Journal of Forensic and Investigative Accounting
Volume 16: Issue 1, January–June 2024

Detecting Financial Statement Fraud: An Alternative Evaluation of Automated Tools Using Portfolio Performance

Amanda Yee
Adrian Gepp
Kuldeep Kumar
James Todd
*Bruce Vanstone**

Introduction

This research combines the fields of accounting fraud detection and investment, the intersection of which is largely unexplored in research, to evaluate financial statement fraud detection models in an investment context and thus complement traditional evaluation. Financial statement fraud involves the intentional publication of false or misleading information in financial statements. Large-scale fraud cases such as Enron, WorldCom, Theranos, and Satyam (India) have made news headlines around the world. Far from being victimless crimes, such frauds leave behind real financial losses borne by individuals. Not only is there a negative effect on the company involved (Dyck et al., 2021), but also on stakeholders such as employees and investors, and further on the whole economy through the misallocation of resources (Kedia et al., 2009). Despite efforts to measure the cost of fraud, its true cost is difficult to determine because much of it goes undetected. However, a staggering value of over $4.7 trillion USD was projected as the total global fraud loss in 2021 (ACFE, 2022). This ACFE study focused on three major categories: asset misappropriation, corruption, and financial statement fraud. Out of the three, financial statement fraud has the lowest frequency, but the highest severity per incident.

Motivated by the costs of fraud and the need for better detection tools, research has aimed to develop fraud detection models to assist auditors (Chiu et al., 2020). While various models have been developed, their comparison and determination of which are most effective has been made challenging by inter-study differences relating to datasets, explanatory variables, time horizons and cost ratios. The evaluation is further hindered by the fact that not all fraud is detected, making it difficult to estimate the accuracy of such models. There is some evidence, however, that there is an effect on financial returns even when fraud is undetected, at least with respect to earnings manipulation (Beneish et al., 2013). To the best of our knowledge, Gepp et al. (2021) have conducted the most extensive comparison thus far, empirically evaluating different techniques used in the literature on the same data from the U.S., finding an ensemble of models afforded the best performance.

Combining the fields of financial statement fraud detection and investment, our primary research question is: "What is the financial impact of investing in companies according to predictions of current fraud detection models?" This question will be evaluated using the established M-Score (Beneish, 1997; Beneish, 1999) and F-Score (Dechow et al., 2011) models, as well as the model developed by Gepp et al. (2021) leveraging both traditional statistical and machine learning techniques, referred to hereafter as the G-Score. The use of three fraud risk models also supports the secondary research question: "Which fraud risk model is most effective for guiding investment decisions?" The three fraud risk models are evaluated within two investment strategies focusing on companies designated as having a high or low relative risk of financial statement fraud. This study's findings will be of relevance to research in financial statement fraud detection and investment, as well as to investors.

The remainder of this article is structured as follows. In Section two, existing literature in both the areas of financial statement fraud and investment are reviewed. Section three presents the research methodology, including the data used, the fraud models and the investment strategies. Section four compares the investment strategies tested and presents the main findings of the research. Section 0 discusses the results presented and provides interpretation, before Section 0 presents the overall conclusions, contributions, and areas for future research.

Amanda Yee, Adrian Gepp, Kuldeep Kumar, James Todd, and Bruce James Vanstone
Centre for Data Analytics, Bond University, Gold Coast, Queensland, Australia
Bangor Business School, Bangor University, Wales, United Kingdom

Summary of Related Research

Related research encompasses research into the prevalence of fraud, detection techniques, and the subsequent effect on investment outcomes and decision making.

Fraud and its Cost

The issue of detection accuracy has historically made it difficult to measure the extent of the financial effect of fraud, but increased attention and greater reporting requirements have improved estimates (Gee et al., 2021). A report published in a collaboration between Crowe UK and the Centre for Counter Fraud Studies at University of Portsmouth estimated the global losses from fraud at $5.38 trillion USD based on a review of 807 loss measurement exercises (Gee et al., 2021). A separate report by the Association of Certified Fraud Examiners (ACFE) estimated $4.7 trillion USD as the global fraud loss in 2021 (ACFE, 2022). Beyond these direct financial costs, fraudulent activity can also reduce the efficiency (Perols, 2011) and liquidity (Dechow et al., 2011) in financial markets.

At the firm level, fraud has direct costs such as regulatory fines and private lawsuits, as well as indirect reputational costs. The reputational costs have been found to be the more substantial of the two (Karpoff et al., 1993), with Armour et al. (2017) finding that reputational costs were almost nine times the size of fines in the United Kingdom. The aggregate cost to overall firm value is substantial; Dyck et al. (2021) found that a fifth of firm value was lost on average after financial statement fraud, a finding made more robust by a similar finding of a 29 percent equity value loss by Karpoff et al. (2008).

The damage fraud does at the individual and social level is also evident. Taking the Enron case as an example, not only did its employees lose their jobs and retirement savings, but many employees at their auditor Arthur Andersen faced similar job and financial losses (Sridharan et al., 2002). Financial losses were also inflicted on the retirement savings of many other people with no direct contact with Enron because they were invested in mutual funds and index funds that in-turn invested in Enron (Sridharan et al., 2002).

Both the magnitude and prevalence of financial fraud appear to have increased over time (Gee et al., 2021; Beasley et al., 2010), exacerbated by increased globalization, complexity of business processes, and the COVID-19 pandemic.

Detecting Financial Statement Fraud

There are research findings that can help to reduce fraud, such as having female representation on company boards (Capezio et al., 2015; Maulidi, 2023), having more independence on audit committees (Khamainy et al., 2022), preventing single entity ownership (Avortri et al., 2020), and having effective corporate gatekeepers (Rezaee et al., 2012). Unfortunately, with perpetrators becoming more advanced and continuously adapting their methods, there are many instances where the prevention mechanism fails.

By analysing the financial and non-financial information of a company, as well as other industry and economic data, many models to detect fraudulent activities have been developed. Due to the large volumes of data available and potentially relevant for fraud detection, statistical and computational methods have been prominent (West et al., 2016). Research over the past two decades has focused on applying a variety of data analytics techniques to detect financial statement fraud. Such models assess the likelihood of fraud having occurred, which can in turn be used to prioritise investigations or potentially make investment decisions.

The most well-known models used for detecting financial statement fraud are the M-Score (Beneish, 1999; Beneish, 1997) and F-Score (Dechow et al., 2011), produced from probit analysis and logistic regression respectively. Both represent established benchmarks against which new models can be compared; see Repousis (2016) as an example. A variety of modelling techniques have been employed to develop new predictive models, including several types of artificial neural networks (Fanning et al., 1998; Hoogs et al., 2007; Huang et al., 2014), support vector machines (Sadasivam et al., 2016), individual and ensembles of decision trees (Perols, 2011), and hybrids of multiple techniques (Chen, 2016). In addition to the modelling techniques themselves, research has explored other aspects of the problem such as parameter estimation with genetic algorithms (Hoogs et al., 2007; Bhattacharya et al., 2011) and inclusion of textual information (Purda et al., 2015). Some evidence indicates that ensembles of techniques offer advantages over individual techniques (McKee, 2009; Whiting et al., 2012), but generalisation of results is made more difficult by differences in data and evaluation methods employed in different studies. Data differences include both covariates considered and the study setting. The U.S. has been the most common setting for studies, with research less often considering other regions such as Britain (Lokanan, 2021), Greece (Kirkos et al., 2007), and China (Ravisankar et al., 2011) with distinct characteristics (e.g., regulatory environments).

In a recent study, Gepp et al. (2021) performed an extensive comparison. Informed by prior research, including the fraud triangle and the many extensions and adaptations since, over thirty models were estimated using an extensive range of both modelling techniques and explanatory variables; all models estimated were then compared on a consistent basis in terms of dataset and evaluation. Consistent with the results of prior work (Whiting et al., 2012; McKee, 2009), the best model overall was found to be an ensemble of five underlying models using different modelling techniques. This best model was shown to substantially outperform the benchmark M-Score and F-Score models in Table 6 of that paper using new holdout data that were not used during model development. Notably, this comparative study focused exclusively on the ability to detect fraud and did not investigate implications for portfolio performance and financial markets in any way.

Factoring in Fraud When Investing

Financial statement fraud has been a regular cause of inefficient investment allocation (Kumar et al., 2009). Investors are deceived by optimistic representations of financial performance and company prospects, and as a result, allocate part of their investment to these misleading companies. Beyond undermining the ability of investors to make informed decisions in portfolio construction, research suggests that fraud reduces firm value whether or not it is detected (Dyck et al., 2021; Beneish et al., 2013), representing additional explicit costs to investors.

Investors aiming to avoid fraudulent firms need to identify these firms prior to the conclusion of a multiple-year investigation by the U.S. Securities and Exchange Commission (SEC) (Gepp et al., 2021). Brazel et al. (2015) consider that investors have several types of fraud red flags available, such as evidence of insider trading or whether an SEC investigation has occurred into the company. By using these red flags to avoid certain companies, a positive impact is made on the investors' portfolio returns.

Summary

Beyond the research papers mentioned in this section, there is little literature in the field connecting fraud and investment performance. The research that has been conducted suggests that the inclusion of fraudulent companies in a portfolio is detrimental to investment performance. While models predicting financial statement fraud have been the subject of research, the effect of using these models when forming investment portfolios remains unknown. This research aims to address this gap by investigating the effect on portfolio performance of using fraud detection models for investment decisions, both to better understand the investment utility of these models and to compare available models.

Materials and Methods

This study incorporates three fraud detection models—the M-Score (Beneish, 1999; Beneish, 1997), the F-Score (Dechow et al., 2011), and the G-Score (Gepp et al., 2021)—into investment strategies by using each to assign a 'fraud score' to companies on the SandP 500 index. The M-Score and F-Score represent well-established tools, whereas the G-Score was recently developed using modern data analytics techniques in an extensive comparison. These scores are used to rank each firm in terms of their likelihood of having committed financial statement fraud. For each fraud model, two portfolios are created by investing only in companies with low (high) fraud propensity scores. The two portfolios are compared based on investment performance to assess whether financial statement fraud risk, whether low or high, affects investment performance. The portfolios are also compared between models to assess the differential value of the models in an investment context. Relevant details of the models used, implementation of strategies, and basis for assessment are detailed in the following subsections.

Fraud Models Recreation

Using static equations, the M-Score and F-Score models use probit and logistic regression respectively to relate company characteristics and balance sheet information to a risk of fraud. Because no modification to the published equations is required, and because the data to construct the variables used in these models is a subset of the information used in the G-Score of Gepp et al. (2021), this latter model's implementation is described here. Readers are directed to the original development of the M-Score (Beneish, 1999; Beneish, 1997) and F-Score (Dechow et al., 2011) for further details of these models.

The G-Score model of Gepp et al. (2021) was developed using data on publicly listed companies in the U.S. from 1998 to 2002 for model training. Due to the nature of the final model, which combined statistical and machine learning techniques in an ensemble, a static equation is not available. Instead, the methodology of Gepp et al. (2021) is mirrored in this work through consideration of companies in the SandP 500 benchmark index and with data from 1998 to 2002 again used for training.

Regarding variables used, the model comprises 50 variables spanning financial, non-financial, comparison between financial and non-financial, macroeconomic, governance and those controlling for company size, age, industry, and listing exchange. Table 3 in Gepp et al. (2021) lists all the variables and their corresponding data source, primarily Capital IQ and Bloomberg. The same data sources were used in this work. These variables included those required for the M-Score and F-Score models.

Regarding model implementation, the original fraud detection model was constructed using multiple proprietary software platforms (Gepp et al., 2021) but was recreated using the freely available R programming language (R Core Team, 2020) and same overall process in this work. The resulting models were checked for similarity with those produced using the original software. The final fraud detection model is an ensemble averaging the fraud-probability outputs of five other models, each using a different technique: Logistic Regression (LR), Linear Discriminant Analysis (DA), Random Forest (RF), Stochastic Gradient Boosting (GBM) and a CART Decision Tree (Tree).

Sample Details

The G-Score is implemented using data from 1998 to 2002. Consequently, all trading strategies are considered from 2003 onwards. Specifically, the investment strategies were implemented using quarterly price data from Bloomberg and were restricted to all companies on the SandP 500 Index from 2003 to 2021. For illustrative equity curves, strategies started with \$1 million in capital. Sub-sample analysis is also done for the Global Financial Crisis (GFC) from 2007 to 2009 and the COVID-19 pandemic for 2020 and 2021.

Investment Strategy Implementation

For each of the three fraud models, two investment strategies were implemented. In the first strategy, the companies considered least likely to be fraudulent by the chosen model are invested in, whereas the second strategy invests in the companies considered most likely to be fraudulent. These differing strategies were used to determine the effect of fraud risk on portfolio performance by incorporating a financial statement fraud model into an investment strategy and the difference in effect for different fraud models.

Every quarter, all companies on the SandP 500 index are ranked according to their assigned fraud score. To allow for wide dissemination of information in the financial statements, information must be available for three months before being used in the calculation of fraud scores. This ranking is used in the two investment strategies, which represent an emphasis on either the Least Likely Fraudulent (LLF) or Most Likely Fraudulent (MLF) companies on the SandP 500 index. The effective trading duration is from the end of March 2003 until the end of December 2021.

Companies are included in the LLF (MLF) portfolio if the company is in the SandP 500 index and was in the bottom (top) decile of fraud scores for the current model in the last 36 months. Rebalancing occurs quarterly, with equal weighting of all companies, and all trades are based on monthly closing prices.

Several decisions that are apparent in the description of the strategies bear justification. First, the 36-month lookback for each company reflects the fact that fraud often goes undetected for months or years, and thus its effect may not be immediately realised. Second, deciles are used, rather than more commonly considered terciles, to restrict the number of positions taken with the goal of concentrating the effects. The persistence of buying signals for 36 months results in more than 10 percent of companies being held at a given point.

The trading strategy implementation did not consider transaction costs or imperfectly divisible units. This decision was made to ensure results reflected the fraud risk of the two portfolios without distortion from market frictions.

Analysis of Investment Performance

The financial performance of the LLF and MLF strategies is assessed to determine the effects of using fraud models to direct investments towards companies that are least or most likely to have fraudulent financial statements. This assessment is conducted for LLF and MLF strategies using each of the three fraud models.

Overall performance is assessed first in terms of each portfolio's annualised returns, standard deviation, and Sharpe ratio. These measures are complemented with equity curves. Next, the risk-adjusted returns for each scenario are evaluated using the Fama-French three-factor model. The Fama-French model regresses returns above the risk-free rate against the broader market, as well as Small Minus Big (SMB) and High Minus Low (HML) factors. The regression intercepts (alphas) represent the returns earned that are not explained by exposures to the three risk factors. The weighting of the risk factors gives insight into the type of companies identified as being least or most likely fraudulent according to the different risk models.

In addition to strategy performance across the entire trading period, the evolution of the trading strategies is also considered in the context of a significant market shock, namely the Global Financial Crisis (GFC). Equity curves associated with the LLF and MLF strategies will be presented for each risk model from 2007 to 2009 to assess the sensitivity of these strategies to such a significant market shock.

Monthly share price data were obtained from Bloomberg for all companies which were in the SandP 500 at any point between 1998 and 2021. Data required to estimate the regression models were obtained from Kenneth French's data library (French, 2022).

Results

Robustness checks were considered for elements of model construction related to minor differences between implementations of GBMs in the proprietary software used by Gepp et al. (2021) and the freely available software used in this work. The sensitivity of findings to strategy details was assessed by reproducing results using terciles instead of deciles. No material differences in results were found with these variations. Throughout the results, the combination of strategy (i.e., LLF or MLF) and risk model (i.e., M-, F-, and G-Score) are referred to in abbreviated form as risk model-strategy (e.g., M-LLF means the LLF strategy with the M-Score).

Analysis if Changes in Equity and Summary Statistics

Figure 1 shows the equity curves of the LLF and MLF strategies from 2003 to 2021, replicated for each risk model. Table 1presents summary statistics for the returns received in each scenario.

Two features are evident in Figure 1, corresponding to the comparison of low-risk and high-risk strategies and the comparison of risk models. First, the peaks and troughs of the LLF and MLF strategies, as well the SandP500 overall, are mirrored across the period as the strategies have similar responses to market conditions, but the LLF strategies are superior regardless of the underlying risk model. Second, the difference in equity curves for the LLF and MLF strategies is most prominent and consistent when the M-Score underpins strategies. The equity of the LLF strategy promptly exceeds that of the MLF strategy, and this differential remains across the 19 years.

Consistent with the equity curves, Table 1 shows that the annualized and risk-adjusted returns of each LLF strategy are greater than the corresponding MLF strategies. Interestingly, and less obvious in the equity curves, the greatest volatility is associated with the LLF strategy underpinned by the M-Score. The LLF strategy underpinned by the G-Score has the greatest Sharpe ratio.

Figure 1: Equity Holding of LLF and MLF Strategies with SandP500 Benchmark

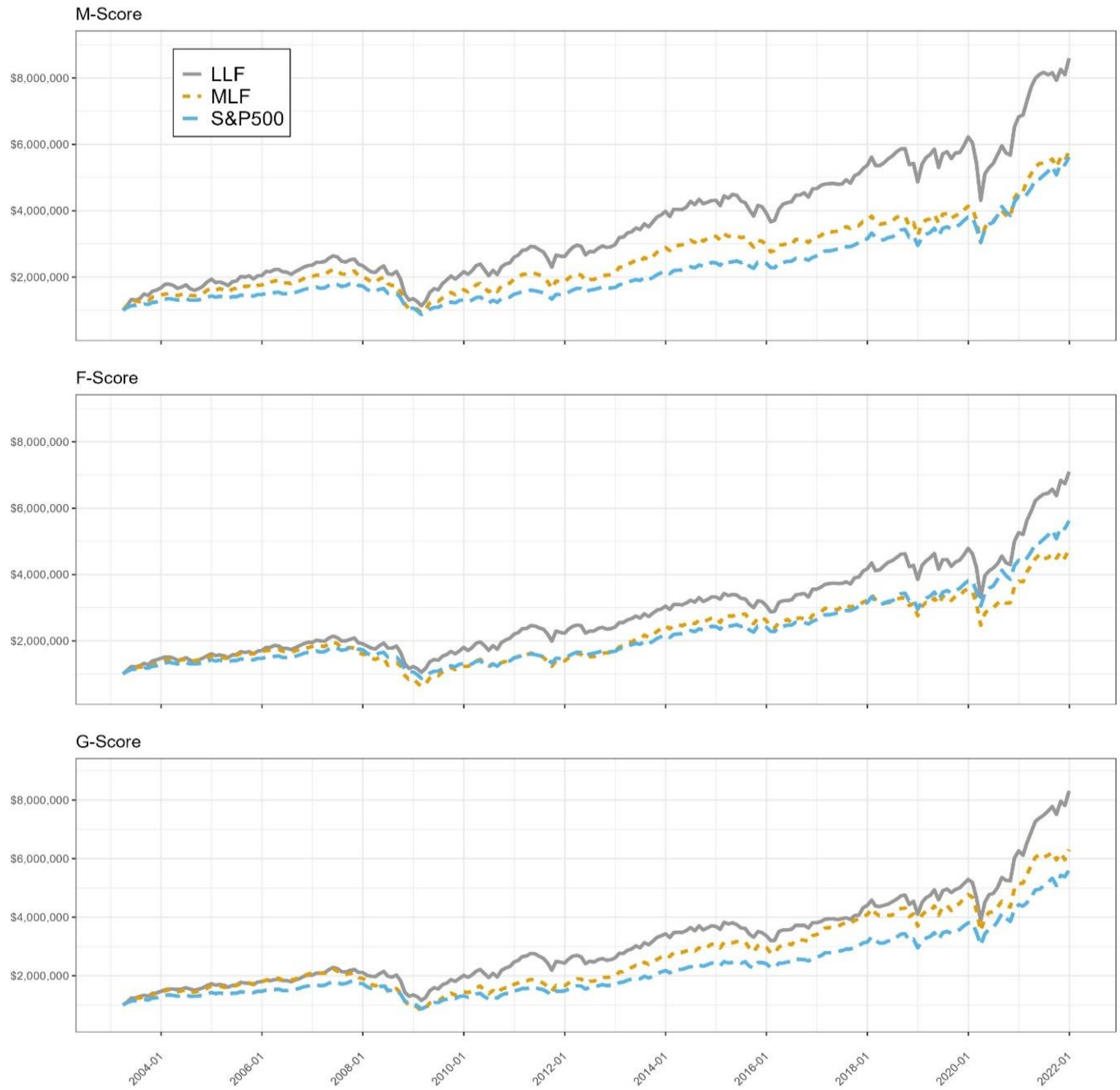


Table 1: Investment Strategy Summary Statistics

	Least Likely Fraudulent			**Most Likely Fraudulent**		
Model	M	F	G	M	F	G
Start Equity	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000	1,000,000
Final Equity	8,596,141	7,098,243	8,313,806	5,804,393	4,774,720	6,307,577
Annualised Return	12.16%	11.02%	11.96%	9.83%	8.70%	10.32%
Annualised Standard Deviation	19.31%	17.99%	16.91%	17.75%	19.00%	17.61%
Annualised Sharpe Ratio	0.5639	0.5426	0.6325	0.4838	0.3926	0.5152
95% Historical Value-at-Risk for Monthly Returns	-8.15%	-7.65%	-6.77%	-7.61%	-7.46%	-7.48%

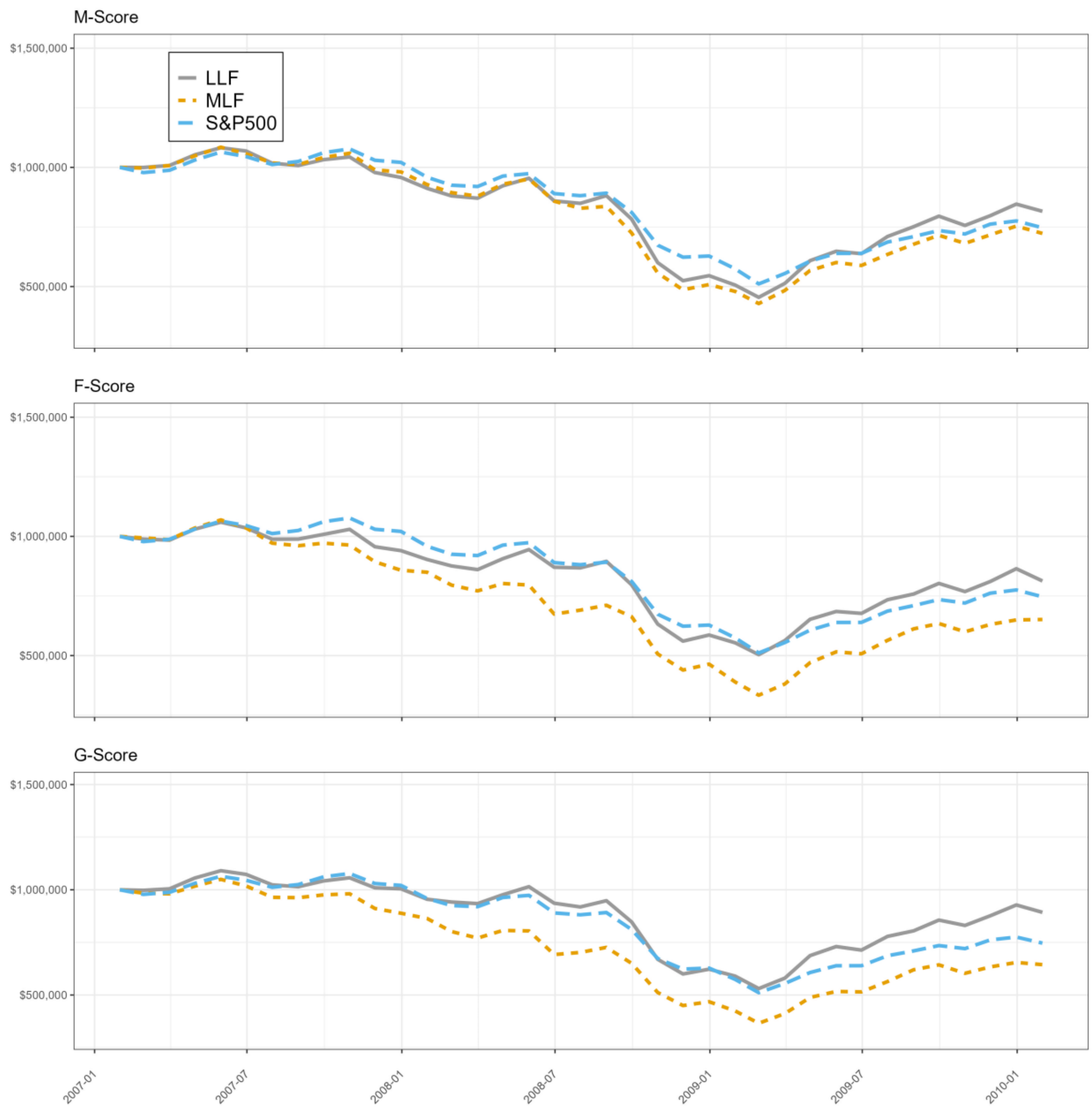
Equity Holdings and Market Shocks

The period from 2007 to 2009 is analyzed separately to explore how the strategies performed during a market-wide crash, namely the GFC. The chosen analysis period encompasses the market crash and short periods immediately preceding and following.

Figure 2 demonstrates that all strategies were negatively affected by the GFC. The LLF strategies appear to perform better than the corresponding MLF strategies, though the difference is exceedingly slight when the M-Score underpins the strategies. The underwhelming differential between M-LLF and M-MLF strategies in this period contrasts with the whole period, where they had the largest differential. The relative superiority of the LLF strategies, albeit slight in the case of the M-Score, is consistent with previous findings that socially responsible funds outperformed conventional funds during the GFC period (Becchetti et al., 2015).

Results over the COVID-19 period mirror those presented here. Equity curves over the COVID-19 period can be found in the Appendix (Section **Error! Reference source not found.**).

Figure 2: Equity Holding Over the Global Financial Crisis



Analysis After Adjusting for Market Performance

The results of the Fama-French three-factor regressions are presented for the LLF and MLF strategy using each risk model in Table 2. For conciseness, results are restricted to the alpha (intercept) and factor loadings (coefficients) for each strategy and risk model, along with the corresponding p-values.

All MLF strategies exhibit negative intercepts that are statistically significant at the 5 percent level. The returns earned by these strategies were thus not commensurate with the corresponding risk exposures. The LLF strategies, on the other hand, are characterized by intercepts that are not statistically different from zero at the 5 percent significance level. Related to the first research question, "What is the financial effect of investing in companies according to predictions of current fraud detection models?" these results suggest an adverse effect of holding companies predicted to have high fraud risk by any of the three models.

Table 2: Fama-French Three-Factor Regression Results

	Least Likely Fraudulent			**Most Likely Fraudulent**		
	M	**F**	**G**	**M**	**F**	**G**
Intercept	-0.1094 (0.3151)	-0.1603 (0.0910)	-0.0581 (0.5228)	-0.2556 (0.0043**)	-0.2867 (0.0085**)	-0.1695 (0.0412*)
Market Premium	1.1493 (0.0000***)	1.0895 (0.0000***)	1.0532 (0.0000***)	1.1135 (0.0000***)	1.1199 (0.0000***)	1.0898 (0.0000***)
SMB	0.2873 (0.0000***)	0.2635 (0.0000***)	0.1563 (0.0001***)	0.1267 (0.0012**)	0.0475 (0.3134)	0.0422 (0.2415)
HML	0.2119 (0.0000***)	0.1514 (0.0000***)	0.1075 (0.0010***)	0.1479 (0.0000***)	0.4565 (0.0000***)	0.3051 (0.0000***)
Statistically significant at the 5% (*), 1% (**) and 0.1% (***) level						

Discussion

Overall, the LLF strategies outperformed the MLF strategies when considering equity growth, accounting for the volatility of returns, and when considering exposure to major risk factors. This performance differential is driven by the negative excess returns of the portfolios investing in companies with the highest fraud risk, irrespective of the risk model used. Considering risk factor exposures, all LLF and MLF strategies had significant and positive sensitivity to market returns and the HML factor. The M-MLF strategy had significant and positive sensitivity to the SMB factor, as did all LLF strategies. The weightings on the SMB factor for the F-MLF and G-MLF strategies were not significantly different from zero. Several elements of the results are discussed in relation to each research question.

What is the Financial Effect of Investing in Companies According to Predictions of Current Fraud Detection Models?

The results broadly indicate that financial statement fraud risk is a relevant consideration in investment decisions. Specifically, investors should be concerned about situations with a high risk of fraud. This finding is also consistent with previous research on the cost of fraud to business value (Dyck et al., 2021) and thus to investors, with a deleterious effect on financial performance from investing in companies with a higher risk of fraud. Relative to LLF strategies, the MLF strategies had reduced capital accumulation throughout the full period and during the GFC. These lower returns were not associated with lower volatility, as the MLF strategies also had lower Sharpe ratios. The MLF strategies also did not adequately compensate for risk exposures and were instead characterised by significant and negative alphas in regression analyses. The financial effect of investing in companies with a lower risk of fraud is less pronounced, with results supporting the view that investors focusing on companies with a low risk of fraud bear no material cost or benefit. Risk-adjusted excess returns for the LLF strategies were not different from zero at the 5 percent level of significance.

It is interesting to consider that the weighting of the SMB factor is consistently statistically significant when investing in stocks that are the least likely to be fraudulent. The positive factor loading appears contrary to the idea that smaller companies have weaker internal controls and are more susceptible to fraud; however, Gepp et al. (2021) point out that their model is trained to detect SEC-accused fraud, and that the SEC might be biased towards catching the largest and most costly fraud cases as a result of their mandate. This characteristic is shared by the M-Score and F-Score models. Thus, while this finding might indicate reduced fraud is associated with smaller public companies, it could also indicate that our ability (because of the regulator's focus) to detect fraud is inferior for smaller public companies relative to larger ones. In contrast, the HML factor loading is positive and statistically significant for all strategies, indicating better results from investing in value stocks (high book-market value) that are not overly affected by a fraud filter on stock selection.

Which Fraud Risk Model is Most Effective for Guiding Investment Decisions?

The results suggest that the choice of fraud risk model, among the three considered, is of little consequence from an overall investment performance perspective. All three models appear to have been effective in differentiating between companies given the consistent contrasting performance of the LLF and MLF strategies. Traditional evaluation of such models is made difficult because not all fraud is detected, and it is thus encouraging that the models appear to be effective when assessed within the less traditional context of investment decisions. While summary statistics provide minor support in favour of the G-Score in terms of the Sharpe ratio, no material differences in investment performance were linked to the underpinning fraud models. The exposure of the different investment strategies to risk factors also does not appear to be substantially affected by the choice of fraud model.

Key Limitations and Future Research Suggestions

The first key limitation of this work is that is only considered companies on the SandP 500 index. The generalisability of findings beyond this index should be of interest and so further investigation of other indices would be valuable. Future research could even extend consideration to all U.S. companies, as well as considering other countries with accompanying differences in factors such as regulatory regimes. Secondly, this research only considered three fraud models (M, F and G scores), but the comparative framework could be used to extend the research to include new promising models as they emerge. Thirdly, the goal of this research was to assess the investment performance of simple strategies focusing on only high or low risk companies, but consideration of fraud risk in more sophisticated strategies may also be of interest. This research direction could remain focused on fraud in isolation, such as weighting investment in companies relative to fraud score rather than using equal weightings or consider fraud risk in combination with other investment preferences.

Conclusion

This article has analysed the financial effect of investing in companies with the highest and lowest propensity for having committed fraud based on the predictions of current fraud detection models. In doing so, it has aimed to contribute to the under-explored intersection of research on financial statement fraud detection and investment performance. Three fraud detection models, consisting of two established models and a more modern data analytics tool, were used to analyse the relationships between financial statement fraud and investment performance. Such fraud models are necessary to assess realistic investment performance as fraud is otherwise only known to have occurred in retrospect.

This work is the first to compare the performance of investment strategies restricted to companies with a low risk versus a high risk of financial statement fraud. The results are important to researchers and the investment community at large because the analysis has focused on the key metric of risk-related results. The findings will also be of particular relevance to investors interested in the effect of fraud risk on their investment portfolios or wishing to avoid unscrupulous companies. Additionally, a contribution is made through the alternative validation of the M-Score (Beneish, 1999; Beneish, 1997), F-Score (Dechow et al., 2011), and G-Score (Gepp et al., 2021) models within the context of an investment strategy.

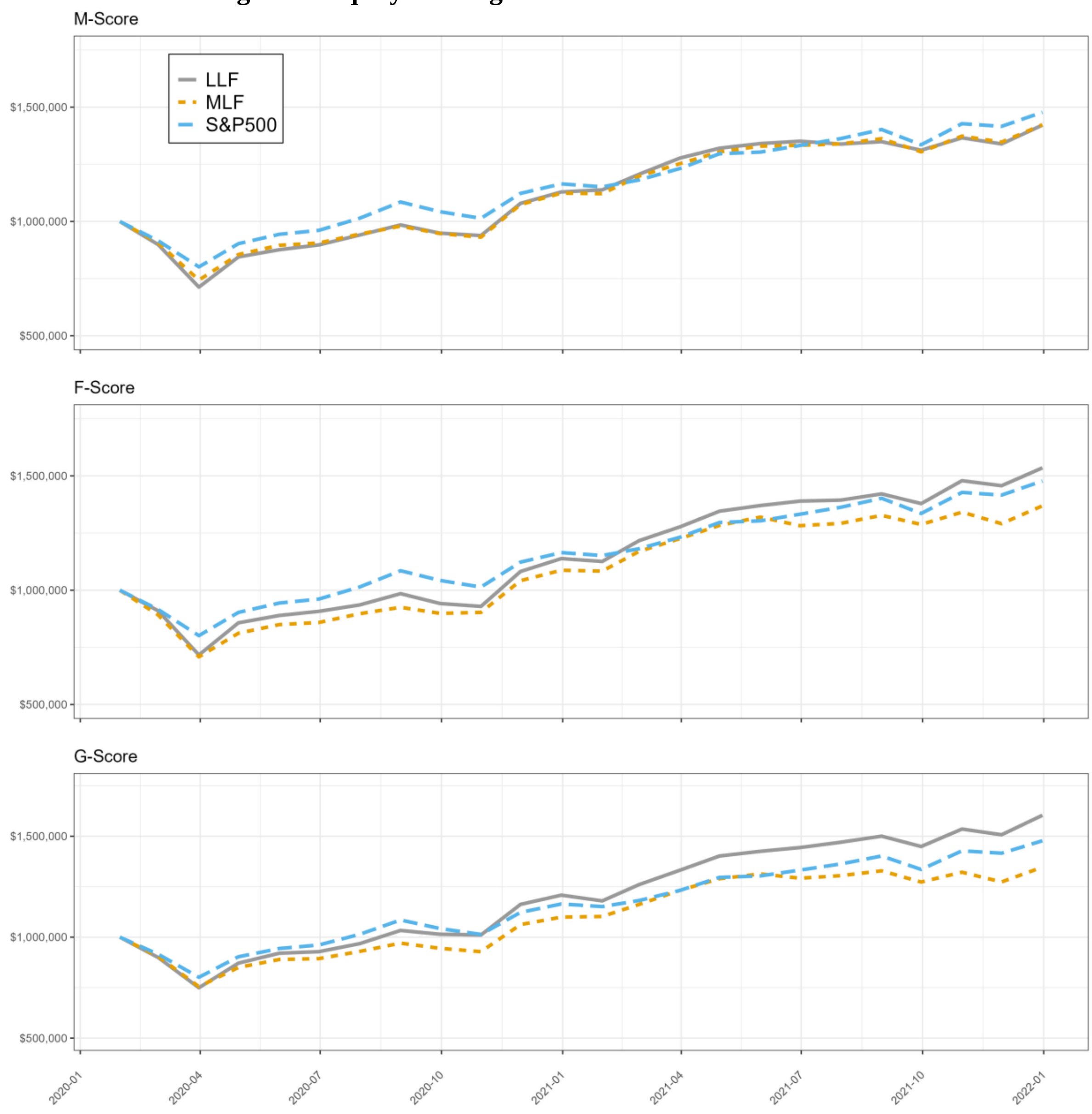
With this research providing an initial investigation into the intersection of the areas of financial statement fraud and investment performance, several promising directions for future research have been identified. These include broadening the investment universe beyond the SandP 500, evaluating additional fraud models, and developing more complex investment signals from fraud model outputs.

Acknowledgements

All authors gratefully acknowledge the valuable feedback from participants at the 9th Australasian Actuarial Education and Research Symposium (AAERS) in 2018.

Appendices

Figure 3: Equity Holdings Over the COVID-19 Pandemic



References

ACFE. 2022. *Report to the Nations* [Online]. Available: https://legacy.acfe.com/report-to-the-nations/2022/ [Accessed].

Armour, J., Mayer, C. and Polo, A. 2017. Regulatory Sanctions and Reputational Damage in Financial Markets. *Journal of Financial and Quantitative Analysis,* 52**,** 1429–1448.

Avortri, C. and Agbanyo, R. 2020. Determinants of management fraud in the banking sector of Ghana: the perspective of the diamond fraud theory. *Journal of Financial Crime,* 28**,** 142–155.

Beasley, M. S., Hermanson, D. R., Carcello, J. V. and Neal, T. L. 2010. *Fraudulent financial reporting: 1998–2007 : an analysis of U.S. public companies* [Online]. The Committee of Sponsoring Organisations of the Tradeway Commission (COSO). Available: www.coso.org [Accessed].

Becchetti, L., Ciciretti, R., Dalò, A. and Herzel, S. 2015. Socially responsible and conventional investment funds: performance comparison and the global financial crisis. *Applied Economics,* 47**,** 2541–2562.

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**,** 271–309.

Beneish, M. D. 1999. The detection of earnings manipulation. *Financial Analysts Journal,* 55**,** 24–36.

Beneish, M. D., Lee, C. M. C. and Nichols, D. C. 2013. Earnings Manipulation and Expected Returns. *Financial Analysts Journal,* 69**,** 57–82.

Bhattacharya, S., Xu, D. and Kumar, K. 2011. An ANN-based auditor decision support system using Benford's law. *Decision Support Systems,* 50**,** 576–584.

Brazel, J. F., Jones, K. L., Thayer, J. and Warne, R. C. 2015. Understanding investor perceptions of financial statement fraud and their use of red flags: evidence from the field. *Review of Accounting Studies,* 20**,** 1373–1406.

Capezio, A. and Mavisakalyan, A. 2015. Women in the boardroom and fraud: Evidence from Australia. *Australian Journal of Management,* 41**,** 719–734.

Chen, S. 2016. Detection of fraudulent financial statements using the hybrid data mining approach. *SpringerPlus,* 5**,** 1–16.

Chiu, T., Wang, Y. and Vasarhelyi, M. A. 2020. The automation of financial statement fraud detection: a framework using process mining. *Journal of Forensic and Investigative Accounting,* 12 (1), 86, 108.

Dechow, P. M., Ge, W., Larson, C. R. and Sloan, R. G. 2011. Predicting Material Accounting Misstatements. *Contemporary Accounting Research,* 28**,** 17–82.

Dyck, I. J. A., Morse, A. and Zingales, L. 2021. How Pervasive is Corporate Fraud? *Rotman School of Management Working Paper 2222608*.

Fanning, K. M. and Cogger, K. O. 1998. Neural network detection of management fraud using published financial data. *Intelligent Systems in Accounting, Finance and Management,* 7**,** 21–41.

French, K. R. 2022. *Data Library* [Online]. Available: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html [Accessed].

Gee, J. and Button, M. 2021. *The financial cost of fraud 2021* [Online]. Crowe. Available: https://www.crowe.com/uk/insights/financial-cost-fraud-data-2021 [Accessed].

Gepp, A., Kumar, K. and Bhattacharya, S. 2021. Lifting the Numbers Game: Identifying key input variables and a best-performing model to detect financial statement fraud. *Accounting and Finance,* 61**,** 4601–4638.

Hoogs, B., Kiehl, T., Lacomb, C. and Senturk, D. 2007. A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud. *Intelligent Systems in Accounting, Finance and Management,* 15**,** 41–56.

Huang, S.-Y., Tsaih, R.-H. and Lin, W.-Y. 2014. Feature extraction of fraudulent financial reporting through unsupervised neural networks. *Neural Network World,* 24**,** 539–560.

Karpoff, J. M., Lee, D. S. and Martin, G. S. 2008. The Cost to Firms of Cooking the Books. *The Journal of Financial and Quantitative Analysis,* 43**,** 581–611.

Karpoff, J. M. and Lott, J. R. 1993. The Reputational Penalty Firms Bear from Committing Criminal Fraud. *The Journal of Law and Economics,* 36**,** 757–802.

Kedia, S. and Philippon, T. 2009. The Economics of Fraudulent Accounting. *The Review of Financial Studies,* 22**,** 2169–2199.

Khamainy, A. H., Ali, M. and Setiawan, M. A. 2022. Detecting financial statement fraud through new fraud diamond model: the case of Indonesia. *Journal of Financial Crime,* 29**,** 925–941.

Kirkos, E., Spathis, C. and Manolopoulos, Y. 2007. Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications,* 32**,** 995–1003.

Kumar, P. and Langberg, N. 2009. Corporate Fraud and Investment Distortions in Efficient Capital Markets. *The RAND Journal of Economics,* 40**,** 144–172.

Lokanan, M. 2021. Applying four quantitative prediction techniques to detect fraud in financial statements. *Journal of Forensic and Investigative Accounting,* 13**,** 362–383.

Maulidi, A. 2023. Gender board diversity and corporate fraud: empirical evidence from U.S. companies. *Journal of Financial Crime,* 30**,** 309–331.

McKee, T. E. 2009. A Meta-Learning Approach to Predicting Financial Statement Fraud. *Journal of Emerging Technologies in Accounting,* 6**,** 5–26.

Perols, J. 2011. Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. *Auditing,* 30**,** 19–50.

Purda, L. and Skillicorn, D. 2015. Accounting variables, deception, and a bag of words: assessing the tools of fraud detection. *Contemporary Accounting Research,* 32**,** 1193–1223.

R Core Team 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Ravisankar, P., Ravi, V., Raghava Rao, G. and Bose, I. 2011. Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems,* 50**,** 491–500.

Repousis, S. 2016. Using Beneish model to detect corporate financial statement fraud in Greece. *Journal of Financial Crime,* 23**,** 1063–1073.

Rezaee, Z. and Kedia, B. L. 2012. Role of corporate governance participants in preventing and participants in preventing and detecting financial statement fraud. *Journal of Forensic and Investigative Accounting,* 4**,** 176–205.

Sadasivam, G. S., Subrahmanyam, M., Himachalam, D., Pinnamaneni, B. P. and Lakshme, S. M. 2016. Corporate governance fraud detection from annual reports using big data analytics. *International Journal of Big Data Intelligence,* 3**,** 51.

Sridharan, U. V., Dickes, L. and Royce Caines, W. 2002. The Social Impact of Business Failure: Enron. *American Journal of Business,* 17**,** 11–22.

West, J. and Bhattacharya, M. 2016. Intelligent financial fraud detection: A comprehensive review. *Computers and Security,* 57**,** 47–66.

Whiting, D. G., Hansen, J. V., McDonald, J. B., Albrecht, C. and Albrecht, W. S. 2012. Machine Learning Methods For Detecting Patterns Of Management Fraud. *Computational Intelligence,* 28**,** 505–527.