The Relationship Between Corporate Governance and Company Performance

New Factors, Models and Approaches to Causality

Conor Reid B.A, B.A.I, H.Dip

A Dissertation submitted to University College Dublin in part fulfilment of the requirements of the degree of M.Sc. in Business Analytics

Michael Smurfit Graduate School of Business, University College Dublin

August, 2018

Supervisors: Dr. James McDermott and Dr. Miguel Nicolau

Head of School: Professor Ciarán Ó hÓgartaigh

Dedication

To my family and friends who supported me throughout the last two years.

Contents

Acknowledgements

Thanks to Dr. James McDermott and Dr. Miguel Nicolau for first proposing this project, and for their subsequent guidance and support throughout the execution. They were always very generous with their time, for which I am very grateful.

Abstract

Corporate governance models, styles and practices vary widely between companies and across regions. There is much debate on the influence of those differing models on company outcomes, as measured by various economic indicators. It is natural then to ask how one optimises corporate governance structures to best influence economic outcomes, and what best characterises those outcomes?

This project aims to address these questions, by first analysing the work of Moldovan and Mutu (2015) and replicating their findings. We then extend those findings, by remodelling the problem and using auxiliary measures of corporate governance and company performance. Most importantly, we introduce modern methods of causal inference to this space in an effort to go beyond simple correlation towards the identification of stronger and more useful casual influences.

Using those correlations and *if-this-then-that* style rules proposed by Moldovan and Mutu (2015), we identify those that have causal merit as well as those that do not. Using the surrounding literature then, we compliment those findings with causal analysis in other areas to more holistically characterise the relationship between corporate governance and economic success.

Chapter 1

Introduction

This chapter forms the introduction to the current study, laying out its motivations and aims. A brief background of the domain is given, followed by the business motivation for this study. This is followed by a summary of the academic contribution, the research goals and scope and ends with an outline of the document structure as a whole.

1.1 Background

This study is primarily concerned with the relationship between corporate governance and company performance, particularly with how the former can be optimised to positively influence the latter. Corporate governance is a widely discussed, debated and researched topic that is as relevant today as it has ever been. A companies governance structures dictates it's policies and actions, ensuring that all stakeholders [1] have input into how the company is run facilitating a shared vision of where the business it is going. Governance policy also acts to mitigate financial and ethical pitfalls by setting clear standards. Thus it is fair to say that corporate governance has a wide ranging influence within and also outside the company.

[1] A stakeholder is someone who has a stake, or a personal interest, in the company. Employees, the local community and the media are all stakeholders.

We frequently see instances of corporate governance failure, which can lead to disastrous consequences both financially and reputationally. Reputation is often of high importance in both the public and private sectors, which motivates more ethical and fair behaviour in order to protect it. Instances where companies fail in this regard often make eye-catching headlines, for example Kirkpatrick (2009), McVeigh (2015) and McLaughlin (2017). In a hyperconnected world where news spreads quickly, the importance of a functioning governance structure is more important than ever. At the time of writing California have imposed a quota on female presence in the board room, outlined by Fox (2018). This highlights the perceived importance of governance practices, and the willingness of legislators to enforce polices they believe to be beneficial to the running of any business in todays society.

The interests of different parties can conflict in any business, such as between the shareholders [2] or directors. There is much debate on how best to align these interests, with suggested initiatives like structuring executive compensation to be at least partly dependant on firm performance. Shareholder interests can also conflict with the interests of the wider public as a whole, or the stakeholders. This is especially true for companies that heavily rely on natural resources as a driver for growth. In this case, sustainability not just of the company but of finite natural resources that the public depend on must be closely governed and managed. This is often the responsibility not just of those within the company, but those outside it too.

It is reasonable to argue that corporate governance influences all aspects of the company, not least its economic success. Moldovan and Mutu (2015) studied this relationship, collecting data on corporate governance and using it to predict corporate success as measured in various ways. They were able to learn models that did this successfully, resulting in a number of rules that inform

[2] A shareholder is often an investor who has equity in the company. They often have no personal interest in the company, solely financial.

corporate governance best practice. The current study uses this as a starting point, and looks to address some limitations within, as outlined in section 1.4. Modern work around causation is also studied, with a view to applying it in this domain. This would act to strengthen previously derived relationships, and presents an opportunity to study the deeper causal influencers of corporate economic outcomes.

1.2 Business Motivation

Moldovan and Mutu (2015) state the conclusions they reached in their research in the form of simple *if-this-then-that* styles rules, and point to the business significance of each. For example, they found that for US based companies the number of women on the board of directors was positively related to company performance. They also found that in Western Europe, companies should employ larger audit teams that has a mitigating effect on the risk of bankruptcy. In Eastern Europe, their main finding is that the presence of an independent chairman best influences economic success.

The business benefit of the above, and other similar findings, is obvious. By deriving a number of relationships between economic success and corporate governance, the authors first prove that a relationship does in fact exist in the first place. That is, elements of high corporate governance performance are strongly associated with company outcomes. Secondly they are able to put forward recommendations for governance best practice and show what elements are most influential, with geographic context. A key element of management is identifying levers with which to effect outcomes in a positive way, providing the core motivation for this type of research.

We look to first verify the above findings, before expanding the scope to include elements of governance not considered in the original study. This would in effect expand the array of tools available to corporations for effecting economic change. We also aim to strengthen these findings by seeking to identify causal mechanisms within this domain, and thus provide a more informed platform for decision makers.

1.3 Academic Contribution

A key element of this study is the exploration of causality research and the application of these techniques in this domain. There is continual active research in this area, with interested parties offering new techniques and methodologies for making the step from correlation to causation in a variety of domains. To our knowledge, casual research has not be applied in the area of corporate governance and its effect on outcomes, and thus would represent a novel endeavour that stands to contribute to the field in a meaningful way.

For example, the rules proposed by Moldovan and Mutu (2015) are backed by strong correlations drawn from highly accurate statistical models. However they make no steps towards estimating a cause and effect element to those relationships, or any other type of deeper analysis. We propose that a significant academic contribution would be to explore how causality is reached and applying that methodology here, to see if more can be said of the aforementioned rules.

As mentioned above, this study plans to expand the work of Moldovan and Mutu (2015) to include other types of company actions and activity. This would help gain a more holistic view of how economic success can be promoted across all company functions.

1.4 Research Goals and Scope

There are a number of key goals that this study aims to achieve. They are presented below, along with a discussion of how success will be measured at each stage.

1. **Reproduce the findings of Moldovan and Mutu (2015).**
   As mentioned, Moldovan and Mutu (2015) made findings that point to interesting relationships between corporate governance and company performance. It would be useful to use similar data to reproduce some of these findings using the same techniques as the authors. Measuring success here is relatively straightforward, facilitied by simple model comparison across a number of performance related metrics.

2. **Remodel the problem**
   Moldovan and Mutu (2015) thresholded on the various measures of company success, that are continuous in nature. This frames the problem as a classification exercise whereas a more natural framing is as a regression exercise. Thus, the study aims to use the same data as before but neglect the thresholding step and perform a regression on the continuous measures of success. By doing so, this study aims to build models that better reflect the true nature of the data. This stage also includes integrating additional independent and dependent variables, to expand the scope and discover new relationships between corporate governance and economic outcomes.

3. **Apply modern work on causality.**
   A number of conclusions on the influence of corporate governance on company performance were reached, using established statistical analysis to subsequently discover various correlations. In order to strengthen these findings and gain deeper insight into the underlying mechanisms of the domain, modern work in causality will be applied. This will involve significant research into the ways in which this can be achieved, including data requirements and any pre-processing steps required. The aim here is to gain a much deeper understanding of the causal structures the effect corporate economic success, to drive best practice and contribute to knowledge base in this area.

It is equally important to discuss what is out of the scope of this study. This study considers only public companies, since the data that has been obtained related to companies listed on various stock exchanges. The study would be complicated by the inclusion of private companies, due to regulation differences and enforced governance structures.

Further, regulatory differences from country to country are not considered. The financial reporting standards are not completely consistent across the regions included in this study, however for the purposes of this study these differences are not accounted for. Some countries may also introduce certain taxation or laws that influence the decisions made by companies in those regions, like a carbon emissions tax that may make companies take their environmental footprint more seriously. Again, no effort to account for these differences is taken.

1.5 Document Outline

This document is laid out as follows. Chapter 2 contains a literature review of this topic including how corporate success can be measured, other predictive corporate features that may be included, a review of other similar studies and concludes with a summary of research in the area of causation. Chapter 3 contains details of this studies methodology, including a summary of the data used and its pre-processing, algorithms used and methodology around applying causal techniques. Chapter 4 contains the results of this study. Included in chapter 5 is a discussion of these results, with some concluding remarks and opportunities for future research outlined in chapter 6.

Chapter 2

Literature Review

2.1 Introduction

It is natural in business to seek to optimise all aspects of corporate activity, with a view to positively effecting economic outcomes. This is reflected in the literature on this domain, which considers the problem in a number of different forms, following a wide range of methodologies and reaching quite different results at times. A review of a snapshot of this literature included here.

First, some of the ways that corporate success can be quantified is presented. This includes how success is defined by Moldovan and Mutu (2015), as well as by others whose approaches may benefit this study. Also included here is a discussion of various fronts on which companies act, such as their corporate social responsibility commitments, with a view to including these aspects in this analysis as independent explanatory features. This is followed by an analysis of existing literature on the relationship between corporate governance and company performance, again including the work of Moldovan and Mutu (2015) as well as other relevant studies. Then, an exploration is carried out of the literature regarding causation and the statistical techniques used to infer causation, with consideration of how these can be realised in this domain. There is also discussion on the issues that arise in attempting to do so, and how they can be addressed. This chapter finishes with a summary of the research gap that this study aims to address.

2.2 Company Performance - Measures and Influencers

2.2.1 Introduction

Among the key aspects of this study is the quantifying of corporate success, in a way that accurately represents *"good"* and *"bad"* corporate performance. There are many ways to do this, one of the most simple of which is to use financial ratios. An uncountable amount of ratios have been developed, each attempting to assign numerical performance ratings to companies and differing from one another by the accounting indictors considered and the aggregation of those indicators. Eidleman (1995) outlines the patterns that ratios tend to follow. He states that these ratios are mostly created by academic researchers, who constantly derive new ways of combining individual metrics together to facilitate meaningful comparison between companies. First, researchers find a sample of companies that meet some predetermined criterion of failure, as well as another sample of comparable firms (size, industry etc) that differ only in financial health. A number of ratios are developed and tested against this dataset to determine which returns values that are consistently and significantly different for each group. Those which do so successfully are kept, the rest are discarded. Weights are assigned to each ratio to reach an aggregate equation. New firms are then scored, and real-world performance recorded to measure how useful the new ratios are in practice.

This study primarily considers corporate governance features as predictors of corporate economic success. This follows the lead of Moldovan and Mutu (2015) who do the same. Having said this, an aim of this study is to expand the range of predictors considered, and to this end we discuss alternative sources of company data that can be incorporated into the model. Perhaps the inclusion of more varied and diverse independent features, such as a companies social responsibility commitments or their impact on the environment can enhance overall understanding in this domain. These areas are discussed in this chapter.

2.2.2 Financial Ratios

Some financial ratios are more useful than others. They vary significantly in complexity, in the aspects of a company they consider and in how they combine those aspects. This section includes a discussion of some of the most commonly used ratios, with comments on how they can best be used and in what context they are most powerful.

One of the indicators of corporate economic performance used by Moldovan and Mutu (2015) is Tobin's Q score. This measure was devised by Tobin (1969) who postulated that the combined market value of a given company should be equal to their replacement costs. When a company's replacement cost is equal to its market value, it is said to be in an ideal state. Any deviation either way (a ratio above or below 1) motivates either future investment or the selling of assets respectively. Moldovan and Mutu (2015) argue that the Q score allows the estimation of intangible assets, and is thus a worthy inclusion as a dependant measure of corporate success.

The use of this measure is well established in the literature, by Chung and Pruitt (1994), Bhagat and Bolton (2008) and Bolton *et al.* (2011). Chung and Pruitt (1994) state that Tobin's Q score plays an important part in financial interactions, and is employed to explain diverse corporate phenomena during the decision making process for investment opportunities. Bolton *et al.* (2011) use Tobin's Q score to propose a model for dynamic investment and risk management and found that investment is best driven by an aggregate model of which the Q score plays an important role.

The formal definition of Tobin's Q is presented by Chung and Pruitt (1994), along side a much more simplified and conservative approximation of the authors making. This less complex definition is seen commonly in literature, for example by Wahba (2008). This formulation is given below as;

$$(Approx.)\ q \ = \ \frac{MVE + PS + DEBT}{TA} \tag{2.2.1}$$

Here, MVE represents the product of a companies share price and count of common outstanding stock shares. PS represents the liquidating value of the companies preferred stock. $DEBT$ represents the companies short-term liabilities minus its assets, also short-term. Finally, TA represents the book value of the total assets of the company. In layman terms, assets that cannot be easily quantified can not always be entered in a companies books, but do always contribute to the share price of that company. Thus, a firm with lots of this type of asset will have a high Q score, and those without will have a low score.

There is debate as to the practically of the Q score. Intuitively, the Q score places a very high importance on one specific aspect of a business. For example, before WhatApp was acquired by Facebook it had very little concrete physical assets (i.e. those that could be recorded in their accounts). Rather, it had a platform with approximately 400million users, and a very high Q score. After purchase, the Q score would have dropped significantly since the amount Facebook paid would become the concrete asset recorded in their books. This is not to say that the Q score is a bad success indicator, but rather it may not fully represent the real value of a company at a given time.

Chung and Pruitt (1994) state in their research that the Q score is often neglected in real-world situations. One of the reasons they give for this is the complexity of the necessary calculations, and a potential unfamiliarity with its operational intricacies. Another reason is the unavailability of relevant data, particularly of sufficiently high accuracy and instantaneous availability. To counteract this, they worked to create and test an accurate approximation of Tobin's Q that utilises only basic financial information, show in equation 2.2.1.

They conclude that their approximation is close enough to the more formal definition to be used where more exhaustive calculations are not possible.

Another measure of corporate success used by Moldovan and Mutu (2015) is the Altman Z score, which can be used as a probabilistic measure of whether a company will fall into bankruptcy within the next two years. It can also be used more generally as a financial distress measure and to predict corporate defaults. The authors point out that there is much advocacy in the literature for using this measure, and this study was unable to find any that strongly reject its usefulness. The Altman Z score is given as;

$$\begin{aligned} Z\ Score \quad = \quad & 1.2\left(\frac{Working\ Capital}{Total\ Assets}\right) + \\ & 1.4\left(\frac{Retained\ Earnings}{Total\ Assets}\right) + \\ & 3.3\left(\frac{Earnings\ before\ Interest\ and\ Tax}{Total\ Assets}\right) + \qquad (2.2.2) \\ & 0.6\left(\frac{Market\ Value\ of\ Equity}{Total\ Liabilities}\right) + \\ & 1.0\left(\frac{Sales}{Total\ Assets}\right) \end{aligned}$$

Among those that support this scores use is Eidleman (1995), who discusses its use in practice. He begins by highlighting Altman's own tests using the Z score which involved accurately predicting 72% of bankruptcies two years prior to the event, although the sample size or companies involved are not mentioned. Eidleman (1995) argues that the Z score is tried and tested, and that it has been demonstrated to be reliable in a variety of contexts and countries.

Eidleman also outlines circumstances that warrant corrections and alterations to equation 2.2.2, in order to generalise it beyond its originally intended means. He argues that before being able to use the Z score, one must ensure the company in question is comparable to those involved in Altman's original study. Altman considered manufacturing and small firms in his original analysis, and thus corrections must be made before scoring companies in different industries. Eidelman points to two specific circumstances here.

The first considers privately held companies, whose stocks are not publicly traded. This meaning that term four of equation 2.2.2 cannot be calculated. To correct for this, the Z score can be re-estimated using book values of equity. In other words, details from balance sheets published by private firms voluntarily can be used rather than details gleamed from the stock market. Certainly a work-around here is to consider solely publicly traded companies. A consequence of this is that such an analysis would only include companies that are bound by the corporate governance code in their jurisdiction, which would need to be taken into account in studies such as this one.

Eidleman's second consideration is for non-manufacturing firms. The fifth term of equation 2.2.2, according to Eidleman, varies significantly by industry. He argues that merchandise firms for example, are significantly less capital intense and thus are much more likely to enjoy higher asset turnover and consequently Z-Scores. Z scores then would be likely to under-predict bankruptcy in these cases. In order to correct for this, a recommendation comes from Altman to eliminate the fifth term and adjust the weights.

The adjusted equation 2.2.3 is shown below;

$$
\begin{aligned}
Z\ Score \quad = \quad & 6.56\left(\frac{Working\ Capital}{Total\ Assets}\right) + \\
& 3.26\left(\frac{Retained\ Earnings}{Total\ Assets}\right) + \\
& 6.72\left(\frac{Earnings\ before\ Interest\ and\ Tax}{Total\ Assets}\right) + \\
& 1.05\left(\frac{Market\ Value\ of\ Equity}{Total\ Liabilities}\right)
\end{aligned}
\tag{2.2.3}
$$

Overall, the Altman Z score seems a highly appropriate indicator of corporate financial strength and thus success, and one that should be considered in this study. Consideration will need to be had for the type of industry included in this analysis, that will inform the exact calculation of the Z score itself.

As mentioned above, both Tobin's Q score and the Altman Z score were used by Moldovan and Mutu (2015), and are to be included in the current studies to emulate their work. However, one of the goals of this study is to extend this work by including other measures of corporate success or health. An exploration of the Beneish M-Score is carried out with this in mind. The Beneish M-Score is an aggregate financial ratio calculated using standard accounting data and acts as a probabilistic measure of intentional manipulation of a companies declared earnings. This score has been used in many studies aiming to detect corporate fraud and deception, or more generally and in combination with other factors to detect financial distress before it happens. Beneish and Nichols (2007) state that although the risk of significant financial loss as a result of financial reporting fraud is huge, investors spend little time utilising publicly available data in an effort to detect that fraud. Thus, this ratio is a worthy addition to this study.

Beneish (1999) calculates the M-Score using the following formula 2.2.4.

$$\begin{aligned} M\ Score \quad = \quad & -4.84 + 0.920(DSRI) + 0.528(GMI) + \\ & 0.404(AQI) + 0.892(SGI) + 0.115(DEPI) - \\ & 0.172(SGAI) + 4.679(TATA) - 0.327(LEVI) \end{aligned} \tag{2.2.4}$$

where

$$DSRI = \left(\frac{Receivables_t}{Sales_t}\right) / \left(\frac{Receivables_{t-1}}{Sales_{t-1}}\right) \tag{2.2.5}$$

The days sales in receivables index (DSRI) measures the amount of days that sales are in accounts receivable in one year compared to the previous year. A ratio of much greater than 1 coupled with a slower growth in sales could be an indication of inflated revenues.

$$GMI = \left(\frac{Sales_{t-1} - CostsofGoodsSold_{t-1}}{Sales_{t-1}}\right) / \left(\frac{Sales_t - CostsofGoodsSold_t}{Sales_t}\right) \tag{2.2.6}$$

The gross margin index (GMI) measures the ratio of current and previous year gross margin. An index much greater than 1 signals that gross margin has declined in the period, which according to Mahama (2015) could provide motivation for manipulation. Lower gross margins are linked to lower future earning, which Beneish and Nichols (2007) state is linked to higher probabilities of manipulation.

$$AQI = \left(1 - \frac{CurrentAssets_t + PPE_t}{TotalAssets_t}\right) / \left(1 - \frac{CurrentAssets_{t-1} + PPE_{t-1}}{TotalAssets_{t-1}}\right) \tag{2.2.7}$$

The asset quality index (AQI) measures asset quality as a ratio of non-current assets (less property plant and equipment) to total assets from one year to the next. A ratio higher than 1 indicates that the company may be involved in cost deferral. A deferred cost is one that in incurred but not charged to expense until a later reporting period, and could be a sign of risky behaviour.

$$SGI = \left(\frac{Sales_t}{Sales_{t-1}}\right) \tag{2.2.8}$$

The sales growth index (SGI) is a simple ratio of sales from year to year. Beneish (1999) states that while growth is not inherently a sign of manipulation, growing firms are viewed as more likely to commit financial reporting fraud due to the pressure placed on managers to reach earnings targets. The authors go on to argue that small, high growth companies stand to lose more in the face of slowing growth figures than their more mature counterparts, leading to potentially risky behaviour.

$$DEPI = \left(\frac{Depreciation_{t-1}}{Depreciation_{t-1} + PPE_{t-1}}\right) / \left(\frac{Depreciation_t}{Depreciation_t + PPE_t}\right) \tag{2.2.9}$$

The deprecation index (DEPI) is a ratio of the rate of deprecation from year to year. A ratio greater than 1 signals that the rate at which assets are losing value has decreased, potentially as a result of management revising upwards their estimates of the viable serviceable lifespan for those assets. Beneish (1999) states that there should be a positive correlation between this ratio and the likelihood of a company manipulating earnings reports.

$$SGAI = \left(\frac{SGAExpenses_t}{Sales_t}\right) / \left(\frac{SGAExpenses_{t-1}}{Sales_{t-1}}\right) \tag{2.2.10}$$

The sales general and administrative expenses index (SGAI) is calculated as the ratio of general sales and administrative expenses to sales from year to year. Beneish (1999) expects a positive correlation between this score and likelihood to commit reporting fraud.

$$TATA = \left(\frac{\Delta WorkingCapital_t - \Delta Cash - \Delta IncomeTaxPayable - Depreciation}{TotalAssets} \right) \tag{2.2.11}$$

Total accruals to total assets (TATA) is a proxy ratio for the extend to which cash underlies reported earnings. Total accruals are calculated as the change in working capital accounts (other than cash) minus depreciation. Higher accruals are thus an indication of less cash, and should be associated with a higher risk of reporting fraud.

$$LEVI = \left(\frac{LTD_t + CurrentLiabilities_t}{TotalAssets_t} \right) / \left(\frac{LTD_{t-1} + CurrentLiabilities_{t-1}}{TotalAssets_{t-1}} \right) \tag{2.2.12}$$

Finally, the leverage index (LEVI) is a ratio of total debt to total assets from year to year, with a value greater than 1 an indicator of an increase in leverage over the period. Beneish (1999) state that this is associated with the probability the company will default, or in other words be unable to make the required payments on their debt obligations.

Mahama (2015) uses the Beneish model, interestingly in conjunction with the Altman Z model seen above to show that the Enron scandal could have been detected as early as 1997, well before the company filed for bankruptcy in 2001. This was done using data from the United States State Examinations Commission (SEC) spanning from 1996 to 2000. As has been mentioned above, an Altman Z score below 1.81 indicates company is at high risk of bankruptcy. Here, an M-Score greater than -2.22 is an indication of manipulated financial reporting. Applying this logic to Enron, the author compiled the below table.

Metric	2001	2000	1999	1998	1997	1996
Z-Score		2.481	3.040	2.029	1.611	1.884
M-Score	-2.358	-0.343	-1.323	-2.426	-2.064	

Table 2.1: Enron Scandal - Mahama (2015)

Immediately obvious here is that Enron was in a state of high bankruptcy risk in 1997, while in subsequent years it hovered above that threshold in a grey area of bankruptcy risk. The M-Score shows us that Enron was likely committing fraud in 1998, a year after experiencing their worst Altman Z score. Mahama (2015) state that being brought so close to bankruptcy in 1997 motivated the manipulation of financial reportings a year later.

Herawati (2015) studies the utility of the Beneish M-Score in detecting financial fraud, specifically using data released by the Financial Services Authority in the United Kingdom relating to companies that are known to have committed fraud during the period of 2011-2014. This amounted to 35 companies, with a further 35 non-fraudulent companies added to the study and picked based on equality of assets and type of industry. The author constructed a logit regression model, using a binary indicator as the dependent variable that represented the presence of fraud or not. The aim then was to find the weightings of each sub-ratio within the M-Score to first analyse its aggregate ability to detect fraud but also the utility of each of the subcomponents. The results of this analysis are shown in figure 2.1.

It's clear from this table that the sub-ratio's GMI, DEPI, SGAI and TATA are all statistically significant and have non-negligible weights. All of DSRI, AQI and LVGI do not and so are less likely to be influential in detecting the presence of financial reporting fraud. Its clear from this study that the Beneish M-Score is a useful metric for analysing the activity of a company and measuring performance on a different dimension than considered here previously.

Variable	Coefficients	Wald	Sig.	
DSRI	0,614	1,945	0,163	
GMI	0,984	4,141	0,042	*
AQI	0,830	2,521	0,112	
DEPI	-1,254	3,266	0,071	**
SGAI	-3,420	5,333	0,021	*
LVGI	-0,029	1,716	0,190	
TATA	-2,592	5,582	0,075	**
Constant	1,356	1,278		
Chi-Square (Hosmer and Lemeshow Test)		5,459	0,708	
-2 Likelihood (Block Number: 0)		97,041		
-2 Likelihood (Block Number: 1)		60,193		
Omnibus Test of Model Coefficients		36,848	0,000	
Nagelkerke R Square		0,546		
Overall Percentage		78,6%		

Where:

* : Significant level 5%

** : Significant level 10%

Figure 2.1: Logit Regression for Beneish M-Score (Source: Herawati (2015)).

Kamal *et al.* (2016) also use the Beneish M-Score to detect earnings manipulation and financial statement fraud in a group of Malaysian public list companies, thus extending its use beyond the context of the United States. Here, the authors gathered data on 17 companies that were, between 1996 and 2014, prosecuted by the Securities Commission Malaysia for commuting such acts of misreporting and intensional obfuscation. They calculated the M-Score for each company in the years prior to each prosecution, and found that in 14 cases (or 82%) the scores were less than the threshold and would have raised alarm at the time. Figure 2.2 shows the numerical findings of Kamal *et al.* (2016).

While the authors here present strong evidence as to the utility of the M-Score in detecting malpractice, it is just as interesting to look at the companies that presented so-called "safe" scores but in actuality proved to be fraudulent. As mentioned previously, a score of less than -2.22 (a greater negative) would classify a company as "safe". There are three such companies here, with the lowest M-Score of -3.76 for a company with the ticker `Mems Tech Bhd`. Interestingly, the LVGI score here is noticeably higher than for all other companies, with the net effect of this component acting to drive the overall M-Score down (since it's multiplier is negative) and reducing the calculated risk. Since we know this company was in fact convicted of fraud, this evidence supports the findings of Herawati (2015) that the LVGI score may not be so useful in overall fraud detection.

Misstatement yr	Company	DSRI	GMI	AQI	SGI	DEPI	SGAI	LVGI	Accruals	MSCORE
2011	Silver Bird Bhd	1.40	1.23	0.93	1.22	0.99	0.80	1.00	0.05	-1.56
2007	Mems Tech Bhd	0.70	1.13	1.36	1.07	1.44	0.77	3.24	-0.13	-3.76
2007	LFE Corp Bhd	1.03	0.40	0.79	1.19	0.54	0.68	0.91	0.03	-2.53
2007	Axis Incorp Bhd	1.02	1.03	0.83	1.26	0.83	0.50	1.04	0.12	-1.65
2007-2006	Oilcorp Berhad	1.50	1.53	0.88	2.90	0.86	0.62	1.16	0.02	-0.02
2007	Talam Corp Bhd	2.60	-0.49	1.02	0.36	0.96	0.75	1.00	-0.03	-2.48
4th quarter 2006	Inix Technologies	0.23	1.19	0.48	6.95	0.42	1.14	0.57	0.16	2.82
dec06/qtrs07	Satang Hold Bhd	1.17	0.91	0.48	1.26	0.52	1.02	1.41	0.33	-1.02
2006	Kosmo Tech Ind Bhd	1.87	1.06	2.27	1.13	0.45	1.78	1.04	0.10	-0.78
2006	Megan Media Hold Bhd	1.65	0.95	0.86	1.14	0.91	0.92	1.04	-0.02	-1.93
4th quarter 2006	Transmile Group Bhd	1.91	0.88	0.89	1.80	1.00	1.43	0.81	0.08	-0.70
2005	Nasioncom Hold Bhd	2.18	1.05	0.78	1.21	0.77	0.98	1.21	0.17	-0.56
2005	Welli Multi Corp Bhd	0.99	1.38	0.92	1.30	0.72	0.70	1.02	0.06	-1.76
2004	United U-Li Corp Bhd	1.09	1.00	1.00	1.23	0.95	0.91	1.13	0.15	-1.53
4th quarter 2004	Goh ban Huat Berhad	0.93	1.60	0.64	0.91	1.13	2.41	0.65	0.33	-1.01
2004	Aktif Lifestyle Berhad	0.84	0.33	26.61	1.01	4.95	1.80	0.10	14.28	73.92
2003	Plymate Hold Bhd	1.33	0.87	0.45	1.21	0.66	0.95	1.05	0.11	-1.81

Figure 2.2: Beneish M-Score for Prosecuted Malaysian Companies (Source: Kamal *et al.* (2016)).

2.2.3 Environmental Considerations

As mentioned previously, there is likely much room for improvement in the range of predictors of success considered in this analysis. One potentially useful area to consider is the environmental performance of the company, studied by Schaltegger and Synnestvedt (2002). The authors primary focus is on environmental management, specifically as a vehicle for corporate success. It is assumed here that environmental management and performance relate to the controlling of costs relating perhaps to the mining of natural resources. This may also refer to fines, penalties or taxes that are incurred due to harmful emissions, which must be carefully managed within a company and process introduced to mitigate against.

The authors discuss two conflicting viewpoints in this space. The first is that improved environmental performance predominantly causes an increase in operating costs due to investment and so on, which in turn negatively effects the profitability of the company. The second viewpoint states the opposite; improving a firms environmental performance in fact induces cost savings, which drives increases in profitability. Interestingly, the authors take the position that environmental investment both (negatively) increases costs and (positively) increases income. Having said this, they argue that such a direct link is not present in practice, and that instead it is the identification and management of the relationship that is more important. The conflicting viewpoints are visualised in figure 2.3.

ES_0 on the vertical axis represents the current level of economic success, described by the authors as a certain shareholder value. According to the pessimistic view of environmental spending, this value decreases as spending in environmental protection increases (through points E and F, to D where non profit can be made). That is, spending in this space reduces profit making ability to eventual zero. The more optimistic view is represented by the path from ES_0 through points A and B to point C (again where no profit is possible). This represents the ideology that some economic gain can be achieved, at

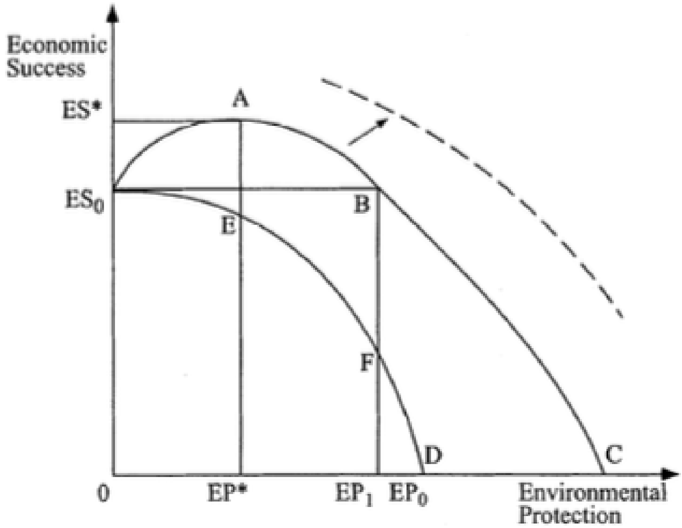


Figure 2.3: Correlation between corporate environmental protection spending and economic success (Source: Schaltegger and Synnestvedt (2002)).

least to some degree before tailing off, by being environmentally conscious. Of course, even in this situation there is an optimal investment level, after which profit making ability decreases.

The authors draw point to two important ideas, derived from figure 2.3. First, they argue that environmental performance can vary at a given level of economic success. It is possible to be equally successful by being either environmental friendly or harmful. This indicates that investment in this space does not necessarily mean poor economic performance. Secondly, the reverse; that economic success can vary for a given level of environmental protection. That is, being environmentally conscious is no guarantee of economic benefit. The authors reiterate that the correlation between economic and environmental performance depends on not just company externalities but internal variables which are influenced by management. It is firm management, who moderate this relationship, that must be optimised in order to gain economically. The authors put this forward as an group of explanatory covariates not considered in the literature to this point.

It is natural then to ask how environmental management can be quantified in data for analysis. Schaltegger and Synnestvedt (2002) suggest two cases for deriving data. The first relates to the firms ability to utilise to the full the economic benefits of environmental protection measures. This may materialise in R&D spending, or other marketing positioning that communicates their efforts and establishes the company as industry leaders. They are seen as quality leaders, with their environmental management policies a significant driver of this. The second case relates to how this reputation is achieved in practice, by realising the optimal environmental performance for maximum economic success. This can take the form of identifying and implementing optimal production processes for example, facilitated by the identification of opportunities and "eco-efficiency potentials".

Wahba (2008) also performs research in this area, studying whether the market values corporate environmental responsibility, specifically in an Egyptian context. Similarly to Schaltegger and Synnestvedt (2002), the author acknowledge divided and inconclusive literature on this topic, stating an aim to present empirical evidence on how engagement with environmental responsibility can positively influence corporate market value. This analysis looks at a sample of 156 firms across 19 industries, using the market value of a firm as the dependant variable (interestingly, quantified by Tobin's Q score mentioned previously). Important to note is the authors use of ISO 14000/14001 certification as a proxy for environmental performance. This is recognised by the authors as non-ideal, however this may still have some utility in future research to address the difficulty in quantifying performance in this area.

An interesting and important consideration is made by Wahba (2008). She states that companies with greater economic success have a greater ability to invest in environmental endeavours, leading to a theory that the two may be jointly determined. This theory is tested before regression analysis run. To do this, the author uses the specification test of Hausman (1978) to test for endogeneity. Endogeneity occurs in econometrics when an explanatory variable correlates with the error term, rather than some other feature. Wahba (2008) states that the result of this analysis is that endogeneity is not occurring, but points to an important consideration that the current study may need to take into account. One of the common causes of this is a causal loop between an independent and a dependant variable, and can obviously lead to wildly inaccurate results.

Wahba (2008) state that the overall finding of this research is that the market does in fact reward firms for their environmental efforts, by positively influencing that firms Tobin's Q score, but do point to similar issues raised by Schaltegger and Synnestvedt (2002). Namely, that a companies decision to reduce their impact on the environment comes with significant cost and that if this is not managed correctly, the economic benefits evaporate.

There is clear support in the literature for the presence of a relationship between corporate economic success and environmental performance. Schaltegger and Synnestvedt (2002) suggest it is the firm management of environmental considerations that influences success, and puts this forward as an explanatory variable not yet considered in literature. This theory is supported by Wahba (2008), who find a strong relationship in this space but with a caveat that management plays a significant role in realising benefits. It remains to be seen exactly how this can be realised in data analysis, and what family of covariates could be incorporated into the model.

2.2.4 Corporate Social Responsibility

Corporate social responsibility, or CSR, is defined by the European Commission as the responsibility of enterprises for their impacts on society. This is similar to being conscious of environmental impact, but is more generalised to include compliance with established ethical standards and national and international norms. Central to CSR is the creation of shared value between themselves, their stakeholders and larger society as well as mitigating possible adverse events with those groups. It is natural to consider elements of CSR then, when discussing predictors of firm success.

Orlitzky *et al.* (2003) perform meta-analysis on the link between corporate social and financial performance. They begin by stating; "Most theorising on the relationship between corporate social / environmental performance (CSP) and corporate financial performance (CFP) assumes that the current evidence is too fractured or too variable to draw any generalisable conclusions." Orlitzky *et al.* (2003)

Their research aims to show this claim is unfounded, and that it is indeed possible to derive insight in this space by providing a more rigorous methodology for drawing such conclusions. They argue that researchers have attempted to find causal relationships between CSP and CSR, but have failed in part due to a failure to see vital differences between theory and operational applications. They also state that an aim of their research is to aggregate knowledge in the area, and highlight important findings they believe to be overlooked. This type of meta-analysis have been proven to provide value by considering a number of disparate and conflicting results holistically to reach an aggregate conclusion.

They present a number of hypotheses about the influence of CSP. They first state that CSP and CFP are positively related, regardless of industry and the context of the given study relating the two. They base this on a number of studies, stating that CSP is a vital form of "good management theory" that boasts competitive advantage by addressing stakeholder concerns quickly and fairly. Secondly, they hypothesise on the temporal nature of this relationship, and further that there is a bidirectional causality between CSR and CFP. That is, there is a circular and virtuous causal relationship that governs performance in each area. This is supported by the idea that prior success in CFP facilitates positive engagement in CSP, due to increased responsibility and freedom at the managerial level. This conclusion mirrors that seen in section 2.2.3, where it was theorised that increased economic success drives better environmental performance.

The third hypothesis put forward by Orlitzky *et al.* (2003) involves the underlying logic behind the correlation between CSP and CFP. The first is that CSP boasts managerial competencies and organisational efficiency, by enabling shared knowledge of the firm's market as well as social and political environments. Secondly they suggest that CSP is a driving factor behind the firm's reputation, and thus elicits significant goodwill from stakeholders.

The fourth and final hypothesis put forward involves the methodology used by previous studies to draw conclusions from. The authors here suggest that the variance in results seen across studies can be explained by sampling or measurement error.

In order to support these claims, Orlitzky *et al.* (2003) perform a meta-analysis involving 52 studies across both CSP and CFP, resulting in a total sample size of 33,878 observations on which they perform their analysis. This analysis involves applying a statistical aggregation technique, which calculates the cumulative correlations across studies and corrects for variable elements of those studies to reach one "true score correlation (ρ)". Using this technique, the authors were able to explain 24% of the cross-study variance in relation to the observed r^2 value, which they suggest is significant. That is, by controlling for sampling and measurement errors across multiple studies, they could reduce the variation in results by 25%. They suggest that this, strengthened with other analysis, supports their collection of hypotheses.

The authors compare their results with others who performed similar research. Key to this comparison is a discussion regarding the linkage between CSP and CFP. That state that other researchers have traditionally been concerned with the "Halo effect". That is, that correlation found (that CSP influences CFP) are due to the experimental procedure or are spuriously perceived. The authors here dismiss this, stating that the only logical halo link would be in the reverse direction. That is, that a company who are high financially successful may be scored highly in CSP regardless of factual information in that regard.

They go further, and state that even this hypothetical (but more logical) argument is debunked by their meta-analysis which shows a strong directional (or temporal) link between CSP and CFP.

As with environmental performance, there is clear evidence that corporate social performance (or perhaps more commonly, responsibility) is a worthy addition to this study. Along side governance features and the aforementioned environmental performance features, there is evidence that CSP can enhance predictive power and understanding in this domain.

2.2.5 ESG Disclosure

Sections 2.2.3 and 2.2.4 contains an analysis of various responsibilities that firms have in relation to the environment as well as more broadly in society. In both cases, the conclusion was that the inclusion of variables that act as a proxy for firms performance in these areas would be a worthy addition to the present study. However, it is often very difficult to acquire this kind of data in sufficient quantities and with sufficient quality to enable this type of research.

Interestingly, Fatemi and Kaiser (2017) carried out research on the effect of ESG investment on firm value and included an analysis of the moderating effect of the disclosure of those investments on investor attitudes. The authors here point to numerous prior studies that examine the relationship between ESG factors and firm valuation, many of such studies are included in the current study. However they state a gap in the literature around the study of the dissemination of information about those factors, and aim to address that in this paper.

The authors use Bloomberg to extract data from 403 U.S. listed companies for a time period between 2006 and 2011. They use Bloomberg's measure of ESG disclosure, which is constructed by considering the extent of each companies disclosure along each component of ESG as derived from those companies own filings. Those filings come in many forms, from dedicated CSR reports to annual reports and corporate website details. The authors here state that this score, which ranges from 0.1 (poor) to 100 (excellent), is likely to reflect all publicly available information in this space. It is also readily available to investors, who play a significant role in valuing a company in the first place. As a measure of firm value, they use the Tobin's Q ratio which is seen numerous times in the current report, and is also available in Bloomberg.

The first result of this study is that strong ESG performance is associated with higher firm valuations, which is in line with research presented up to this point. Interestingly, they also find that when ESG disclosure is isolated from ESG performance features, it is found to be negatively correlated with firm valuation. The authors point out that the most valuable insights from this paper lie at the intersection of ESG factors and disclosure. They found that in the case of strong ESG performance, high disclosure ultimately weakens the increase in valuation. Markets may interpret such an increase in communication as an attempt by the company to justify significant investment in the space, rather than coming from an altruistic motivation. High levels of disclosure were also shown to mitigate the negative effects on valuation in the face of ESG concerns. The authors states that disclosure in this case could be perceived as a sign to investors that the company is making efforts to overcome their deficiencies by putting in place new operations or initiatives. This leads to the interesting and counterintuitive conclusion that poorly performance companies in this space should disclosure details at a higher rate then very strongly performing companies.

It is clear that the effects of investment in corporate social responsibility and environmental preservation can, an aggregate, lead to increases in firm valuation. The effect is complicated by how firms decide to communicate this to investors and the wider public, perhaps uncovering an intrinsic mistrust between both parties. In any case, the inclusion of disclosure scores to this present analysis could enable some interesting insights.

2.2.6 Executive Compensation

An interesting and much studied element of corporate structure and governance is executive compensation. Core and Larcker (1999) look at chief executive officer (CEO) compensation as an element of corporate governance, and its effect on firm performance. They highlight a common argument in the literature that since the board of directors are influenced by the CEO, they do not structure CEO compensation to maximise shareholder value. They study the potential association between CEO pay and the quality of firm's corporate governance, before relating that quality to future economic performance.

Firstly, the authors find that the characteristics of the board of directors composition and the ownership features of the firm (in other words, corporate governance features) have a strong correlation with the level of CEO compensation. They control here for standard economic predictors of compensation to isolate the effect of corporate governance. Perhaps unsurprising, they find that compensation is higher when in the case of CEO duality, but also when the board is larger in size and is composed of a greater number of independent directors. Of course, these latter features could be associated themselves with large companies with inherently more complex CEO demands and thus higher rewards.

Indeed, the authors present two opposing explanations for the observed variation in compensation using governance features. The first explanation is that they inadequately constructed a model for compensation outside of governance features, and thus the inclusion of those features would provide no useful insight. The second is that the observed variation is due to the presence of unresolved agency problems. Agency problems involve conflicts of interest where one party is expected to act in the others best interests rather than their solely their own. In this contest, such problems manifest in boards awarding disproportionate levels of compensation that do not relate to future economic success.

Core and Larcker (1999) state the need to address this problem, and so by carrying out a regression analysis involving the economic success factors relating to the firm, and the proportion of CEO compensation that is explained by its corporate governance. By doing so, they can examine whether excess CEO pay can be explained by greater levels of stock related performance which would eliminate the possibility of any agency problems. They find a negative association here, suggesting that CEO's are not necessarily compensated in line with firm performance (at least as quantified by stock returns). Instead, the authors argue that the weightings attached to governance features in the compensation equation are related to the effectiveness of the governance structure rather than legitimate or traditional proxies for the CEO's equilibrium wage. Core and Larcker (1999) thus conclude that firms with weaker corporate governance have greater agency problems, and underperform on the stock market. In addition, CEO's at such firms are more likely to be rewarded with greater levels of compensation.

2.3 Corporate Governance and Company Performance

2.3.1 Introduction

In this chapter, an exploration is included of research that directly attempts to answer similar questions as the current study. That is, other works that look at the relationship between corporate governance and company performance, and attempt to characterise and model that relationship. This includes the work of Moldovan and Mutu (2015), whose research forms the basis for the current study. An analysis is made of their methodology and conclusions, as well as a deeper analysis of what areas can be improved upon.

2.3.2 Existing research

Moldovan and Mutu (2015) made an attempt to find relationships between how a company governs and its economic success, using data mining techniques. They derived a dataset from the Bloomberg financial system, containing 50 independent corporate governance features on areas such as board fo directors structure and the companies yearly tax and interest liabilities. They considered three stock indexes; the S&P 500 (a collection of 500 American companies), the STOXX Europe 600 (a collection of 600 of Europes top performing companies) and the STOXX Eastern Europe 300 (a collection of 300 of Eastern Europes top performing companies). They complimented this with two dependant variables which numerically measures corporate success. Those measures are Tobin's Q score and the Altman Z score, the details of which are outlined in section 2.2. Using the aforementioned governance variables, they built statistical models to predict *"good"* and *"bad"* company performance.

Interestingly, Moldovan and Mutu (2015) decided to discretise Tobin's Q into two classes, either *"good"* for high scores or *"bad"* for low scores. The threshold used here was the median Q score of the entire dataset. They carry out similar preprocessing on the Altman Z score, creating three classes and allocating observations to each based on performance before using a classification algorithm for learning purposes. While the authors cite this methodology in previous literature, it may prove interesting to carry out regression on the real-valued scores as well. Such alternative methodology may yield positive results, that would perhaps be more nuanced and more appropriate for this data.

Using a number of different algorithms, Moldovan and Mutu (2015) were able to predict corporate outcomes using governance features with high accuracy. Interestingly all algorithms performed approximately equally well for American and European datasets, indicating a low level of model dependance here. Results from the Eastern European dataset were less consistent across algorithms, attributed to missing data on the corporate governance side.

As mentioned previously, the authors here were able to present some simple rules based on their models for corporate governance best practice. For example, they conclude that there is a positive correlation between a female presence on the board of directors in American companies, thresholding on 20% presence for the benefits to be detected. An independent lead director in the same dataset was shown to incur a higher risk of bankruptcy as measured by the Altman Z score. In Europe, the authors came to a similar conclusion, stating that the presence of an independent lead director or former CEO on the board of directors resulted in poorer Tobin Q scores. Interestingly and contrary to the case in America, the presence of women on the board was negatively linked to performance. Finally, in Eastern Europe Moldovan and Mutu (2015) found that a smaller age range within the companies directorship was positively related with performance. Altman Z scores improve significantly with an independent chairman, or a female CEO.

A significant piece of research in this area is the work of Bebchuk and Ferrell (2005), who examine what features of corporate governance are most influential on firm valuation. They use the 24 provisions followed by the Investor Responsibility Research Center (IRRC) as the basis for their analysis, aiming to create a so-called *Entrenchment Index* that constitutes the provisions that best correlate with company valuation as measured by Tobin's Q score. Previous research has included all 24 provisions and found a negative association, however the authors here state that there is no reason why all 24 are equally useful and so they aim to refine this research and present a simplified model.

Bebchuk and Ferrell (2005) propose 6 provisions for their composite index, having discarded the others as uncorrelated or insignificant. They list these 6 as; staggered boards, limits to shareholder bylaw amendments, poison pills, golden parachutes, and supermajority requirements for mergers and charter amendments. Poison pills and golden parachutes are referenced as two of the most salient and well know measures taken in preparation for a hostile takeover. A poison pill is a measure companies take to prevent such a takeover, and involving making the companies stock unfavourable to the acquiring firm. This is facilitated by allowing shareholders to purchase shares at a discount, leading to a dilution of shares held by the acquiring company which in turn makes the takeover more expensive. Golden parachutes on the other hand refer to substantial benefits awarded to executives if the company is acquired and they are let go. Bebchuk and Ferrell (2005) state that this insulates management from the economic costs they would otherwise incur after losing their role. This provision is included by the authors due to these insulating effects and the agency problems they may induce, as well as an expressed desire by shareholders to limit their use during the period of this study.

The remaining provisions included in the index revolve around limitations imposed on shareholders voting rights, which are their most important source of power. Firms that implement a staggered board maintain a number of classes of directors, where only one class is re-elected in a given year. This makes wide spread change at the desire of shareholders very difficult, and is considered a key defence against challenges to company control. Evidence is also presented in this study of a negative correlation between the presence of a staggered board in isolation and the Tobins Q score from Bebchuk and Cohen (2005), as well as evidence that announcements that a staggered board will be dismantled lead to significant increases in stock returns from Guo and Nohel (2008). For these reasons the provision is included in the *Entrenchment Index*.

Included also in the index are provisions that make it difficult for shareholders to exert control over events such as by-law and charter amendments and mergers. These are; limits on by-law amendments, supermajority requirements for mergers and supermajority provisions for charter amendments. The authors mention that in addition to theoretical justification for including these provisions, the parties interviews as part this study were unified in their opposition to them.

Having constructed their *Entrenchment Index*, Bebchuk and Ferrell (2005) then performed a regression analysis using Tobin's Q as a proxy for form value. They found a significant negative correlation here, highlighting the indexes (and its constituent parts) negative effect on overall stock performance. The authors state that this study focuses attention more acutely to the features of corporate governance that most influence a firms valuation, as opposed to what they call the "kitchen sink approach". Interestingly, especially so in the context of the current study, the authors make reference to the issue of inferring causation from mere correlation. Rather than having identified such a causal link, the author state that the observed correlation could be due to a tendency of managers of lower value or lower potential firms to adopt entrenching provisions. They then note that even if this behaviour tendency was provable, the identified correlation would still indicate that entrenchment is a key factor in those low value firms remaining in that low value state. Ultimately, it is not possible to definitively prove the direction of influence, whether it be from high entrenchment to low value or the other way around.

2.4 Inferring Causation

2.4.1 Introduction

It is stated ad nauseam in scientific and popular literature that "Correlation does not imply causation" or similar. In statistical analysis, it is tempting to infer causal relationships between features where only correlation has been proven and indeed a significant amount of literature seems to do exactly this. For example, the study of Moldovan and Mutu (2015) on which the current study is based, makes strong claims as to the relationship between corporate governance and company success. Here, the authors make a statement (among others) that Western European companies can expect to increase their Z-Score by employing a large audit committee, backed by a positive relationship between the two. In reality, the authors perform insufficient analysis to support causal inference, instead finding correlations worthy of further investigation.

Interestingly, there is a significant body of research that argues that it is impossible to prove causation in any case. Intuitively, one must consider in ones analysis all variables that could possibly prove to be causal, least a conclusion be drawn that leaves out some feature that in fact is the underlying causal influencer for some observed phenomenon. Of course in practical studies this is almost never fully achieved due to real-world complexity, and so it may be that advanced techniques are required to move towards this special case.

One of the major issues with applying causality research in this domain is the study design. Many studies, particularly in the medical field, are able to take advantage of robust experimental design standards that facilitate a deep and accurate exploration of results. They are able to control for unobserved covariates using randomised trials, and generally have a large degree of control over the statistical parameters of the study. Study methodology here often follows a procedure of maintaining a treatment and control group, which are large enough to safely assume that they are characteristically equivalent. Researchers can then estimate the presence and magnitude of any observed difference in outcome, and infer it to be caused by the randomly assigned treatment since that treatment is under the control of the researcher.

Outside of such highly controlled experimental environments, causal inference becomes more difficult since we begin to deal with observational studies. The work of Moldovan and Mutu (2015) is such a study, where the authors extract historical data that was generated outside of their control and try to uncover relationships within. Esarey (2015) identifies an interesting problem with this type of work. The author argues that often observations (whether they be people, organisations or events of some kind) self-select to be treated. This has a significant effect on outcomes. He gives the example of education; those who choose to complete higher education may be those who stand to gain from it the most, and so it is difficult to estimate educations effect on income. If only those who really need education attend, then its positive influence on outcomes (a persons salary for example) would be skewed. Similarly in the study of Moldovan and Mutu (2015), it is difficult to assess the benefit of various elements of corporate governance on firm performance due to a potential self-selecting process undertaken by those who perform well in the former.

This is a highly complex space, and one that certainly calls for advanced techniques that can mitigate the issues outlined above. The remainder of this section is dedicated to some of these techniques, with discussion of their technicalities and practical applications.

2.4.2 Matching

One approach to bridging the gap between experimental and non-experimental (or observational) studies is matching, outlined by Stuart (2010) who considers studies that use observational data that can be divided into treated and non-treated groups. Matching is then used to study the effects of this treatment on some outcome, in a very similar way to standard experimental trials in the medical field for example. He describes first how one of the biggest benefits of randomised experimental studies is that the treated and un-treated groups are guaranteed to be randomly different (as opposed to uniformly different, which introduces bias) from one another, on both observed and unobserved covariates (or features that may influence the outcome). That is, such experiments are able to control for factors that have not been explicitly designed for in the experiment. Statistical matching aims to imitate this for observational studies in which such design is very difficult, by balancing the distribution of potentially useful features in the treated and control groups. This is achieved by identifying observations that differ only in treated status, facilitating the analysis of the causal effect of that treatment. In effect this ignores un-observed features, and aims to reduce bias in the distribution of observed features as much as possible. The concept of *strong ignorability* is heavy relied upon here, which is to say that it is assumed that all features that may influence the outcome are being considered. As mentioned previously, this is very difficult to achieve in practice and often must be relaxed to achieve any results at all.

Stuart (2010) states that matching is a potentially useful mechanism in supervised learning, where the outcome is known and the goal is to estimate its effect. Thus, matching is considered in this study where the treatment can be quantified as fluctuations in various explanatory corporate governance variables (among others) which in turn influences a firms economic outcome.

Stuart (2010) sets out four key stages in the matching process. They are;

1. Defining the measure of *closeness* between observations.
2. Choosing an appropriate method for matching observations.
3. Quality assessment of matched samples, returning to Step 1 depending on results.
4. Treatment analysis, given the results of Step 3.

A measure of closeness quantifiably determines whether an observation is a good match for another. This is a crucial aspect to matching, and can be subdivided into two parts. The first involves pruning the dataset features for those to include, taking into consideration the concept of *strong ignorability* mentioned earlier. Stuart (2010) points out that poor results are expected from using small sets of features, particularly those that pertain solely to a narrow view of that observation (for example, demographic details of individuals). He states that there is little disadvantage in including features that are not actually associated with outcome, albeit a slight increase in variance is expected. Conversely, neglecting a feature that is associated with outcome is very costly, and so it is recommended to include as many features as is practical as a precaution. The author also recommends carrying this out without relying on observed outcomes, and instead make decisions based on domain knowledge. This represents a trend in causal analysis, in that domain knowledge is heavily relied upon to strengthen correlation's rather than immediately turning to deeper statistical analysis.

The second aspect to closeness is the measure of distance itself, or the similarity between two observations in the data. There are many ways to do this, that vary in the exactness of the match required. Stuart (2010) argues that exact matching is ideal, but very often unattainable especially with high dimensional data. Requiring a very high degree of exactness leads to observations remaining unmatched which then fall out of consideration, a phenomenon that in turn can lead to more bias than if the matching measure required less exactness.

One particular measure of closeness is the propensity score used in propensity score matching (PSM). The propensity score is used to account for covariates that predict receiving the treatment in the first place, which mitigates the issues of treatment self-selection mentioned above. This technique was first introduced by Rosenbaum and Rubin (1983) and has been frequently shown to lead to optimised matching. For example, PSM is discussed extensively by Dehejia and Wahba (2002), who explore its use in addressing the various issues that arise in trying to achieve satisfactory matching between treated and control groups in non-experimental datasets across a high dimensional set of characteristics. They propose propensity score matching for this, and apply that method to a sample dataset to illustrate its effectiveness. Dehejia and Wahba (2002) detail how this method succeeded in their experiments to focus attention on a small subset of observations that are comparable across treatment and control groups which mitigates the bias inherent in non-experimental studies. The authors here describe how it is not actually possible to find the true propensity score, so instead a best estimate is derived. In most of the statistics literature, the logit model is used which is defined by Dehejia and Wahba (2002) as

$$Pr(T_i = 1|X_i) = \frac{e^{\lambda h(X_i)}}{1 + e^{\lambda h(X_i)}} \tag{2.4.1}$$

where T_i is the treatment assignment status and $h(X_i)$ constitutes linear and higher order terms of the covariates with which to condition on to enable the researcher to ignore the distribution of treatment assignment.

Dehejia and Wahba (2002) go on to discuss how PSM is essentially a method of assigning weights on comparison units (pairs of control and treatment observations) when estimating the effect of the treatment on outcomes. The define this in a mathematical sense as;

$$\tau|_{T=1} = \frac{1}{\mid N \mid} \sum_{i \in N} \left(Y_i - \frac{1}{\mid J_i \mid} \sum_{j \in J_i} Y_j \right) \tag{2.4.2}$$

where $\tau|_{T=1}$ is the effect of the treatment, $\mid N \mid$ denotes the number of observations with the treatment, J_i is set of comparison observations matched to treatment unit i and $\mid J_i \mid$ denotes the number of comparison units in J_i.

Dehejia and Wahba (2002) raise an interesting point here, on the exact implementation of the matching process which strongly influences the ability to achieve optimal closeness between control and treated observations. They discuss the decision to match with and without replacement, and compare both methods theoretically and experimentally with a sample dataset. Using replacement means that each treated observation can be matched to the nearest control observation even if that control observation is matched more than once. This, they argue, minimises the propensity score distance between the comparison observations and this in turn drives down the bias.Eenforcing no replacement on the other hand may lead to sub-optimal matching since we may have many treated observations that are not similar to the control observations. While this increases bias, the precision of the estimated treatment effect could be enhanced. The authors also point out that without replacement, the order in which observations are matched can significantly alter the quality of the matching process. They analyse this difference experimentally using data from the National Supported Work program in the United States (a program which aims to provide work experience to individuals with economic or social difficulties). They ultimately find that when the characteristic overlap between control and treated groups was high, matching without replacement was optimal. Matching with replacement is a better choice when overlap is minimal.

The next stage in the matching process is the choice of matching method, which uses the closeness distance to create the matches themselves. The motivation behind using one method over another lies in the number of observations that remain after matching has taken place, and the relative weights that different observations receive. One such method is nearest neighbour matching (NNM), which is stated as the most common, most understandable and easiest to implement methods available by Stuart (2010). Dehejia and Wahba (2002) also use this method in their case study, showing that is yields better results than others. In essence, this method couples a treated observation with its nearest un-treated neighbour when plotted in space, minimising the distance between the two. Controls can be put in place to dictate the exactness of each match, discarding cases if a suitable match is not found. This helps to prevent bad matches, but leads to difficulties in interpreting results. What results from NNM is a dataset of similar dimension to the amount of treated cases, which arguably reduces the power of the data. Stuart (2010) states in response to this that model precision is effected most by the smaller group size in any dataset, and so balancing observations down to this smaller group size should not in fact dramatically reduce it power.

The third stage of the general matching process is a quality assessment of the matched samples, which is the most important step according to Stuart (2010). The aim here is to rate how balanced the matches set is, where balance refers to the similarity of feature distributions, and the independence of features and treatment status. Poor results here calls for alternative distance measures and matching methods, and so iteration is often required to find the optimal methodology. Stuart (2010) proposes numerical diagnostics to achieve this step. This involves the inspection of the difference in means of each feature, divided by the standard deviation which gives the *standard bias*. This is performed for each feature, as well as their two-way interactions and squares. The author discards other common tests here such as hypothesis tests, due to contextual issues and how balance is interpreted by those tests.

The fourth and final step of the matching process as outlined by Stuart (2010) is outcome analysis. It should be noted that matching is not actually a tool used for inferring causation, but rather presents a new dataset that is treated as if sourced through randomised methods. Analysis is then doen on this new dataset to uncover and identity patterns.

King *et al.* (2014) also characterise the trade-off between matched sample sizes and the balance between classes into the matched subset, identified above by Stuart (2010). While Stuart (2010) argued that any negative effects of sample reduction were offset by the increased balance between groups, King *et al.* (2014) argues the opposite. They claim that practitioners often do see sample reduction as an issue, citing manual tweaking that research carry out in an effort to optimise sample size as well as balance, or their tendency to settle for suboptimal solutions. King *et al.* (2014) argue that optimising only one of these parameters is not a viable solution nor a necessary one, but that current solutions available for optimising both require significant manual intervention which is time consuming and usually suboptimal. In response to this, they propose a new approach that they claim address a number of issues.

The so called *matching frontier* is a methodology that the authors claim fully characterises the trade-off between dataset imbalance and matched sample size, allowing researchers to visually inspect where the optimal solution lies for a given dataset. Each location along the frontier is denoted by the resultant matched sample. Moving along this frontier (i.e. varying sample size), the frontier returns a data subset such that no other subset of the same size has more optimal class balance characteristics. That is, the returned matched dataset is optimally balanced for its size. The implications of this for researchers are obvious. Using this method, one has much finer control over the matching process than is apparent in the work of Stuart (2010). The latter provides a framework that involves significant manual iteration, which is shown to be suboptimal and unnecessary by King *et al.* (2014).

2.4.3 Minimal-Model Semantics

Pearl and Verma (1995) present a highly influential theory of causation, and provide guidelines on how to make the step from strong correlation towards inferring causal relationships. This is certainly one of the more influential studies in this space, and approaches the problem of causation slightly differently to matching laid out in section 2.4.2. The authors here propose what they refer to as *a minimal-model semantics of causation*, which they claim debunks the myth that causal influences cannot be distinguished from illegitimate covariation. They argue such a distinction is made possible using inductive reasoning.

They begin by stating generalities of proving causation, and of causal systems. Firstly, they state that any intelligent system that aims to learn about their environment and act on that knowledge cannot rely solely on preprogrammed causal knowledge (which would be derived from human knowledge and experience and encoded into the system). Rather, it must be able to transform observable phenomenon into cause and effect relationships. This is a key point, and speaks to the statements of other authors who state domain knowledge is vital for causal inference. They argue further that when causal relationships are stated in ordinary conversation, they reflect probabilities of event occurring rather than absolute truths. Thus, probability theory should be sufficient to identify such relationships. It is clear that the authors place a large degree of faith in the abilities of ordinary people, going as far as to say of peoples ability to perceive causal relationships "..we must find a computational model that emulates this perception".

Key to the model proposed by Pearl and Verma (1995) is the notion of a directed acyclic graph (DAG). The authors theorise that fundamentally, all processes in nature are controlled by causal mechanisms that govern how observable and unobservable variables interact. In general, they state that "A causal model of a set of variables U is a directed acyclic graph (DAG), in which each node corresponds to a distinct element of U." Pearl and Verma (1995)

Variables are represented as nodes, with edges representing causal influences between those variables. Using this model, it becomes clear how the influence of parent variables on child variables can be found. One of the issues that has been raised in this chapter previously is that of the influence of unobservable factors, which are impossible to eliminate in observational studies in practice.

The authors here model these factors as probabilistic disturbances to the DAG, that perturbs the relationships within. By stating assumption of independence of parent and child nodes (or features), the authors state that the disturbances are local to that pairing. This is certainly a novel approach, and presents a theoretical framework for dealing with inevitable externalities.

Pearl and Verma (1995) on go to discuss the question of model structure and choice. Logically, since the model U is not bounded by any predetermined constraint (i.e. inputted causal knowledge or otherwise) there in an infinite amount of models that could be fitted to a given distribution. Each would create different causal relationships with a different set of probabilistic disturbances. The authors call on inductive reasoning here, arguing that any model can be removed if there is a more simple alternative that is as consistent with the data. Models not removed are referred to as *minimal models*, and are used to reach the authors definition of inferred causation. They state that "A variable X is said to have a causal influence on a variable Y if a strictly directed path from X to Y exists in every minimal model consistent with the data." Pearl and Verma (1995).

Moving on, the authors define a latent structure here as a pair $L =< D, O >$ where D is a causal model over all available variables and O is a subset of those variables. One such structure is preferred over another if its can be said to be more simple, as gauged by its expressive power rather than syntax. For example, if one latent structure that calls a great deal more parameters than another may still be preferable, if the second enables a greater range of probability distributions over the variables and is thus less precise or informative. Simpler models are more refined and contained, and are thus more easily falsified. Simpler models that aren't easily falsified then gained greater credibility.

Developing the theory of latent structures further, the authors define a minimal latent structure as well as a condition for structures to be consistent with a given distribution $\hat{P}$ before stating the definition for inferred causation. They state that given a distribution $\hat{P}$, a variable C has a causal influence on another variable E if and only if there is a directed path $C \rightarrow E$ in every minimal latent structure consistent with that distribution. The authors argue this can not be considered a guaranteed method of identifying physical processes in the real world, but rather it aids the examination of processes that can be plausibly inferred from non-experimental-data.

2.5 Research Gap

Over the course of this literature review, a snapshot of existing studies in this domain has been presented as well as a discussion of the implications for the current research. It is clear that a large body of work exists in this space, that has facilitated a deep and wide understanding of the drivers of economic success in corporations. Having said this, there are certainly gaps in this research left to be explored.

Firstly, it is clear that there is a wide range of plausible contributing factors to corporate economic success. Many of these were neglected by Moldovan and Mutu (2015) in their original study, which stands to benefit from their inclusion. Perhaps the more significant issue that the current study aims to address is that of the identification of causal relationships. Much work has gone into establishing strong correlations between governance structures and company outcomes, which are often reenforced with domain specific knowledge or simply accepted due to their plausibility by practitioners. The current research aims to apply modern work on statistics based causal estimation in order to determine which of the above findings have causal merit and which do not. Thus the potential contribution to the field is significant.

Chapter 3

Methodology

3.1 Introduction

This chapters outlines the methods followed during this analysis, with a view to explaining the steps taken in detail and to aid replication. The Knowledge Discovery in Databases (KDD) process as outlined by Fayyad *et al.* (1996) was followed where possible. This is a well established end-to-end framework for deriving knowledge from raw data. To that end, data acquisition and the raw data characteristics are discussed first. This is followed by a discussion of the preprocessing and reduction required to make this data useable, including how missing data and outliers are handled. An outline of the data mining techniques employed is given, including analysis of the advantages and disadvantages of the various statistical methods and associated software packages available. Following this, the steps taken for each element of this study are outlined in detail. That is, the replication of and expansion on the work by Moldovan and Mutu (2015) and the application of causal research in this domain. An analysis of the methods for interpreting results is given, including what measures of algorithmic success should be used and related matters. Overall, this chapter represents the technical aspect of this study and aims to facilitate the replication and expansion of this analysis by future researchers.

Git and GitHub was used as a version control and task tracking tool. All code modules written in fulfilment of this study, including all source files of this report, are available at `https://github.com/ReidConor/dissertation`. Using source control has various uses, not least acting as a cloud storage mechanism in case of local machine failure. In addition, changes to the project over time can be much more easily managed, including changes to the datasets used. Making this code available on GitHub also facilitates collaboration and the communication of ideas between collaborators.

3.2 Data Acquisition

3.2.1 Core Data

The primary source of data for the current analysis comes from the authors of the paper it extends. Darie Moldovan and Simona Mutu were kind enough to provide the data they used in their analysis, providing the complete dataset and granting permission to use it here. This is highly beneficial for a number of reasons. Firstly, any results derived here can be placed in a much clearer context, since we can directly and numerically compare the findings of this study to the original and identify areas of achieved improvement. Secondly, being granted access to a purpose built dataset prior to undertaking this analysis represents a significant catalyst for progress and expedites the process of gaining greater understanding in this domain. The statements made by the authors based on identified correlations can also be used in the causal inference stage of this study as the basis for the formulation of research questions.

Three datasets where provided by Darie Moldovan and Simona Mutu, each covering 52 features for three distinct stock indexes. They are; the S&P 500 based in the United States, the STOXX 600 based in Europe and the STOXX 300 based in Eastern Europe. Combined, these datasets total 1400 records of companies from the year 2014. The authors scrapped this data using the Bloomberg financial data repository, which contains a vast amount of historical financial data on companies across the world. In their study, the authors decided to analyse each market in isolation rather than in combination, inferring that the relationship between corporate governance and performance is characteristically different between markets.

3.2.2 New Factors - Independent

Part of this study is the exploration of new factors that could be introduced into the analysis to better explain corporate success. In other words, new independent variables to append onto the core dataset that extend the original research. A Bloomberg terminal was used to acquire all data outlined in this section due to the ease at which features could be found, extracted and integrated with the original dataset.

Discussed in section 2.2.3 is the importance of environmental performance and its influence on overall economic health. To this end an exploration of the data available in Bloomberg was conducted, however it was found that the amount of missing data for the majority of environment-related features in the year in question was prohibitive to their inclusion. Thus, it was decided to use a propriety score formulated by Bloomberg themselves as a proxy for performance in this area. The justification for this is outlined in section 2.2.5.

Another new independent feature introduced here is total CEO compensation for each company in this study, as discussed in section 2.2.6. CEO compensation is readily available in Bloomberg for the year under study, and so it is relatively easy to extract and append this measure onto the core dataset.

3.2.3 New Factors - Dependent

Another goal of this study is to explore other dependant variables, or in other words auxiliary indicators that characterise company outcomes. A single new dependent variable is included, namely the Beneish M Score as outlined in section 2.2.2. The M Score uses an aggregate of various company-specific financial ratios to calculate the probability of that company having intentionally manipulated it's reported earnings. There are two variations on this score, one using a combination of five financial ratios and the other adding an additional three. All variables were derived from Bloomberg for the appropriate years, and appended to the original dataset.

3.3 Data Pre-Processing and Reduction

3.3.1 Missing Values

Moldovan and Mutu (2015) decided to remove any observations in their data that had missing values in the dependant variable, or not enough information to calculate those values. It could be argued that these emissions are justified, since incomplete data could unfairly skew the properties of that observation and misrepresent it in the data. Any conclusions that were made using these observations could be fundamentally flawed. Below is table outlining the degree of missing dependant variables per dataset.

Dataset	Row Count	Missing Tobins Q Score	Missing Altman Z Score
SPX	500	4	81
SXXP	600	4	127
EEBP	300	3	65

For the classification stage of the current study, where an attempt is made to replicate the findings of the original authors, these rows will be removed in the same way.

However when it comes to the regression and causal estimation stage, missing data represents a more complex issue. Horton and Kleinman (2007) state that it is critically important to address missing data, particularly in observational analysis with may predictors (as in the current study), as it arises frequently in almost all investigations using real world data. There are a number of reasons for the presence of missing data, from randomly missing data points (i.e. the propensity to be missing is independent of the value itself) to non-randomly missing data points (where the true value influences the propensity for it be to missing). The table below outlines the degree of missing values in each dataset by listing the dimensions of each if complete-case analysis were carried out.

Dataset	Row Count	Complete Cases
SPX	500	56
SXXP	600	2
EEBP	300	0

It is clear that complete case analysis is infeasible here, and so some other method of handling incomplete data is required. Horton and Kleinman (2007) discuss a range of methods for addressing this issue, with the specific aim of enabling the fitting of a logistic regression model on a sample dataset. One such method is multiple imputation, which they describe as a multi-step approach to estimating incomplete data that relies on an assumption that values are missing at random. First, the missing entries are filled in m times. These new values are drawn from a distribution that is different for each entry and variable. There is then an analysis stage where the m completed data sets are studied in isolation. Finally, the m datasets are pooled into a final result. Rubin (2004) states that if the imputation method is correctly implemented, then the resulting dataset is valid for statistical modelling.

The companies represented in these datasets are public, and so are responsible for accurately reporting along a number of dimensions like company directorship and board composition, as well as financial statements which are audited by a third party. However this does not cover all features involved in this study, which may be optional for reporting purposes. Regulation in this regard also differs between markets involved in this analysis, making it difficult to deduce whether data is missing at random or not. Jakobsen *et al.* (2017) state that in this case, multiple imputation may be suitable.

A popular implementation of this technique is the `MICE` package in `R`, as outlined by van Buuren and Groothuis-Oudshoorn (2011). `MICE` imputes using chained equations, which involves specifying the imputation model on a variable by variable basis and using the other variables as predictors. At each step in the algorithm, an imputed value is generated and used in the imputation of the next variable. This process is repeated for each iteration, until convergence is reached as specified by the Gibbs sampling procedure, outlined in more detail by Yildirim (2012). This method can handle both continuous and discrete variables, as is required with the present data. `MICE` will be used to prepare each dataset for the regression as well as causal inference stages of the current study.

3.3.2 Outliers

Moldovan and Mutu (2015) state that they remove outliers in the data, citing a desire to remove *"data errors"*. They do not provide an explanation of what characterises an outlier or a data error, nor do they present any evidence that outliers are fair omissions. It is generally accepted that outliers must be proven to be mistakes at the data collection stage, or invalid in some other way to justify leaving them out of the analysis. Without such justification, outliers are valid data points and may prove crucial to the formulation of a faithful model. While they state that removal only discounts 122 observations in total, this amounts to roughly 9% of the original dataset.

Since outlier detection and omission can have a significant impact on model performance, as shown for example by Pollet and van der Meij (2017) and Zijlstra *et al.* (2011), this study will conduct identical analysis with and without the presence of outliers in order to assess the utility of their inclusion. As mentioned above, the original study neglects to detail how the authors characterised or identified outliers, and so some methodology for doing so must be chosen. Cousineau and Chartier (2010) reviews several different methods of outlier detection, one of which is Cook's distance originally proposed by Cook (1977). Cook's distance considers the influence of a given case i on all n fitted values in a regression analysis, and is calculated as

$$D_i = \frac{e_i^2}{pMSE}\Big(\frac{h_i i}{(1-h_i i)^2}\Big) \tag{3.3.1}$$

where e_i is the i^{th} element of the residual vector, $h_i i$ is known as the leverage and MSE is the mean square error. As noted before, both original dependent variables (the Tobins Q score and Altman Z score) are continuous in nature. Moldovan and Mutu (2015) threshold on this value to transform the problem to a classification task, whereas the current study both replicates this and performs regression on the original values. For this reason, a method to identify outliers that relies on regression analysis is chosen, and is used before any thresholding takes place.

3.3.3 Thresholding

In their analysis, Moldovan and Mutu (2015) threshold on both the Tobins Q score and Altman Z score to create classes from the continuous measures. This frames the problem as a classification task rather than regression which might be a more natural framing. Similar classes are created here, to facilitate a comparison of results. The original continuous measures are retained to facilitate regression analysis.

Tobins Q is discretised using the median to split observations into two classes, as per the original authors who cite Creamer and Freund (2010) as suggesting such a split. This is suitable for both the classification stage and causal estimation stage. The discretisation of the Altman Z score however is more involved. Moldovan and Mutu (2015) create three classes here, as suggested by Altman (1968). Those classes are listed as "distress", "grey" and "safe" referring to the company's risk of bankruptcy. The following values are used to create these classes.

This is suitable for the classification stage of the current study. However, the causal estimation stage requires a binary class as the dependant variable and so

Threshold	Class
AZS >2.99	Safe
2.99 >AZS >1.81	Gray
1.81 >AZS	Distress

Table 3.1: Altman Z score classes

the "grey" and "distress" classes are merged into one. Causal results referring to the Altman Z score thus refer to estimating the effect of some treatment on a "safe" or "not safe" level of bankruptcy risk.

3.4 Data Mining, Algorithms and Software

As referenced numerous times above, there are three main stages to this study; classification, regression and causal estimation. Each of these steps requires a distinct approach and choice of toolset and software. Here, a discussion on these choices is included with justification for each.

3.4.1 Classification

Moldovan and Mutu (2015) in the original study approached the research question as a classification problem, thresholding on the continuous dependant variables and using appropriate algorithms to achieve that goal. For each dataset and measure of success, they implemented four distinct classification algorithms and compared performance between each using identical metrics. This study implements a subset of these algorithms to facilitate a limited verification and comparison. Since the main objectives of this study lie elsewhere, this study neglects to include each and every algorithm used originally.

The `Adaboost M1` algorithm proved to be one of the highest performing implementations in the original study, and so is included here also. The `adabag` package in `R` is used, which implements the algorithm as proposed by Freund and Schapire (1996). As per the name, this algorithm uses *boosting*, which the authors state can be used to significantly reduce the error of weak learners by unifying them in a weighted sum that represents the final output of the boosted classifier. In this sense, the performance of each individual weak learner can be poor, however as long as each is better than a random guess the final result converges to what is known as a *strong learner*. `Adaboost` represents an improvement on bagging, attempting to build multiple models using randomly chosen training instances and eventually combining into a single model with improved accuracy. Adaboost adds an adaptive layer to this, by disproportionally weighting up poorly modelled instances (those with higher error) in subsequent models.

The next most performant algorithm used by Moldovan and Mutu (2015) was `J48`, which is a `Java` implementation of the `C4.5` decision tree algorithm. This algorithm builds decision trees from training data using information entropy, which represents the expected value of a random variable that describes the amount of information contained in a particular split or location in a decision tree. In this way the algorithm choses attributes of the data that most effectively and efficiently splits the samples into subsets enriched in one class. Since the current study uses `R` rather than `Java`, an translated implementation is required. The `C5.0` algorithm was chosen, which is a next generation version of the `C4.5` algorithm, implementing various runtime efficiencies in terms of speed and memory usage.

3.4.2 Regression

For the regression stage of this analysis, both the Tobins Q score and the Altman Z score remain as continuous variables which replace the thresholded values used above. A standard linear model is of the form

$$Y_i = X_i^T \beta + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2),\ i = 1, ...n \tag{3.4.1}$$

The optimal linear unbiased estimator for β is found by solving

$$(X^T X)\beta = X^T y \tag{3.4.2}$$

For equation 3.4.2 to be solvable, the matrix $X^T X$ must be invertible which is not the case when the number of independent variables is much larger than the number of observations in the dataset or if there is collinearity between variables that are previously believed to be independent. In the current dataset, the first condition is not met although the ratio of observations to features is as low as approximately 6:1 for the STOXX 300 dataset and so may still be a valid concern. Collinearity within this dataset is certainly a source of concern, due to the large number of features included. Analysis showed that of 1830 possible pairings of features in the in all three datasets, there were 796, 944 and 392 pairings with a statistically significant collinearity coefficient respectively. A high degree of collinearity between independent variables can cause the regression coefficients to become very sensitive to small changes in the model. It can also reduce the precision of the estimate coefficients, reducing the statistical power of the model. It is clear then that a regression methodology capable of dealing with this is required, that uses an alternative estimator for β.

Hoerl and Kennard (1970) proposed a modification of equation 3.4.2 based on a perturbation, denoted by λ, to the matrix $(X^T X)$ making it invertible. The equation thus becomes

$$(X^T X + \lambda I)\beta = X^T y \tag{3.4.3}$$

where I is the identity matrix. The estimator of β thus becomes

$$\hat{\beta}(\lambda) = (X^T X + \lambda I)^{-1} X^T y \tag{3.4.4}$$

and is called the *ridge estimator*. Hoerl and Kennard (1970) state that there exists some constant λ that leads to a mean square error less than that achieved by ordinary least squares regression. This holds even when the original matrix in equation 3.4.2 is in fact invertible. The question now arises about how to chose the optimal value for λ. There is much research on this topic, however for the current study the software will be computing values for λ itself taking into account the fact that we require the MSE to be as small as possible.

A shortcoming of ridge regression is that is does not inherently include any variable selection stage, in that it estimates coefficients but does not act to reduce any to exactly zero and thus disregard them entirely. This can be considered suboptimal when only a few predictors are likely to be influential or where model interpretability is important. Lasso regression (least absolute shrinkage and selection operator) aims to address these issues, by performing both the regularisation step carried out by ridge regression as well as a variable selection stage. Fonti and Belitser (2017) state that feature selection is a vital step in data analysis, and can act to simplify the final model by removing features that are not important as well as reducing the size of the problem to enable other algorithms to operate more quickly. Lasso works to minimise the sum of the squared errors with an upper bound on the total sum of the (absolute) values of the actual model parameters.

The current study implements both lasso and ridge regression under the general umbrella of regularised linear regression on all three dependent variables (Tobins Q and Altman Z scores, as well as the new Benish M-Score where available). The `glmnet` package in `R` is used to achieve this, which fits a generalised linear model using a regularisation parameter `alpha`. `Alpha` dictates where along the spectrum between lasso and ridge the penalty factor lies. A value of 1 leads the algorithm to use lasso, a value of 0 leads the algorithm to use ridge. Values inbetween lead to elastic net regression, described by Fonti and Belitser (2017) as a combination of the two. For each dataset and dependent variable, a series of models are built using a range of `alpha` values between 0 and 1 in 0.1 increments. Analysis is then carried out to see which implementation minimises the MSE.

3.4.3 Causal Estimation

For the causal estimation portion of this study, a propensity score matching algorithm is used as discussed in section 2.4. In order to achieve this, a module named `causality` written in `python` and made available in GitHub is used. This project is available at `https://github.com/akelleh/causality`. Among other pieces of functionality, this module allows the direct formulation and execution of propensity score matching and provides an interface for graphically judging the quality of the matching process. `causality` was installed using `pip`, and interfaced with using custom written `R` and `python` scripts that prepared the data, executed the module and collected the results.

The first step in this stage of the study is the formulation of research questions, many of which are derived from the correlations and *'if-this-then-that'* style conclusions of Moldovan and Mutu (2015). Many of those conclusions are market specific. However in this study each statement is tested across all markets in order to verify the applicability of each on a global scale. To those statements, this study adds new research questions motivated by existing research in this domain outlined in the literature review. A list and discussion of those questions can be found in section 3.5.

Each research question informs the choice of *treatment* and *outcome* pairs. Treatments are variables contained in the data already and represent some condition that may cause a variation in the outcome. The outcome then is another term for a dependant variable, as used in both the classification and regression stages of this study. We use both the Tobin Q score and Altman Z score, as well as the Benish M-Score where possible.

For each pair, a collection of variables to control for are selected. Note that propensity score matching requires the control on variables that characterise whether a unit receives the treatment or not, so that the effect of the treatment on the outcome can be isolated. Thus, a classifier was built in a very similar manner to the method outlined in section 3.4.1 to model the allocation of the treatment in each case. The variables that are reported as important by that model are then selected as the variables to control for in the matching algorithm. The algorithm implemented in this study requires a binary outcome variable. Thus, when Tobin's Q score is used as the outcome, the binary class as per the original paper by Moldovan and Mutu (2015) is used. When the Altman Z score is used as the outcome, the "distress" and "grey" risk bands are merged into one yielding a binary class that can be interpreted as "safe" and "not safe" risk levels.

3.5 Causal Estimation - Motivating Statements

Moldovan and Mutu (2015) make eight statements regarding the influences of corporate governance on corporate success. To these, the current study adds two that are based on findings contained in the literature that warrant further investigation. All are listed below, each referring to the market that the finding was original attributed to. Moldovan and Mutu (2015) can be assumed to be the authors unless otherwise stated.

1. For the American companies inside the S&P 500 index, we found a positive correlation between the percentage higher than 20pct of women in the board and the Tobins Q ratio.

2. For the American companies inside the S&P 500 index, we found...the presence of an independent lead director in the company along with a financial leverage higher than 2.5 incur a higher risk of bankruptcy.

3. When analysing the Eastern European companies data, we found that a smaller age range for the board members is positively related with the companies performance.

4. When analysing the Eastern European companies data, we found...that a financial leverage less than 4 is needed in order to be on the upper side of the Tobins Q ratio.

5. When analysing the Eastern European companies data, we found...to be on the safe zone of the Altman Z-score it is important to have an independent chairperson or even a woman as CEO.

6. For the Western European companies....the presence of an independent lead director or a former CEO in the board could be a sign of weaker performances, being negatively correlated with Tobins Q

7. For the Western European companies.... a large percentage of women in the board could also affect negatively the performance.

8. For the Western European companies....for the companies with large financial leverage in order to be in the “safe” zone of the Altman Z-score it could be a good idea to adopt an Auditing Committee with more than four people.

9. Core and Larcker (1999) state that firms with weaker corporate governance underperform on the stock market, and are more likely to rewarded CEO with greater levels of compensation.

10. Fatemi and Kaiser (2017) state that strong ESG (Environmental, Social and Governance) performance is associated with higher firm valuations.

3.6 Interpretation

3.6.1 Dependent Variables

The dependent variables used in this study have been mentioned frequently up until this point, without reference to what it means to optimise those measures. A brief explanation is included below on a variable basis about what it means to increase or decrease these measures in real world terms.

The Tobin’s Q score is calculated as the market value of a company divided by the replacement value of a firm’s assets. In theory, a value between 0 and 1 infers that the stock is undervalued since the cost to replace that companies assets is greater than the value of its stock. Such companies would then be attractive to buyers, who might purchase the company instead of establishing their own similar business. A Tobin’s Q score over 1 infers that the company is overvalued, because the stock is more expensive than the replacement cost of its assets. These companies then might experience increased levels of competition from new businesses, who seek to capture some of the market share on offer and in turn reduce the Q score of the original company. It is clear then that a high Q score is neither inherently a good or bad thing, but rather a simple indicator whose significance relies on the interests of the interpreter. From an objective point of view, an optimal value is 1 and any extremes away from this neutral point can be considered suboptimal. Having said this, a large quantity of the literature refers to increases in, or large values of, the Q score as advantageous or desirable. This includes the original paper that this study is based on, and so a similar style is adopted here.

The Altman Z score is more straightforward. The Z score, originally designed for firms in the manufacturing industry, gauges a companies likelihood of entering bankruptcy. This score can easily be discretised, since officially a score below 1.8 indicates a high risk of bankruptcy while a score above 3 indicated a low risk. Scores in-between cover a grey area of risk. Thus, optimising the Z score means increasing it towards 3.

Finally, the Benish M score measures the likelihood that the reported earnings of a company have been intentionally manipulated. Overall, lowering the M score is optimal since a score below -2.22 (a greater negative) indicates the firm is unlikely to be manipulating earnings.

3.6.2 Classification

Moldovan and Mutu (2015) used a number of metrics for measuring performance. They first show the accuracy of each model, which is simply the number of correctly identified observations in the data. Next they show the precision for each class, which is the number of correctly identified instances over the total predicted instances for that class. In the literature, these measures are also referred to the sensitivity and specificity.

Finally, they list the area under the receiver operator characteristic (ROC) curve. A ROC curve is created by plotting the sensitivity against the fall-out (or 1 - specificity) at various discrimination thresholds. The discrimination threshold describes the cutoff imposed on the predicted probabilities required for assigning an observation to a given class. The area under the curve (AUC) then, as described by Flach (2007) is a single numerical measure of the model's performance, equivalent to the probability that a uniformly drawn random positive is ranked before a random negative.

3.6.3 Regression

As mentioned above, regularised linear regression is use in this study and varies with the penalty applied from ridge to lasso regression. The r^2 for each model is presented, as an indication of the quantity of the variance in that data that is described by the model. Alongside this, the root mean square error (RSME) is included as an measure of the difference between the values predicted by the model and the actual values observed. Important to note here is that the RMSE is in the same units as the dependent variable it describes.

3.6.4 Causal Estimation

Section 3.5 outlines the research questions proposed for this section of the current study, and will form the basis for a number of result sets. For each question (and thus treatment and outcome pair), a number of performance related metrics are presented.

Each result set is presented following a uniform template. A header is given as an indication of the type of corporate governance features involved. The market for which the results relate is listed, followed by the dependent variable (or the outcome). The treatment is given in the form:

$$condition \quad ? \quad value\ if\ true \ : \ value\ if\ false \tag{3.6.1}$$

The *condition* here is the specific combination of independent variables that dictate whether an observation is in the treated or control group.

After this the motivating statement from section 3.5 is given if applicable. In brackets, a flag is given for whether the results agree with the statement or not as well as whether this statement was originally made about this particular market or not (*original* or *not original*). This is purely to aid quick interpretation.

The causal estimate as per the matching process is then given, as a 95% confidence interval. These figures represent the average treatment effect (ATE) of the intervention on the outcome. The ATE measures the difference in mean outcomes between the treated and control group, or in other words the average gain from treatment for units who were actually treated. A positive ATE indicates the treatment acted to increase the magnitude of the outcome measure, a negative the opposite. This portion of the study uses a binary outcome variable, and so the ATE can be interpreted as the percentage difference in the probability of either of the outcomes for a given unit. For example, an ATE of 15% translates to an increase of 15% in the probability of outcome 1. A -15% ATE translates to an increase of 15% in the probability of obtaining outcome 0. For the outcome as measured by Tobin's Q score, a class of 1 translates to a higher than average value (or a inference that the stock is overvalued). For the outcome as measured by the Altman Z score, an outcome of 1 translates to a "safe" risk status.

Finally a series of plots are included showing the quality of that matching process on a covariate basis. Important here is an overlap on the x-axis between the two distributions. Significant overlap shows that the matching process was successful, finding equivalent observations in both the treated and control group to compare.

Chapter 4

Results

4.1 Introduction

This chapter outlines the results of each stage of this study. First, results from the replication of the work of Moldovan and Mutu (2015) are shown, who modelled the problem as a classification exercise by thresholding on the continuous, success measuring dependant variables. Next, regression results are presented which are obtained by neglecting that thresholding. Finally, causal estimation results are presented which are obtained by implementing propensity score matching, using those findings as the basis for research questions.

Below, some summary statistics are presented on the dependent variables used in this study. This is so as to provide a reference point for analysing the results below, particularity the regression results which list the RMSE (root mean square error) which is in the same units as the dependant variable involved.

Tobin's Q Score						
Market	**Min**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
S&P 500	0.8268	1.2670	1.7490	2.1760	2.5180	9.6360
STOXX Europe 600	0.7437	1.0630	1.4300	1.9250	2.0810	72.290
STOXX Eastern Europe 300	0.3399	0.9335	1.0590	1.3540	1.4320	7.8870

Altman Z Score						
Market	**Min**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
S&P 500	-9.896	2.450	4.026	4.562	5.372	28.480
STOXX Europe 600	-0.6806	2.2150	3.3420	4.3980	5.0350	46.510
STOXX Eastern Europe 300	-1.994	1.880	2.816	3.447	4.225	24.250

Benish M Score							
Market	**Version**	**Min**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
S&P 500	Eight Var	-8.905	-2.829	-2.634	-2.660	-2.471	0.000
S&P 500	Five Var	-13.720	-3.013	-2.918	-2.903	-2.775	0.525

4.2 Classification

For the classification portion of this study, two of the most performant algorithms used by Moldovan and Mutu (2015) are implemented. For each dataset and algorithm, the original results are presented alongside the results of this study as a direct comparision. Include also are results obtained from an identical analysis but with outliers left in the dataset rather than removing them. This is discussed in section 3.3.2. For the S&P, results are also presented that relate to identical analysis but with extra covariates included, as outlined in section 3.2.2.

S&P 500

Below, the results for the S&P market are shown. Note that the original study removed outliers before modelling, and so it should be assumed that the current study did likewise unless otherwise stated.

Dependent Variable : Tobin's Q Score						
Algorithm	**Study**	**Modifier**	**Accuracy (%)**	**Precision Class 0**	**Precision Class 1**	**ROC**
Adaboost	M&M	-	89.7177	0.890	0.905	0.957
Adaboost	Current	-	92.1213	0.948	0.897	0.922
Adaboost	Current	With Outliers	93.3734	0.950	0.915	0.933
Adaboost	Current	Extra Variables	94.5783	0.962	0.931	0.947
J48	M&M	-	85.2823	0.841	0.866	0.854
J48 (C5.0)	Current	-	87.9519	0.886	0.871	0.879
J48 (C5.0)	Current	With Outliers	87.9518	0.840	0.930	0.885
J48 (C5.0)	Current	Extra Variables	89.7590	0.951	0.847	0.899

Table 4.1: Classification Results - S&P 500, Tobin's Q Score

			Dependent Variable : Altman Z Score				
Algorithm	**Study**	**Modifier**	**Accuracy (%)**	**Precision Class 0**	**Precision Class 1**	**Precision Class 2**	**ROC**
Adaboost	M&M	-	82.944	0.843	0.725	0.865	0.942
Adaboost	Current	-	83.3334	0.818	0.348	0.957	0.879
Adaboost	Current	With Outliers	84.892	0.818	0.529	0.910	0.861
Adaboost	Current	Extra Variables	88.4892	0.882	0.6	0.925	0.871
J48	M&M	-	73.6086	0.722	0.569	0.816	0.843
J48 (C5.0)	Current	-	78.985	0.629	0.400	0.871	0.802
J48 (C5.0)	Current	With Outliers	81.2949	0.591	0.529	0.910	0.832
J48 (C5.0)	Current	Extra Variables	76.978	0.6	0.485	0.928	0.831

Table 4.2: Classification Results - S&P 500, Altman Z Score

Shown below is a sample decision tree, constructed by the C5.0 (implementation of J48) for the Altman Z score. This was constructed using the auxiliary features included in this study. Although those variables do not drastically influence the aggregate model performance (as shown by the tables above), the social disclosure score interestingly appears near the bottom and is assigned an importance of 5% by the algorithm.

```
Asset <= 0.4571385:
:...Fincl..l <= 1.938166:
:   :...P.B <= 2.182174: grey (8)
:   :   P.B > 2.182174: safe (6.1/0.1)
:   Fincl..l > 1.938166:
:   :...X..Indep.Directors <= 84.62:
:       :...EPS > 4.76: grey (2)
:       :   EPS <= 4.76:
:       :   :...P.B <= 2.246135: distress (5)
:       :       P.B > 2.246135: safe (7.2/0.2)
:       X..Indep.Directors > 84.62:
:       :...Dvd.Yld > 2.724871: distress (23)
:           Dvd.Yld <= 2.724871:
:           :...Oper.ROE <= 9.037546: distress (7)
:               Oper.ROE > 9.037546:
:               :...Fincl..l <= 5.621416: grey (6)
:                   Fincl..l > 5.621416: distress (2)
Asset > 0.4571385:
:...Interest <= 74.38541:
    :...Board.Mtgs.. <= 6: safe (5/1)
    :   Board.Mtgs.. > 6:
    :   :...Sz.Aud.Cmte <= 3: distress (2.1/1)
    :       Sz.Aud.Cmte > 3:
    :       :...X..Feml.Execs.1 <= 0: distress (2.1/0.1)
    :           X..Feml.Execs.1 > 0: grey (8/1)
    Interest > 74.38541:
    :...Fincl..l <= 3.202927:
        :...X5Yr.Avg.Adj.ROE > 13.8612: safe (116.1/1.6)
        :   X5Yr.Avg.Adj.ROE <= 13.8612:
        :   :...X..Wmn.on.Bd <= 3: safe (32.4/4.5)
        :       X..Wmn.on.Bd > 3: grey (2)
        Fincl..l > 3.202927:
        :...Asset <= 0.6752533:
            :...OPM.T12M <= 18.86604: distress (5)
            :   OPM.T12M > 18.86604: grey (3.1)
            Asset > 0.6752533:
            :...EV.EBITDA.T12M <= 4.797605: distress (2)
                EV.EBITDA.T12M > 4.797605:
                :...Indep.Directors <= 9: safe (19.8/0.1)
                    Indep.Directors > 9:
                    :...SOCIAL_DISCLOSURE_SCORE <= 28.0702: grey (3.2)
                        SOCIAL_DISCLOSURE_SCORE > 28.0702: safe (9/1)
```

Figure 4.1: J48 - S&P with Extra Variables for Altman Z Score - Decision Tree

STOXX Europe 600

Below, the results for the STOXX Europe 600 index are shown. Note that the auxiliary variables mentioned in section 3.2.2 and 3.2.3 were not gathered for this index.

Dependent Variable : Tobin's Q Score						
Algorithm	**Study**	**Modifier**	**Accuracy (%)**	**Precision Class 0**	**Precision Class 1**	**ROC**
Adaboost	M&M	-	88.2353	0.891	0.874	0.946
Adaboost	Current	-	94.4723	0.961	0.929	0.945
Adaboost	Current	With Outliers	90.9547	0.904	0.916	0.910
J48	M&M	-	87.395	0.871	0.877	0.874
J48 (C5.0)	Current	-	89.9497	0.915	0.884	0.900
J48 (C5.0)	Current	With Outliers	87.940	0.887	0.873	0.879

Table 4.3: Classification Results - STOXX Europe 600, Tobin's Q Score

Dependent Variable : Altman Z Score							
Algorithm	**Study**	**Modifier**	**Accuracy (%)**	**Precision Class 0**	**Precision Class 1**	**Precision Class 2**	**ROC**
Adaboost	M&M	-	76.2681	0.804	0.522	0.864	0.769
Adaboost	Current	-	82.1656	0.793	0.692	0.887	0.895
Adaboost	Current	With Outliers	77.0709	0.758	0.591	0.869	0.874
J48	M&M	-	74.2754	0.788	0.531	0.803	0.856
J48 (C5.0)	Current	-	66.8789	0.75	0.369	0.804	0.798
J48 (C5.0)	Current	With Outliers	69.4267	0.72	0.4318	0.8182	0.8243

Table 4.4: Classification Results - STOXX Europe 600, Altman Z Score

STOXX Eastern Europe 300

Finally, the results for the STOXX Eastern Europe 300 index are shown below. Again, the auxiliary variables mentioned in section 3.2.2 and 3.2.3 were not gathered for this index.

Dependent Variable : Tobin's Q Score						
Algorithm	**Study**	**Modifier**	**Accuracy (%)**	**Precision Class 0**	**Precision Class 1**	**ROC**
Adaboost	M&M	-	81.8182	0.823	0.813	0.889
Adaboost	Current	-	84.8488	0.88	0.816	0.848
Adaboost	Current	With Outliers	82.8282	0.826	0.829	0.828
J48	M&M	-	77.1044	0.783	0.760	0.836
J48 (C5.0)	Current	-	78.7879	0.7021	0.865	0.783
J48 (C5.0)	Current	With Outliers	81.8182	0.826	0.808	0.818

Table 4.5: Classification Results - STOXX Eastern Europe 300, Tobin's Q Score

Dependent Variable : Altman Z Score							
Algorithm	**Study**	**Modifier**	**Accuracy (%)**	**Precision Class 0**	**Precision Class 1**	**Precision Class 2**	**ROC**
Adaboost	M&M		63.7602	0.746	0.496	0.675	0.771
Adaboost	Current	-	58.974	0.50	0.36	0.714	0.781
Adaboost	Current	With Outliers	75.641	0.500	0.482	0.936	0.821
J48	M&M	-	65.6676	0.79	0.615	0.598	0.8
J48 (C5.0)	Current	-	62.821	0.334	0.447	0.813	0.723
J48 (C5.0)	Current	With Outliers	70.5122	0.334	0.650	0.746	0.879

Table 4.6: Classification Results - STOXX Eastern Europe 300, Altman Z Score

4.3 Regression

There are no results from the original study to directly compare the regression results to, and so they are presented stand-alone. Tables are broken down by index and dependant variable. An indication is given, in the form of yellow highlighting, of the optimal *alpha* value and thus the optimal mix of ridge and lasso regression for each results set.

S&P 500

Shown below are regression results for the S&P 500 stock index.

Dependent Variable - Tobin's Q Score		
Alpha	**r^2**	**RMSE**
0 (Ridge)	0.514504662287613	1.54335249821869
0.1	0.736922216718774	0.752106800916883
0.2	0.723096373541166	0.918016920932065
0.3	0.736780891632018	0.729876793419644
0.4	0.728597775405239	0.802518515140072
0.5	0.736648154658892	0.736084477965634
0.6	0.728370638237572	0.729268748938413
0.7	0.736782224916274	0.740975835009645
0.8	0.737003712724359	0.757117048022052
0.9	0.143531162654549	1.94074213436316
1.0 (Lasso)	0.724748692727598	0.730134780410399

Table 4.7: Regression Results - S&P 500, Tobin's Q Score

Dependent Variable - Altman Z Score		
Alpha	**r^2**	**RMSE**
0 (Ridge)	0.372331817825629	13.7865694258794
0.1	0.512946605245459	10.1747805657968
0.2	0.478411885246279	10.4174275138483
0.3	0.512484854172808	10.3869701107608
0.4	0.478650059026158	9.85821851442949
0.5	0.512396355841667	9.61495426389283
0.6	0.513129590937884	10.3182884065172
0.7	0.512360136424729	10.2731129074996
0.8	0.51296402797329	10.0251122703722
0.9	0.158067206245976	16.6297402971145
1 (Lasso)	0.481696182985345	9.7934982676994

Table 4.8: Regression Results - S&P 500, Altman Z Score

The Benish MScore, a probabilistic indictor of intentional financial reporting fraud, was scraped for the S&P 500 set of companies, the regression results for which are presented below. As mentioned previously, both eight variable and five variable versions were included in this analysis. Each is shown separately below.

Dependent Variable - Benish M-Score			
Version	**Alpha**	r^2	**RMSE**
Eight Var	0 (Ridge)	0.251555481058366	0.484013829347385
Eight Var	0.1	0.252608488775166	0.46535791024093
Eight Var	0.2	0.24789844766805	0.468175452095253
Eight Var	0.3	0.2198137766366	0.49415213706807
Eight Var	0.4	0.249189763200139	0.465779130278631
Eight Var	0.5	0.24549798240231	0.500453741652562
Eight Var	0.6	0.249004530587714	0.470463109817926
Eight Var	0.7	0.243925438107858	0.49064305638191
Eight Var	0.8	0.249749421198673	0.481038667522551
Eight Var	0.9	0.250291037515949	0.48584148244414
Eight Var	1 (Lasso)	0.246306522467078	0.477759843928307
Five Var	0 (Ridge)	0.291732059704306	0.962571803318562
Five Var	0.1	0.290985985078737	0.934920221885852
Five Var	0.2	0.289997856384489	0.922163141720248
Five Var	0.3	0.288128362377427	0.959393842374148
Five Var	0.4	0.290435845389172	0.936954320855632
Five Var	0.5	0.285369018902806	0.969059899849606
Five Var	0.6	0.290276479666761	0.927261185661328
Five Var	0.7	0.289285439066985	0.982050720561862
Five Var	0.8	0.290876191983076	0.94214790269637
Five Var	0.9	0.290922502141949	0.937690067080429
Five Var	1 (Lasso)	0.290548584733936	0.949720674526324

Table 4.9: Regression Results - S&P 500, Benish M-Score

STOXX Europe 600

Shown below are regression results for the STOXX Europe 600 stock index.

Dependent Variable - Tobins Q Score		
Alpha	**r^2**	**RMSE**
0 (Ridge)	0.924245520228615	4.52752573635043
0.1	0.936820396581318	5.61364933389961
0.2	0.936272434248921	6.12198520884377
0.3	0.936184193929519	6.60733408752952
0.4	0.936486005016966	6.78628962380346
0.5	0.937847903413279	6.655449299569
0.6	0.935338676848096	6.88804707065998
0.7	0.937887391702568	7.08560490201231
0.8	0.935720858099178	7.04925327857539
0.9	0.935982694015804	7.14379996319223
1 (Lasso)	0.935805655675326	7.19289948947667

Table 4.10: Regression Results - STOXX Europe 600, Tobin's Q Score

Dependent Variable - Altman Z Score		
Alpha	**r^2**	**RMSE**
0 (Ridge)	0.493776589092134	12.4040377499728
0.1	0.492009638562525	13.0187735752629
0.2	0.492487747960643	12.571232251759
0.3	0.493364660095417	12.8447464487333
0.4	0.495255261342842	12.8572572045948
0.5	0.443363098262155	16.167925121637
0.6	0.49250064188208	12.408424237533
0.7	0.410969802370212	14.5362919598181
0.8	0.493903828085151	12.4931721421183
0.9	0.494432810764225	12.128503187072
1 (Lasso)	0.494019676145083	12.1087103638653

Table 4.11: Regression Results - STOXX Europe 600, Altman Z Score

STOXX Eastern Europe 300

Shown below are regression results for the STOXX Eastern Europe 300 stock index.

Dependent Variable - Tobins Q Score		
Alpha	**r^2**	**RMSE**
0 (Ridge)	0.704355214812389	0.498974752786726
0.1	0.694006623409445	0.53354739142002
0.2	0.727564731166982	0.47859949878705
0.3	0.69625780677679	0.475553525491744
0.4	0.707544799542626	0.478664221197991
0.5	0.656609682843319	0.546502619727653
0.6	0.68326568451999	0.50410936532166
0.7	0.667251282563606	0.590747734679593
0.8	0.668367543398841	0.462386508939621
0.9	0.699087289610586	0.454557500864114
1 (Lasso)	0.660904967486424	0.462774482588657

Table 4.12: Regression Results - STOXX Eastern Europe 300, Tobin's Q Score

Dependent Variable - Altman Z Score		
Alpha	**r^2**	**RMSE**
0 (Ridge)	0.497359901145225	7.4448320397538
0.1	0.521420102039222	7.63377981958334
0.2	0.479465764190985	8.27754797472766
0.3	0.503750477243543	7.70658821833145
0.4	0.503990128494823	7.86100761373032
0.5	0.524032769294246	8.25112976717522
0.6	0.512202382858808	8.15220512063965
0.7	0.507649931007144	8.02131671759577
0.8	0.493904840655251	8.1278950336226
0.9	0.495564417663372	8.2173172609159
1 (Lasso)	0.480490621105762	8.19196705962938

Table 4.13: Regression Results - STOXX Eastern Europe 300, Altman Z Score

4.4 Causal Estimation

This section deals with results pertaining the the causal estimation stage of the current study. As mentioned before, causal estimation here involves the selection of treatment and effect pairs, as well as covariates to control on. Results are presented with reference to these pairs and covariates, and reference back to the motivating statement made in the original study where appropriate (see section 3.5 for more detail).

Note that while all original findings are tied to specific markets, those findings were tested for all markets in this study. Therefore the motivating statement listed for a particular set of results below may not match the dataset to which it was originally assigned. In some cases it was not possible to achieve a sufficient matching rate for particular datasets, in others the estimates were insignificant or the 95% CI crossed zero. In cases of the latter, results are omitted because the aggregate direction of the effect cannot be determined.

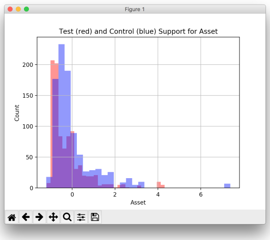
Each result set is presented following a uniform template, as per section 3.6.4.

4.4.1 S&P 500

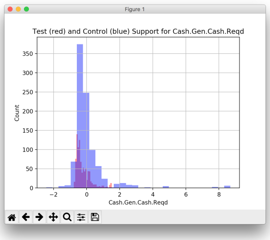
4.4.1.1 Independent Director & Financial Leverage

Market	S&P 500
Outcome	Altman Z Score
Treatment	(Indep Lead Dir & Financial Leverage > 2.5)? 1 : 0
Motivating Statement	Statement 2 [agrees, original]
Estimate ($\%\Delta, 95\%\ CI$)	(-0.15 , -0.12 , -0.09)

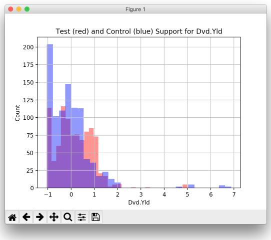
Matching Plots



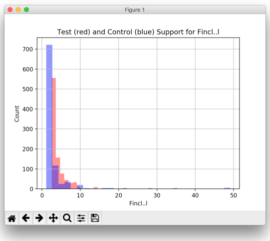
(a) Asset



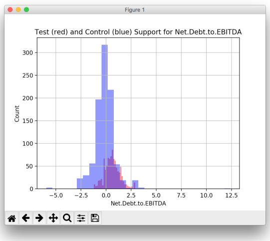
(b) Cash Gen/Cash Reqd



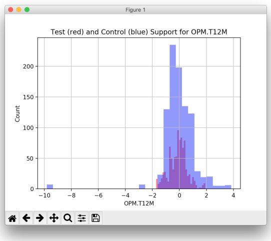
(c) Dvd Yld



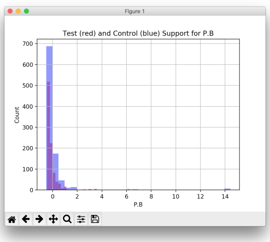
(d) Fin Lev



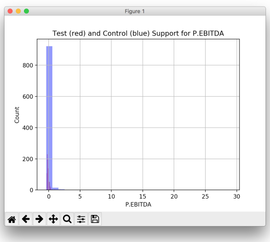
(e) Net Debt/EBITDA



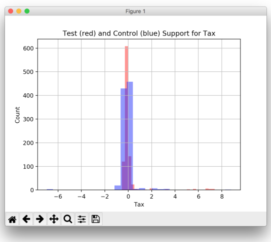
(f) OPM 12Mth



(g) P/B



(h) P/EBITDA



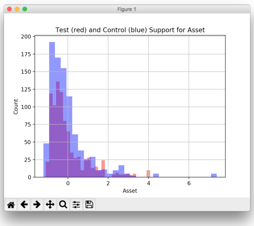
(i) Tax

Figure 4.2: Causal Estimation - Independent Director & Financial Leverage

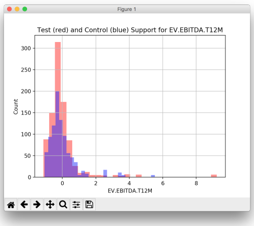
4.4.1.2 CEO Compensation

Market	S&P 500
Outcome	Tobins Q Score
Treatment	CEO Comp > median(CEO Comp)? 1 : 0
Motivating Statement	Statement 9 [agrees, original]
Estimate ($\%\Delta, 95\%\ CI$)	(-0.11, -0.085, -0.06)

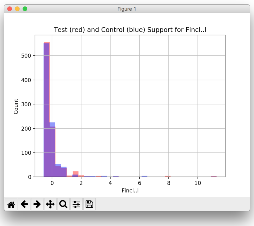
Matching Plots



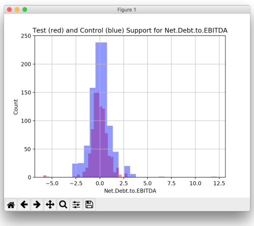
(a) Assets



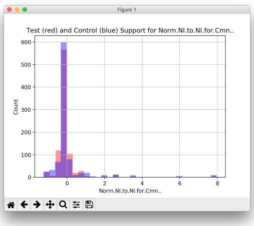
(b) EBITDA 12Mth



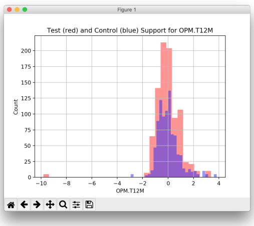
(c) Fin Lev



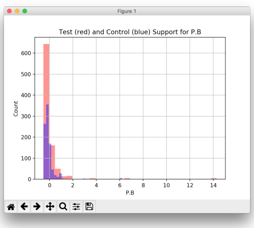
(d) Net Debt/EBITDA



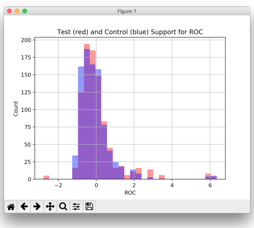
(e) Norm NI to NI for Cmn



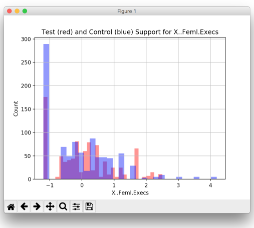
(f) OPM 12Mth



(g) P/B



(h) ROC



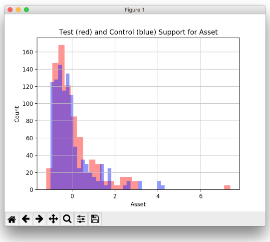
(i) % Female Executives

Figure 4.3: Causal Estimation - CEO Compensation

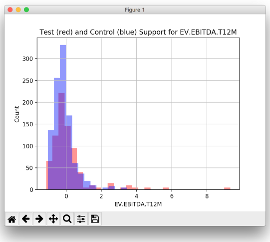
4.4.1.3 Board of Directors Average Age

Market	S&P 500
Outcome	Tobins Q Score
Treatment	B.O.D Avg Age < mean(B.O.D Avg Age)? 1 : 0
Motivating Statement	3 [disagrees, not original]
Estimate ($\%\Delta, 95\%\ CI$)	(-0.10, -0.08 , -0.06)

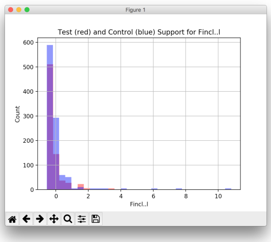
Matching Plots



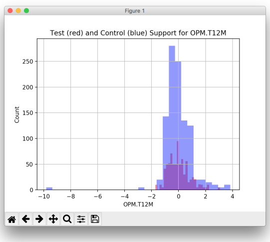
(a) Asset



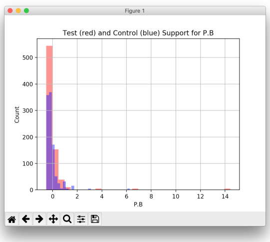
(b) EBITDA 12Mth



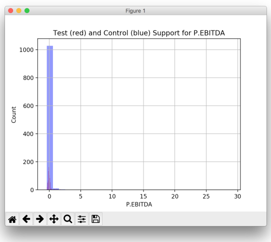
(c) Fin Lev



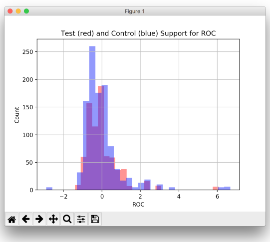
(d) OPM 12Mth



(e) P/B



(f) P/EBITDA



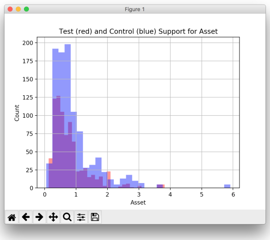
(g) ROC

Figure 4.4: Causal Estimation - Board of Directors Average Age

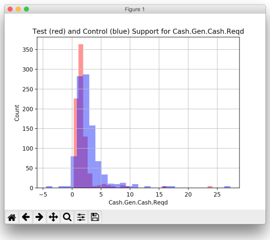
4.4.1.4 Social Disclosure Score

Market	S&P 500
Outcome	Altman Z Score
Treatment	Social Disclosure > mean(Social Disclosure) ? 1 : 0
Motivating Statement	Statement 10 [agrees, NA]
Estimate ($\%\Delta, 95\%\ CI$)	(0.04 , 0.09 , 0.14)

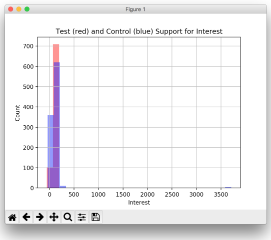
Matching Plots



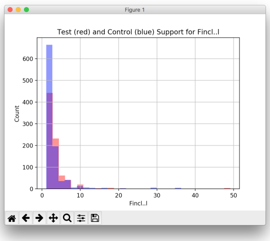
(a) Asset



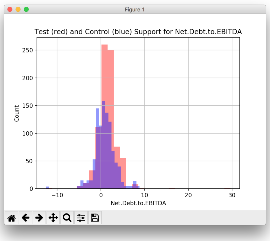
(b) Cash Gen/Cash Reqd



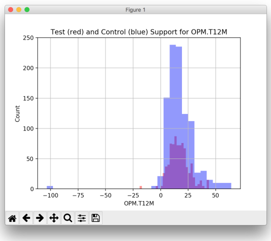
(c) Interest



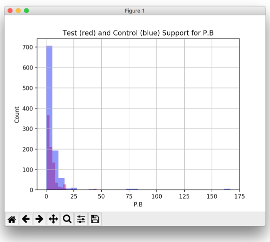
(d) Fin Lev



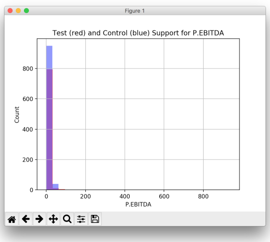
(e) Net Debt/EBITDA



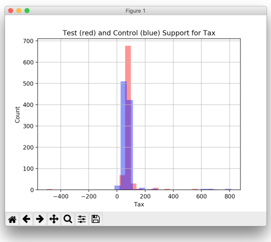
(f) OPM 12Mth



(g) P/B



(h) P/EBITDA



(i) Tax

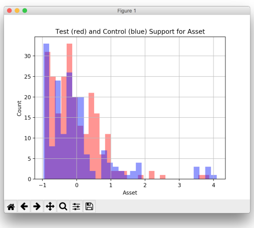
Figure 4.5: Causal Estimation - Social Disclosure

4.4.2 STOXX Europe 600

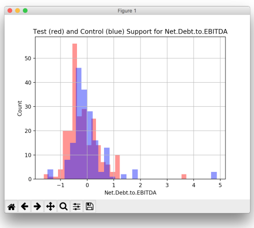
4.4.2.1 Female Board Membership

Market	STOXX Europe 600
Outcome	Tobin's Q Score
Treatment	% Female's on Board > 20%? 1 : 0
Motivating Statement	Statements 7 [disagrees, original], 1 [agrees, not original]
Estimate $(\%\Delta, 95\%\ CI)$	(0.06, 0.1, 0.14)

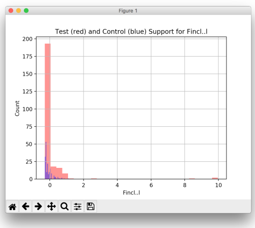
Matching Plots



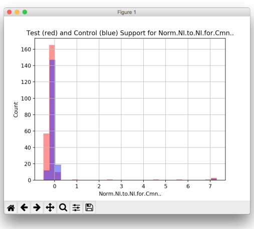
(a) Asset



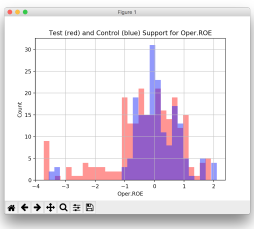
(b) Net Debt/EBITDA



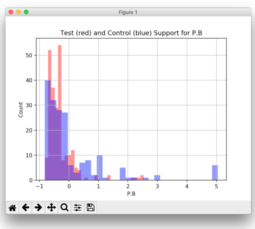
(c) Fin Lev



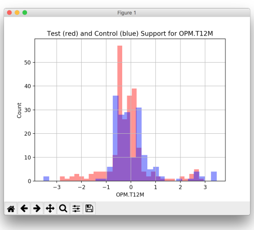
(d) Norm NI to NI for Cmn



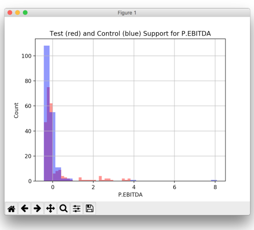
(e) Oper ROE



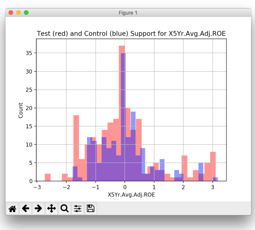
(f) P/B



(g) OPM 12Mth



(h) P/EBITDA



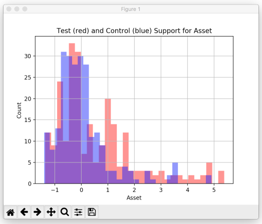
(i) AVG Adj ROE

Figure 4.6: Causal Estimation - Female Board Membership

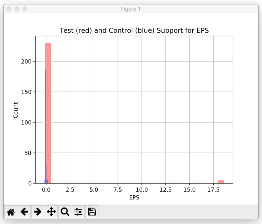
4.4.2.2 Independent Director or Former CEO on the Board

Market	STOXX Europe 600
Outcome	Tobin's Q Score
Treatment	(Indep Director | Former CEO on Board)? 1 : 0
Motivating Statement	Statement 6 [disagrees, original]
Estimate (%Δ, 95% CI)	(0.30 , 0.35 , 0.40)

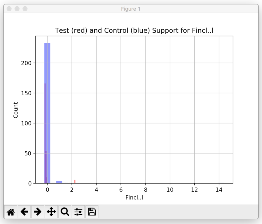
Matching Plots



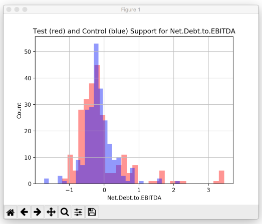
(a) Asset



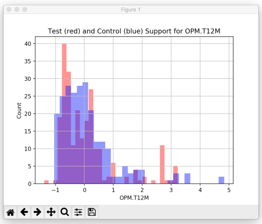
(b) EPS



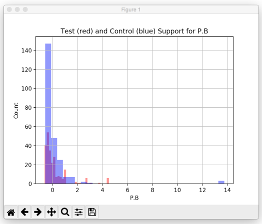
(c) Fin Lev



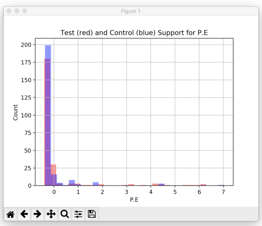
(d) Net Debt/EBITDA



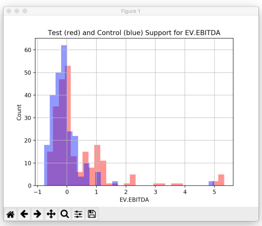
(e) OPM 12Mth



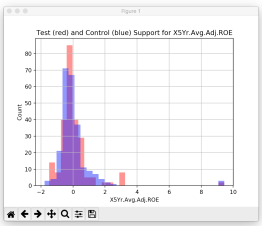
(f) P/B



(g) P/E



(h) EV/EBITDA



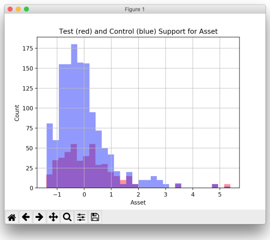
(i) Avg Adj ROE

Figure 4.7: Causal Estimation - Indep Director or Former CEO on the Board

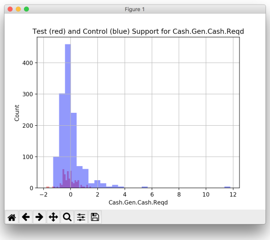
4.4.2.3 Independent Director & Financial Leverage

Market	STOXX Europe 600
Outcome	Altman Z Score
Treatment	(Indep Lead Dir & Financial Leverage > 2.5)? 1 : 0
Motivating Statement	Statement 2 [agrees, not original]
Estimate ($\%\Delta, 95\%\ CI$)	(-0.30 , -0.27 , -0.24)

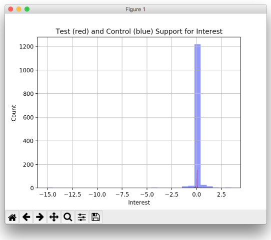
Matching Plots



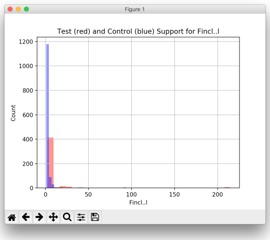
(a) Asset



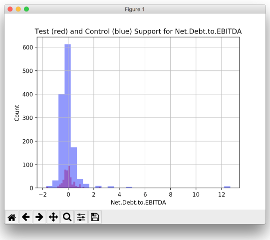
(b) Cash Gen/Cash Reqd



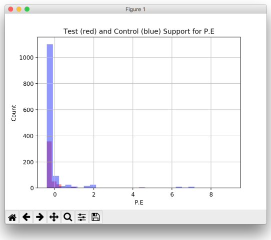
(c) Interest



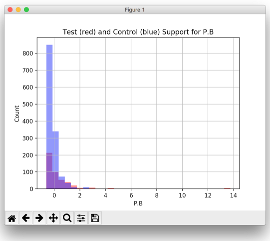
(d) Fin Lev



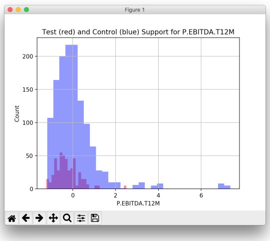
(e) Net Debt/EBITDA



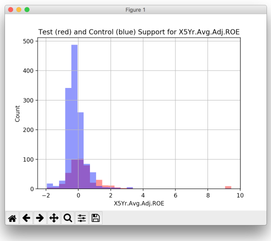
(f) P/E



(g) P/B



(h) P/EBITDA 12Mth



(i) Avg Adj ROE

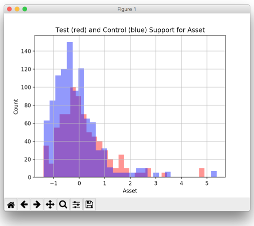
Figure 4.8: Causal Estimation - Independent Director & Financial Leverage

4.4.3 STOXX Eastern Europe 300

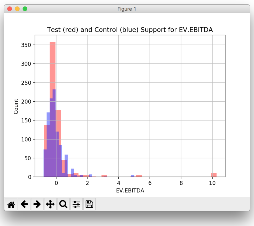
4.4.3.1 Independent Chairperson or Female CEO or Equivalent

Market	STOXX Eastern Europe 300
Outcome	Altman Z Score
Treatment	(Indep Chairman | Female CEO)? 1 : 0
Motivating Statement	Statement 5 [agrees, original]
Estimate ($\%\Delta, 95\%\ CI$)	(0.24, 0.29, 0.34)

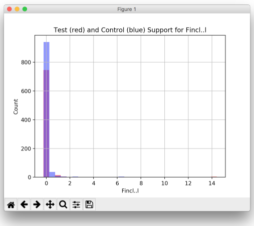
Matching Plots



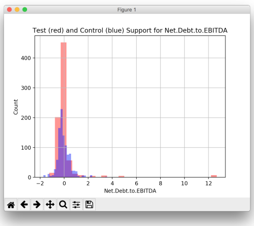
(a) Asset



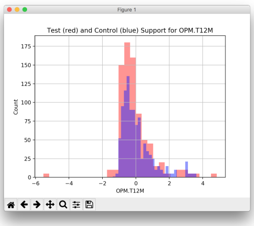
(b) EV/EBITDA



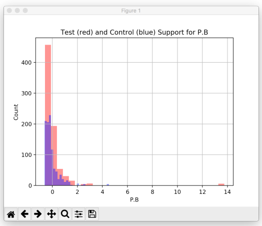
(c) Fin Lev



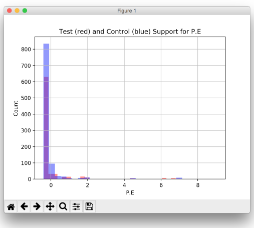
(d) Net Debt/EBITDA



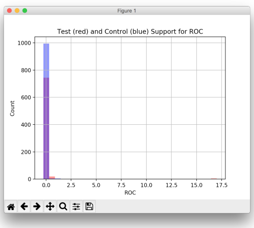
(e) OPM 12Mth



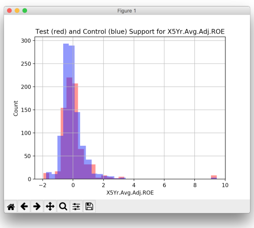
(f) P/B



(g) P/E



(h) ROC



(i) Avg Adj ROE

Figure 4.9: Causal Estimation - Indep Chairperson or Female CEO (or Equiv)

Feature Descriptions

The following table contains long form descriptions for some of the main variables used in this study. This table is reproduced from the work of Moldovan and Mutu (2015).

Variable	**Description**
Tax	Tax burden for the last 12 months
Interest	Interest burden for the last 12 months
Asset	Amount of sales or revenues generated per dollar of assets. The ratio is an indicator of the efficiency with which a company is deploying its assets
Fincl	Financial leverage. Measures the average assets to average equity
Oper ROE	Normalized ROE. Returns on Common Equity based on net income excluding one-time charges
Dvd P/O	Dividend Payout Ratio. Fraction of net income a firm pays to its shareholders in dividends, in percentage
Board Size	Number of Directors on the company?s board
% Non Exec Dir on Bd	Percentage of the board of directors that is comprised of non-executive directors
% Indep Directors	Independent directors as a percentage of total board membership
CEO Duality	Indicates whether the company?s Chief Executive Officer is also Chairman of the Board, as reported by the company
Indep Chrprsn	Indicates whether the company chairperson was independent as of the fiscal year end
Indep Lead Dir	Indicates whether the company has an independent lead director within the board of directors
Frmr CEO or its Equiv on Bd	Indicates whether a former company chief executive officer (CEO) or person with equivalent role has been a director on the board
% Women on Board	Percentage of Women on the Board of Directors

Variable	Description
Ind Dir Db Mtg Att	Percentage of board meetings attended by independent directors
Unit or 2 Tier Bd sys	Indicates whether the company?s board has a Unitary (1) or Two Tier (2) system. Marked 2 when board system has separate boards for Supervisory/Commissioner board and Management board
Prsdg Dir	Indicates whether the company has a presiding director in its board of directors
% Feml Execs Feml CEO or Equiv	Number of female executives, as a percentage of total executives. Indicates whether the company Chief Executive Officer (CEO) or equivalent is female
Age Young Dir	Age of the youngest director on the company board in years
BOD Age Rng	Age range of the members of the company board in years, calculated by subtracting the age of the youngest director on the company board from the age of the oldest director on the company board
Age Old Dir	Age of the oldest director on the company board in years Average age of the members of the board
Bd Avg Age	Average age of the memebers of the board
Board Duration	Length of a board member?s term, in years
Board Mtgs #	Total number of corporate board meetings held in the past year
Exec Dir Bd Dur	Length of an executive director board member?s term, in years

Program code

All code written in the fulfilment of this project is available at `https://github.com/ReidConor/dissertation`. Contained in that repository are all scripts and files used to preprocess the various datasets involved, run the numerous algorithms and finally compile the results. Note that this code frequently interacts with a local instance of MySQL, for data control and preservation purposes. The code required to establish this infrastructure is not included. The various files that make up this document are stored there also.

Bibliography

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

Bebchuk, C. A., L. and A. Ferrell. 2005. What matters in corporate governance? *Harvard Law School John M. Olin Center Discussion*.

Bebchuk, L. and A. Cohen. 2005. The costs of entrenched boards. *Journal of Financial Economics*, pages 409–433.

Beneish, M. 1999. The detection of earnings manipulation. *Financial Analyst Journal*, pages 24–36.

Beneish, M. and D. Nichols. 2007. The predictable cost of earnings manipulation. *SSRN-id1006840*.

Bhagat, S. and B. Bolton. 2008. Corporate governance and firm performance. *Journal of corporate finance*, **14**(3): 257–273.

Bolton, P., H. Chen and N. Wang. 2011. A unified theory of tobin's q, corporate investment, financing, and risk management. *The journal of Finance*, **66**(5): 1545–1578.

Chung, K. H. and S. W. Pruitt. 1994. A simple approximation of tobin's q. *Financial management*, pages 70–74.

Cook, R. D. 1977. Detection of influential observation in linear regression. *Technometrics*, **19**(1): 15–18.
URL `http://www.jstor.org/stable/1268249`

Core, R. H., J. and D. Larcker. 1999. Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics*, pages 371–406.

Cousineau, D. and S. Chartier. 2010. Outliers detection and treatment: a review. *International Journal of Psychological Research*, **3**(1): 58 – 67. ISSN 20112084.
URL `https://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=61167387&site=ehost-live`

Creamer, G. and Y. Freund. 2010. Learning a board balanced scorecard to improve corporate performance. *Decision Support Systems*, **49**(4): 365–385.

Dehejia, R. H. and S. Wahba. 2002. Propensity score matching methods for nonexperimental causal studies. *The Review of Economics and Statistics*, **84**(1): 151–161.

Eidleman, G. J. 1995. Z scores-a guide to failure prediction. *The CPA Journal*, **65**(2): 52.

Esarey, J. 2015. Causal inference with observational data.

Fatemi, G. M., A. and S. Kaiser. 2017. Esg performance and firm value: The moderating role of disclosure. *Global Finance Journal*.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth. 1996. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, **39**(11): 27–34.

Flach, P. 2007. Putting things in order. on the fundamental role of ranking in classification and probability estimation.

Fonti, V. and E. Belitser. 2017. Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*.

Fox, B. 2018. California prepares to impose female quota on company boards.
URL `https://www.ft.com/content/14d3aea0-ac82-11e8-94bd-cba20d67390c`

Freund, Y. and R. Schapire. 1996. Experiments with a new boosting algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference*, page 148–156.

Guo, K. T., R. and T. Nohel. 2008. Undoing the powerful anti-takeover force of staggered boards. *Journal of Corporate Finance*, pages 274–288.

Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, pages 1251–1271.

Herawati, N. 2015. Application of beneish m-score models and data mining to detect financial fraud. *Procedia - Social and Behavioral Sciences*, pages 924–930.

Hoerl, A. E. and R. W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**(1): 55–67. ISSN 00401706. URL `http://www.jstor.org/stable/1267351`

Horton, N. and K. Kleinman. 2007. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *American Statistician*, **61**: 79–90.

Jakobsen, J. C., C. Gluud, J. Wetterslev and P. Winkel. 2017. When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts. *BMC medical research methodology*, **17**(1): 162.

Kamal, M., M. Salleh and A. Ahmad. 2016. Detecting financial statement fraud by malaysian public listed companies: The reliability of the beneish m-score model. *Jurnal Pengurusan*, pages 23 – 32.

King, G., C. Lucas and R. A Nielsen. 2014. The balance-sample size frontier in matching methods for causal inference. *American Journal of Political Science*.

Kirkpatrick, G. 2009. The corporate governance lessons from the financial crisis. *OECD Journal: Financial Market Trends*, **2009**(1): 61–87.

Mahama, M. 2015. Detecting corporate fraud and financial distress using the altman and beneish models. *International Journal of Economics, Commerce and Management*, **III**.

McLaughlin, E. C. 2017. Man dragged off united flight has concussion, will file suit, lawyer says.
URL `http://edition.cnn.com/2017/04/13/travel/united-passenger-pulled-off-flight-lawsuit-family-attorney-speak/`

McVeigh, I. 2015. Volkswagen scandal: Bad governance is often a sign of trouble ahead.
URL `http://www.telegraph.co.uk/finance/comment/11893886/Volkswagen-scandal-Bad-governance-is-often-a-sign-of-trouble-ahead.html`

Moldovan, D. and S. Mutu, 2015. Learning the relationship between corporate governance and company performance using data mining. In: *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pages 368–381. Springer.

Orlitzky, M., F. L. Schmidt and S. L. Rynes. 2003. Corporate social and financial performance: A meta-analysis. *Organization studies*, **24**(3): 403–441.

Pearl, J. 2009. *Causality*. Cambridge university press.

Pearl, J. and T. S. Verma. 1995. A theory of inferred causation. *Studies in Logic and the Foundations of Mathematics*, **134**: 789–811.

Pollet, T. V. and L. van der Meij. 2017. To remove or not to remove: the impact of outlier handling on significance testing in testosterone data. *Adaptive Human Behavior and Physiology*, **3**(1): 43–60. ISSN 2198-7335. doi:10.1007/s40750-016-0050-z.
URL `https://doi.org/10.1007/s40750-016-0050-z`

Rosenbaum, P. R. and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, pages 41–55.

Rubin, D. B. 2004. *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

Schaltegger, S. and T. Synnestvedt. 2002. The link between green and economic success: environmental management as the crucial trigger between environmental and economic performance. *Journal of environmental management*, **65**(4): 339–346.

Stuart, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, **25**(1): 1.

Tobin, J. 1969. A general equilibrium approach to monetary theory. *Journal of money, credit and banking*, **1**(1): 15–29.

van Buuren, S. and K. Groothuis-Oudshoorn. 2011. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, **45**(3): 1–67.
URL `https://www.jstatsoft.org/v45/i03/`

Wahba, H. 2008. Does the market value corporate environmental responsibility? an empirical examination. *Corporate Social Responsibility and Environmental Management*, **15**(2): 89–99.

Yildirim, I. 2012. Bayesian inference: Gibbs sampling.

Zijlstra, W. P., L. A. van der Ark and K. Sijtsma. 2011. Outliers in questionnaire data: Can they be detected and should they be removed? *Journal of Educational and Behavioral Statistics*, **36**(2): 186–212.