicting financial statement fraud: The effectiveness of the fraud triangle and SAS No. 99. *Advances in Financial Economics* 13 (Corporate Governance and Firm Performance):53-81.

Song, X.-P., Z.-H. Hu, J.-G. Du, and Z.-H. Sheng. 2014. Application of Machine Learning Methods to Risk Assessment of Financial Statement Fraud: Evidence from China. *Journal of Forecasting* 33 (8):611-626.

Spathis, C. 2002. Detecting false financial statements using published data: some evidence from Greece. *Managerial Auditing Journal* 17 (4):179-191.

Spathis, C., M. Doumpos, and C. Zopounidis. 2002. Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques. *The European Accounting Review* 11 (3):509-535.

Srivastava, R. P., T. J. Mock, and J. L. Turner. 2009. Bayesian Fraud Risk Formula for Financial Statement Audits. *Abacus: A Journal of Accounting, Finance and Business Studies* 45 (1):66-87.

Steinberg, D. 2015. *CART vs. The Clones*. Salford Systems [cited January 2015]. Available from https://www.salford-systems.com/blog/dan-steinberg/item/9-cart%C2%AE-vs-the-clones.

———. 2014. *Controlling Node Sizes in a CART Tree*. Salford Systems [cited September 2014]. Available from https://www.salford-systems.com/blog/dan-steinberg/item/5-controlling-node-sizes-in-a-cart-tree.

———. 2012. *A Reminder About Missing Values In Data Mining*. Salford Systems [cited December 2012]. Available from https://www.salford-systems.com/blog/dan-steinberg/item/2-a-reminder-about-missing-values.

Stice, J. D. 1991. Using Financial and Market Information to Identify Pre-Engagement Factors Associated with Lawsuits against Auditors. *The Accounting Review* 66 (3):516-533.

Sudjianto, A., M. Yuan, D. Kern, S. Nair, A. Zhang, and F. Cela-Diaz. 2010. Statistical methods for fighting financial crimes. *Technometrics* 52:5-19.

Summers, S. L., and J. T. Sweeney. 1998. Fraudulently Misstated Financial Statements and Insider Trading: An Empirical Analysis. *The Accounting Review* 73 (1):131-146.

Sutherland, E. H. 1940. White-Collar Criminality. *American Sociological Review* 5 (1):1-12.

Sutton, C. D. 2005. Classification and Regression Trees, Bagging, and Boosting. In *Handbook of Statistics*, edited by E. J. W. C.R. Rao and J. L. Solka: Elsevier, 303-329.

TheLawDictionary.org. 2015. *The Law Dictionary Featuring Black's Law Dictionary Free Online Legal Dictionary 2nd Ed.* [cited January 2015]. Available from http://thelawdictionary.org/about/.

Trotman, K., and M. Gibbins. 2005. *Financial accounting: an integrated approach*. Third ed. Melbourne, Australia: Thomson Learning Australia.

Uzun, H., S. H. Szewczyk, and R. Varma. 2004. Board Composition and Corporate Fraud. *Financial Analysts Journal* 60 (3):33-43.

Wang, I., R. Radich, and N. Fargher. 2011. An analysis of financial statement fraud at the audit assertion level. In *AFAANZ conference 2009 and the American Accounting Association annual meeting 2010*: Retrieved from http://www.uow.edu.au/content/groups/public/@web/@commerce/@econ/documents/doc/uow120444.pdf.

Wang, T. Y. 2013. Corporate Securities Fraud: Insights from a New Empirical Framework. *Journal of Law, Economics, and Organization* 29 (3):535-568.

Wang, T. Y., A. Winton, and X. Yu. 2010. Corporate Fraud and Business Conditions: Evidence from IPOs. *The Journal of Finance* 65 (6):2255-2292.

Warren, J. 2012. *The Most Effective Controls in Limiting Fraud Losses*. ACFE Insights [cited December 2012].

Weisbach, M. S. 1988. Outside Directors and CEO Turnover. *Journal of Financial Economics* 20:431-460.

Welch, O. J., T. E. Reeves, and S. T. Welch. 1998. Using a genetic algorithm-based classifier system for modeling auditor decision behavior in a fraud setting. *Intelligent Systems in Accounting, Finance & Management* 7 (3):173-186.

Whiting, D. G., J. V. Hansen, J. B. McDonald, C. Albrecht, and W. S. Albrecht. 2012. Machine Learning Methods For Detecting Patterns Of Management Fraud. *Computational Intelligence* 28 (4):505-527.

Wolfe, D. T., and D. R. Hermanson. 2004. The Fraud Diamond: Considering the Four Elements of Fraud. *The CPA Journal* 74 (12):38-42.

Yang, F.-M. L. a. C.-H. 2008. Predicting business failure under the existence of fraudulent financial reporting. *International Journal of Accounting and Information Management* 16 (1):74-86.

Yu, F., and X. Yu. 2011. Corporate Lobbying and Fraud Detection. *Journal of Financial and Quantitative Analysis* 46 (6):1865-1891.

Yuan, J., C. Yuan, and X. Deng. 2008. The Effects of Manager Compensation and Market Competition on Financial Fraud in Public Companies: An Empirical Study in China. *International Journal of Management* 25 (2):322-335.

Zhou, W., and G. Kapoor. 2011. Detecting evolutionary financial statement fraud. *Decision Support Systems* 50 (3):570-575.

Never tell me the odds.

- Han Solo in Star Wars: Episode V

Appendices

Appendix A. The Weighted Error Cost (WEC) for all the models on the holdout data.

	Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model	**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
Naïve	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
F-score	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.077	1.049	0.964	1.000	1.041	1.000	1.000	1.000	1.000	1.000	1.000
M-score	2.479	1.592	1.148	1.059	1.030	1.020	1.015	1.012	1.024	0.941	1.083	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DA_All	2.444	1.556	1.112	1.154	1.219	1.065	0.970	0.917	0.892	0.787	1.213	1.438	1.101	1.189	1.237	1.089	1.124	1.077
DA_U15%	0.964	0.964	0.964	1.071	1.154	1.065	1.074	1.000	0.951	0.858	1.284	1.148	1.130	1.101	1.118	1.136	1.172	1.296
DA_U15%LR	3.947	2.172	1.284	1.107	1.189	1.187	1.089	1.071	1.022	0.828	1.296	0.982	1.089	1.107	1.130	1.154	1.201	1.367
DA_U1	3.923	2.148	1.260	1.231	1.195	1.079	1.053	0.976	0.925	0.757	1.024	1.302	1.095	1.112	1.148	1.183	1.254	1.272
DA_Step	0.988	0.988	0.988	1.260	1.083	1.024	0.994	0.988	0.949	0.787	1.130	1.408	1.083	1.118	1.118	1.142	1.189	1.166
LR_All	3.93	2.16	1.27	1.11	1.20	1.04	0.96	1.05	0.97	0.76	1.09	1.30	1.08	1.12	1.17	1.21	1.29	1.58
LR_U15%	0.97	0.97	0.97	0.97	1.08	1.05	1.05	0.99	0.95	0.85	1.27	1.02	1.15	1.21	1.27	1.14	1.17	1.30
LR_U15%LR	3.95	2.17	1.28	1.11	1.21	1.11	1.12	1.05	0.98	0.82	1.27	1.03	1.08	1.11	1.14	1.16	1.21	1.37
LR_U1	2.45	1.56	1.12	1.06	1.25	1.09	1.12	1.01	0.94	0.75	0.93	1.18	1.05	1.09	1.12	1.14	1.02	1.07
LR_Step	0.98	0.98	0.98	1.20	1.08	1.04	1.03	0.93	0.89	0.77	1.11	1.23	1.08	1.11	1.14	1.17	1.23	1.00
LR_MS_S	2.444	1.556	1.112	1.024	1.083	1.034	1.130	1.030	1.002	0.805	1.290	1.231	1.089	1.118	1.148	1.178	1.237	1.444
LR_MS_F	8.349	3.911	1.692	1.249	1.160	1.152	1.038	0.970	0.925	0.822	1.024	1.059	1.154	1.219	1.059	1.083	1.130	1.296
NN_BK	5.420	2.757	1.426	1.160	0.994	1.276	1.195	1.124	1.051	0.852	1.272	1.089	1.266	1.355	1.030	1.036	1.047	1.089
NN_GA_1	3.935	2.160	1.272	1.095	1.036	1.016	1.006	1.041	1.037	1.006	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NN_GA_5	43.615	40.621	10.444	5.148	2.751	1.953	1.553	1.314	1.154	0.846	1.195	1.497	2.148	2.473	2.799	3.124	3.775	5.669
DT_Suite	1.000	1.000	1.000	1.000	1.225	1.030	1.337	1.124	1.134	0.669	1.160	1.160	1.112	1.142	1.172	1.201	1.000	1.000
DT_One	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.953	0.858	0.669	1.053	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RF_8	2.456	1.568	1.124	1.036	1.053	0.982	0.864	0.840	0.791	0.686	0.905	1.089	1.320	1.473	1.308	1.402	0.947	0.941
RF_66	2.426	1.538	1.095	1.006	0.964	0.913	0.997	0.982	0.880	0.675	1.012	0.905	1.083	1.172	1.260	1.349	0.876	0.876
TN	1.000	1.000	1.000	1.036	1.000	0.986	0.905	0.834	0.779	0.651	0.917	1.124	1.112	0.899	0.917	0.935	0.970	0.964

Appendix A. The Weighted Error Cost (WEC) for all the models on the holdout data.

	Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model	**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
NN_LR	3.929	2.154	1.266	1.266	1.089	1.053	0.985	0.899	0.842	0.787	1.225	1.071	1.053	1.083	0.947	0.947	0.947	0.947
DT_LR	33.166	13.639	7.633	3.284	1.834	1.351	1.109	0.964	0.868	0.675	1.059	1.444	2.095	2.450	2.805	3.160	3.870	6.355
DTnode_LR_Step	65.491	26.438	6.911	3.225	1.805	1.331	1.095	0.953	0.858	0.651	1.018	1.385	2.118	2.231	2.550	2.870	3.432	5.586
NN-DTnode_LR_Step	66.964	27.024	7.053	3.059	1.864	1.371	1.124	0.976	0.878	0.669	1.041	1.414	2.160	2.533	2.905	3.278	3.775	6.178
Vote5	0.988	0.988	0.988	0.988	0.988	0.933	0.837	0.787	0.748	0.639	0.970	0.976	1.012	0.994	0.947	0.947	0.959	0.988
Vote3_RF_TN_DT	1.000	1.000	1.000	1.047	1.053	0.994	0.879	0.763	0.720	0.580	0.888	1.000	1.071	0.899	0.929	0.947	0.905	0.970
Vote3_RF_TN_DA	0.982	0.982	0.982	1.024	0.988	0.947	0.834	0.799	0.773	0.657	0.941	1.136	1.118	1.059	1.065	1.036	0.888	0.970
Vote3_RF_TN_NN	0.988	0.988	0.988	1.047	1.030	0.966	0.864	0.834	0.789	0.692	0.893	1.012	1.118	1.118	0.929	0.947	0.899	0.970
AV5_NoNN	0.988	0.988	0.988	0.976	0.935	0.923	0.858	0.828	0.797	0.592	0.834	0.953	1.000	1.006	1.018	0.941	0.959	0.988
AV2_RF_TN	2.473	1.586	1.142	1.036	0.982	0.937	0.891	0.811	0.759	0.669	0.852	1.018	1.166	1.018	1.000	1.041	0.935	0.959
AV3_RF_TN_DT	2.473	1.586	1.142	1.041	0.953	0.957	0.805	0.781	0.753	0.609	0.858	0.959	1.059	1.000	0.976	1.000	0.982	0.982
DT_One_DA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.947	0.856	0.669	1.041	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Simpler Models (discussed in Chapter 6)																		
TN_90%_6	0.976	0.976	0.976	0.976	0.947	1.041	1.086	0.970	0.893	0.740	0.923	1.041	1.278	1.154	1.213	1.272	1.391	1.089
TN_4	0.994	0.994	0.994	0.994	1.485	1.258	1.157	1.172	1.081	0.929	1.166	1.172	1.101	1.124	1.148	1.172	1.219	1.385
LR_Step_11	0.988	0.988	0.988	1.148	1.083	1.073	1.038	0.982	1.002	0.710	1.036	1.213	1.077	0.994	0.994	0.994	0.994	0.994
LR_Int_11	0.988	0.988	0.988	1.130	1.089	1.075	1.050	0.982	0.996	0.817	1.172	0.935	1.012	1.036	1.095	1.112	1.148	1.272
LR_TN_6	0.970	0.970	0.970	0.970	0.970	1.000	0.985	0.976	0.980	0.817	0.994	1.077	1.166	1.213	1.260	1.308	1.160	1.284

Appendix B. The Weighted Error Cost (WEC) for all the models on the training data.

	Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model	**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
Naïve	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
F-score	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.980	0.954	0.814	0.949	0.969	1.000	1.000	1.000	1.000	1.000	1.000
M-score	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.881	0.969	0.997	0.997	0.997	0.997	0.997	0.997	0.997
DA_All	0.925	0.925	0.925	0.912	0.854	0.756	0.703	0.671	0.625	0.485	0.671	0.827	0.905	0.919	0.925	0.932	0.939	0.942
DA_U15%	0.949	0.949	0.949	0.942	0.905	0.871	0.815	0.759	0.722	0.563	0.803	0.912	0.959	0.966	0.966	0.966	0.966	0.966
DA_U15%LR	0.959	0.959	0.959	0.959	0.946	0.873	0.817	0.783	0.755	0.590	0.803	0.905	0.936	0.942	0.942	0.942	0.942	0.942
DA_U1	0.973	0.973	0.973	0.956	0.902	0.828	0.758	0.708	0.668	0.546	0.698	0.824	0.902	0.908	0.912	0.915	0.922	0.932
DA_Step	0.993	0.993	0.993	0.929	0.861	0.838	0.827	0.786	0.745	0.597	0.759	0.847	0.915	0.922	0.929	0.932	0.939	0.942
LR_All	0.939	0.939	0.939	0.929	0.868	0.783	0.727	0.675	0.620	0.478	0.654	0.776	0.861	0.864	0.868	0.871	0.878	0.902
LR_U15%	0.953	0.953	0.953	0.953	0.915	0.873	0.815	0.763	0.727	0.573	0.793	0.898	0.925	0.939	0.953	0.959	0.959	0.959
LR_U15%LR	0.963	0.963	0.963	0.963	0.959	0.880	0.822	0.783	0.750	0.580	0.776	0.908	0.936	0.942	0.942	0.942	0.942	0.942
LR_U1	0.963	0.963	0.963	0.956	0.895	0.816	0.761	0.692	0.644	0.508	0.695	0.820	0.892	0.905	0.919	0.932	0.936	0.936
LR_Step	0.990	0.990	0.990	0.953	0.919	0.868	0.797	0.732	0.681	0.539	0.783	0.888	0.925	0.932	0.939	0.946	0.959	0.973
LR_MS_S	0.895	0.895	0.895	0.895	0.868	0.834	0.768	0.705	0.660	0.522	0.702	0.861	0.912	0.912	0.912	0.912	0.912	0.912
LR_MS_F	0.912	0.912	0.912	0.912	0.912	0.806	0.732	0.685	0.653	0.532	0.729	0.841	0.885	0.902	0.919	0.922	0.929	0.953
NN_BK	0.959	0.959	0.959	0.959	0.936	0.902	0.832	0.769	0.728	0.603	0.844	0.895	0.942	0.966	0.983	0.983	0.983	0.983
NN_GA_1	0.936	0.936	0.936	0.936	0.936	0.936	0.936	0.936	0.918	0.864	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
NN_GA_5	0.759	0.475	0.220	0.119	0.085	0.073	0.068	0.064	0.062	0.054	0.085	0.102	0.136	0.153	0.169	0.186	0.220	0.332
DT_Suite	1.000	1.000	1.000	1.000	1.000	0.899	0.215	0.366	0.408	0.034	0.542	0.827	0.912	0.912	0.912	0.912	1.000	1.000
DT_One	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.041	0.038	0.034	0.061	1.000	1.000	1.000	1.000	1.000	1.000	1.000
RF_8	0.017	0.017	0.017	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.058
RF_66	0.047	0.047	0.047	0.047	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.068	0.068
TN	0.993	0.993	0.993	0.217	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.085	0.085	0.085	0.085	0.593

Appendix B. The Weighted Error Cost (WEC) for all the models on the training data.

	Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model	**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
NN_LR	0.932	0.932	0.932	0.905	0.861	0.783	0.700	0.634	0.590	0.468	0.664	0.817	0.885	0.898	0.898	0.898	0.898	0.898
DT_LR	0.292	0.292	0.092	0.041	0.024	0.018	0.015	0.014	0.012	0.010	0.017	0.024	0.037	0.041	0.044	0.047	0.054	0.078
DTnode_LR_Step	0.108	0.108	0.108	0.064	0.047	0.042	0.039	0.037	0.036	0.031	0.051	0.071	0.112	0.132	0.142	0.153	0.169	0.217
NN-DTnode_LR_Step	0.054	0.054	0.054	0.054	0.051	0.045	0.041	0.037	0.035	0.031	0.044	0.058	0.085	0.098	0.112	0.125	0.139	0.186
Vote5	0.980	0.980	0.980	0.868	0.600	0.495	0.447	0.014	0.014	0.007	0.010	0.393	0.746	0.756	0.847	0.847	0.847	0.895
Vote3_RF_TN_DT	0.993	0.993	0.993	0.217	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.085	0.085	0.085	0.085	0.593
Vote3_RF_TN_DA	0.922	0.922	0.922	0.207	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.081	0.081	0.081	0.081	0.586
Vote3_RF_TN_NN	0.953	0.953	0.953	0.214	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.078	0.085	0.085	0.085	0.593
AV5_NoNN	0.878	0.878	0.878	0.746	0.397	0.312	0.203	0.119	0.095	0.027	0.071	0.169	0.498	0.637	0.658	0.685	0.725	0.854
AV2_RF_TN	0.190	0.190	0.190	0.041	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.003	0.003	0.014	0.156
AV3_RF_TN_DT	0.231	0.231	0.231	0.078	0.027	0.027	0.027	0.027	0.027	0.007	0.007	0.007	0.007	0.014	0.014	0.014	0.031	0.220
DT_One_DA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.027	0.027	0.034	0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Simpler Models (discussed in Chapter 6)																		
TN_90%_6	0.925	0.925	0.925	0.925	0.454	0.156	0.017	0.017	0.017	0.017	0.169	0.169	0.169	0.468	0.468	0.468	0.468	0.980
TN_4	0.993	0.993	0.993	0.993	0.512	0.501	0.427	0.285	0.280	0.136	0.315	0.454	0.993	0.997	1.000	1.003	1.010	1.034
LR_Step_11	0.986	0.986	0.986	0.956	0.936	0.901	0.878	0.831	0.766	0.607	0.759	0.844	0.969	0.969	0.969	0.969	0.969	0.969
LR_Int_11	0.976	0.976	0.976	0.959	0.939	0.905	0.863	0.807	0.767	0.631	0.803	0.908	0.946	0.959	0.963	0.963	0.963	0.963
LR_TN_6	0.969	0.969	0.969	0.969	0.969	0.966	0.941	0.925	0.914	0.834	0.905	0.929	0.956	0.963	0.969	0.976	0.986	0.986

Appendix C. The percentage accuracy of all models on the holdout data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
Naïve	**O**	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%
F-score	**O**	50%	50%	50%	50%	50%	50%	50%	50%	50%	52%	52%	52%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	0%	9%	9%	31%	96%	96%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	92%	92%	72%	8%	8%	0%	0%	0%	0%	0%	0%
M-score	**O**	50%	50%	50%	50%	50%	50%	50%	50%	49%	53%	51%	50%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	0%	0%	1%	56%	90%	100%	100%	100%	100%	100%	100%	100%
	L	99%	99%	99%	99%	99%	99%	99%	99%	98%	50%	12%	0%	0%	0%	0%	0%	0%	0%
DA_All	**O**	51%	51%	51%	53%	59%	59%	59%	59%	61%	61%	61%	61%	56%	52%	52%	52%	52%	51%
	F	4%	4%	4%	8%	28%	29%	30%	30%	38%	50%	57%	67%	95%	95%	95%	98%	98%	99%
	L	99%	99%	99%	98%	90%	89%	89%	89%	83%	72%	65%	56%	17%	9%	9%	5%	5%	2%
DA_U15%	**O**	52%	52%	52%	52%	53%	53%	57%	57%	57%	57%	58%	54%	54%	49%	49%	49%	49%	49%
	F	4%	4%	4%	5%	11%	11%	30%	30%	30%	49%	56%	89%	95%	98%	98%	98%	98%	98%
	L	100%	100%	100%	99%	95%	95%	85%	85%	85%	66%	59%	19%	14%	1%	1%	1%	1%	1%
DA_U15%LA	**O**	50%	50%	50%	50%	55%	54%	54%	54%	54%	59%	59%	56%	55%	51%	51%	51%	51%	51%
	F	1%	1%	1%	1%	17%	21%	21%	22%	22%	49%	53%	95%	95%	98%	98%	98%	98%	98%
	L	99%	99%	99%	99%	93%	88%	88%	85%	85%	69%	64%	16%	15%	4%	4%	4%	4%	4%
DA_U1	**O**	51%	51%	51%	52%	54%	57%	58%	59%	59%	62%	63%	63%	54%	53%	53%	53%	53%	51%
	F	4%	4%	4%	7%	16%	24%	30%	33%	33%	59%	72%	72%	96%	96%	96%	96%	96%	98%
	L	99%	99%	99%	97%	93%	91%	86%	85%	85%	65%	54%	53%	11%	10%	10%	10%	10%	3%

Appendix C. The percentage accuracy of all models on the holdout data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
DA_Step	**O**	51%	51%	51%	53%	53%	53%	53%	57%	57%	61%	57%	57%	53%	53%	51%	51%	51%	51%
	F	1%	1%	1%	9%	9%	9%	9%	25%	25%	50%	72%	72%	96%	96%	98%	98%	98%	99%
	L	100%	100%	100%	96%	96%	96%	96%	88%	88%	71%	43%	43%	9%	9%	5%	5%	5%	4%
LR_All	**O**	51%	51%	51%	53%	59%	59%	60%	59%	59%	62%	63%	64%	54%	54%	54%	54%	54%	54%
	F	2%	2%	2%	7%	28%	28%	30%	42%	42%	55%	64%	71%	96%	96%	96%	96%	96%	96%
	L	99%	99%	99%	98%	91%	91%	89%	76%	76%	69%	62%	57%	12%	12%	12%	12%	12%	12%
LR_U15%	**O**	51%	51%	51%	51%	54%	54%	58%	57%	57%	57%	58%	55%	54%	54%	54%	49%	49%	49%
	F	3%	3%	3%	3%	12%	13%	29%	29%	29%	49%	57%	94%	94%	94%	94%	98%	98%	98%
	L	100%	100%	100%	100%	96%	95%	86%	86%	86%	66%	59%	15%	14%	14%	14%	1%	1%	1%
LA_U15%LA	**O**	50%	50%	50%	50%	55%	55%	54%	54%	57%	59%	59%	55%	56%	50%	50%	50%	50%	50%
	F	1%	1%	1%	1%	17%	18%	22%	22%	30%	54%	54%	93%	95%	98%	98%	98%	98%	98%
	L	99%	99%	99%	99%	92%	91%	86%	86%	83%	64%	64%	17%	16%	3%	3%	3%	3%	3%
LR_U1	**O**	51%	51%	51%	52%	57%	57%	58%	60%	60%	62%	66%	66%	54%	54%	54%	53%	51%	51%
	F	3%	3%	3%	6%	25%	25%	34%	40%	40%	56%	75%	75%	96%	96%	96%	97%	99%	99%
	L	99%	99%	99%	99%	90%	90%	82%	80%	80%	69%	56%	56%	12%	12%	12%	9%	4%	4%
LR_Step	**O**	51%	51%	51%	51%	51%	54%	58%	61%	61%	62%	61%	58%	52%	52%	52%	52%	52%	50%
	F	2%	2%	2%	4%	4%	14%	30%	36%	36%	47%	67%	80%	97%	97%	97%	97%	97%	100%
	L	100%	100%	100%	98%	98%	95%	87%	85%	85%	76%	55%	36%	7%	7%	7%	7%	7%	0%
LR_MS_S	**O**	51%	51%	51%	51%	52%	52%	59%	59%	58%	60%	58%	59%	51%	51%	51%	51%	51%	51%
	F	4%	4%	4%	4%	7%	7%	37%	37%	40%	49%	56%	80%	97%	97%	97%	97%	97%	97%
	L	99%	99%	99%	99%	97%	97%	80%	80%	76%	71%	60%	37%	6%	6%	6%	6%	6%	6%

Appendix C. The percentage accuracy of all models on the holdout data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
LR_MS_F	**O**	51%	51%	51%	51%	53%	58%	58%	58%	58%	59%	64%	57%	55%	55%	54%	54%	54%	54%
	F	5%	5%	5%	5%	11%	30%	30%	30%	30%	51%	70%	90%	93%	93%	98%	98%	98%	98%
	L	97%	97%	97%	97%	95%	86%	86%	86%	86%	66%	58%	24%	17%	17%	11%	11%	11%	11%
NN_BK	**O**	50%	50%	50%	50%	54%	54%	54%	55%	55%	57%	55%	54%	54%	54%	50%	50%	50%	50%
	F	2%	2%	2%	2%	9%	24%	28%	31%	31%	56%	63%	91%	91%	91%	99%	99%	99%	99%
	L	98%	98%	98%	98%	98%	85%	81%	78%	78%	59%	47%	18%	18%	18%	1%	1%	1%	1%
NN_GA_1	**O**	51%	51%	51%	51%	51%	51%	51%	50%	51%	50%	50%	50%	50%	50%	50%	50%	50%	50%
	F	2%	2%	2%	2%	2%	2%	2%	5%	10%	15%	100%	100%	100%	100%	100%	100%	100%	100%
	L	99%	99%	99%	99%	99%	99%	99%	95%	92%	85%	0%	0%	0%	0%	0%	0%	0%	0%
NN_GA_5	**O**	56%	61%	61%	58%	58%	58%	58%	58%	58%	58%	58%	58%	58%	58%	58%	58%	58%	58%
	F	28%	62%	62%	64%	64%	64%	64%	64%	64%	65%	65%	67%	67%	67%	67%	67%	67%	70%
	L	83%	60%	60%	52%	52%	52%	52%	52%	52%	50%	50%	48%	48%	48%	48%	48%	48%	46%
DT_Suite	**O**	50%	50%	50%	50%	52%	51%	58%	59%	55%	67%	57%	51%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	10%	3%	49%	48%	47%	62%	70%	91%	97%	97%	97%	97%	100%	100%
	L	100%	100%	100%	100%	93%	98%	67%	70%	64%	72%	44%	12%	4%	4%	4%	4%	0%	0%
DT_One	**O**	50%	50%	50%	50%	50%	50%	50%	67%	67%	67%	67%	50%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	0%	62%	62%	62%	62%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	72%	72%	72%	72%	0%	0%	0%	0%	0%	0%	0%
RF_8	**O**	51%	51%	51%	51%	52%	55%	64%	65%	65%	66%	64%	64%	65%	65%	63%	63%	58%	53%
	F	2%	2%	2%	2%	7%	14%	37%	46%	46%	67%	82%	82%	85%	85%	91%	91%	99%	100%
	L	99%	99%	99%	99%	98%	96%	91%	85%	85%	64%	46%	46%	45%	45%	36%	36%	17%	6%

Appendix C. The percentage accuracy of all models on the holdout data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
RF_66	**O**	52%	52%	52%	52%	54%	61%	64%	66%	66%	66%	67%	64%	64%	64%	64%	64%	56%	56%
	F	5%	5%	5%	5%	9%	28%	46%	63%	63%	63%	66%	91%	91%	91%	91%	91%	100%	100%
	L	99%	99%	99%	99%	99%	94%	82%	69%	69%	69%	67%	36%	36%	36%	36%	36%	12%	12%
TN	**O**	50%	50%	50%	54%	58%	58%	64%	67%	67%	67%	64%	64%	64%	59%	59%	59%	59%	52%
	F	0%	0%	0%	8%	21%	23%	39%	50%	50%	55%	79%	79%	90%	98%	98%	98%	98%	100%
	L	100%	100%	100%	99%	96%	93%	88%	83%	83%	80%	50%	50%	39%	21%	21%	21%	21%	4%
NN_LR	**O**	51%	51%	51%	53%	54%	60%	64%	64%	64%	61%	60%	58%	56%	55%	53%	53%	53%	53%
	F	3%	3%	3%	9%	12%	30%	44%	44%	44%	51%	57%	89%	96%	96%	100%	100%	100%	100%
	L	99%	99%	99%	96%	96%	89%	83%	83%	83%	70%	63%	27%	15%	13%	5%	5%	5%	5%
DT_LR	**O**	62%	62%	66%	66%	66%	66%	66%	66%	66%	66%	66%	66%	66%	66%	66%	66%	66%	66%
	F	38%	38%	62%	62%	62%	62%	62%	62%	62%	62%	62%	62%	64%	64%	64%	64%	64%	64%
	L	87%	87%	71%	71%	71%	71%	71%	71%	71%	71%	71%	71%	68%	68%	68%	68%	68%	68%
DTnode_LR_Step	**O**	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	68%	68%	68%	67%	67%
	F	60%	60%	60%	62%	62%	62%	62%	62%	62%	63%	63%	63%	63%	68%	68%	68%	69%	69%
	L	74%	74%	74%	72%	72%	72%	72%	72%	72%	72%	72%	72%	72%	69%	69%	69%	64%	64%
NN-DTnode_LR_Step	**O**	67%	67%	67%	67%	66%	66%	66%	66%	66%	67%	67%	67%	67%	67%	67%	67%	66%	66%
	F	60%	60%	60%	60%	62%	62%	62%	62%	62%	63%	63%	63%	63%	63%	63%	63%	66%	66%
	L	73%	73%	73%	73%	70%	70%	70%	70%	70%	70%	70%	70%	70%	70%	70%	70%	66%	66%
Vote5	**O**	51%	51%	51%	51%	53%	55%	60%	67%	67%	68%	64%	62%	57%	55%	53%	53%	52%	51%
	F	1%	1%	1%	1%	7%	11%	24%	45%	45%	60%	75%	89%	96%	98%	100%	100%	100%	100%
	L	100%	100%	100%	100%	99%	99%	97%	88%	88%	76%	54%	34%	17%	11%	5%	5%	4%	1%

Appendix C. The percentage accuracy of all models on the holdout data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
Vote3_RF_TN_DT	**O**	50%	50%	50%	50%	52%	54%	61%	68%	68%	71%	65%	64%	64%	59%	59%	59%	57%	51%
	F	0%	0%	0%	1%	7%	12%	28%	50%	50%	66%	80%	86%	91%	98%	98%	98%	99%	100%
	L	100%	100%	100%	99%	98%	96%	93%	87%	87%	76%	50%	43%	37%	21%	20%	20%	15%	3%
Vote3_RF_TN_DA	**O**	51%	51%	51%	51%	55%	57%	64%	66%	66%	67%	63%	63%	64%	59%	59%	59%	58%	51%
	F	2%	2%	2%	4%	13%	17%	34%	44%	44%	60%	79%	80%	90%	95%	96%	97%	99%	100%
	L	100%	100%	100%	99%	98%	96%	93%	88%	87%	75%	47%	47%	38%	22%	22%	20%	17%	3%
Vote3_RF_TN_NN	**O**	51%	51%	51%	50%	53%	57%	63%	65%	65%	65%	65%	65%	64%	59%	59%	59%	58%	51%
	F	1%	1%	1%	1%	9%	17%	34%	44%	44%	60%	81%	85%	90%	94%	98%	98%	99%	100%
	L	100%	100%	100%	99%	98%	96%	92%	86%	86%	70%	49%	45%	38%	24%	20%	20%	16%	3%
AV5_NoNN	**O**	51%	51%	51%	51%	54%	56%	59%	62%	63%	70%	66%	63%	55%	53%	53%	53%	52%	51%
	F	1%	1%	1%	2%	9%	14%	20%	30%	33%	75%	85%	89%	98%	99%	99%	100%	100%	100%
	L	100%	100%	100%	100%	99%	98%	98%	93%	92%	66%	46%	37%	12%	7%	7%	6%	4%	1%
AV2_RF_TN	**O**	50%	50%	50%	51%	56%	58%	64%	67%	67%	67%	66%	66%	64%	62%	62%	62%	59%	52%
	F	1%	1%	1%	2%	14%	20%	39%	50%	50%	57%	83%	83%	89%	95%	96%	96%	99%	100%
	L	99%	99%	99%	99%	98%	96%	89%	85%	85%	76%	48%	48%	40%	30%	29%	29%	18%	4%
AV3_RF_TN_DT	**O**	50%	50%	50%	51%	55%	56%	62%	65%	65%	70%	64%	63%	61%	59%	58%	58%	56%	51%
	F	1%	1%	1%	2%	11%	14%	28%	38%	38%	78%	85%	89%	93%	96%	98%	98%	99%	100%
	L	99%	99%	99%	99%	99%	97%	96%	92%	92%	62%	44%	36%	30%	21%	19%	19%	14%	2%
DT_One_DA	**O**	50%	50%	50%	50%	50%	50%	50%	66%	66%	67%	66%	50%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	0%	60%	60%	62%	64%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	73%	73%	72%	68%	0%	0%	0%	0%	0%	0%	0%

Appendix C. The percentage accuracy of all models on the holdout data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
Simpler Models (discussed in Chapter 6)																			
TN_90%_6	**O**	51%	51%	51%	51%	56%	58%	63%	63%	63%	63%	60%	60%	60%	57%	57%	57%	57%	50%
	F	2%	2%	2%	2%	14%	25%	49%	49%	49%	49%	88%	88%	88%	94%	94%	94%	94%	99%
	L	100%	100%	100%	100%	98%	91%	77%	77%	77%	77%	31%	31%	31%	20%	20%	20%	20%	1%
TN_4	**O**	50%	50%	50%	50%	53%	53%	54%	55%	55%	54%	52%	54%	50%	50%	50%	50%	50%	50%
	F	1%	1%	1%	1%	20%	20%	24%	37%	37%	48%	79%	88%	98%	98%	98%	98%	98%	98%
	L	100%	100%	100%	100%	86%	86%	84%	73%	73%	59%	25%	20%	2%	2%	2%	2%	2%	2%
LR_Step_11	**O**	51%	51%	51%	51%	51%	51%	51%	57%	58%	64%	57%	57%	51%	50%	50%	50%	50%	50%
	F	1%	1%	1%	3%	4%	7%	7%	25%	40%	63%	82%	82%	98%	100%	100%	100%	100%	100%
	L	100%	100%	100%	98%	98%	96%	96%	88%	76%	66%	32%	32%	4%	1%	1%	1%	1%	1%
LR_Int_11	**O**	51%	51%	51%	51%	51%	52%	58%	58%	57%	59%	56%	56%	54%	54%	51%	51%	51%	51%
	F	1%	1%	1%	5%	6%	8%	29%	29%	32%	57%	70%	98%	98%	98%	98%	98%	98%	98%
	L	100%	100%	100%	98%	97%	95%	86%	86%	81%	61%	42%	14%	11%	11%	3%	3%	3%	3%
LR_TN_6	**O**	51%	51%	51%	51%	51%	52%	52%	52%	52%	59%	54%	54%	51%	51%	51%	51%	50%	50%
	F	3%	3%	3%	3%	3%	6%	6%	6%	6%	79%	92%	92%	95%	95%	95%	95%	98%	98%
	L	100%	100%	100%	100%	100%	98%	98%	98%	98%	40%	17%	17%	7%	7%	7%	7%	2%	2%

Appendix D. The percentage accuracy of all models on the training data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
Naïve	**O**	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%
F-score	**O**	50%	50%	50%	50%	50%	50%	50%	55%	55%	59%	54%	54%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	1%	18%	18%	41%	98%	98%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	92%	92%	78%	9%	9%	0%	0%	0%	0%	0%	0%
M-score	**O**	50%	50%	50%	50%	50%	50%	50%	50%	50%	56%	54%	50%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	0%	0%	2%	71%	94%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	99%	41%	15%	0%	0%	0%	0%	0%	0%	0%
DA_All	**O**	54%	54%	54%	57%	69%	69%	70%	70%	74%	76%	75%	73%	59%	56%	56%	55%	55%	53%
	F	7%	7%	7%	16%	43%	45%	46%	46%	64%	77%	83%	86%	98%	99%	99%	100%	100%	100%
	L	100%	100%	100%	99%	94%	94%	94%	94%	84%	75%	67%	59%	20%	12%	12%	9%	9%	6%
DA_U15%	**O**	53%	53%	53%	54%	59%	59%	68%	68%	68%	72%	70%	58%	55%	52%	52%	52%	52%	52%
	F	5%	5%	5%	9%	20%	20%	46%	46%	46%	74%	79%	96%	98%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	98%	98%	89%	89%	89%	69%	61%	21%	13%	3%	3%	3%	3%	3%
DA_U15%LA	**O**	52%	52%	52%	52%	61%	64%	64%	65%	65%	71%	70%	56%	56%	53%	53%	53%	53%	53%
	F	4%	4%	4%	4%	26%	35%	35%	39%	39%	73%	80%	98%	99%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	96%	93%	93%	92%	92%	68%	60%	15%	13%	6%	6%	6%	6%	6%
DA_U1	**O**	51%	51%	51%	57%	62%	65%	69%	71%	71%	73%	72%	71%	56%	55%	55%	55%	55%	53%
	F	3%	3%	3%	15%	27%	35%	49%	54%	54%	78%	87%	88%	99%	100%	100%	100%	100%	100%
	L	100%	100%	100%	99%	97%	95%	90%	88%	88%	68%	56%	54%	13%	11%	11%	11%	11%	7%

Appendix D. The percentage accuracy of all models on the training data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
DA_Step	**O**	50%	50%	50%	60%	60%	60%	60%	67%	67%	70%	66%	66%	56%	56%	55%	55%	55%	53%
	F	1%	1%	1%	21%	21%	21%	21%	46%	46%	71%	91%	91%	99%	99%	100%	100%	100%	100%
	L	100%	100%	100%	99%	99%	99%	99%	87%	87%	70%	42%	42%	12%	12%	9%	9%	9%	6%
LR_All	**O**	53%	53%	53%	57%	67%	67%	69%	74%	74%	76%	75%	73%	58%	58%	58%	58%	58%	58%
	F	6%	6%	6%	14%	39%	39%	45%	65%	65%	78%	84%	88%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	99%	95%	95%	93%	84%	84%	74%	66%	59%	16%	16%	16%	16%	16%	16%
LR_U15%	**O**	52%	52%	52%	52%	59%	59%	67%	67%	67%	71%	70%	57%	56%	56%	56%	52%	52%	52%
	F	5%	5%	5%	5%	20%	22%	45%	45%	45%	73%	80%	98%	99%	99%	99%	100%	100%	100%
	L	100%	100%	100%	100%	98%	97%	89%	89%	89%	70%	60%	15%	14%	14%	14%	4%	4%	4%
LA_U15%LA	**O**	52%	52%	52%	52%	61%	62%	65%	65%	67%	71%	71%	57%	56%	53%	53%	53%	53%	53%
	F	4%	4%	4%	4%	26%	28%	37%	37%	46%	80%	80%	98%	99%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	96%	95%	92%	92%	87%	62%	62%	16%	13%	6%	6%	6%	6%	6%
LR_U1	**O**	52%	52%	52%	55%	65%	65%	70%	73%	73%	75%	72%	72%	58%	58%	58%	55%	53%	53%
	F	4%	4%	4%	11%	34%	34%	52%	59%	59%	76%	87%	87%	99%	99%	99%	100%	100%	100%
	L	100%	100%	100%	99%	95%	95%	89%	86%	86%	73%	56%	56%	18%	18%	18%	9%	6%	6%
LR_Step	**O**	51%	51%	51%	55%	55%	63%	69%	71%	71%	73%	69%	65%	55%	55%	55%	55%	55%	51%
	F	1%	1%	1%	12%	12%	30%	51%	57%	57%	71%	84%	91%	99%	99%	99%	99%	99%	100%
	L	100%	100%	100%	99%	99%	95%	88%	85%	85%	76%	53%	40%	11%	11%	11%	11%	11%	3%
LR_MS_S	**O**	55%	55%	55%	57%	61%	61%	71%	71%	72%	74%	74%	65%	55%	54%	54%	54%	54%	54%
	F	11%	11%	11%	14%	23%	23%	55%	55%	60%	70%	83%	92%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	98%	98%	87%	87%	84%	78%	64%	37%	11%	9%	9%	9%	9%	9%

Appendix D. The percentage accuracy of all models on the training data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
LR_MS_F	**O**	54%	54%	54%	54%	61%	70%	70%	71%	71%	73%	71%	62%	59%	59%	55%	55%	55%	55%
	F	9%	9%	9%	9%	24%	49%	49%	51%	51%	76%	85%	96%	98%	98%	100%	100%	100%	100%
	L	100%	100%	100%	100%	97%	91%	91%	91%	91%	71%	58%	28%	20%	20%	11%	11%	11%	11%
NN_BK	**O**	52%	52%	52%	52%	57%	64%	66%	68%	68%	70%	68%	58%	58%	58%	51%	51%	51%	51%
	F	4%	4%	4%	4%	15%	37%	41%	48%	48%	75%	80%	98%	98%	98%	100%	100%	100%	100%
	L	100%	100%	100%	100%	98%	92%	90%	87%	87%	64%	55%	18%	18%	18%	2%	2%	2%	2%
NN_GA_1	**O**	53%	53%	53%	53%	53%	53%	53%	54%	56%	57%	50%	50%	50%	50%	50%	50%	50%	50%
	F	6%	6%	6%	6%	6%	6%	6%	10%	17%	26%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	98%	95%	87%	0%	0%	0%	0%	0%	0%	0%	0%
NN_GA_5	**O**	62%	93%	93%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	97%	94%
	F	24%	86%	86%	95%	95%	95%	95%	95%	95%	97%	97%	98%	98%	98%	98%	98%	98%	99%
	L	100%	100%	100%	99%	99%	99%	99%	99%	99%	98%	98%	95%	95%	95%	95%	95%	95%	90%
DT_Suite	**O**	50%	50%	50%	50%	58%	55%	91%	86%	83%	98%	76%	61%	54%	54%	54%	54%	50%	50%
	F	0%	0%	0%	0%	20%	11%	84%	80%	76%	97%	93%	98%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	96%	100%	98%	92%	90%	99%	60%	24%	9%	9%	9%	9%	0%	0%
DT_One	**O**	50%	50%	50%	50%	50%	50%	50%	98%	98%	98%	98%	50%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	0%	97%	97%	97%	97%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	99%	99%	99%	99%	0%	0%	0%	0%	0%	0%	0%
RF_8	**O**	99%	99%	99%	99%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	97%
	F	98%	98%	98%	98%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	94%

Appendix D. The percentage accuracy of all models on the training data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
RF_66	**O**	98%	98%	98%	98%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	97%	97%
	F	95%	95%	95%	95%	99%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	93%	93%
TN	**O**	50%	50%	50%	89%	100%	100%	100%	100%	100%	100%	100%	100%	100%	96%	96%	96%	96%	70%
	F	1%	1%	1%	78%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	92%	92%	92%	92%	41%
NN_LR	**O**	53%	53%	53%	58%	59%	68%	75%	75%	75%	77%	76%	65%	59%	58%	55%	55%	55%	55%
	F	7%	7%	7%	16%	19%	42%	63%	63%	63%	75%	82%	94%	98%	99%	100%	100%	100%	100%
	L	100%	100%	100%	99%	99%	94%	87%	87%	87%	78%	69%	37%	20%	18%	10%	10%	10%	10%
DT_LR	**O**	85%	85%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%
	F	71%	71%	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	98%	98%	98%	98%	98%	98%
DTnode_LR_Step	**O**	95%	95%	95%	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%	96%	96%	96%	95%	95%
	F	89%	89%	89%	97%	97%	97%	97%	97%	97%	98%	98%	98%	98%	99%	99%	99%	99%	99%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	99%	99%	99%	99%	93%	93%	93%	90%	90%
NN-DTnode_LR_Step	**O**	97%	97%	97%	97%	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%	96%	96%
	F	95%	95%	95%	95%	97%	97%	98%	98%	98%	98%	99%	99%	99%	99%	99%	99%	99%	99%
	L	100%	100%	100%	100%	100%	100%	99%	99%	99%	99%	98%	98%	98%	98%	98%	98%	93%	93%
Vote5	**O**	51%	51%	51%	57%	70%	75%	78%	99%	99%	100%	100%	80%	63%	62%	58%	58%	58%	55%
	F	2%	2%	2%	13%	40%	51%	55%	99%	99%	99%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	61%	25%	24%	15%	15%	15%	11%

Appendix D. The percentage accuracy of all models on the training data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
Vote3_RF_TN_DT	**O**	50%	50%	50%	89%	100%	100%	100%	100%	100%	100%	100%	100%	100%	96%	96%	96%	96%	70%
	F	1%	1%	1%	78%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	92%	92%	92%	92%	41%
Vote3_RF_TN_DA	**O**	54%	54%	54%	90%	100%	100%	100%	100%	100%	100%	100%	100%	100%	96%	96%	96%	96%	71%
	F	8%	8%	8%	79%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	92%	92%	92%	92%	41%
Vote3_RF_TN_NN	**O**	52%	52%	52%	89%	100%	100%	100%	100%	100%	100%	100%	100%	100%	96%	96%	96%	96%	70%
	F	5%	5%	5%	79%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	92%	92%	92%	92%	41%
AV5_NoNN	**O**	56%	56%	56%	63%	80%	84%	90%	94%	95%	99%	96%	92%	75%	68%	67%	66%	64%	57%
	F	12%	12%	12%	25%	60%	69%	80%	88%	91%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	98%	93%	83%	50%	36%	34%	32%	27%	15%
AV2_RF_TN	**O**	91%	91%	91%	98%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	99%	92%
	F	81%	81%	81%	96%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	99%	84%
AV3_RF_TN_DT	**O**	88%	88%	88%	96%	99%	99%	99%	99%	99%	100%	100%	100%	100%	99%	99%	99%	98%	89%
	F	77%	77%	77%	92%	97%	97%	97%	97%	97%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	99%	99%	99%	99%	99%	99%	99%	97%	78%
DT_One_DA	**O**	50%	50%	50%	50%	50%	50%	50%	99%	99%	98%	99%	50%	50%	50%	50%	50%	50%	50%
	F	0%	0%	0%	0%	0%	0%	0%	97%	97%	97%	99%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	99%	99%	0%	0%	0%	0%	0%	0%	0%

Appendix D. The percentage accuracy of all models on the training data. **F**/**L**/**O** = Percentage of **F**raudulent/**L**egitimate/**O**verall statements correctly classified.

		Prior-adjusted Relative Cost of Missing Fraud ($PaRC_{IF}$)																	
Model		**0.004**	**0.01**	**0.04**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**1**	**2**	**3**	**5**	**6**	**7**	**8**	**10**	**17**
Simpler Models (discussed in Chapter 6)																			
TN_90%_6	**O**	54%	54%	54%	54%	77%	92%	99%	99%	99%	99%	92%	92%	92%	77%	77%	77%	77%	51%
	F	7%	7%	7%	7%	55%	84%	98%	98%	98%	98%	100%	100%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	83%	83%	83%	53%	53%	53%	53%	2%
TN_4	**O**	50%	50%	50%	50%	76%	76%	79%	86%	86%	93%	85%	78%	51%	51%	51%	51%	51%	51%
	F	1%	1%	1%	1%	52%	52%	59%	74%	74%	90%	99%	99%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	99%	99%	99%	99%	99%	96%	70%	57%	2%	2%	2%	2%	2%	2%
LR_Step_11	**O**	51%	51%	51%	54%	55%	58%	58%	65%	69%	70%	66%	66%	52%	52%	52%	52%	52%	52%
	F	1%	1%	1%	8%	12%	19%	19%	42%	62%	77%	92%	92%	100%	100%	100%	100%	100%	100%
	L	100%	100%	100%	100%	99%	97%	97%	87%	77%	62%	41%	41%	5%	3%	3%	3%	3%	3%
LR_Int_11	**O**	51%	51%	51%	54%	56%	58%	65%	65%	67%	68%	67%	57%	55%	55%	52%	52%	52%	52%
	F	2%	2%	2%	7%	13%	20%	42%	42%	50%	74%	85%	98%	99%	99%	100%	100%	100%	100%
	L	100%	100%	100%	100%	99%	97%	89%	89%	84%	63%	50%	15%	12%	12%	4%	4%	4%	4%
LR_TN_6	**O**	52%	52%	52%	52%	52%	55%	55%	55%	55%	58%	56%	56%	54%	54%	54%	54%	51%	51%
	F	3%	3%	3%	3%	3%	14%	14%	14%	14%	86%	98%	98%	99%	99%	99%	99%	100%	100%
	L	100%	100%	100%	100%	100%	97%	97%	97%	97%	31%	14%	14%	8%	8%	8%	8%	1%	1%